



# EXTERNAL QUALITY ASSESSMENT (EQA) REPORTS AND THEIR INTERPRETATION

*Dr Jocelyn Naicker  
National Health Laboratory Service (NHLS)  
and University of the Witwatersrand  
SAACB Education Day 2010*

# EXTERNAL QUALITY ASSESSMENT SCHEME (EQAS)

- External agency
- Evaluation is performed on several Labs that analyse the same specialised samples
- Performed at regular intervals
- Frequency may vary between schemes
- Retrospective assessment

- Compares performance between labs
- National/regional basis – several:  
NHLS EQAS, Thistle, UKNEQAS, BIORAD EQAS,  
RIQAS, CAP, RCPA etc

“Blind” samples



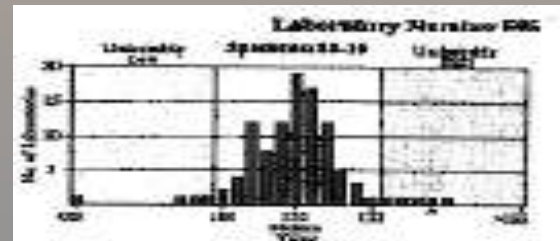
Analysis same as patient samples



results are returned to EQAS  
organiser



Statistical analysis



## NHLS EQAS (2001)

NHLS labs, public sector, Southern African countries

Clinical Chemistry

Blood gases

TDM



monthly

Target values for most

# **THISTLE – SA**

National and International

Central values – Consensus means

EQAS of European countries vary widely in the frequency and type of specimen analysed, target values/consensus means as well as in the criteria for acceptable performance

## **USA (CAP)**

Proficiency testing became a legal requirement in 1988 when the CLIA (Clinical Laboratories Improvement Amendments) were approved

- Frequency of sampling (up to 5 specimens 3x per annum)



# EQA GOALS

1. An objective external (independent) assessment of the lab's long term performance
2. To alert the lab of problems with accuracy (aimed to at preventing the reporting of incorrect patient results. (failure of internal alerts)

3. To provide Management and clients with the confidence that the on-going processes for monitoring analytical processes is sound
4. Identification of group problems – (affecting specific analysers/ reagents/ kits)

5. Complimentary to the IQC which monitors daily accuracy of the lab's methods

# BASIC STATISTICS

- **Mean** =  $\frac{\sum x}{n}$

is the sum of all the measurements  
divided by the number of measurements

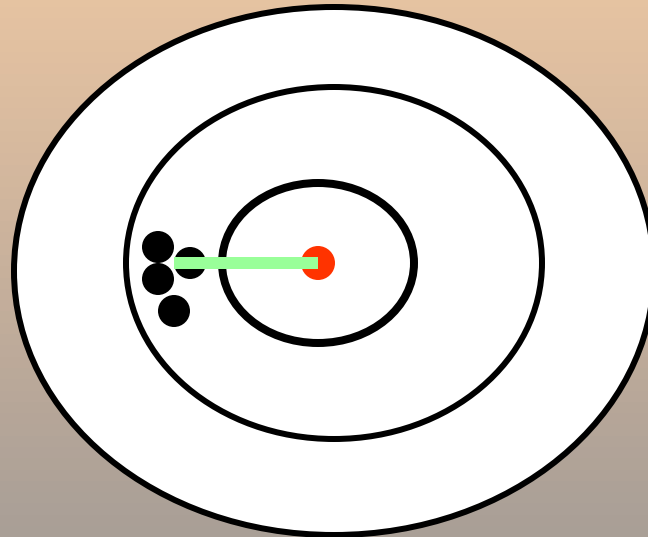
- **Median** is the point on the scale that has an equal number of observations above and below.
- **Mode** is the most frequently occurring result.

- **TRUENESS**

defined as the **closeness** of the agreement between the mean of a large series of measured results and the true concentration

# BIAS

Is the mathematical **difference** between the mean value of a series of measurements and the true value



and is a measurement of  
**SYSTEMATIC ERROR**

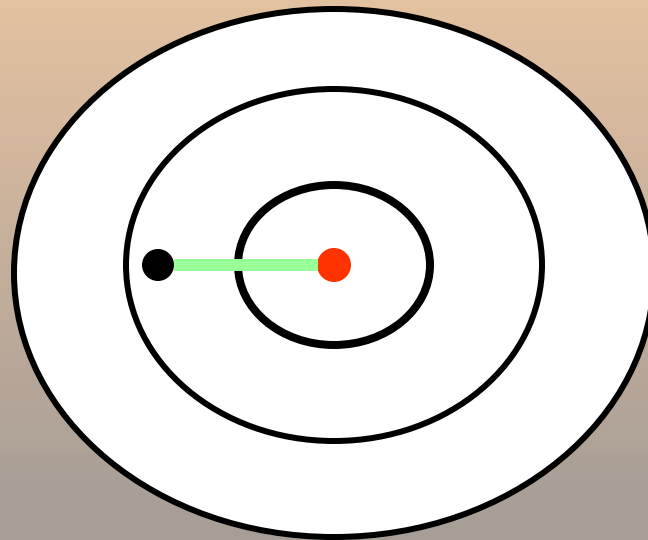
% Bias

$$= \frac{\text{Result} - \text{True value}}{\text{TV}} \times 100$$



# ACCURACY

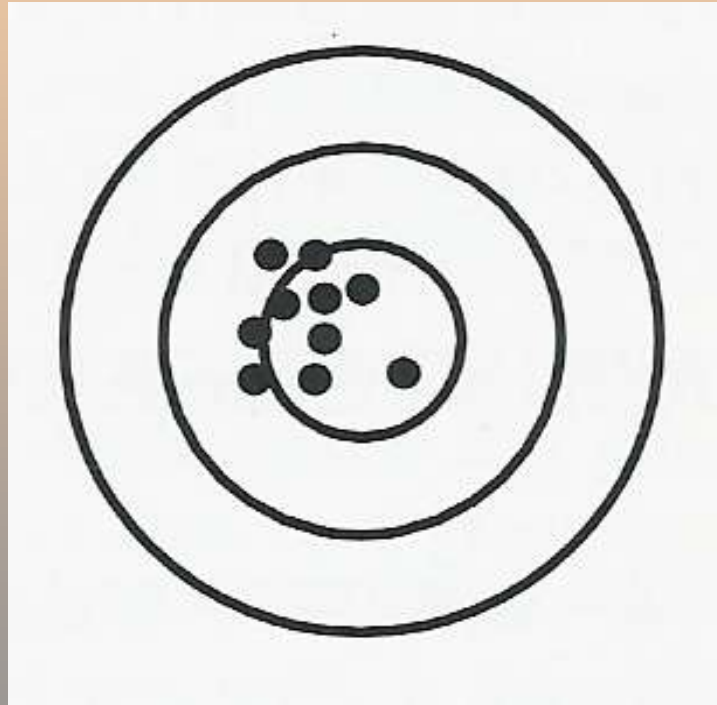
defined as the **closeness** of the agreement between a **SINGLE** observed measurement and the true concentration and is inversely related to the UOM



**Influenced by both bias and imprecision**  
and is therefore an expression of Total Error

# PRECISION

refers to the **closeness** of agreement of a series of independent measurements



# **PRECISION**

divided into:

## **1. Repeatability**

constant conditions (“within run”)

## **2. Reproducibility**

changed conditions:

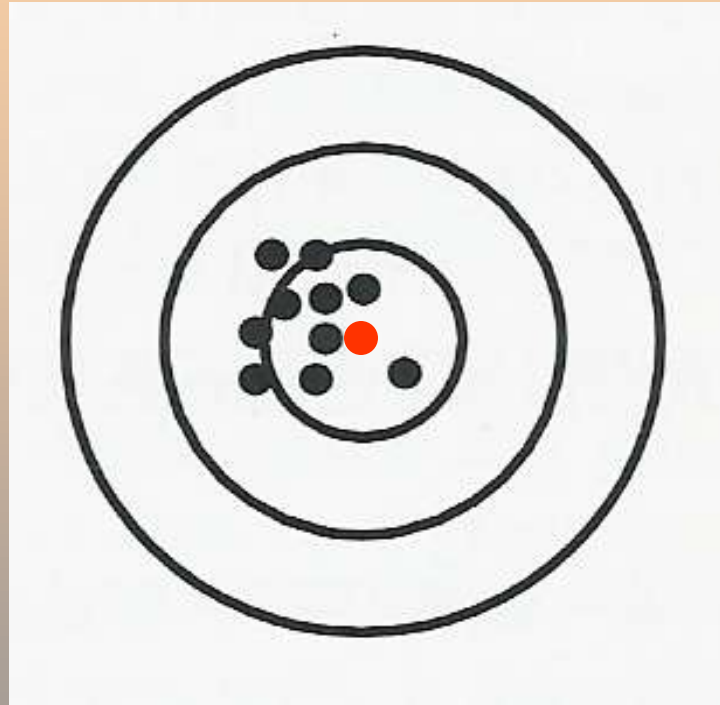
- Total/“Between run” precision  
(intermediate precision)
- Interlaboratory precision (EQAS)

# **IMPRECISION**

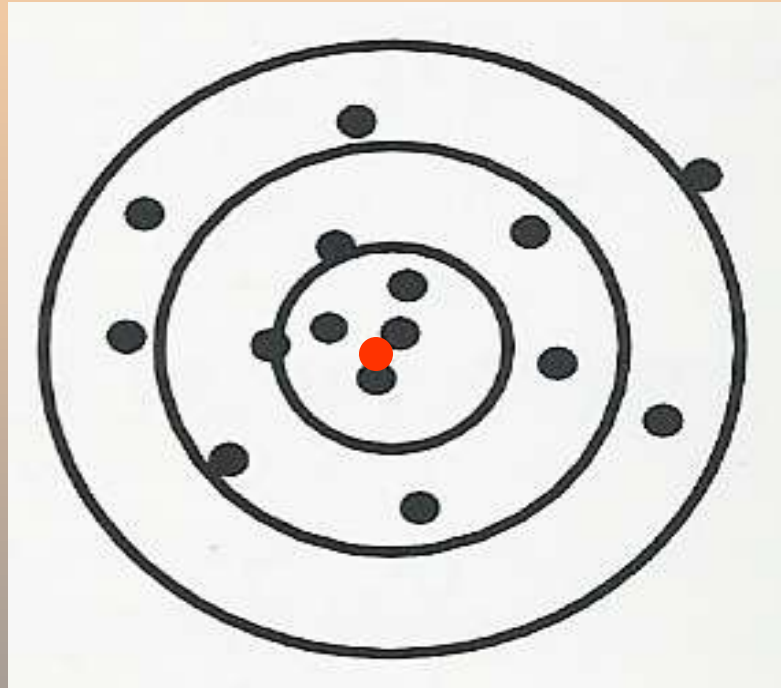
the inverse relationship of precision

is only related to **RANDOM ERRORS**  
of measurements  
expressed as:  
SD and CV

