
 **UK NEQAS FOR CLINICAL CHEMISTRY**
UNITED KINGDOM NATIONAL EXTERNAL QUALITY ASSESSMENT SCHEMES

Developments in harmonised scoring
systems and data presentation
'The ABC of EQA'

Dr David G Bullock

*Director, Wolfson EQA Laboratory
Organiser, UK NEQAS for Clinical Chemistry*

Finlay MacKenzie, Jonathan Middle and Jane French

 **W H O COLLABORATING CENTRE
FOR RESEARCH & REFERENCE SERVICES IN CLINICAL CHEMISTRY**

Wolfson EQA Laboratory
Queen Elizabeth Medical Centre
Birmingham B15 2UE
U K

Scheme Design, Report Formats and the ABC of EQA

- All well designed EQA Schemes are trying to do the same thing and the introduction of the ABC of EQA scoring system simply reinforces this
- This push towards harmonisation is for **your** benefit!
- We want to present the data in best way possible and in the same way for all UK NEQAS schemes
- This will allow you to assimilate data simply and will ease comprehension
- Pre-digestion of data and allocation of 'scores' **does** work and it **does** encourage improvement

Scheme Design, Report Formats and the ABC of EQA

- The 'ABC of EQA' is more than just the A, B and C statistics themselves.
- The 'ABC of EQA' is a package that comprises scheme design, statistical data processing, graphical and tabulated data presentation.

Scheme design

What do you want from UK NEQAS Birmingham?

- Good scheme design
- Good quality specimens
- Fast turn-round of high quality personalised reports
- Scoring systems that inform and allow focused remedial action

At the analyte/specimen level

- We send you a specimen which you analyse
- You tell us what result you got
- We tell you what the target value for that specimen was
- We tell you what percentage out, either positive or negative, your result was compared to the target
- We tell you for that concentration of analyte and for that percentage your result was out, how this compares with what a typical average laboratory should get

At the rolling time window level

- We summarise all your results for the last 6 months
- We tell you, on average, how biased you are, and if this bias is consistent
- We tell you, on average, how your overall performance compares to the current state-of-the-art
- We tell you, on average, how your overall performance compares to the average performance of other analytes

Method comparisons and trend data

- We let you see how other labs using your method are performing and also how other methods are performing
- The graphs we provide you with will show you trend data on how you are performing over time, letting you see if you are getting better or worse

Basic Scheme Statistics and Nomenclature

There are:

- Specimen level statistics
- Rolling time window level statistics

- Laboratory specific
- Method specific
- Overall [all labs, all methods]

Basic Scheme Statistics and Nomenclature

The **analyte** specimen level statistics are many and varied and include:-

- Target values
- All Laboratory Trimmed Mean **ALTM**
(*the overall consensus mean*)
- Grouped Laboratory Trimmed Mean **GLTM**
- Method Laboratory Trimmed Mean **MLTM**
- Medians *etc*

Basic Scheme Statistics and Nomenclature

The **laboratory analyte** specimen-level statistics for each are:-

- [Your result]
- Specimen %bias
- Specimen transformed bias
- Specimen Accuracy Index

Basic Scheme Statistics and Nomenclature

The **laboratory-analyte** rolling time window statistics are:-

- A score
- B score
- C score

The 'ABC of EQA' scoring system

The A, B and C Scores

There are three **scores**

(by convention, we are defining here that 'scores' are calculated over a rolling time window, combining information derived from many specimens)

A is for **Accuracy** (total error)

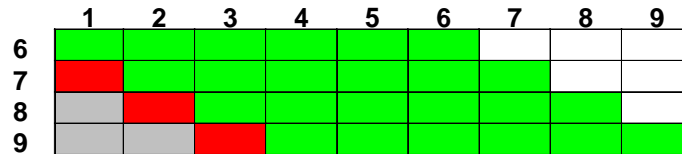
B is for **Bias**

C is for **Consistency of bias**

Every laboratory will have an A, B and C score for each analyte they measure

The Basics of the ABC Scoring system

- Each of these 3 scores is calculated over a rolling time window
- Each 'rolling time window score' comprises data (results) from many specimens
 - *they are always being updated with fresh current data, and at the same time historical data drops out of the 'time window':*



The Basics of the ABC Scoring system

- The time window we have employed has been set at 6 distributions (equivalent to 6 months) for 'standard schemes'
- In order to obtain sufficient data from less frequently assessed assays, a period of 12 months may be required

[Though the details are dependent on individual schemes, the principle holds true for any UK NEQAS (Birmingham) scheme]

The full definitions are as follows

- All of our individual 'specimen statistics' and their (running) average counterparts are based on the single simple and intuitive % **bias calculation** shown below.

$$\text{Specimen \% bias} = \frac{(\text{result-target})}{\text{target}} * 100\%$$

The Basics of the ABC Scoring system

- The time window we have employed has been set at 6 distributions (equivalent to 6 months) for 'standard schemes'
- In order to obtain sufficient data from less frequently assessed assays, a period of 12 months may be required

[Though the details are dependent on individual schemes, the principle holds true for any UK NEQAS Birmingham scheme]

THE A SCORE

- **The Accuracy A score has been transformed**
 - transformation uses a 'degree of difficulty' factor
 - A scores are broadly comparable across analytes
 - *for example:*
 - A score of 85 for TSH
 - A score of 85 for sodium
 - performing both analytes equally well, on average
- **The A score is a 'screen', normalised to the state of the art**

THE B SCORE

- **The Bias B score has not been transformed**
 - B scores are expressed as percentages
 - bias and consistency of bias relate to actual bias
 - *for example:*
 - B score of 5% for TSH
 - performance is quite acceptable
 - B score of 5% for sodium
 - urgent need to review the assay!
- **The B score assesses your bias relative to the target value**

THE C SCORE

- **The Consistency of bias C score has also not been transformed**
 - it complements the B score (*cf* BIAS and VAR)
 - do you usually have the same bias:
 - are results imprecise?
 - does bias differ according to concentration?
 - does bias depend on specimen matrix?
 - has the bias changed during the rolling time window?
- ***Poor consistency of bias is not the same as imprecision***

CALCULATION - RAW SCORES

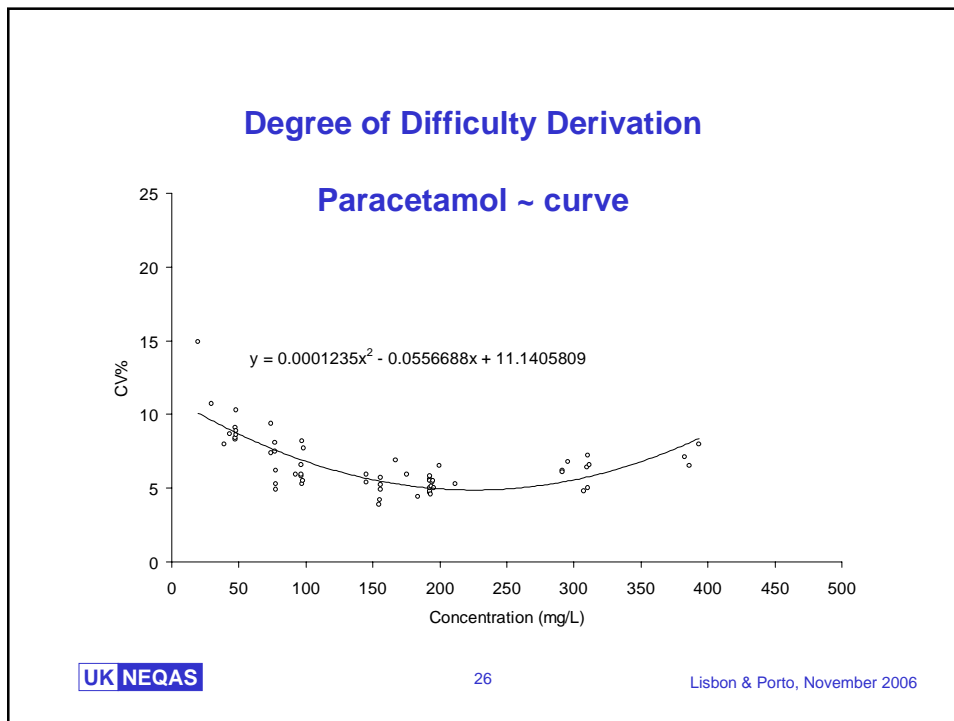
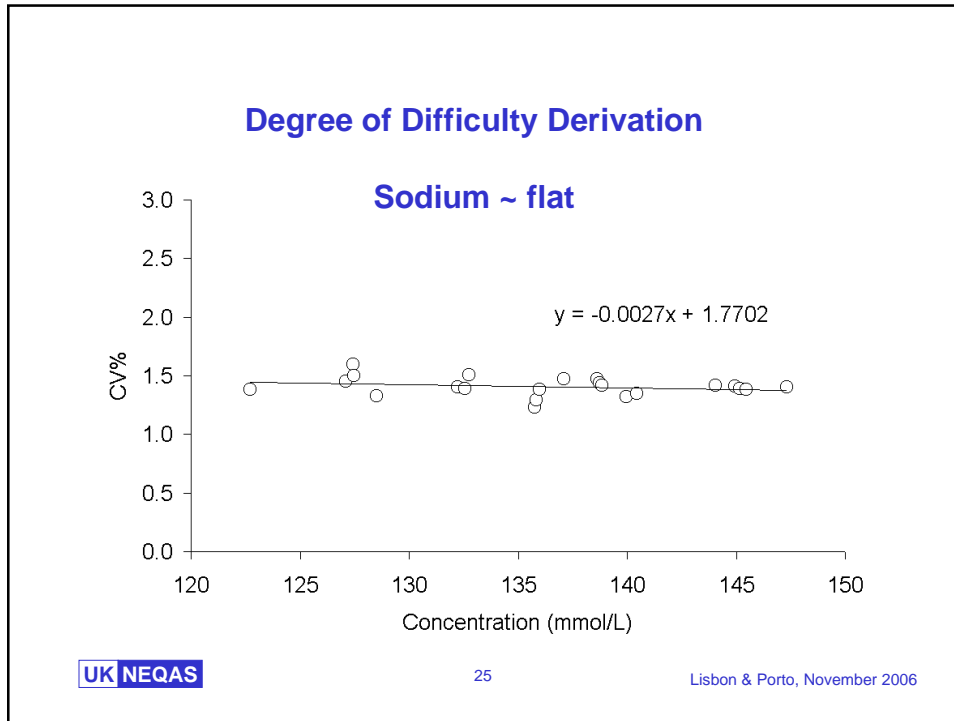
- **These use all the % bias figures in the rolling time window:**
- **the B score is the trimmed mean of the % biases**
- **the C score is the trimmed SD of this calculation**
- ***the A score could be the trimmed mean ignoring sign***

CALCULATION - TRANSFORMED SCORES - 1

- **These use a transformation of the % bias:**
 - a 'degree of difficulty' factor is used
 - factors can be established in several ways
 - a fixed factor (*cf* the CCV in Variance Index system)
 - concentration-dependent factors
 - based on modelling
 - based on biological variation
 - based on clinical needs ('fit for purpose')
- **'Fit for purpose' preferable, but difficult to define**
- **We therefore favour the modelling approach**

CALCULATION - TRANSFORMED SCORES - 2

- **We use the modelling approach**
- **This is a two stage process:**
 - elimination of concentration-dependence:
 - examine the SD v ALTM relationship during 1998 & 1999
 - derive an equation for concentration-dependent factors
 - normalisation to the state of the art:
 - adjust the factors, yielding
 - **a median A score of 100 at January 2000**



We calculate the A Score as follows :

- we take each **Specimen % bias** in the time window and transform it by the 'degree of difficulty' factor to get a **Specimen transformed bias** [this can be positive or negative]
- we take the modulus of each 'Specimen transformed bias'; each of these is called a **Specimen Accuracy Index** [as it is a modulus it has no sign]
- finally, we calculate the '**A score**' as the trimmed mean of all of these **Specimen Accuracy Indices**

Scores: What is good and what is bad?

- For all UK NEQAS Birmingham Schemes, all our scores have been set such that a **low** score is '**good**' and a **high** score is '**bad**'.


Report formats

REPORT STRUCTURING

- **Participants differ in their needs**
 - some want minimal data
 - some want lots of data
- **Different staff have different needs**
 - laboratory Directors want summary data
 - analysts want detailed information
- **Laboratories' needs change**
 - limited data when all is OK
 - detailed examination for problems

REPORT STRUCTURING ADDRESSES THESE NEEDS

Distribution Summary Page



UK NEQAS for Lead & Cadmium in Blood		Laboratory
Distribution : 573 Date : 1-Nov-2000		Page 1 of 5
Distribution Summary		

This new scoring system (the ABC of EQA) was introduced at the last distribution.


In summary:
 - A is for Accuracy (total error), and the A score replaces MBVS but is transformed in a concentration-dependent manner and normalised to a median score of 100 at January 2005; the previous acceptable limits are now 180 for Lead and 250 for Cadmium.
 - B is for Bias and C is for Consistency of Bias, and the B and C scores replace MBDS and MBSS but are expressed as plus percentages for they are not transformed.
 At all times, the inclusion of 'unassigned results' does not affect the target values and, hence, the A, B and C scores of previous distributions.