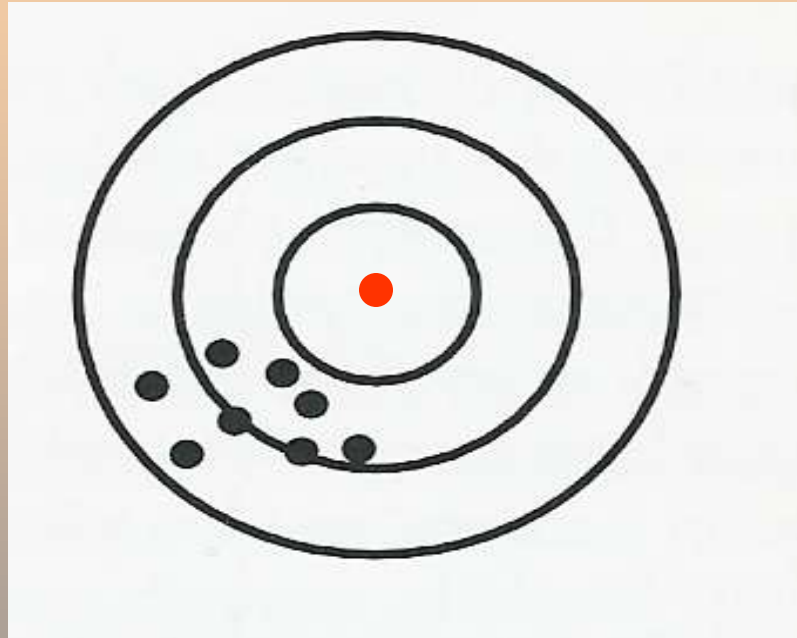
Imprecision LOW



Imprecision **HIGH**

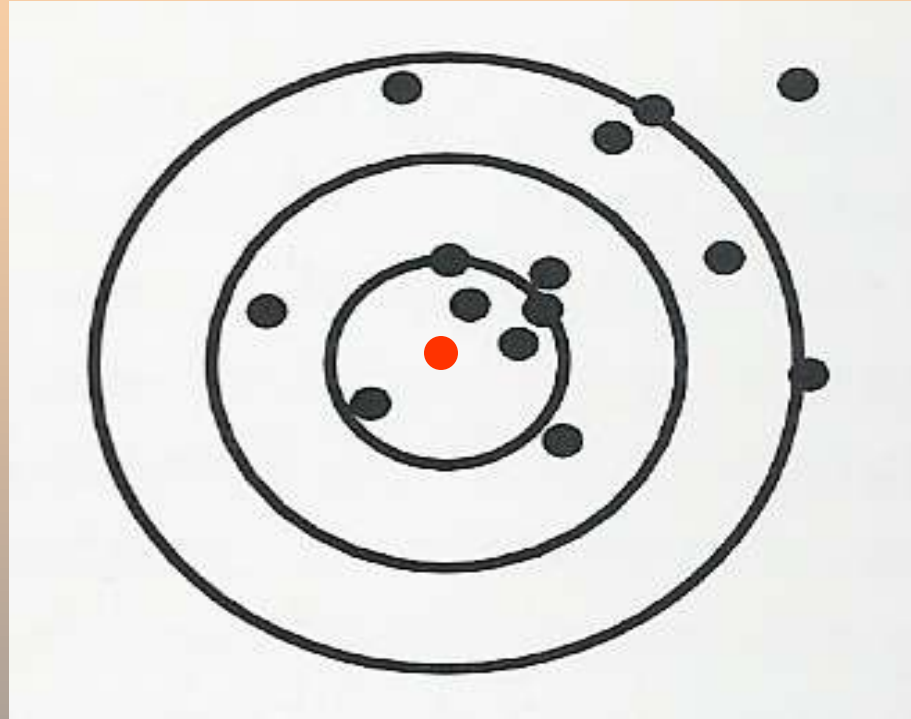


Imprecision LOW



Imprecision **HIGH**

Bias **HIGH**





# Standard Deviation (SD)

SD describes scatter around the mean  
(expressed in the units being measured)

# Standard Deviation (SD)

$$SD = \sqrt{\frac{\sum (X_1 - X)^2}{N - 1}}$$

Mean absolute deviation

# Coefficient of Variation (CV)

$$CV(\%) = \frac{\text{SD}}{\text{Mean}} \times 100$$

The smaller the CV the more reproducible the results

# Performance Criteria

Which results are Acceptable?

A simple method of assessing performance

to indicate whether:

- persistent dissatisfactory performance
- deterioration in performance
- any improvement
- casual errors

# **STATISTICAL METHODS**

Statistical Methods are used to assist in determining whether performance is acceptable by analysing, summarising and reporting on “observations”

In assessing performance using these Statistical methods one should always consider performance at the important **Medical Decision levels** and the clinical impact of unacceptable results

## How is Objectivity achieved?

To enable Clinical relevance of the results obtained the results should be within a **certain %** of **Assigned central values**

# ASSESSMENT OF PERFORMANCE

Based on

1. Central Reference value

1. Acceptable Range of performance  
(Allowable limits of performance)



## EQAS Assignment of Analyte dependant

### **Central Reference Value:**

1. overall mean/median of all participating labs are used to indicate the ‘True value’

Or

2. Set Target values

(specified by each EQAS’s program information)

## How are TARGET VALUES set?

1. Reference labs using Reference methods
2. Labs using “Higher order” methods
  - precision, trueness and analytical specificity that is  $>$  that in routine labs

3. SRM samples by selected Labs

4. Weighed - in values e.g. drugs

- EQA samples supplemented with known weights of and analyte

# ALLOWABLE LIMITS OF PERFORMANCE (ALP)

- is unique for each analyte
- considers unavoidable imprecision of the method
- and normal diurnal variations (BV)
- TAE calculation

# ALP

Determines the **level of error** that is required to minimise the risk of adverse outcome when applied to the reporting of patient results

ALP

includes **RANDOM ERRORS** due to:

1. Within and between run variability
2. between calibration variability
3. lot to lot variability of reagent/calibrator
4. between analyser variability

ALP

In addition there may be  
**SYSTEMATIC ERRORS** due to:

1. Calibration standards used
2. Method / Instrument specific factors  
(Calcium – arsenazo vs cresolphthalein complex)

# ALP

Is expressed as :

**$\pm x$  from the TV**

$x$  may be in absolute units

or in %

or in an absolute value up to a certain limit

and a % above a certain limit



# Standard Deviation Index (SDI)

$$\text{SDI} = \frac{\text{Lab result} - \text{Group Mean}}{\text{SD of Group}}$$

SDI > 2

Lab not in agreement  
with labs within the Group

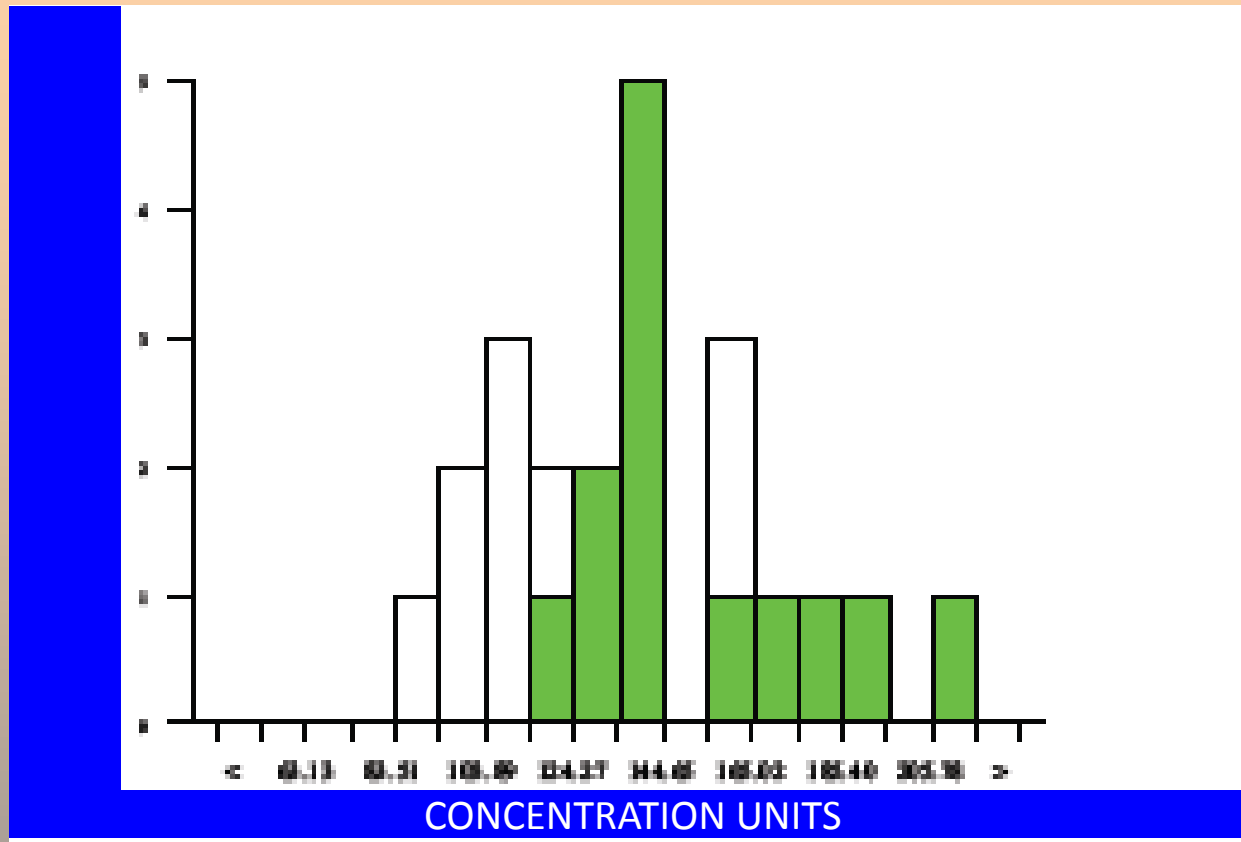
## **COEFFICIENT of VARIATION INDEX (CVI)**

$$= \frac{\text{Lab CV}}{\text{Group CV}}$$

an index of imprecision

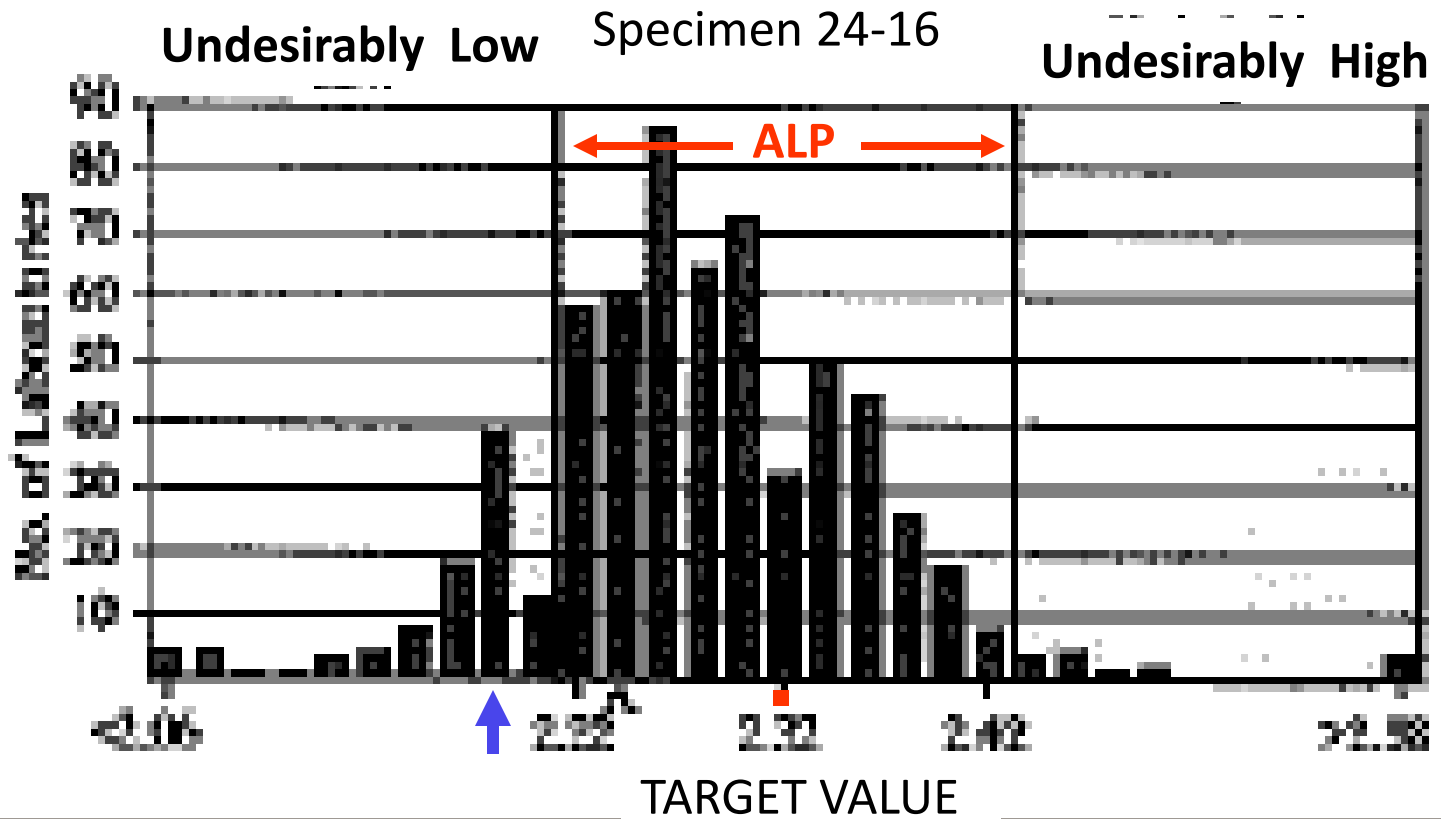
# BAR GRAPHS/HISTOGRAMS

NO. OF LABS



Qualitative results Histogram allows quick visualisation if your lab's result falls within the overall results for all methods as well as for your method group

# HISTOGRAMS

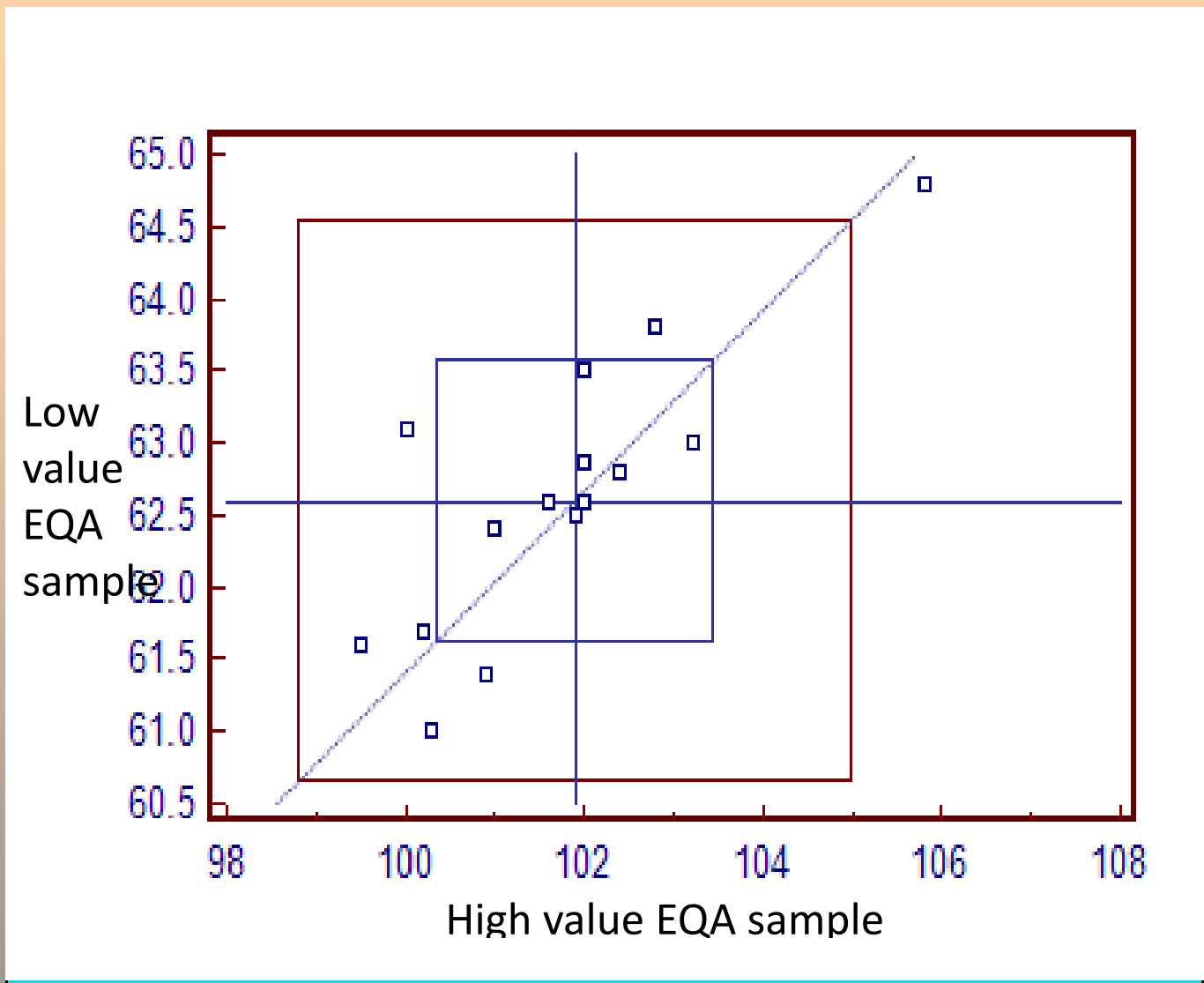


## **YOUDEN PLOTS**

when 2 EQA samples – high and low  
are analysed by each lab (same method  
analytical principle, instrument, reagent)

observed results plotted

Allows comparison of the relationship of  
each level's value to the group's  
performance



Blocks = SDs

Horizontal and vertical median reference lines have equal no. of observations on either side