	Specimen	Part	Result	Target	Specimen Bias	Specimen transformed bias	Specimen Accuracy Index	A score	B score	C score
Lead (total/L)	573A	117	2.17	2.362	-8.1	-229	278	48	-9.1	1.7
	573B	119	0.47	0.512	-8.1	-121	122			
Cadmium (total/L)	573A	117	31.4	33.8	-5.0	-41	61	95	-1.1	0.5
	573B	119	20.0	26.4	-26.6	-211	111			

31

Lisbon & Porto, November 2006

Distribution Summary Page Detail



UK NEQAS for Lead & Cadmium in Blood		Laboratory
Distribution : 573 Date : 1-Nov-2000		Page 1 of 5
Distribution Summary		

This new scoring system (the ABC of EQA) was introduced at the last distribution.

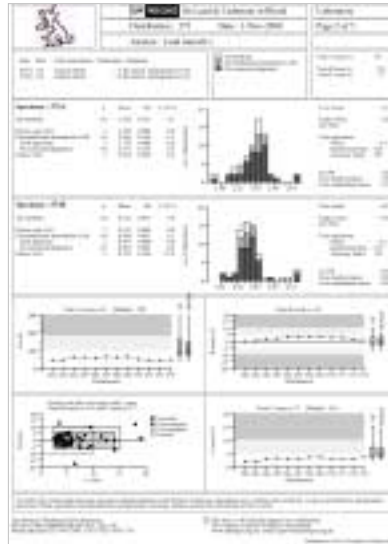
In summary:
 - A is for Accuracy (total error), and the A score replaces MBVS but is transformed in a concentration-dependent manner and normalised to a median score of 100 at January 2005; the previous acceptable limits are now 180 for Lead and 250 for Cadmium.
 - B is for Bias and C is for Consistency of Bias, and the B and C scores replace MBDS and MBSS but are expressed as plus percentages for they are not transformed.
 At all times, the inclusion of 'unassigned results' does not affect the target values and, hence, the A, B and C scores of previous distributions.

	Specimen	Part	Result	Target	Specimen Bias	Specimen transformed bias	Specimen Accuracy Index	A score	B score	C score
Lead (total/L)	573A	117	2.17	2.362	-8.1	-229	278	48	-9.1	1.7
	573B	119	0.47	0.512	-8.1	-121	122			
Cadmium (total/L)	573A	117	31.4	33.8	-5.0	-41	61	95	-1.1	0.5
	573B	119	20.0	26.4	-26.6	-211	111			

32

Lisbon & Porto, November 2006

Analyte Page 1

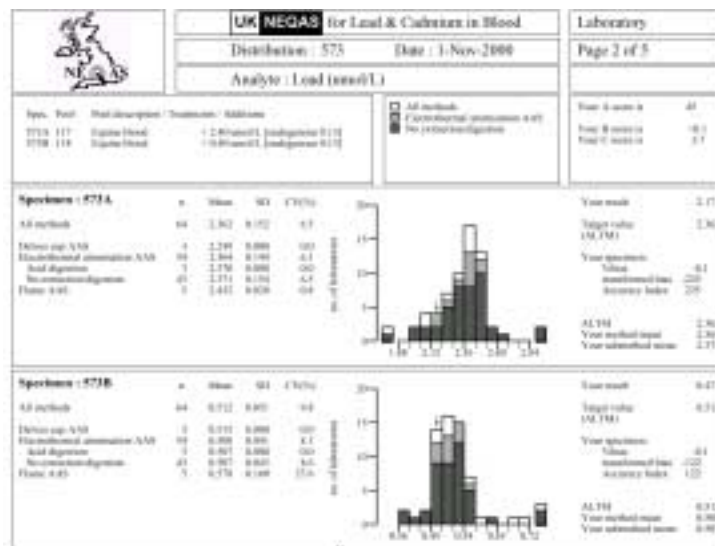


UK NEQAS

33

Lisbon & Porto, November 2006

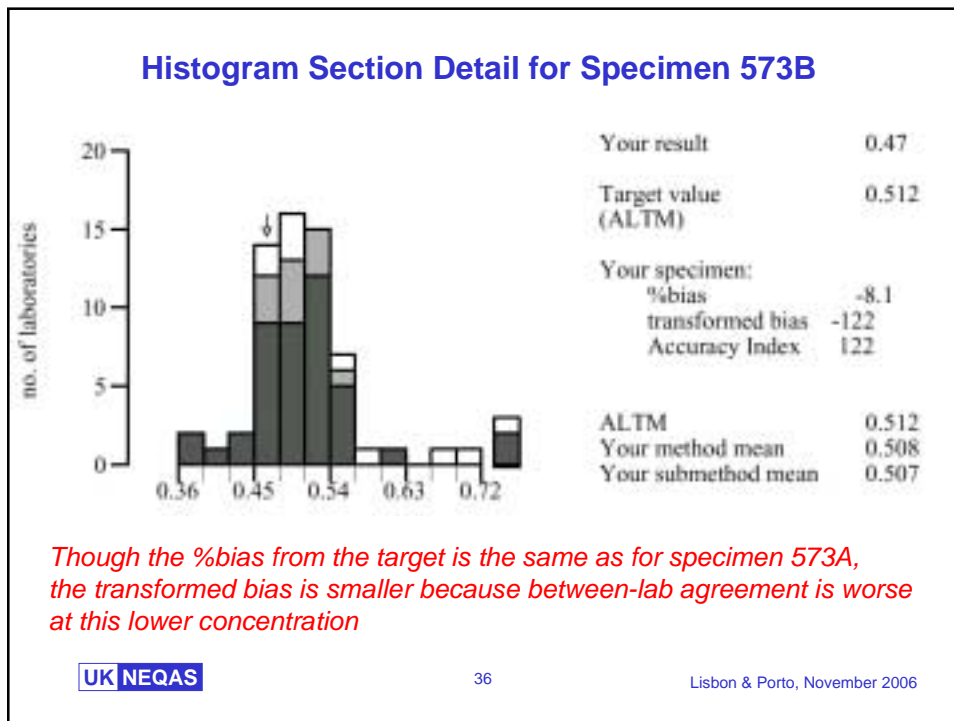
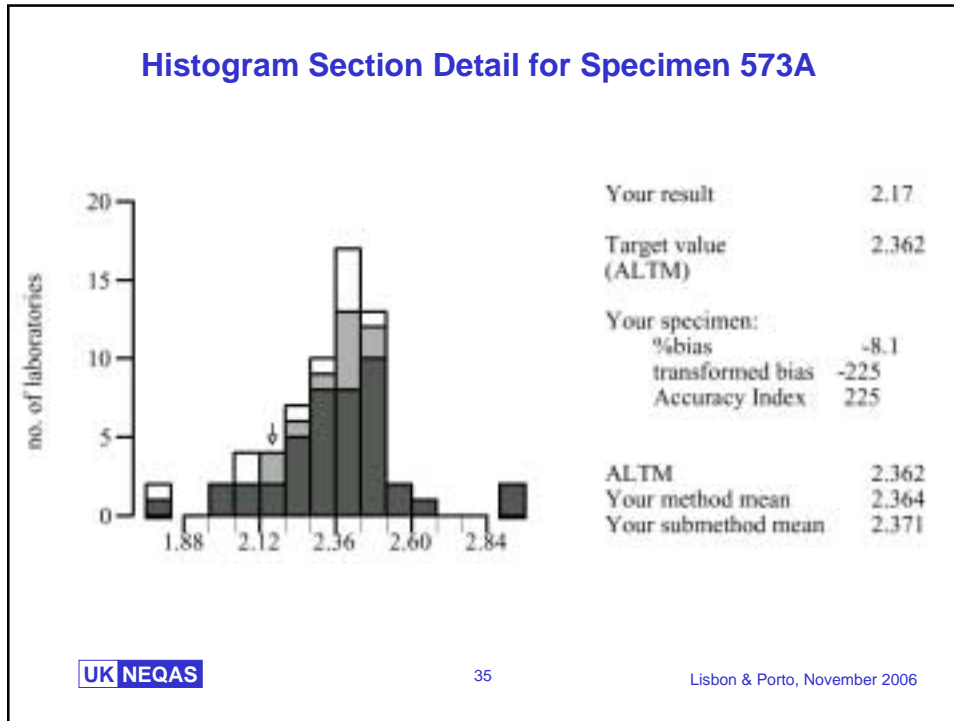
Analyte Page 1 Detail

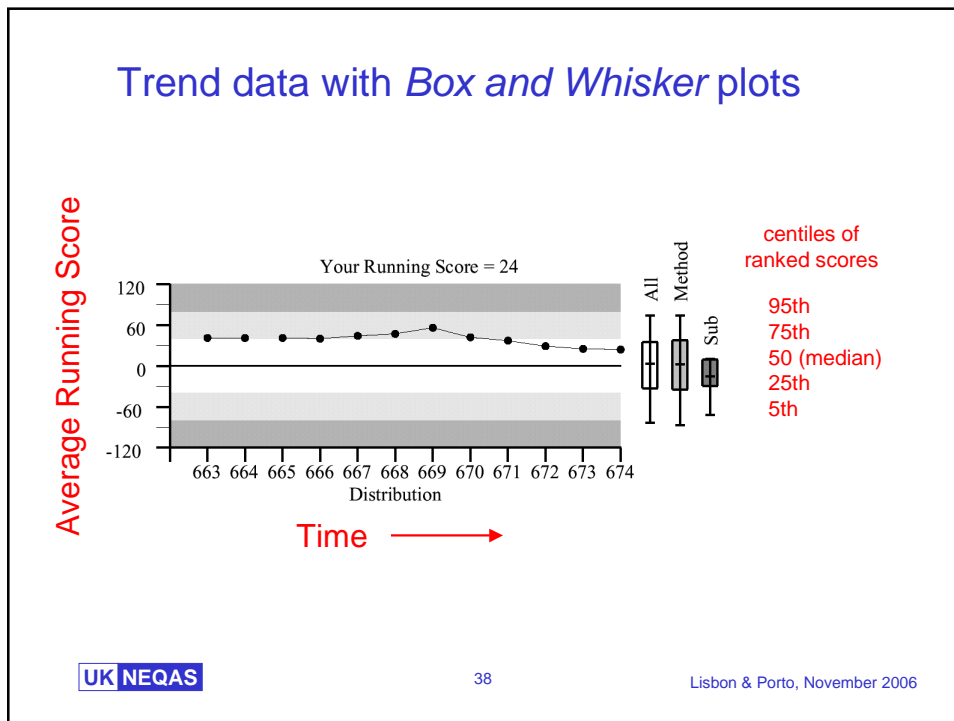
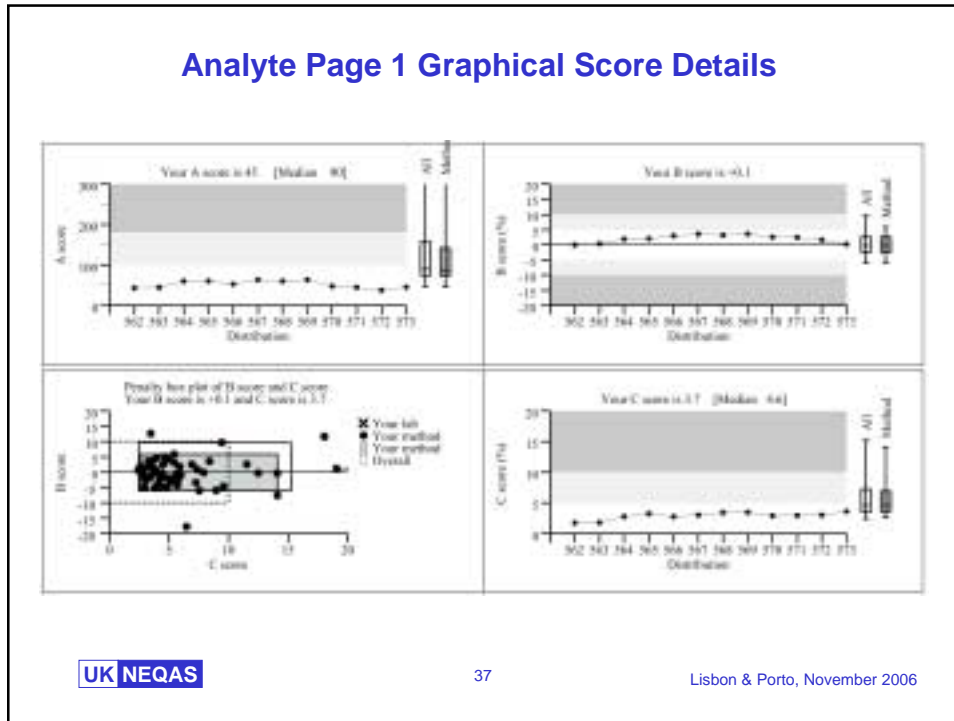