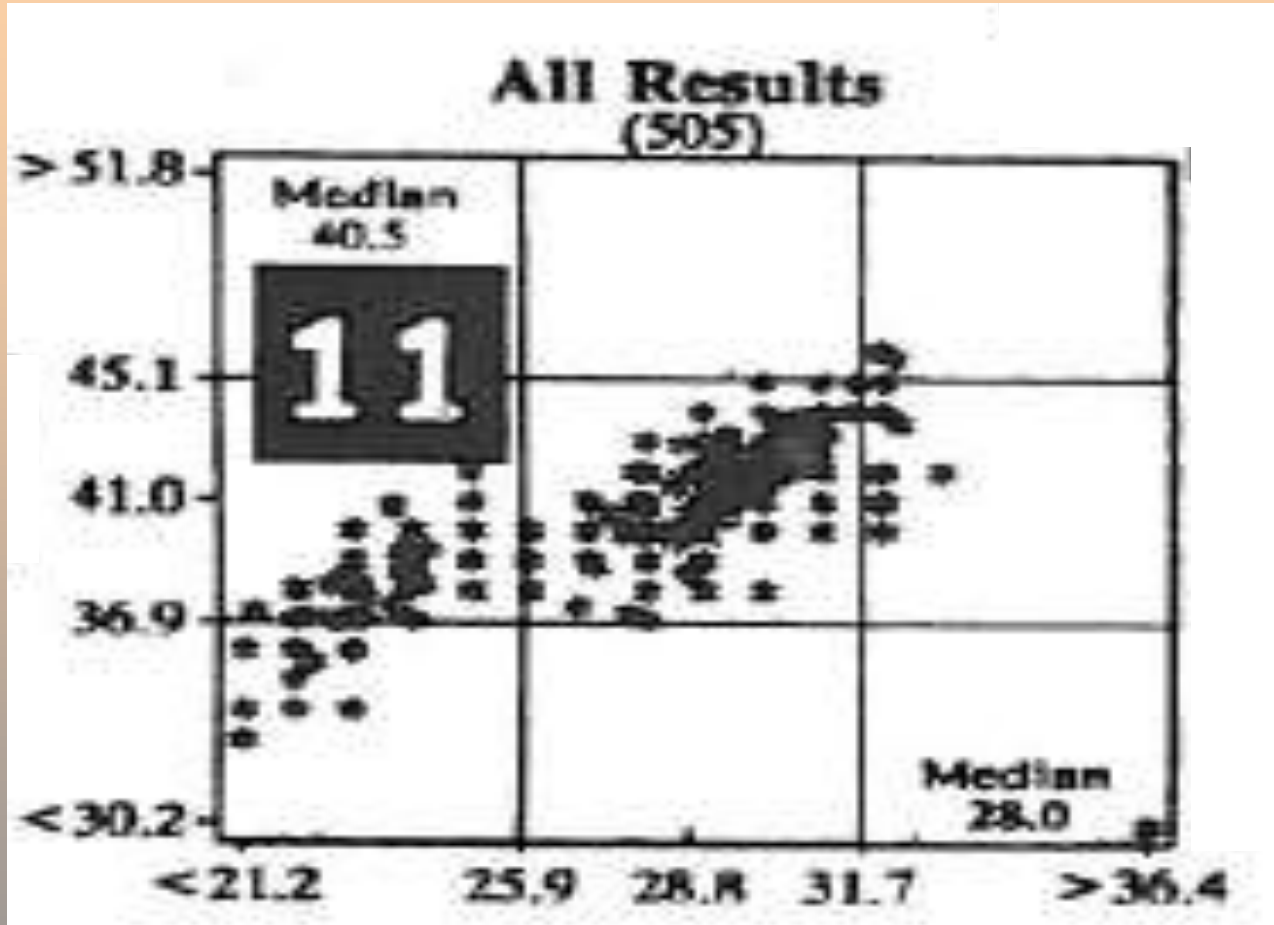
Central intersection of the median lines = Manhattan median