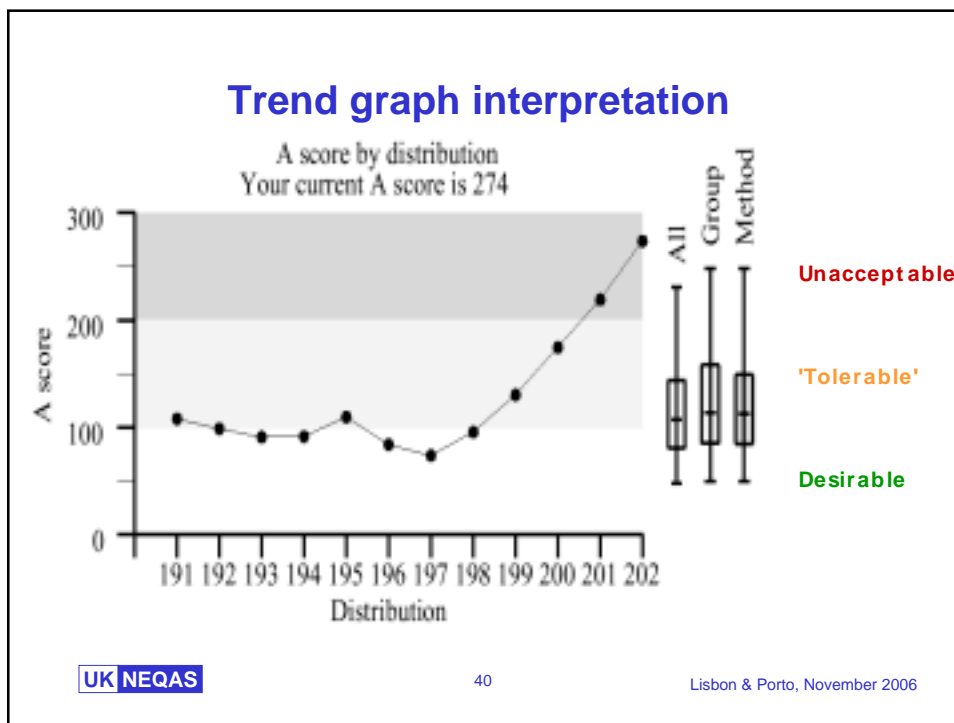
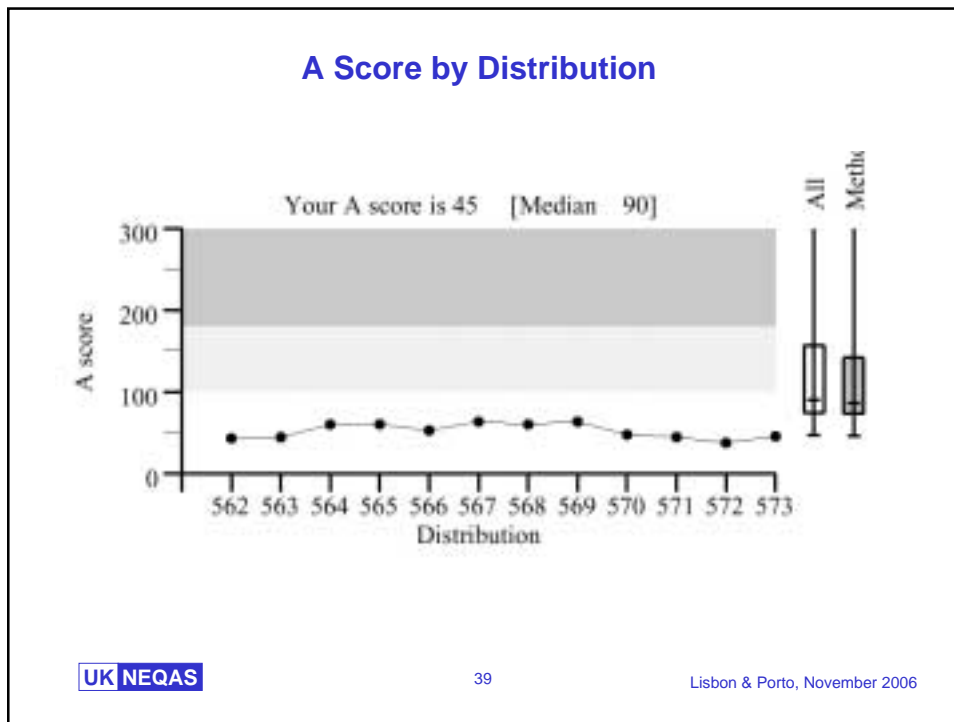
UK NEQAS

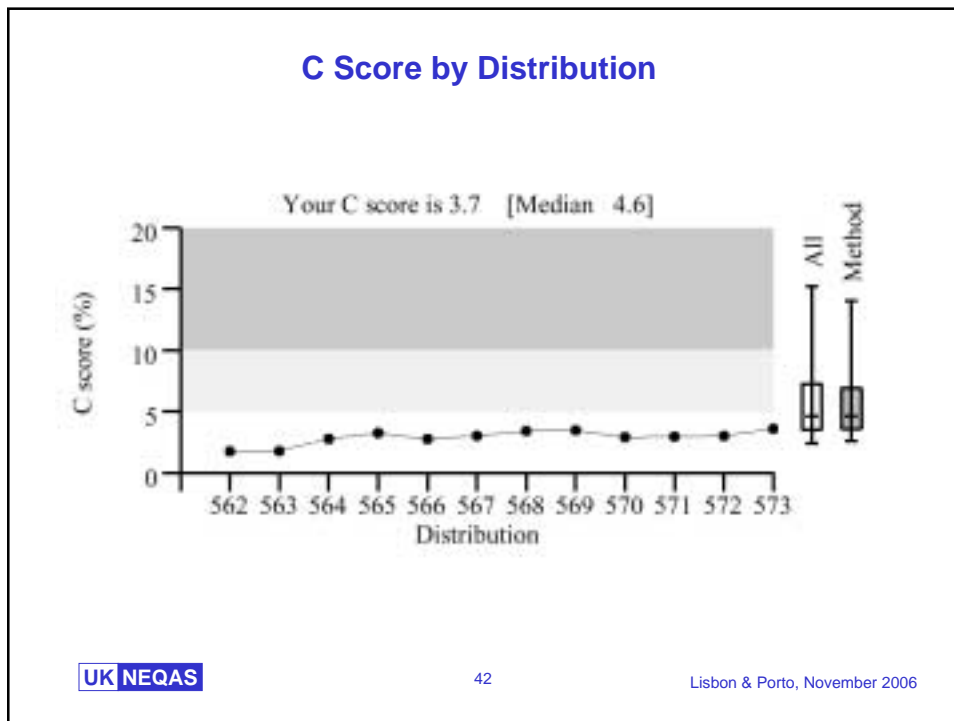
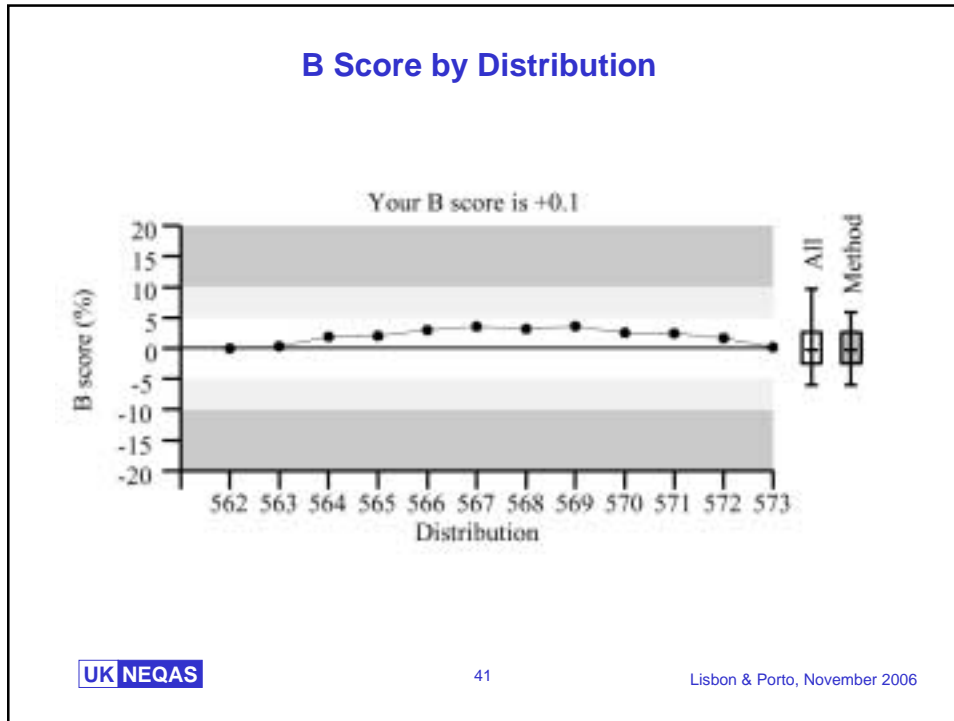
34

Lisbon & Porto, November 2006

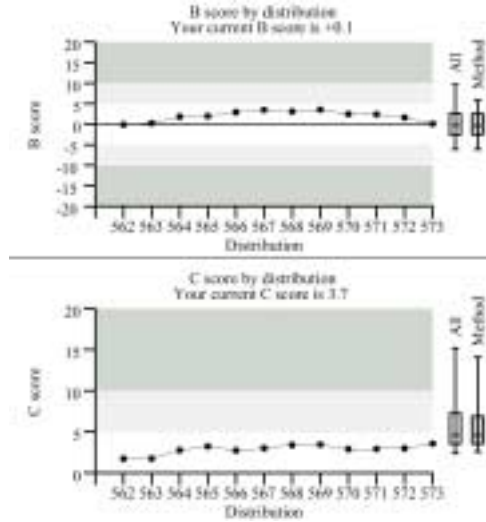








Detail From Report / Snapshot Page

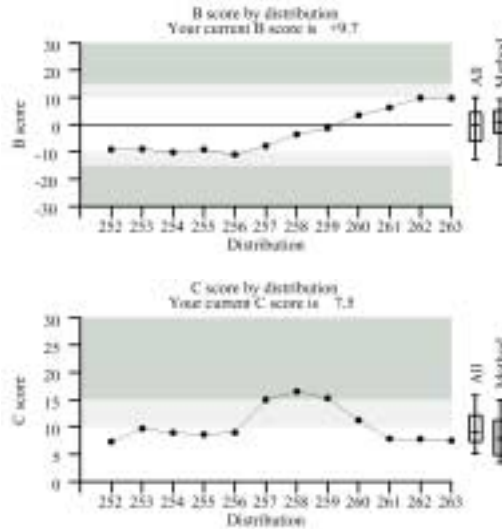


UK NEQAS

43

Lisbon & Porto, November 2006

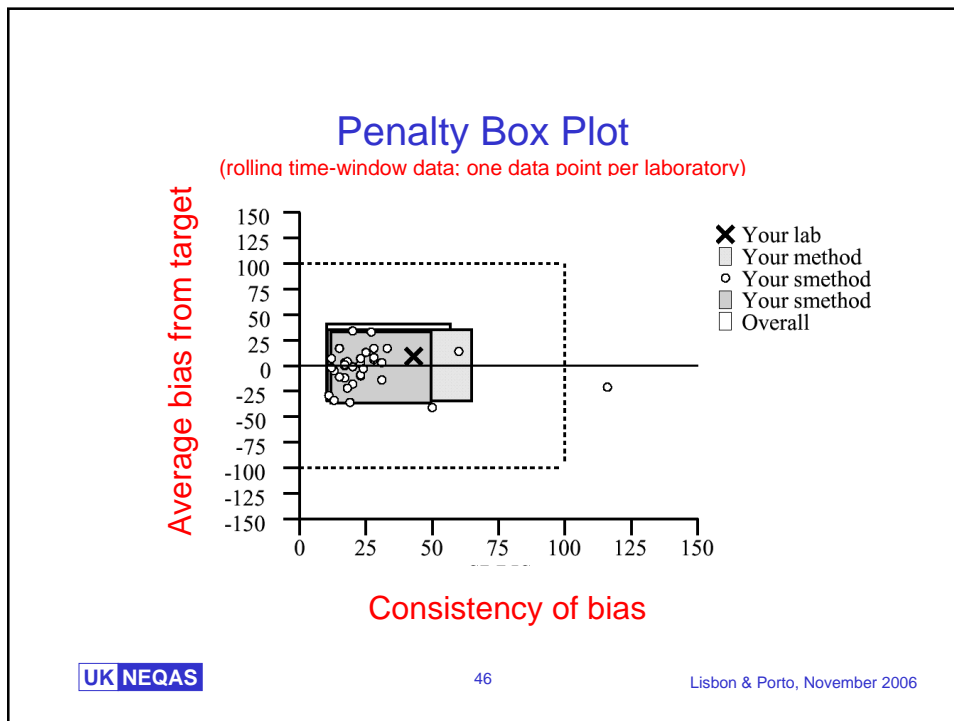
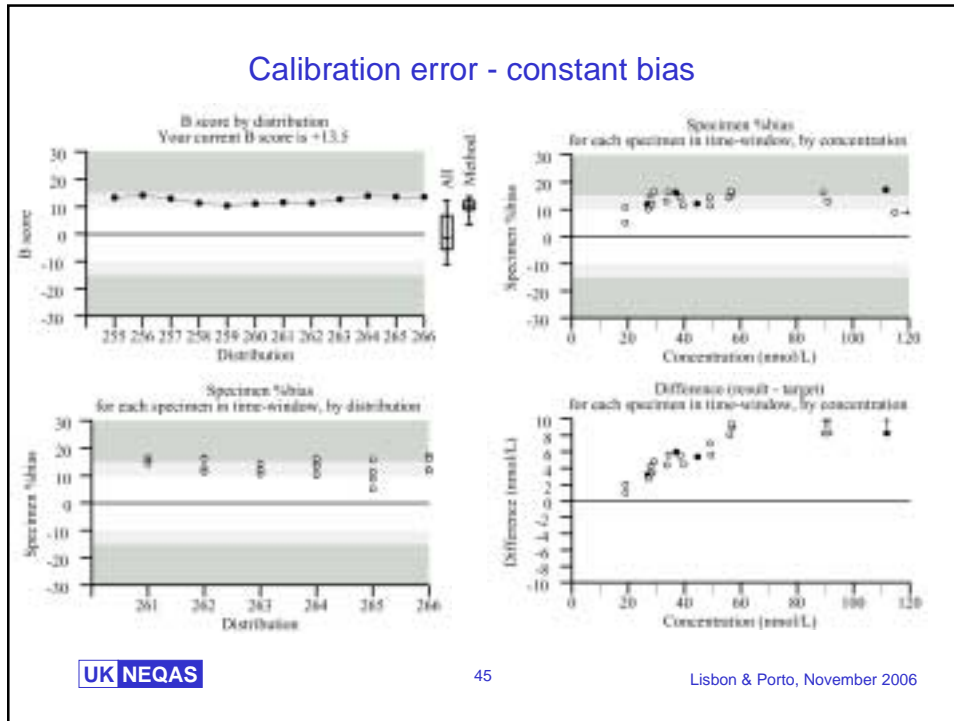
Change in calibration

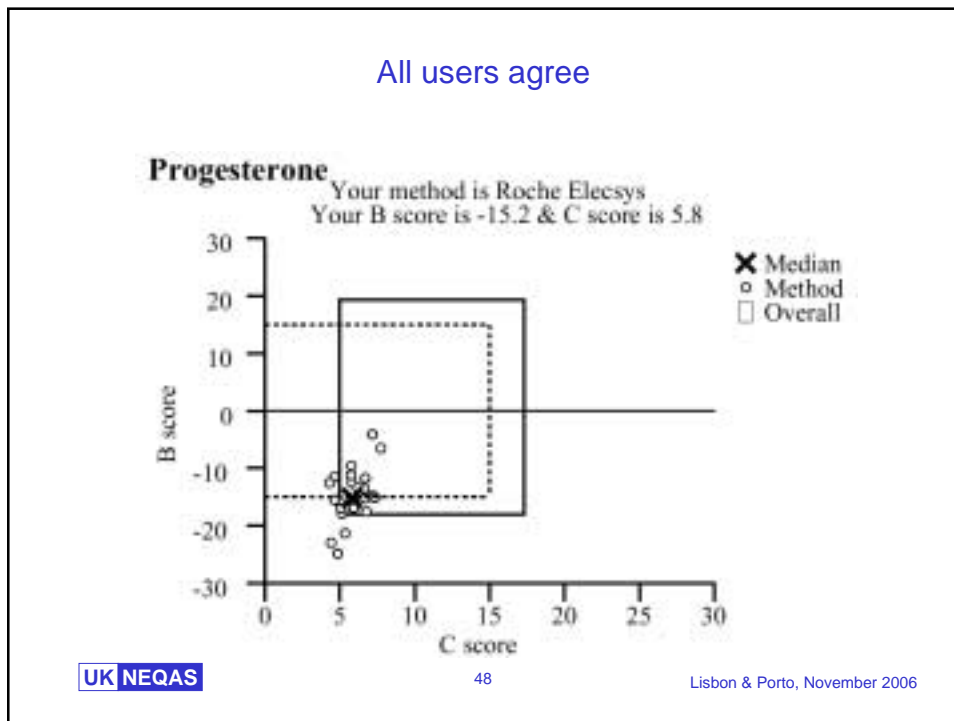
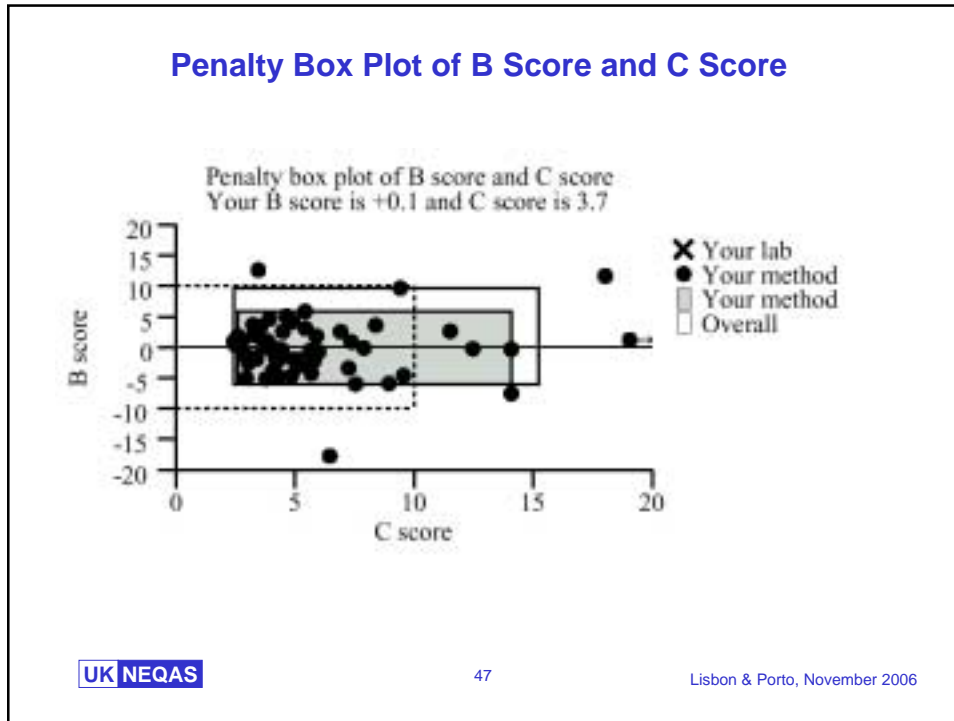


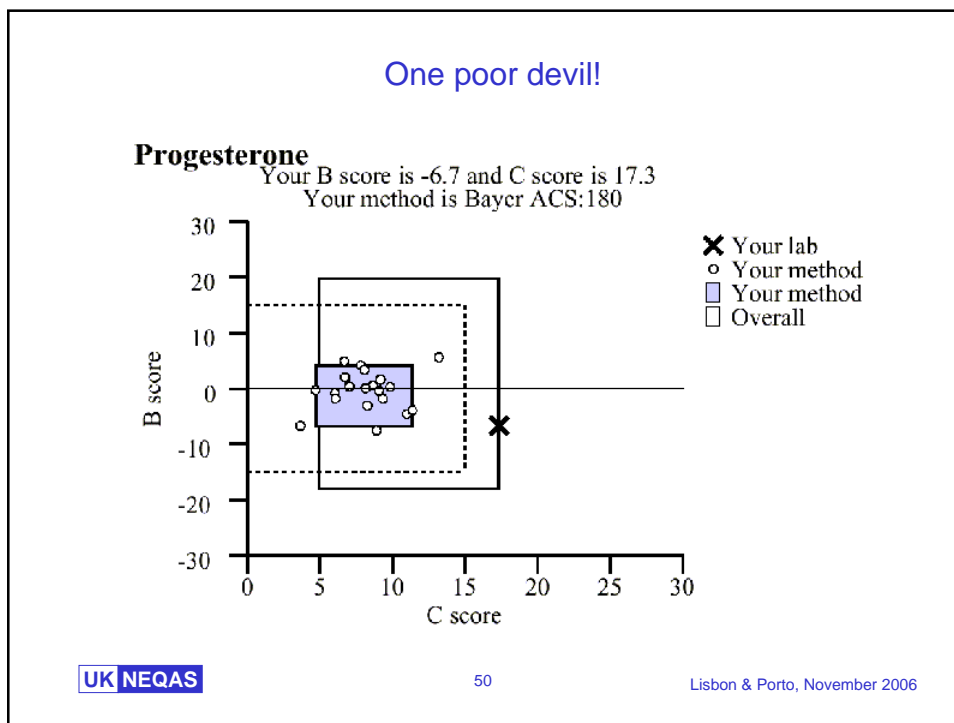
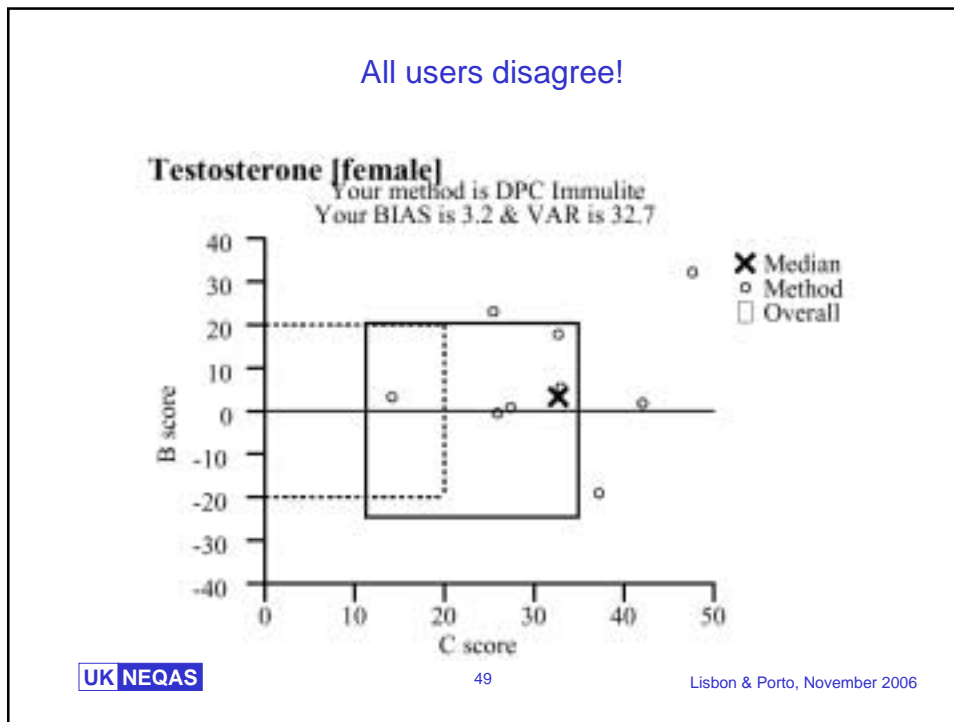
UK NEQAS

44

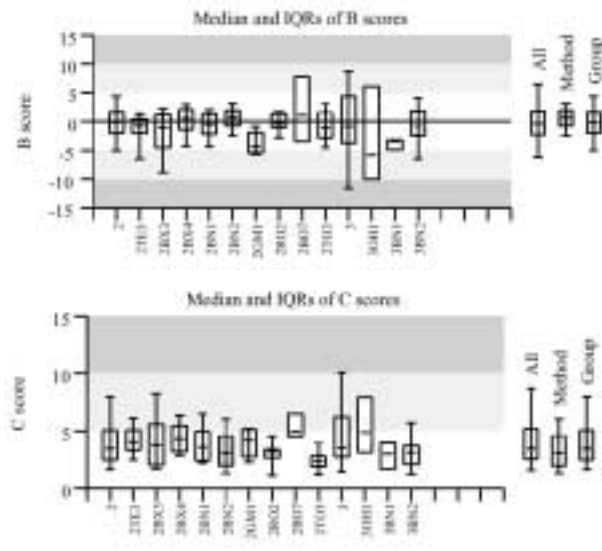
Lisbon & Porto, November 2006



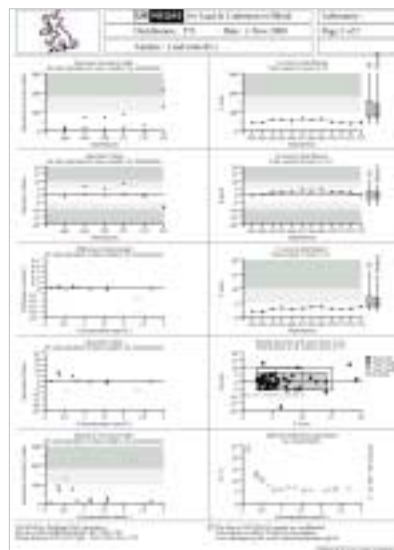




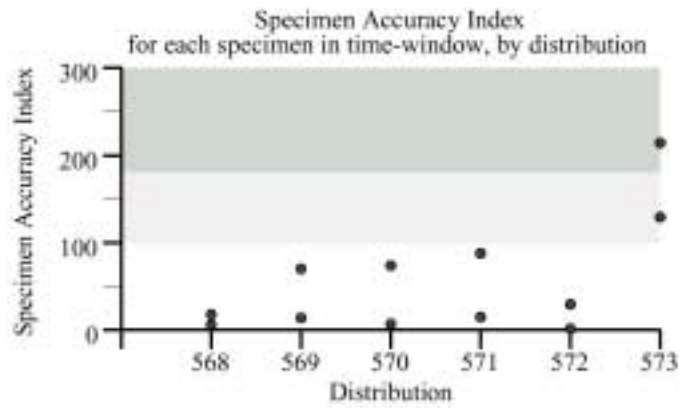
Method Box and Whisker Plots of B and C scores