Far away from MM but close to ref line + Systematic error

Far from reference lines = large random errors

# YOUDEN PLOT

HIGH  
VALUE  
EQA

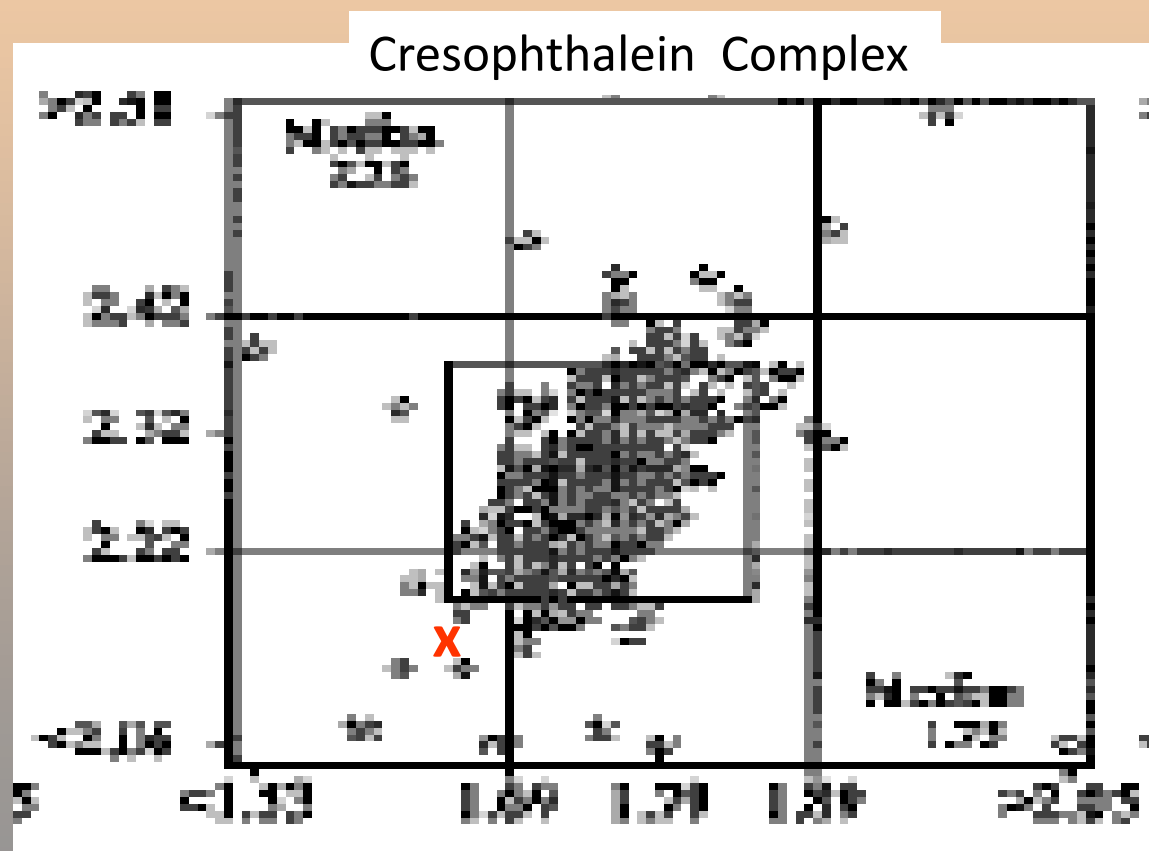


LOW VALUE EQA

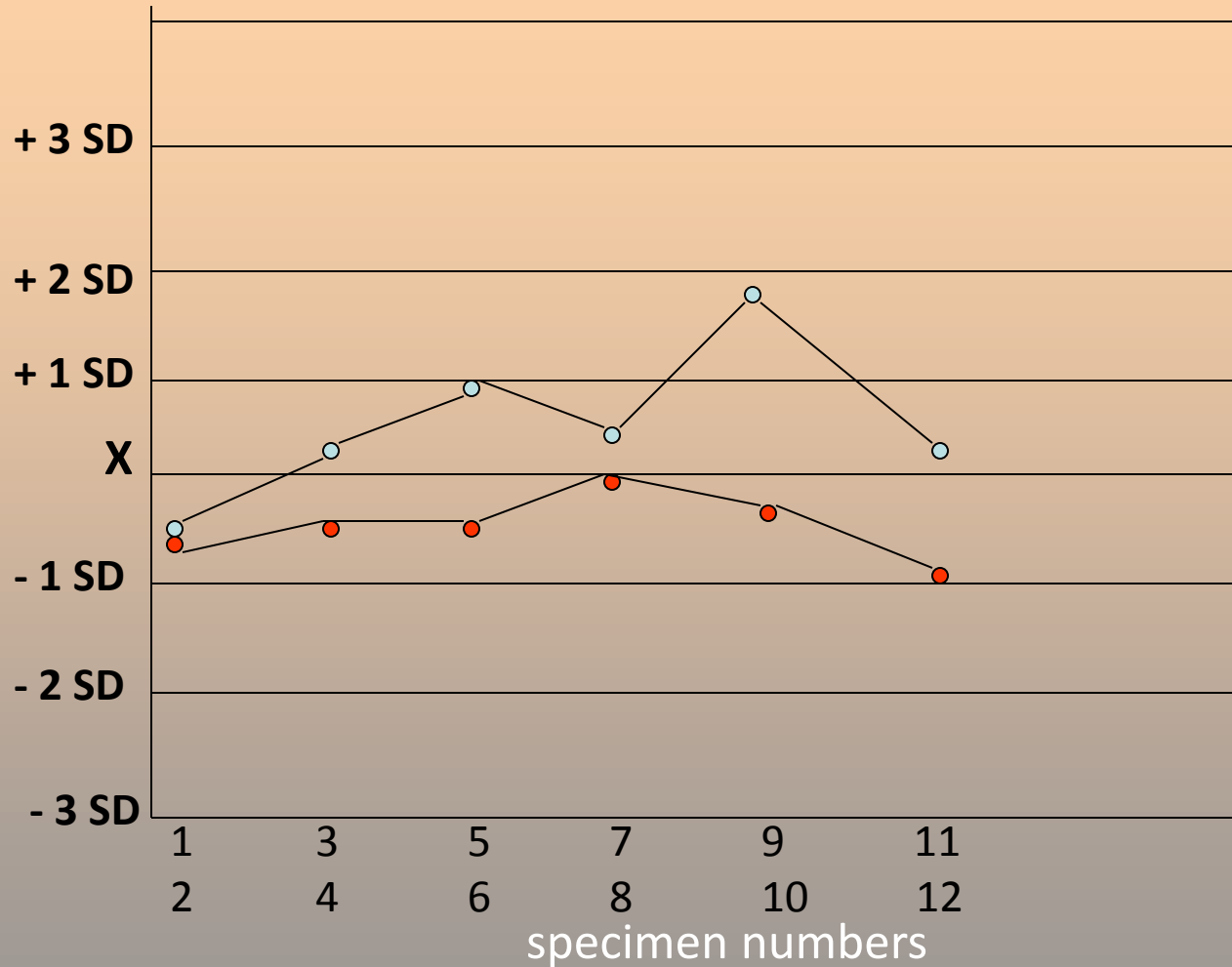
Albumin



# PEER REVIEW BOXES (Subgroups)

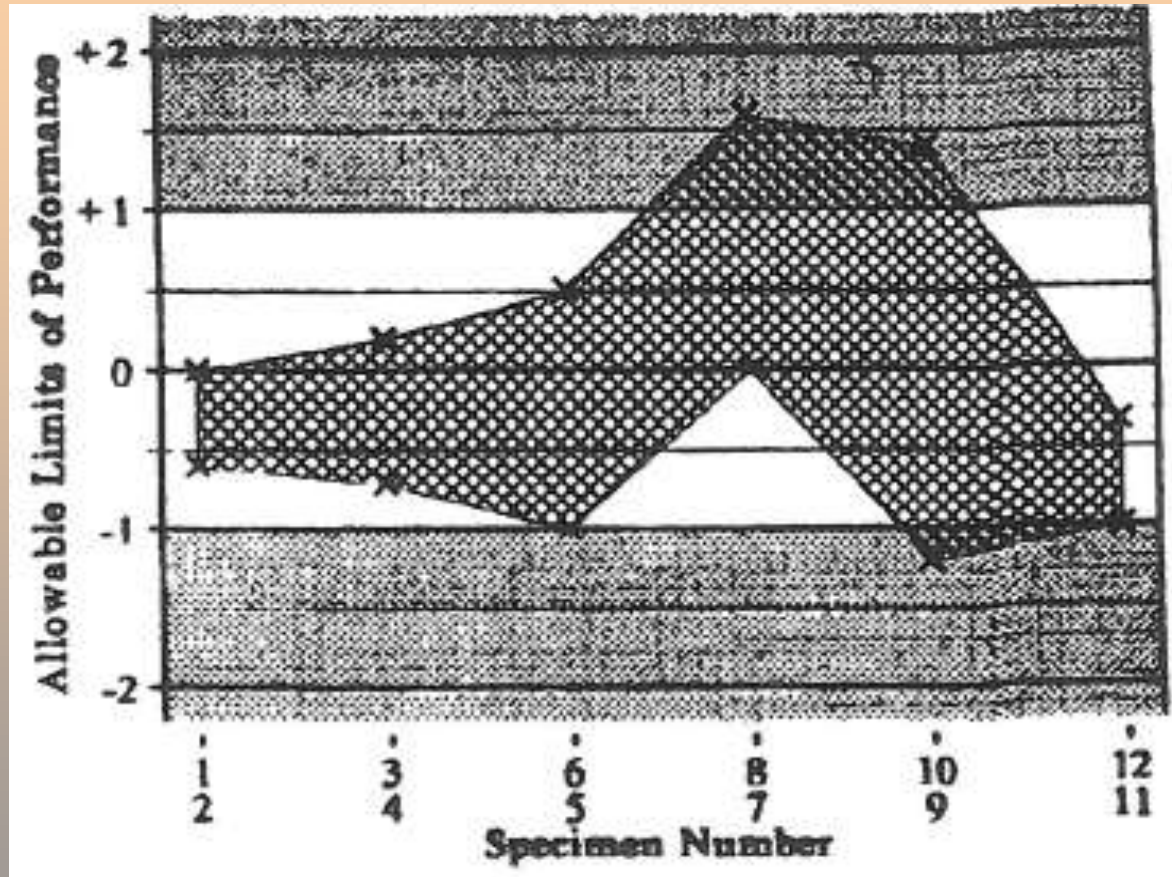


ALP

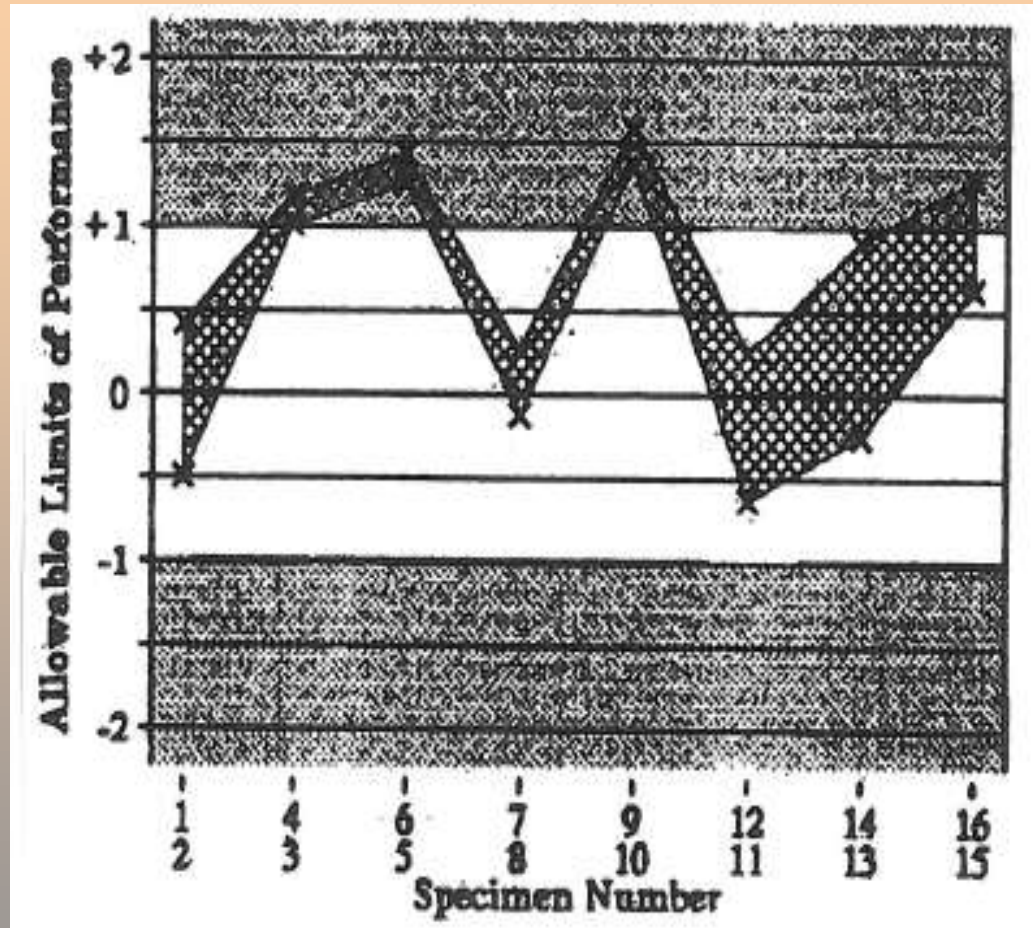


LEVEY – JENNINGS style of reports

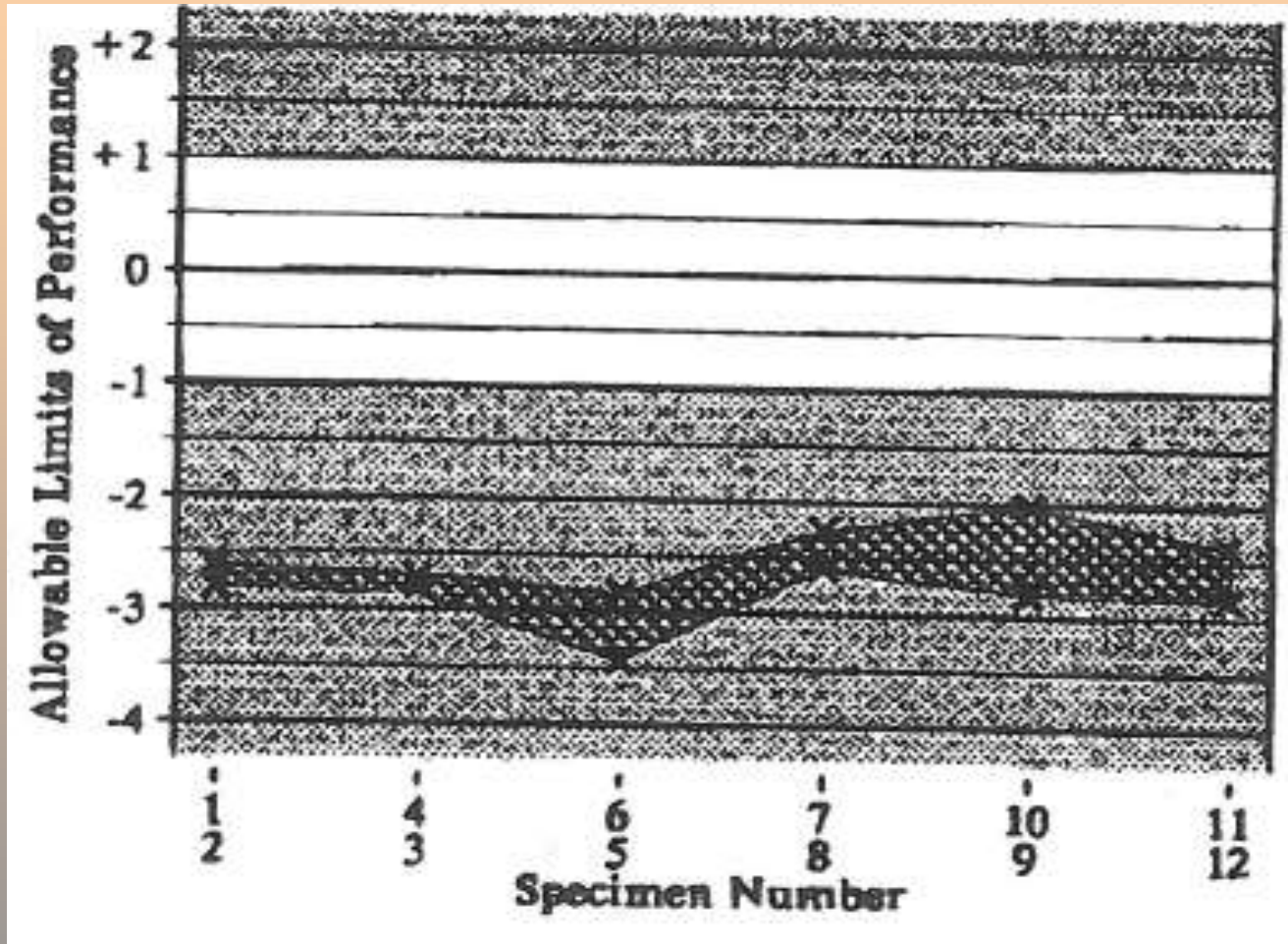
# Levey-Jennings-Like Plots



poor precision – within run