'Snapshot' Page



Specimen Accuracy Index for Each Specimen in Time-window, by Distribution

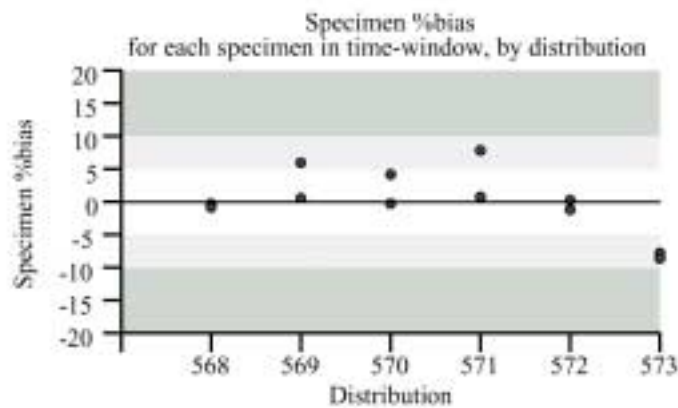


UK NEQAS

53

Lisbon & Porto, November 2006

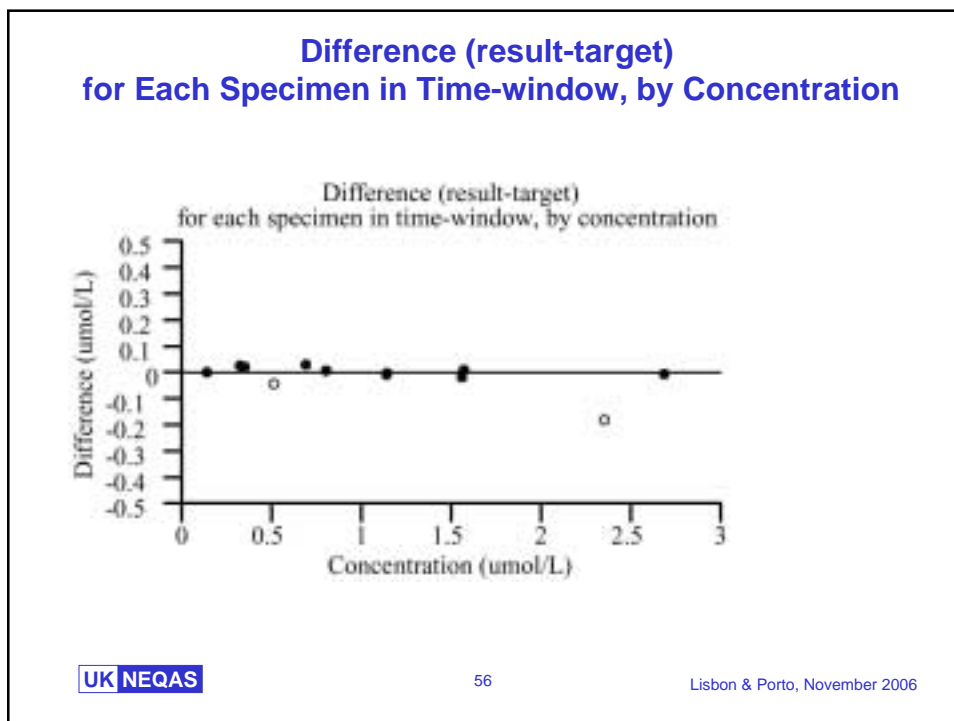
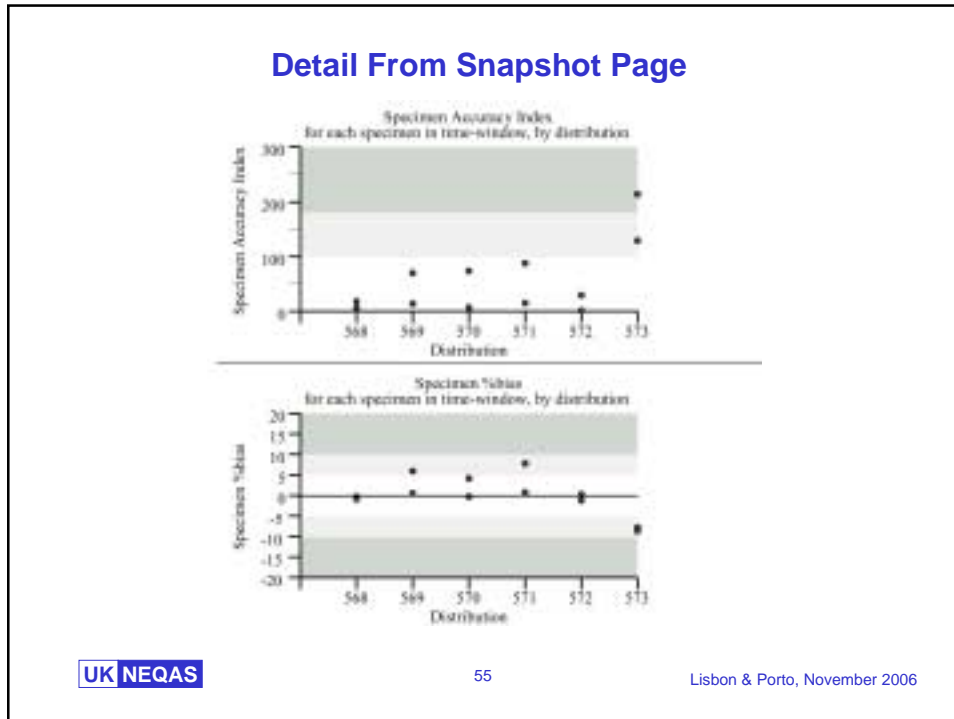
Specimen %bias for Each Specimen in Time-window, by Distribution

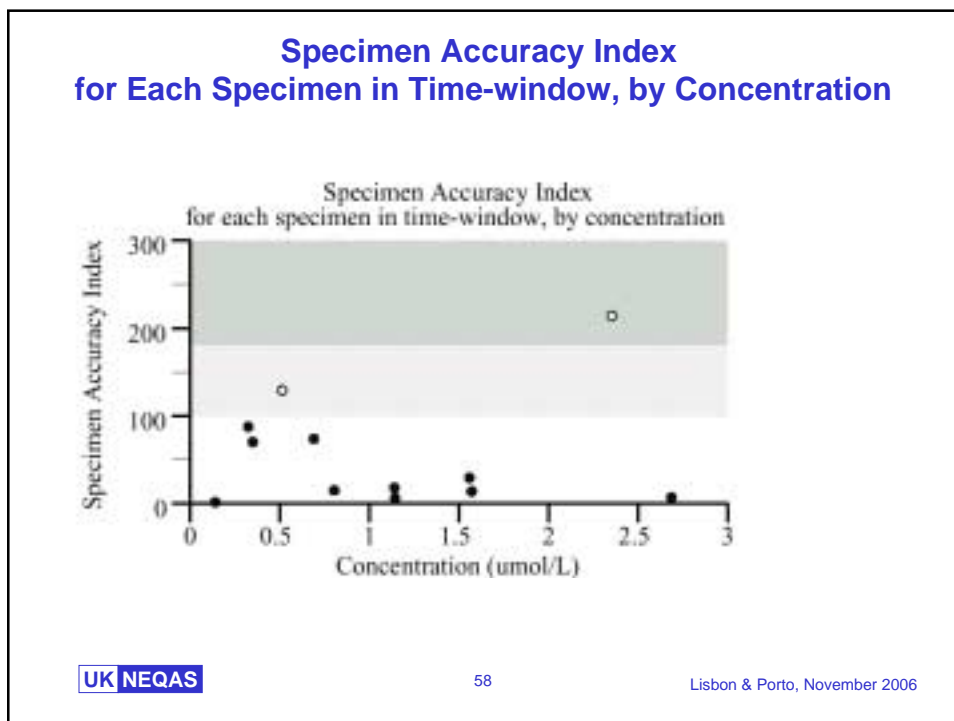
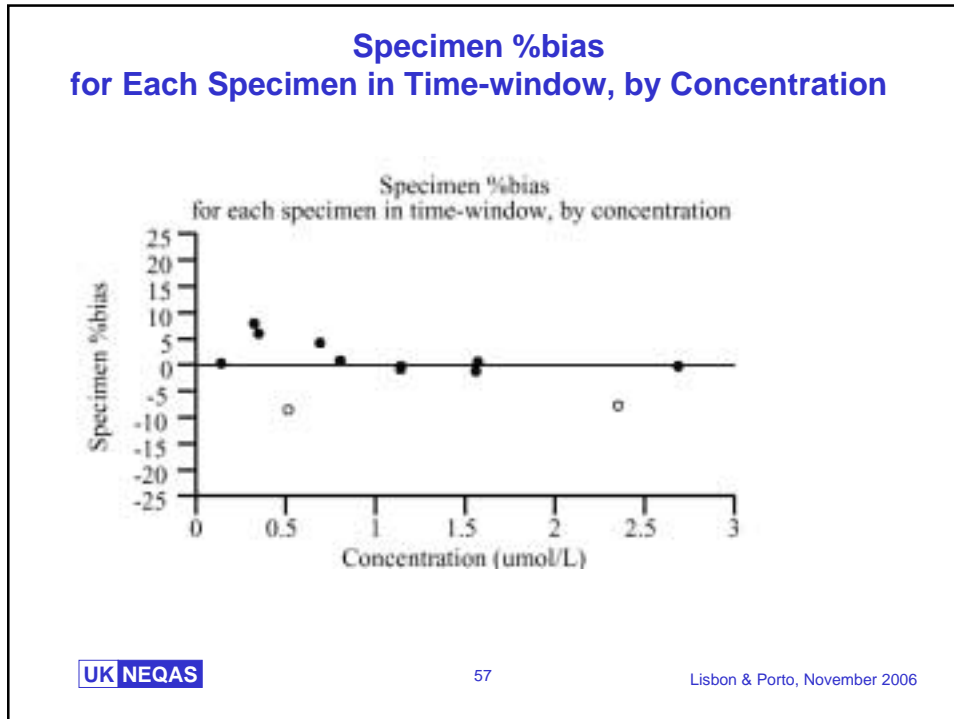


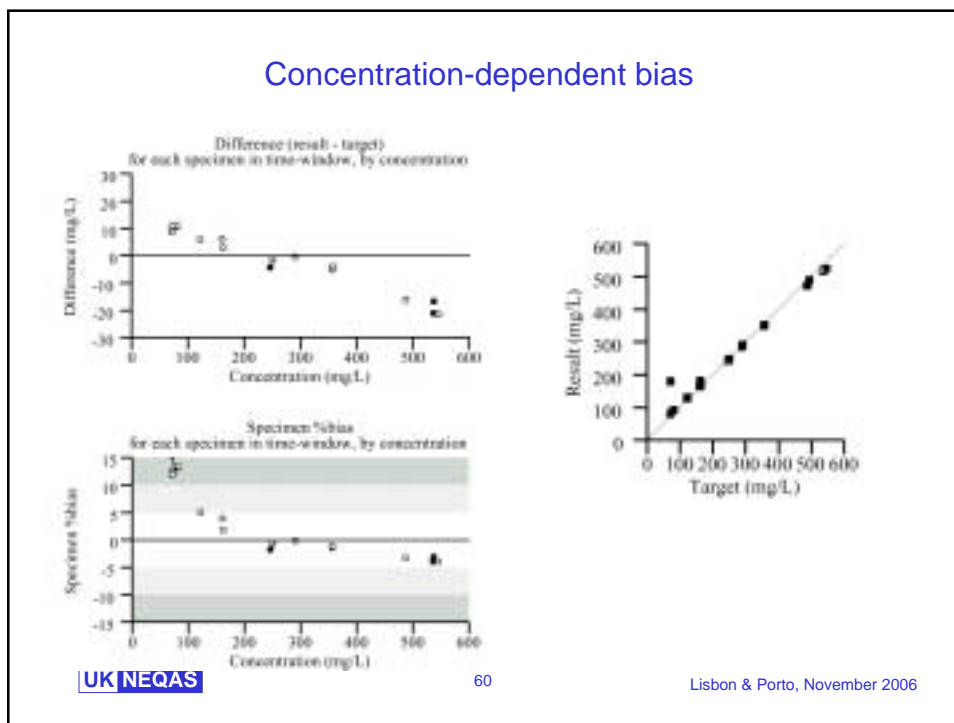
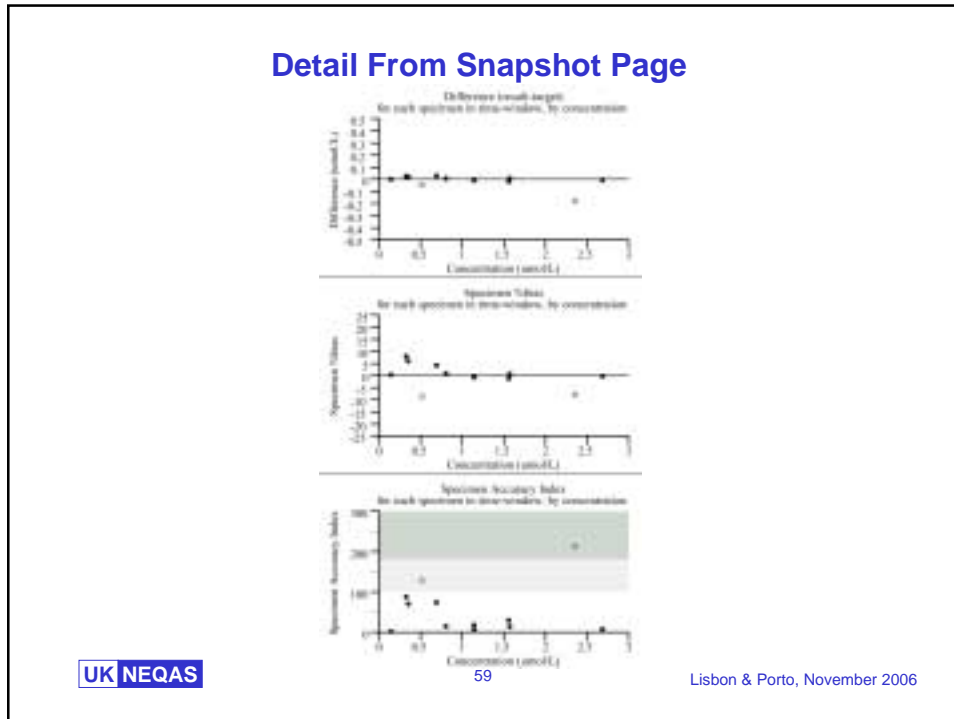
UK NEQAS

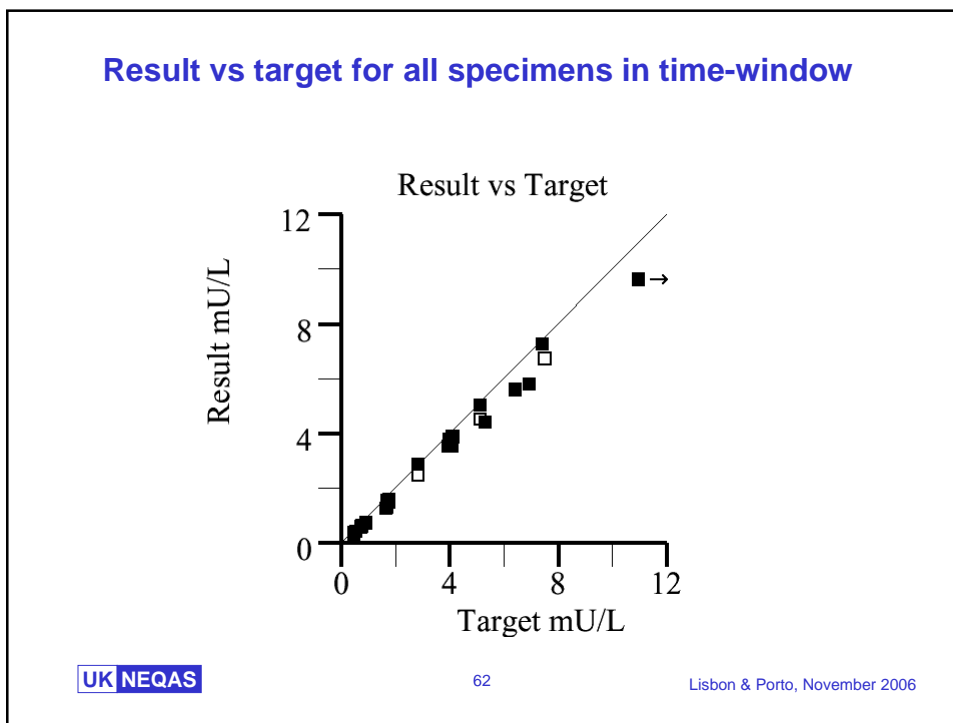
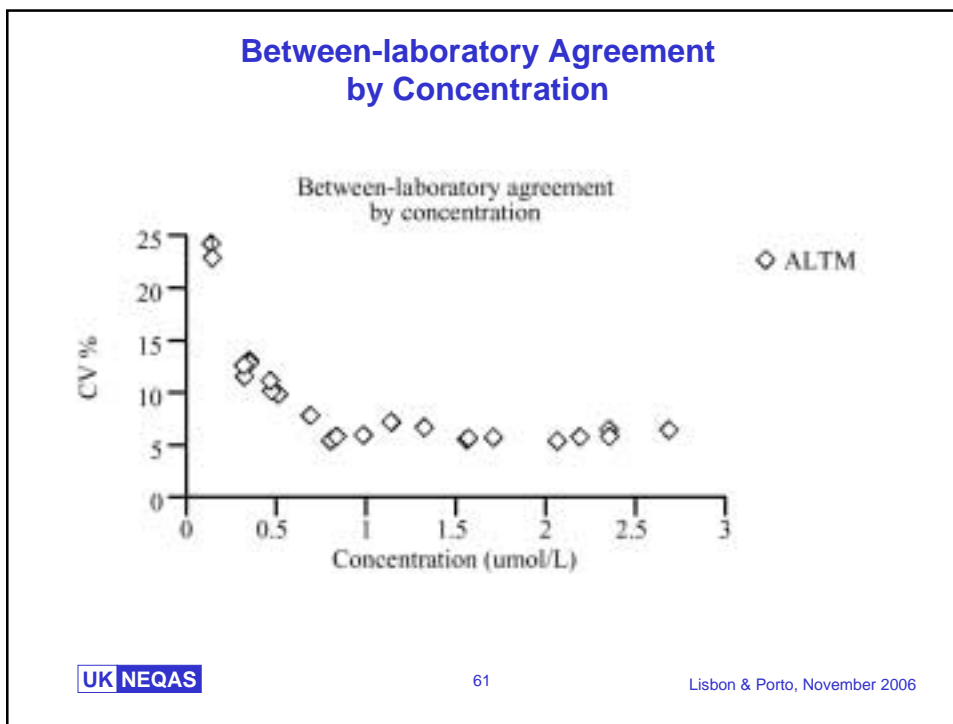
54

Lisbon & Porto, November 2006









INTERPRETATION - 2

A score	B score	C score	
Small	Small	Small	Satisfactory
Large	Large	Small	Proportional bias
Large	Small	Large	Variability*
Large	Large	Large	Bias + variability*

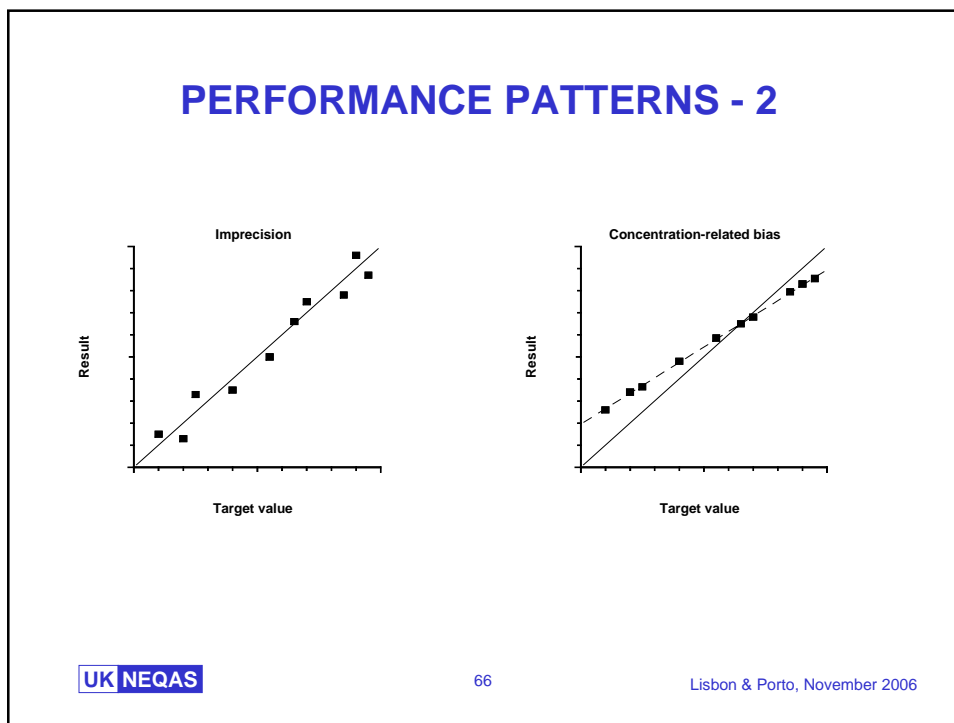
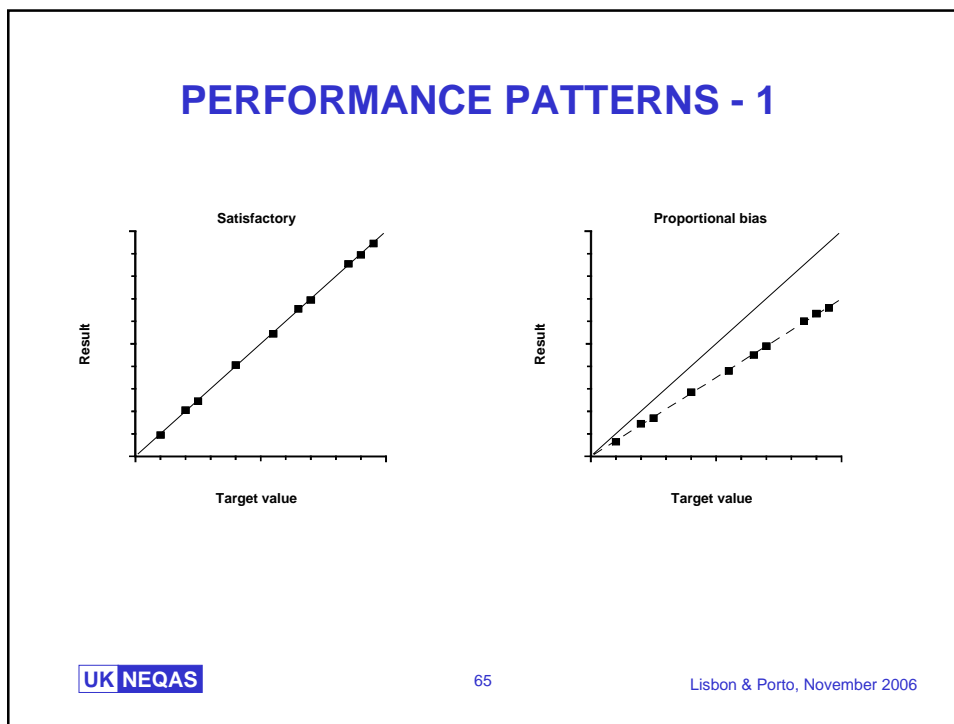
* variability is not the same as imprecision