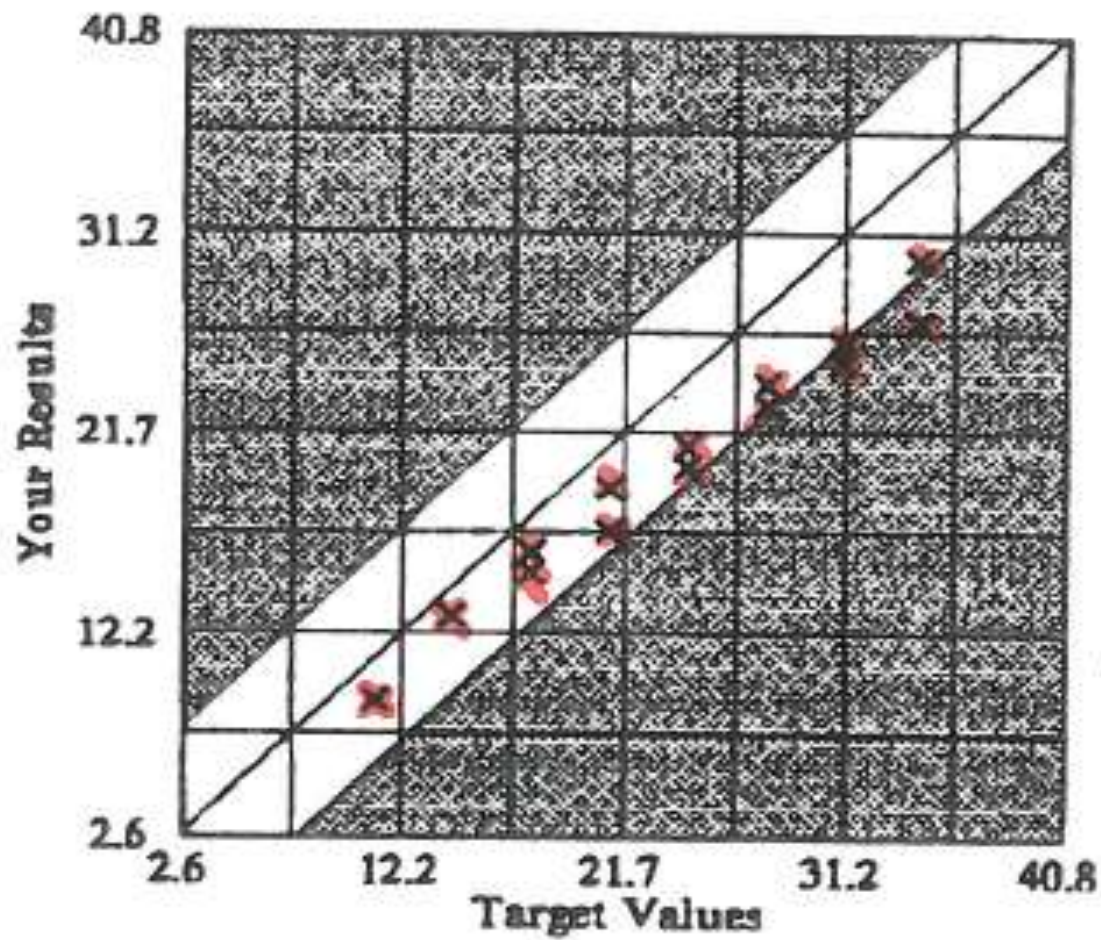
poor precision – between run



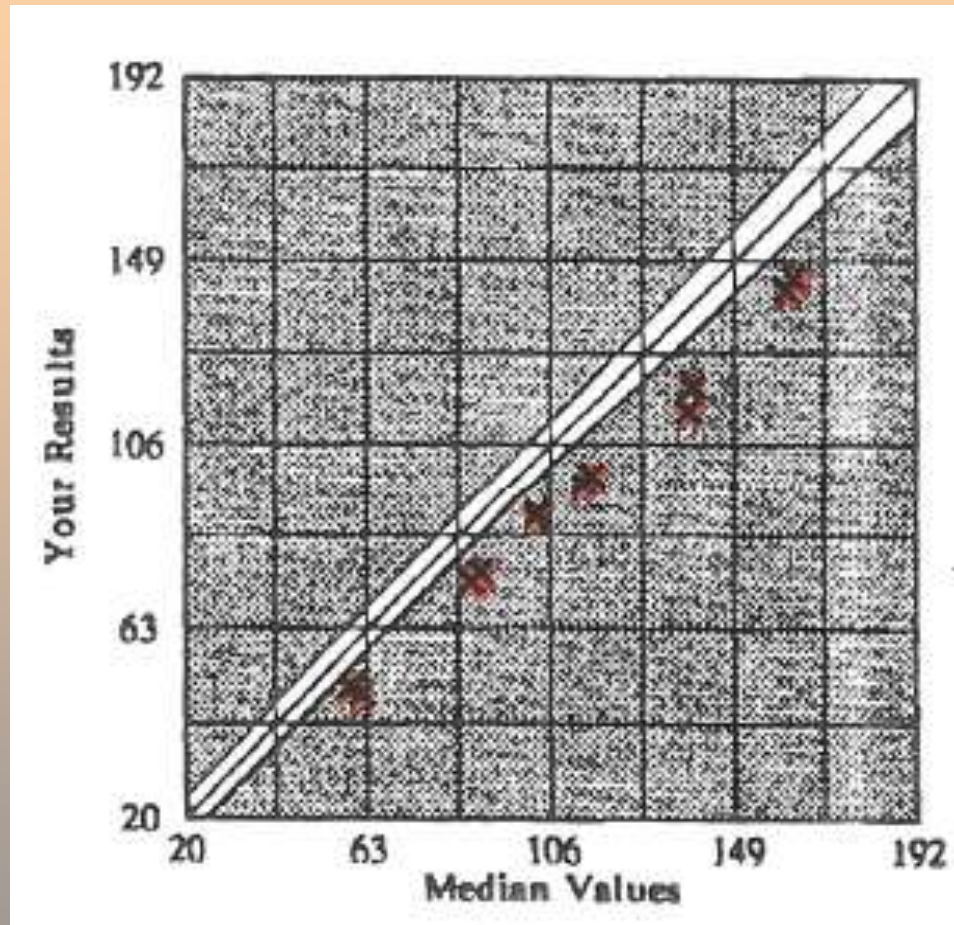
Biased results

# **LINEARITY GRAPHS**

Linearity refers to the relationship between measured and expected values over the analytical range  
visualise whether relationship is linear or not

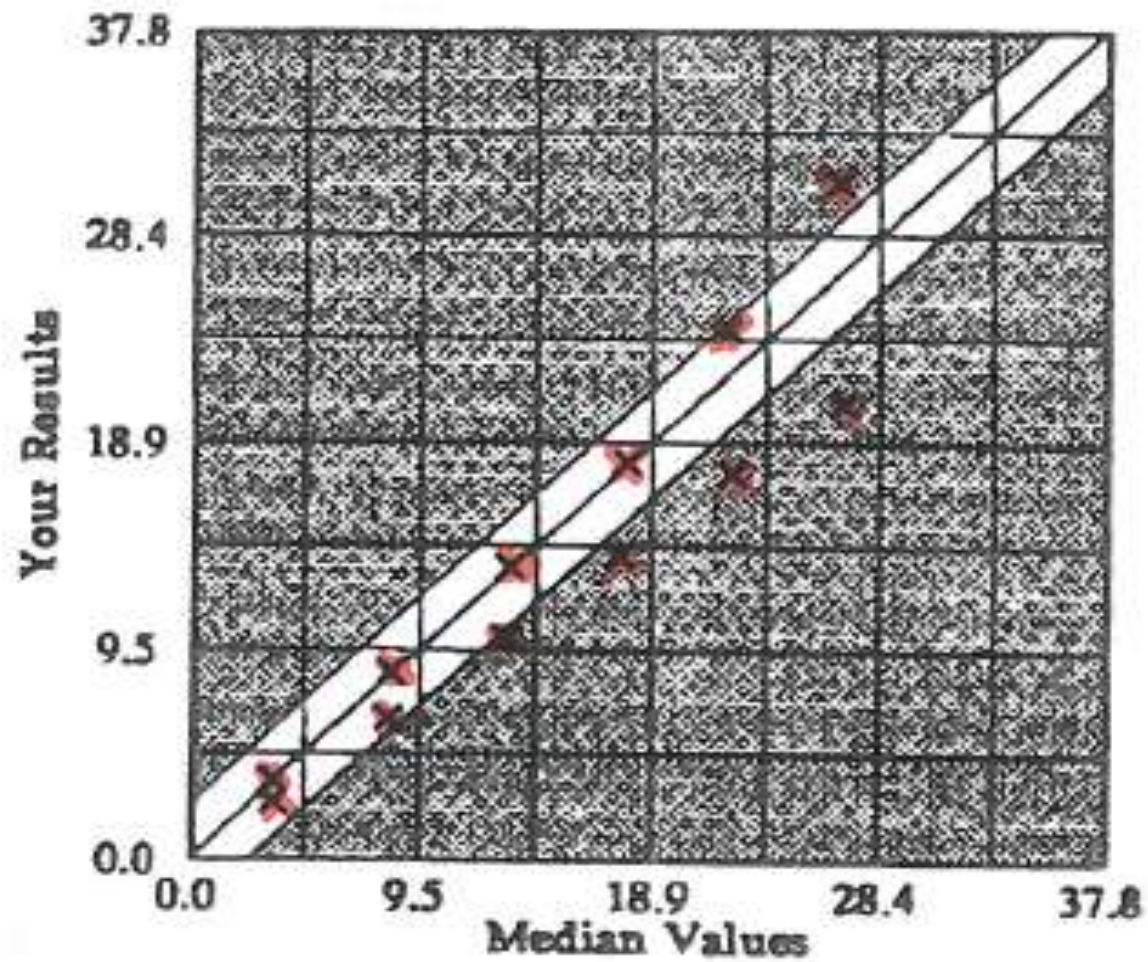


Non - Linear



Proportional Bias





Imprecision

# BLAND ALTMAN PLOTS

Differences between the lab's results  
and the Group (consensus) means (y-axis)

plotted against

Group means (x-axis)