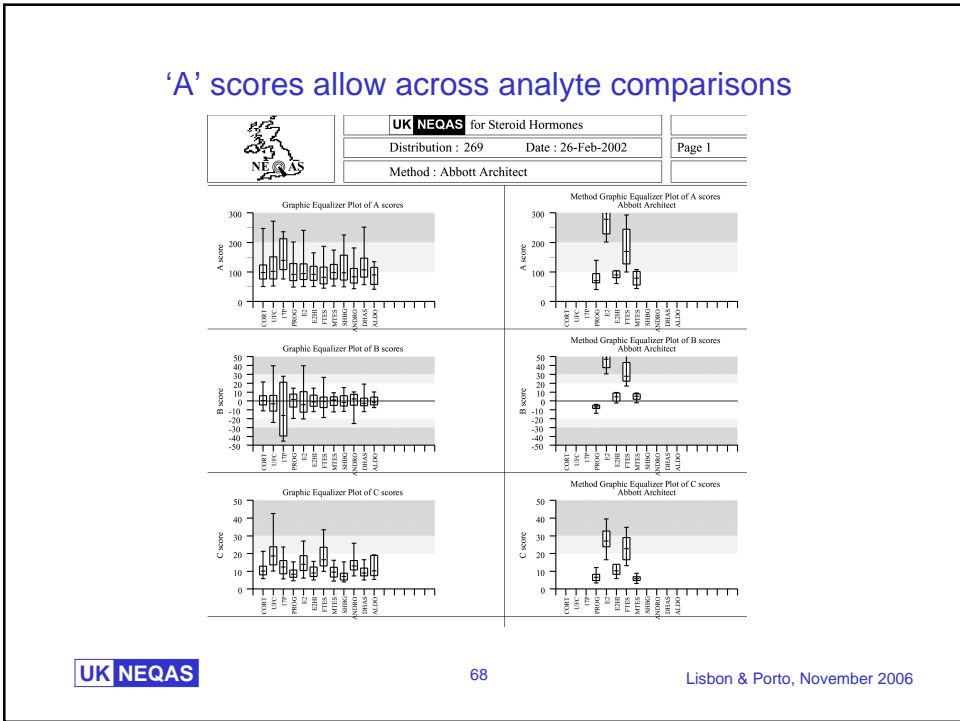
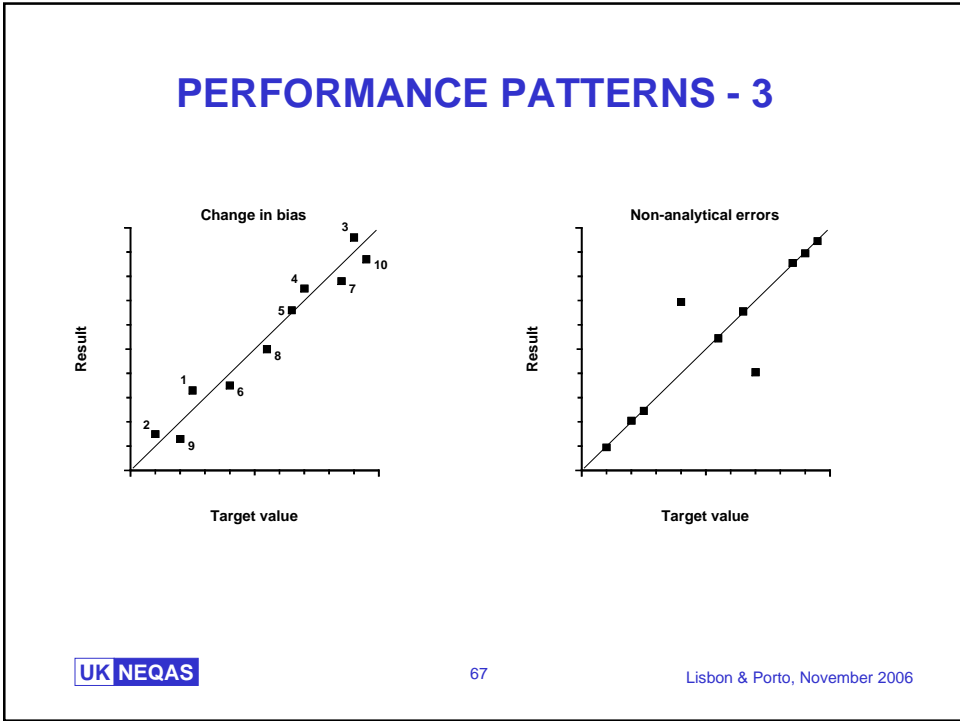
INTERPRETATION - 3

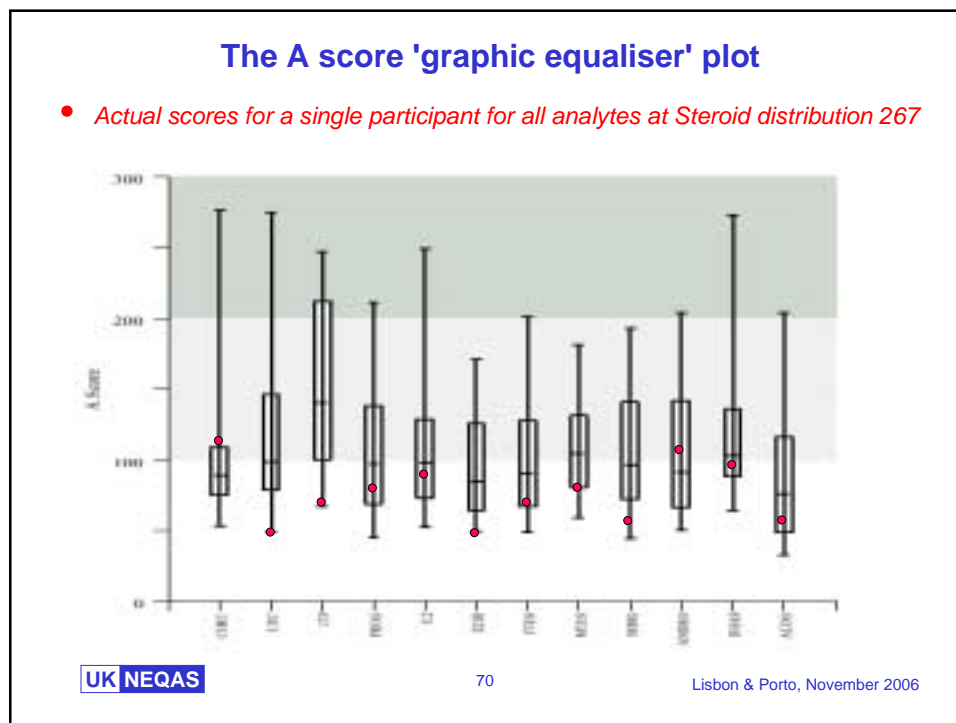
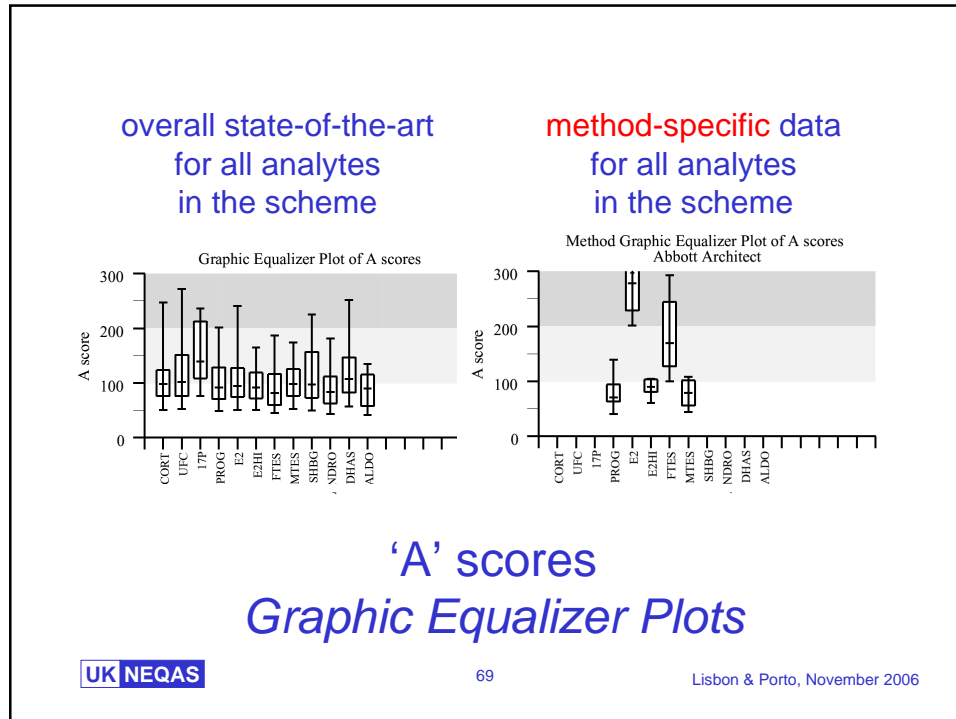
Sources of variability (lack of consistency):

- imprecision
 - *[short term changes in bias]*
- concentration-related bias
 - *non-linearity*
- bias changes with time
- non-analytical errors

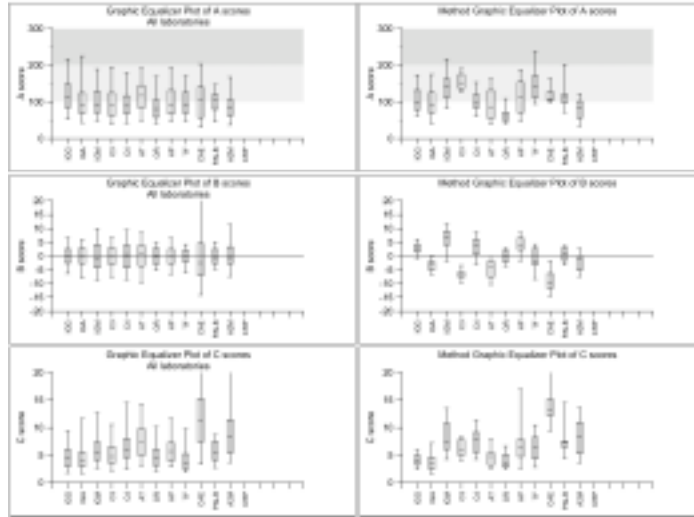
VARIABILITY IS NOT THE SAME AS IMPRECISION







Dade Behring (nephelometry) - March 2006

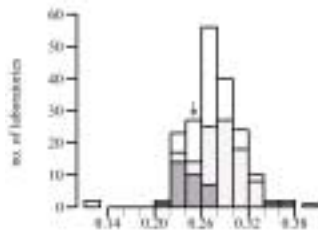


UK NEQAS

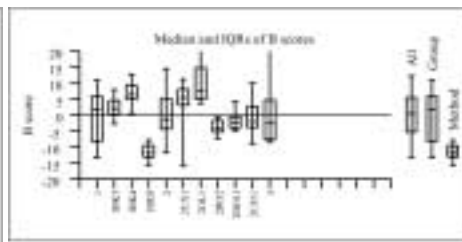
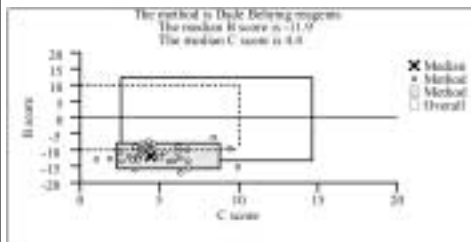
71

Lisbon & Porto, November 2006

Dade Behring C4 - February 2003



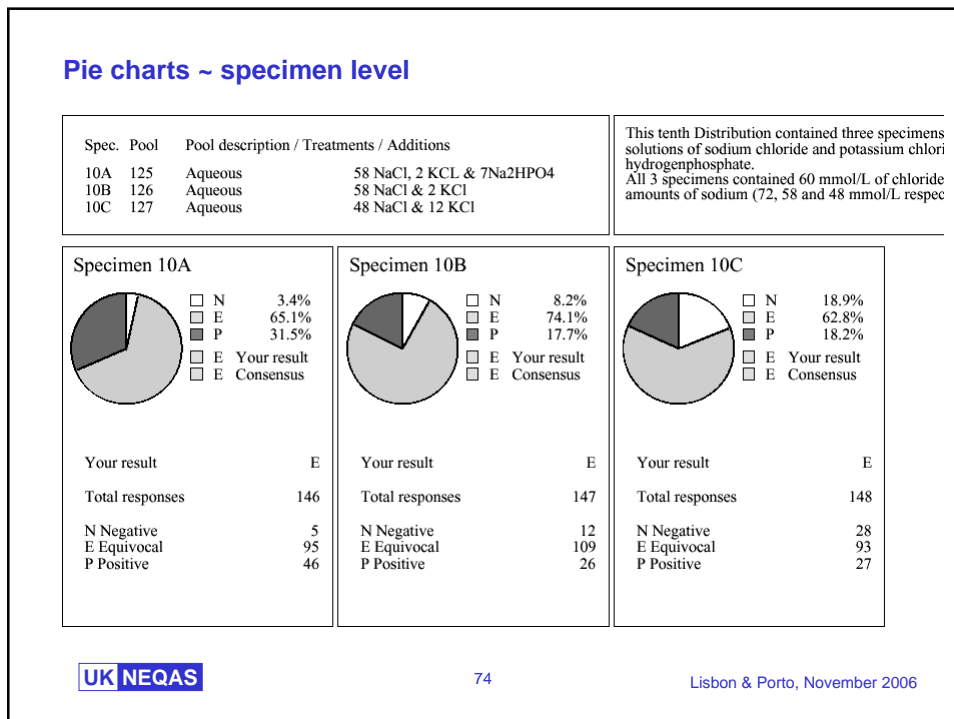
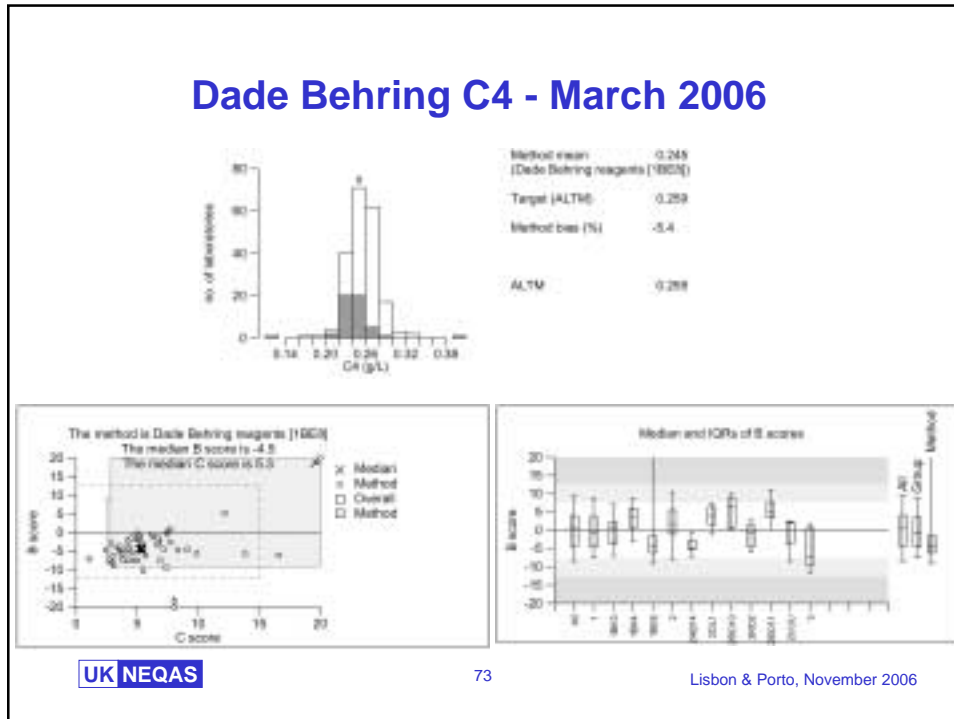
Method mean (Dade Behring reagent) 8.246
 Target (ALTM) 8.276
 Method bias (%) -11.8
 ALTM 8.276



UK NEQAS

72

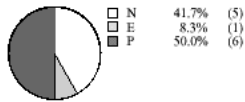
Lisbon & Porto, November 2006



Pie charts ~ rolling time window level

What colour are your pies?

Specimens distributed in each category



The first pie chart displays the proportion of negative [N], equivocal [E] and positive [P] specimens sent out over the last 4 distributions (as this is the seventh distribution this includes distributions 4 to 7).

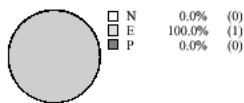
The proportion you reported in agreement with the consensus from other laboratories is shown below. If you are in complete agreement your Negative pie should be white, your Equivocal pie light grey and your Positive pie dark grey. As the number of data points increases this will allow you to judge if you are more stringent or less stringent than the average.

Your interpretation for each category

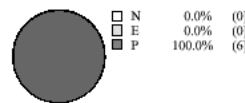
Negative



Equivocal



Positive



Pie charts ~ rolling time window level

What colour are your pies?

Specimens distributed in each category

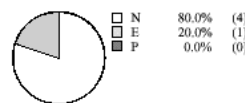


The first pie chart displays the proportion of negative [N], equivocal [E] and positive [P] specimens sent out over the last 4 distributions (as this is the seventh distribution this includes distributions 4 to 7).

The proportion you reported in agreement with the consensus from other laboratories is shown below. If you are in complete agreement your Negative pie should be white, your Equivocal pie light grey and your Positive pie dark grey. As the number of data points increases this will allow you to judge if you are more stringent or less stringent than the average.

Your interpretation for each category

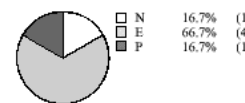
Negative



Equivocal



Positive



Take-home messages

- Good scheme design underpins any statistics and graphics
- Use graphical output wherever possible
- Have the supporting raw data available on request
- Give the data in a structured format
- Only drill down as far as you need
- Act on your EQA data ~ it contains a mass of information that is impossible to get from any other source

ANY QUESTIONS ? . . .

www.ukneqas.org.uk

email ClinChem@ukneqas.org.uk

FAX +44 121 414 1179

phone +44 121 414 7300

Appendix (not included in presentation)

**Worked example of
A, B and C Scores**

example results for a lab

This is an example calculation using the results for a typical 2-specimen and monthly distribution for a fictitious scheme.

In Distribution 573 in the UK NEQAS for Vegetables, Laboratory 12345 obtained the following results for *Serum Rhubarb*:-

2.17 units/litre for specimen 573A and
0.47 units/litre for specimen 573B

We will deal in detail with the results from the first specimen

Laboratory 12345 obtained a result of 2.17 units/litre on specimen 573A, which had a target value of 2.362 units/litre.

Specimen % bias

So, the Specimen %bias for Laboratory 12345 on specimen 573A is calculated as:-

$$\text{Specimen \%bias} = \frac{(\text{result} - \text{target})}{\text{target}} * 100$$

$$\text{Specimen \%bias} = \frac{(2.17 - 2.362)}{2.362} * 100$$

$$\text{Specimen \%bias} = -8.1$$

results for specimen B

On the second specimen in the distribution, specimen 573B, Laboratory 12345 got a result of 0.47 units/litre for analyte *Serum Rhubarb*.

The target was 0.512 units/litre and so the Specimen %bias was, again, coincidentally -8.1.

$$\text{Specimen \% bias} = \frac{(0.47 - 0.512)}{0.512} * 100$$

$$\text{Specimen \% bias} = -8.1$$

Specimen transformed bias

When we take a **Specimen %bias** and transform it by the concentration dependent 'degree of difficulty' factor we get a **Specimen transformed bias** [this can be positive or negative].

The Specimen transformed bias gives an indication in a 'common currency' (allowing for degree of difficulty) how much higher or lower than the target value your result was. This is calculated as follows:-

$$\text{Specimen transformed bias} = \frac{\text{Specimen \%bias}}{\text{the degree of difficulty}}$$

Specimen Accuracy Index

When we take the modulus of this 'Specimen transformed bias', this gives an indication as to the distance (ignoring direction) of the result is from its target.

Again, this remains in the same common currency of degree of difficulty as the Specimen transformed bias and is called the **Specimen Accuracy Index** [since it is a modulus it does not carry a sign].

$$\text{Specimen Accuracy Index} = |\text{Specimen transformed bias}|$$

and is used as an indicator of 'total error' and is in the same 'currency' as the 'A' score

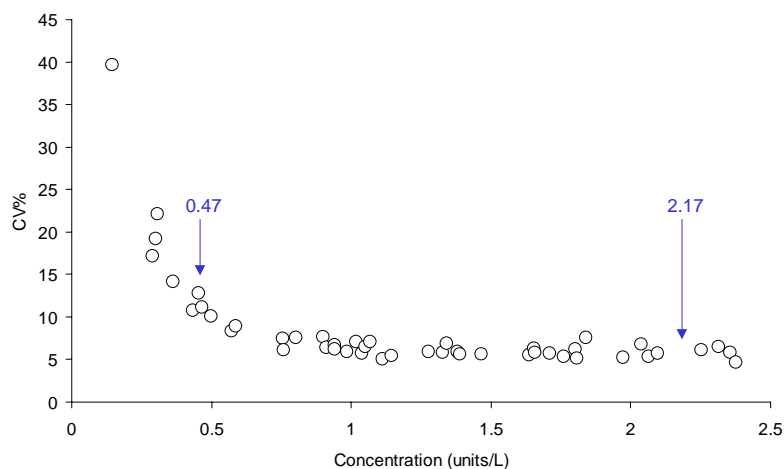
Degree of Difficulty

It is a well known fact that it is more difficult to measure low levels of *Serum Rhubarb* than it is to measure high levels of *Serum Rhubarb*.

The between laboratory agreement is poorer at low levels than it is at higher levels.

This is shown in the graph here. We have used the curve as the basis of the determination of the 'degree of difficulty' when measuring *Serum Rhubarb*.

Between-laboratory agreement for serum rhubarb



Specimen B was better than Specimen A

So, despite the fact that Laboratory 12345 was the same percentage away from the target for each specimen, because we take into account this degree of difficulty, the result on specimen 573B (the low concentration) is 'comparatively better' than the result on specimen 573A (the high specimen).

This is because it should be 'comparatively easier' to measure *Serum Rhubarb* at high levels than it is at low levels.

A Specimen bias of -8% at an 'easy' to measure 2, is worse than a Specimen bias of -8% at a 'difficult' to measure 0.5

summary

In summary, Laboratory 12345 has the following

	Specimen 573A	Specimen 573B
Laboratory result	2.17	0.47
Target value	2.362	0.512
Specimen %bias	-8.1	-8.1
Specimen transformed bias	-225	-122
Specimen Accuracy Index	225	122

But all of this is at the single specimen level,
what about the rolling time window 'trend data' picture?

How do we calculate the A, B and C scores?

All specimens in the time window contribute to the A, B and C scores.

The time window for the UK NEQAS for Vegetables is 6 Distributions, and since there are two specimens per distribution, there are **twelve** specimens' worth of data to be considered.

specimen % bias in specimen order

Consider all the results from Laboratory 12345 in the time window up to and including Distribution 573.

Distribution	Specimen	Result	Target	Specimen %bias	Specimen transformed bias	Specimen accuracy index
573	573B	0.47	0.512	-8.14	-122	122
573	573A	2.17	2.362	-8.14	-225	225
572	572B	1.54	1.559	-1.20	-30	30
572	572A	0.14	0.140	+0.31	+2	2
571	571B	0.35	0.325	+7.85	+88	88
571	571A	0.81	0.804	+0.80	+15	15
570	570B	0.72	0.691	+4.19	+74	74
570	570A	2.68	2.687	-0.26	-7	7
569	569B	1.58	1.571	+0.57	+14	14
569	569A	0.37	0.349	+5.97	+70	70
568	568B	1.13	1.139	-0.82	-18	18
568	568A	1.14	1.143	-0.29	-6	6

specimen % bias ranked % bias

If we rank the Specimen %biases in ascending order we get

Distribution	Specimen	Result	Target	Specimen %bias
573	573B	0.47	0.512	-8.14
573	573A	2.17	2.362	-8.14
572	572B	1.54	1.559	-1.20
568	568B	1.13	1.139	-0.82
568	568A	1.14	1.143	-0.29
570	570A	2.68	2.687	-0.26
572	572A	0.14	0.140	+0.31
569	569B	1.58	1.571	+0.57
571	571A	0.81	0.804	+0.80
570	570B	0.72	0.691	+4.19
569	569A	0.37	0.349	+5.97
571	571B	0.35	0.325	+7.85

To get a more robust estimate of the B score, we trim away extreme data points to lessen any undue influence that they might have.

trimming

When we trim out the extreme Specimen %biases at each end, we 'lose' specimen 573B's -8.14% and specimen 571B's +7.85

Distribution	Specimen	Result	Target	Specimen %bias
573	573B	0.47	0.512	-8.14
573	573A	2.17	2.362	-8.14
572	572B	1.54	1.559	-1.20
568	568B	1.13	1.139	-0.82
568	568A	1.14	1.143	-0.29
570	570A	2.68	2.687	-0.26
572	572A	0.14	0.140	+0.31
569	569B	1.58	1.571	+0.57
571	571A	0.81	0.804	+0.80
570	570B	0.72	0.691	+4.19
569	569A	0.37	0.349	+5.97
571	571B	0.35	0.325	+7.85

B and C definition

The average of the remaining 10 Specimen %biases is +0.1%. This is your B score.

The standard deviation of the remaining 10 Specimen %biases is 3.7% This is your C score.

ranking by Specimen Accuracy Indices

To get the A score, we follow a similar procedure. For Laboratory 12345, the Specimen Accuracy Indices for the last 12 specimens in the time window (up to and including Distribution 573) are ranked.

Distribution	Specimen	Result	Target	Specimen Accuracy Index
572	572A	0.14	0.140	2
568	568A	1.14	1.143	6
570	570A	2.68	2.687	7
569	569B	1.58	1.571	14
571	571A	0.81	0.804	15
568	568B	1.13	1.139	18
572	572B	1.54	1.559	30
569	569A	0.37	0.349	70
570	570B	0.72	0.691	74
571	571B	0.35	0.325	88
573	573B	0.47	0.512	122
573	573A	2.17	2.362	225

calculating the A score

When we trim out the extreme Specimen Accuracy Indices at each end, we 'lose' specimen 572A's 2 and we 'lose' specimen 573A's 225.

Distribution	Specimen	Result	Target	Specimen Accuracy Index
572	572A	0.14	0.14	
568	568A	1.14	1.143	6
570	570A	2.68	2.687	7
569	569B	1.58	1.571	14
571	571A	0.81	0.804	15
568	568B	1.13	1.139	18
572	572B	1.54	1.559	30
569	569A	0.37	0.349	70
570	570B	0.72	0.691	74
571	571B	0.35	0.325	88
573	573B	0.47	0.512	122
573	573A	2.17	2.362	

The average of the remaining 10 Specimen Accuracy Indices, **45**, is your A score



HbA1c Ready-Reckoner

	conc											
	4	5	6	7	8	9	10	11	12	13	14	
%bias	1	18	22	26	30	34	36	36	34	30	26	22
	2	36	43	52	60	68	72	72	68	60	52	44
	3	54	65	77	90	102	108	108	102	91	78	65
	4	72	87	103	120	135	144	145	136	121	104	87
	5	90	108	129	150	169	181	181	170	151	130	109
	6	108	130	155	181	203	217	217	204	181	156	131
	7	127	152	181	211	237	253	253	238	211	181	153
	8	145	173	206	241	271	289	289	272	242	207	174
	9	163	195	232	271	305	325	325	305	272	233	196
	10	181	217	258	301	338	361	362	339	302	259	218
	15	271	325	387	451	500	500	500	500	453	389	327