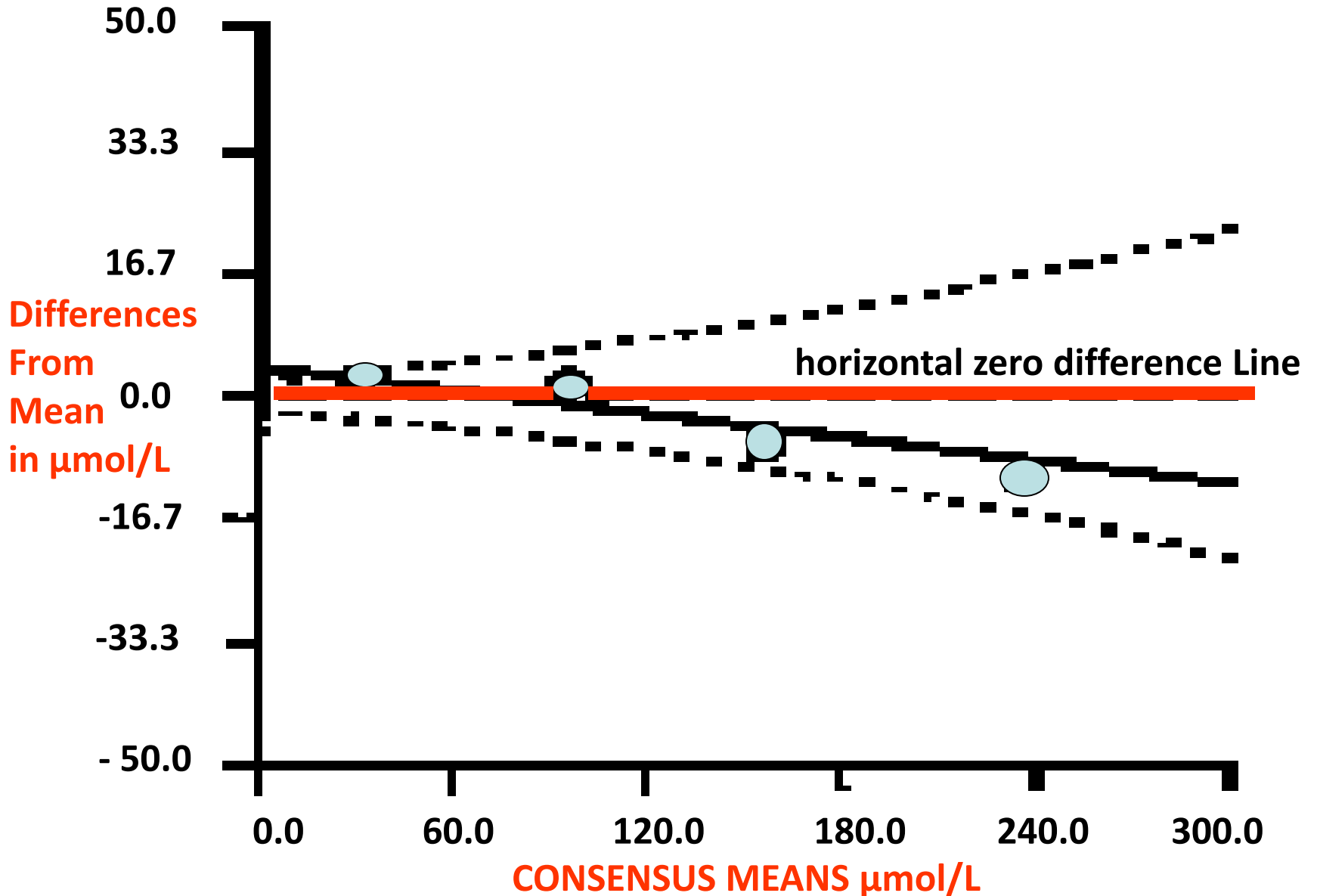
# PHENOBARBITAL

UKNEQAS

EXPECTED LINE

FITTED LINE

95 % CONFIDENCE LIMITS



Results of Survey number 1 of 2010/11

Date of Report: 12 May 2010

Defined Target

LABORATORY: 003

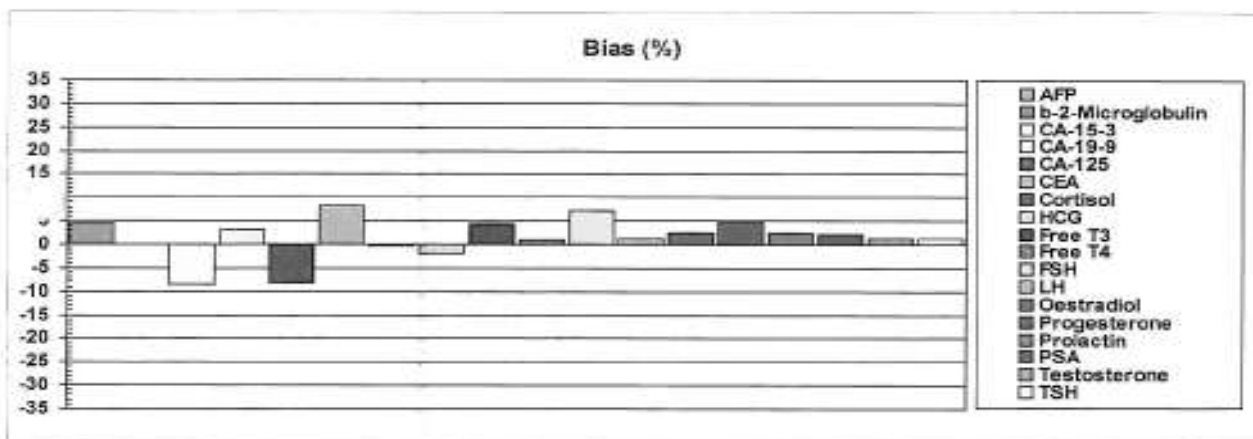
Hospital x

REGION: |

INSTRUMENT:

CENTAUR/IMMULITE

TEST	RESULTS	METHOD	FLAG	BIAS %	ACCEPTANCE LIMITS			ALL LAB MEAN	No.
					LOW	HIGH	TARGET		
a-Fetoprotein kU/L	190.00	Centaur		4.40	155.00	209.00	182.00	190.000	4
b-2-Microglobulin mg/L		Pro Spec		NR					
CA-15-3 kU/L	90.60	Centaur		-8.48	84.20	114.00	99.00	97.233	3
CA-19-9 kU/L	199.00	Centaur		3.11	184.00	222.00	193.00	104.155	2
CA-125 kU/L	167.00	Centaur		-8.24	195.00	208.00	182.00	166.000	3
CEA ug/L	62.20	Centaur		8.30	41.00	55.40	48.20	51.575	4
Cortisol nmol/L	1074.00	Centaur		-0.28	897.00	1257.00	1077.00	1113.000	3
HCG U/L	550.00	Immufite		-1.96	488.00	653.00	561.00	550.000	1
Free T3 pmol/L	23.50	Centaur		4.44	19.10	25.90	22.50	23.100	4
Free T4 pmol/L	37.20	Centaur		1.09	31.30	42.30	36.80	36.560	5
FSH U/L	75.30	Centaur		7.11	59.80	80.80	70.30	74.675	4
LH U/L	81.10	Centaur		1.33	50.80	69.80	60.30	58.800	4
Oestradiol pmol/L	1578.00	Centaur		2.54	1153.00	1921.00	1537.00	1633.333	3
Progesterone nmol/L	105.00	Centaur		4.95	79.80	122.00	101.00	102.675	4
Prolactin ug/L	47.00	Centaur		2.40	39.10	52.80	45.90	45.150	4
PSA ug/L	21.90	Centaur		2.34	18.20	24.60	21.40	22.300	5
Testosterone nmol/L	37.30	Immufite		1.38	28.60	45.00	36.80	37.300	1
TSH mU/L	23.10	Centaur		1.32	19.40	26.20	22.80	24.120	6



Target values are not available for all instruments for all of the analytes. The % Bias cannot be calculated for those results.

21/05/10

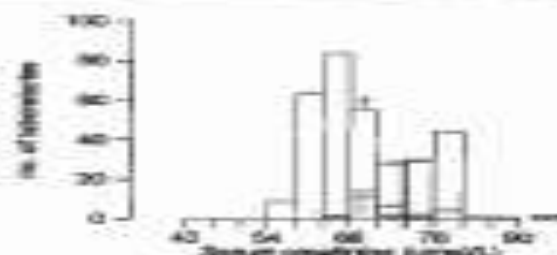


Spec	Pool	Pool description / Treatment / Addition
486	120	Pool: human serum (Pool 121) diluted 4:1 with P60
488	122	Pool: 121 with 10 umol/L added creatinine
489	149	Pool: 148 (400) with 80 umol/L added creatinine

- All methods
- Dry slide
- CCD (J&J) Gen 80+ (1402)

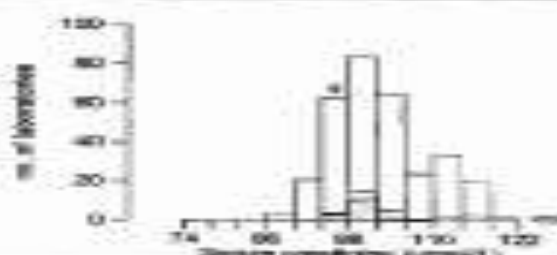
Pool A score is	
Pool B score is	-4.8
Pool C score is	2.1

Specimen :	n	Mean	SD	CV(%)
All methods	313	68.3	7.7	11.3
Dry slide	30	71.4	4.8	6.6
CCD (J&J) Gen 80+ (1402)	15	69.7	1.4	2.0
CCD (J&J) original (140)	15	74.8	8.8	11.8
Compartek Kinetic Jaffe	148	64.3	3.5	5.5
Beckman reagents (1000)	32	63.3	4.0	6.3
Roche Moduler reagents (1000)	74	65.8	2.8	4.2
Traditional Kinetic Jaffe	80	71.8	9.8	13.6
Abbott reagents (1148)	27	63.3	5.1	8.1
Olympus reagents (1104)	41	78.2	2.3	2.9
O'Leary	19	77.3	5.8	7.5
Enzymatic	20	69.7	2.2	3.2



Your result	67
Target value (ALTM)	68.3
Your specimen's bias	-1.3
Normalized bias	-1.9
Accuracy Index	18
True value	68
CCD (J&J) Gen 80+ (1402) (reagent)	69.7

Specimen :	n	Mean	SD	CV(%)
All methods	318	102.8	7.2	7.0
Dry slide	30	102.8	5.5	5.4
CCD (J&J) Gen 80+ (1402)	15	102.8	1.8	1.8
CCD (J&J) original (140)	15	100.8	7.8	7.7
Compartek Kinetic Jaffe	148	100.8	4.5	4.5
Beckman reagents (1000)	32	98.3	3.8	3.8
Roche Moduler reagents (1000)	74	102.0	3.8	3.8
Traditional Kinetic Jaffe	80	108.4	10.0	9.4
Abbott reagents (1148)	27	98.8	8.3	8.5
Olympus reagents (1104)	41	113.7	3.8	3.3
O'Leary	19	107.0	4.3	4.0
Enzymatic	20	98.8	3.0	3.1



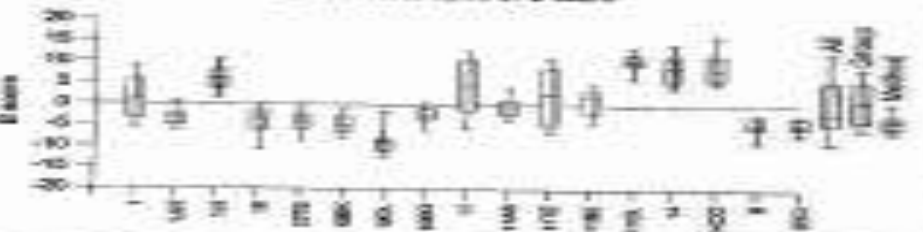
Your result	99
Target value (ALTM)	102.8
Your specimen's bias	-4.7
Normalized bias	-4.7
Accuracy Index	47
True value	97
CCD (J&J) Gen 80+ (1402) (reagent)	102.8

Specimen :	n	Mean	SD	CV(%)
All methods	318	57.7	6.5	11.3
Dry slide	30	67.1	7.8	11.6
CCD (J&J) Gen 80+ (1402)	15	60.8	4.1	6.6
CCD (J&J) original (140)	15	72.3	7.8	10.8
Compartek Kinetic Jaffe	148	52.5	3.2	6.1
Beckman reagents (1000)	32	53.1	4.4	8.2
Roche Moduler reagents (1000)	74	52.8	2.5	4.8
Traditional Kinetic Jaffe	80	67.4	9.0	13.1
Abbott reagents (1148)	27	53.7	4.8	9.4
Olympus reagents (1104)	41	60.3	3.1	5.1
O'Leary	19	63.8	3.8	6.1
Enzymatic	20	60.8	2.3	3.8



Your result	60
Target value (ALTM)	57.7
Your specimen's bias	+2.3
Normalized bias	+3.9
Accuracy Index	39
True value	60
CCD (J&J) Gen 80+ (1402) (reagent)	60.8

Median and IQRs of B score



Median and IQRs of C score



## **Mayo Clinic study**

Proficiency testing programs are far from ideal monitors of lab performance

50% errors - problems with the survey itself  
only 28 % related to specific analytical problems

## **TROUBLESHOOTING**

1. Incorrect classification
2. Incorrect units / conversion
3. Incorrect sample tested
4. Technical errors – reconstitution/dilution  
inadequate mixing etc
5. Transcription errors
6. Relook at the IQC data
7. Are there trends? High/low bias etc?
8. Change in reagents/calibrants/standardisation?

# ROYAL COLLEGE OF PATHOLOGISTS OF AUSTRALASIA (RCPA)

## METHOD CLASSIFICATION SYSTEM

The method classification system is designed to comprehensively describe your analytical system for each analyte. The method classification system is broken down into four categories that describe analytical principle, instrument, reagent manufacturer and calibrator. This breakdown facilitates a high level of peer review, enabling your data to be shown in relation to your immediate peers using the same analytical systems.

RCPA QAP provides all participants with a list of codes for each of the four categories (if applicable) for every analyte. Participants need to review their analytical system and provide a breakdown of their method using these codes. An example of the method classification process is provided below.

### General Serum Chemistry Program: Performing albumin on a Roche Hitachi Modular using Roche reagents & calibrator

The list of options for each category for this analyte is consulted:

Category 1 Analytical Principle (Program Specific)	Category 2 Measurement System Examples (See Master List-All Programs)	Category 3 Reagent Source Examples (See Master List-All Programs)	Category 4 Calibrator (Program Specific)
<b>Dye Binding</b>	21H Roche Diagnostics Hitachi 917	000 Own Preparation	<b>A</b> Analyser Specific Calibrator supplied by your manufacturer.
<b>A</b> Bromocresol Green Succinate Buffer	21J Roche Diagnostics Hitachi 912	016 Beckman Coulter	<b>Other Calibrator</b>
<b>B</b> Bromocresol Green – Citrate Buffer	21K Roche Diagnostics Hitachi 902/904	019 Biomerieux	<b>B</b> Other Standard not manufacturer supplied.
<b>E</b> Bromocresol Purple	21L Roche Diagnostics Hitachi Modular	020 Bio-Rad	<b>C</b> Bovine commercially assayed
<b>G</b> Bromocresol Green – Unspecified Buffer	21M Roche Diagnostics Cobas c501/c 311/c 111	021 Roche Diagnostics (Hitachi)	<b>D</b> Human commercially assayed
<b>H</b> Bromocresol Green – Malate Buffer		069 Roche Diagnostics (Integra)	<b>E</b> Bovine Material – independently calibrated
		098 Siemens (Bayer)	<b>F</b> Human Material – independently calibrated
		125 Randox	<b>I</b> International Protein Standard calibrated vs. BCR CRM 470



Due Date : 21/04/2008

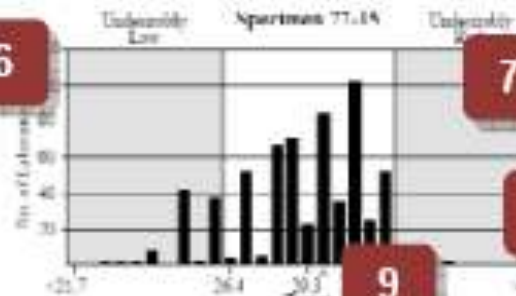
1

Albumin (g/L)

2

Laboratory Number

6



7

**YOUR DATA**

Result (°) for 77-15 = 30.0 g/L  
Result (°) for 77-16 = 40.0 g/L

3

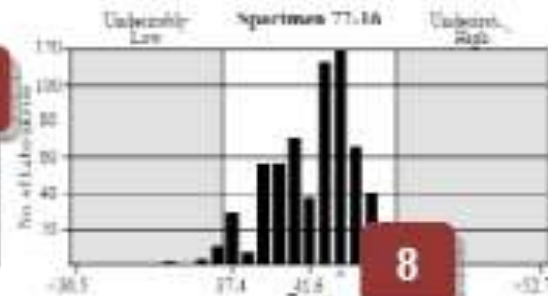
Your Method Classification: B 21L 021 A

B DCG - Citrate Buffer  
21L Roche Diagnostics Hitachi Modular  
021 Roche Diagnostics (Hitachi)  
A Analyser Specific Inhibit Suppl

4

5

Allowable Limits of Performance  
± 0.4 g/L (2007) ± 0.5 g/L



8

All Results (62)

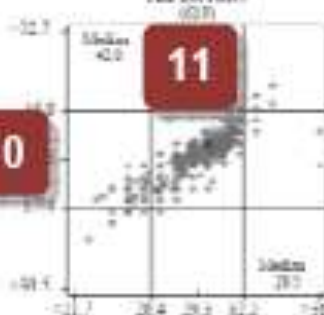
Your Method B 21L 021 (90)

BCC Citrate Buffer (217)

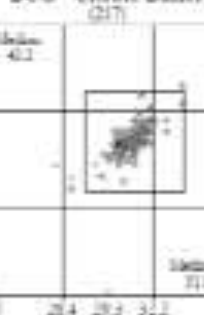
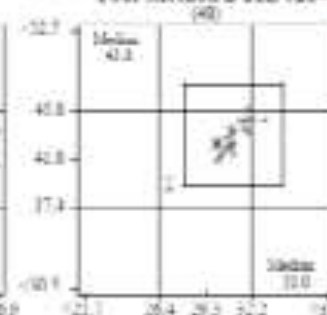
Roche Diagnostics Hitachi Modular (77)

Roche Diagnostics (Hitachi) Reagent (316)

10



11



12



Current Data for Cycle 77

Spec	Method	Target	Result
77-01	B 21L 021 A	33.0	36.0
77-02		25.1	27.0
77-03	B 21L 021 A	37.5	40.0
77-04		21.7	28.0
77-05	B 21L 021 A	33.4	34.0
77-06		31.8	32.0
77-07	B 21L 021 A	40.0	52.0
77-08		28.5	31.0
77-09	B 21L 021 A	35.7	31.0
77-10		37.5	47.0
77-11	B 21L 021 A	40.0	31.0
77-12		35.0	31.0
77-13	B 21L 021 A	35.1	36.0
77-14		39.8	36.0
77-15	B 21L 021 A	30.0	30.0
77-16		47.8	44.0

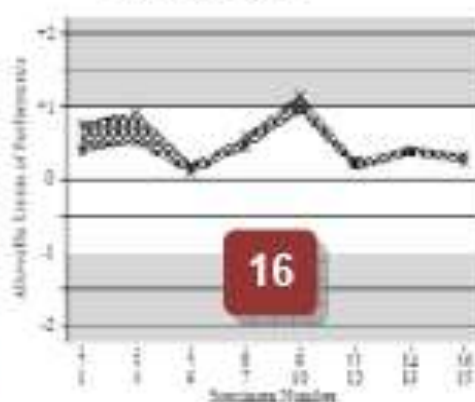
13

14

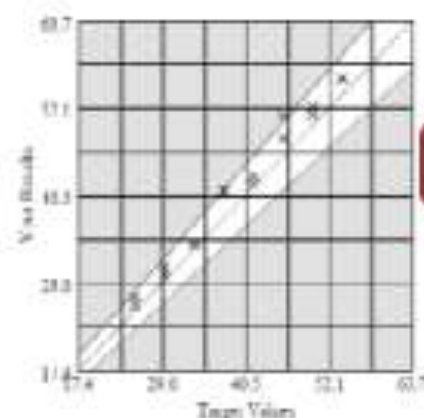
High

15

**SUMMARY DATA**



16



17