

***Proficiency Testing –  
Are we fully utilising a valuable resource ?***

***3<sup>rd</sup> AOAC Europe Eurachem Symposium  
March 3/4<sup>th</sup> 2005 – Brussels***

**John Gilbert – Science Director  
Central Science Laboratory**



## Introduction

- ◆ What do participants want from PT
- ◆ What does the analytical community want from PT

## Limitations of proficiency testing

- ◆ Choice of assigned value
- ◆ Bump hunting
- ◆ Selection of  $\sigma$ -value

## Method performance – examples of insights from PT data

- ◆ Tin in vegetable puree
- ◆ Chloramphenicol in prawns
- ◆ Acrylamide in cooked foods
- ◆ Deoxynivalenol in wheat

## Conclusions



## ***What do participants want from PT***

Satisfactory z-score based on:-

- ◆ Accurate assigned value without method dependence
- ◆  $\sigma$ -value decided on best practice
- ◆ Relevant test material
- ◆ Relevant analyte at relevant concentration

## ***What does the analytical community want from PT?***

- ◆ Progressive improvement in performance of laboratories globally
- ◆ Insight into any analyte method dependence
- ◆ Insight into method performance



## *Assigned values and target $\sigma$ values*

- ◆ 'True' value and 'assigned value'
  - the 'true' value is an ideal
  - the 'assigned value' is the best estimate of the determinand
- ◆ Target Standard Deviation
  - The target standard deviation sets the limits of satisfactory performance in the PT



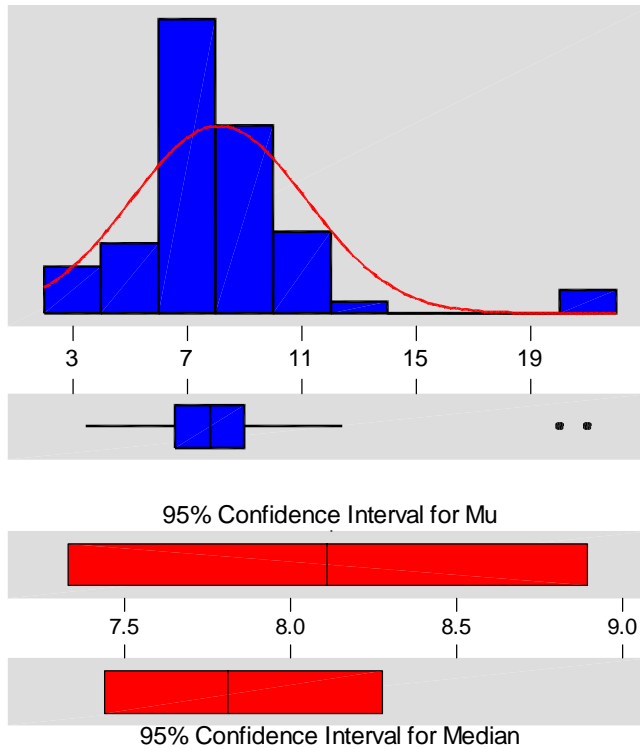
## *The Assigned Value*

- ◆ FAPAS<sup>®</sup> derives the assigned value from the most appropriate measure of central tendency:
  - robust mean
  - median
  - mode



# Simple Mean v Robust Mean

## Descriptive Statistics



Variable: ass. value

Anderson-Darling Normality Test

A-Squared: 2.450  
P-Value: 0.000

Robust Mean 7.82879

Mean 8.11113  
StDev 3.06012  
Variance 9.36431  
Skewness 2.08479  
Kurtosis 7.53429  
N 61

Minimum 3.4700  
1st Quartile 6.5900  
Median 7.8100  
3rd Quartile 9.0000  
Maximum 21.0000

95% Confidence Interval for Mu  
7.3274 8.8949

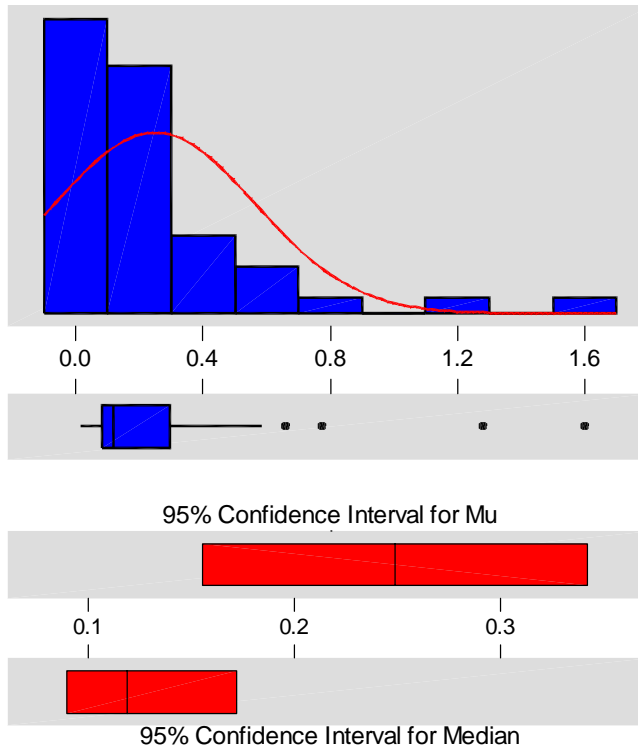
95% Confidence Interval for Sigma  
2.5972 3.7255

95% Confidence Interval for Median  
7.4405 8.2757



# Robust Mean and Median Limitations

## Descriptive Statistics



Variable: afm1

Anderson-Darling Normality Test

A-Squared: 5.339      Robust Mean 0.184686  
P-Value: 0.000

Mean 0.248691  
StDev 0.314717  
Variance 9.90E-02  
Skewness 2.78245  
Kurtosis 8.74819  
N 46

Minimum 0.01700  
1st Quartile 0.08100  
Median 0.11900  
3rd Quartile 0.29500  
Maximum 1.60000

95% Confidence Interval for Mu  
0.15523      0.34215

95% Confidence Interval for Sigma  
0.26104      0.39639

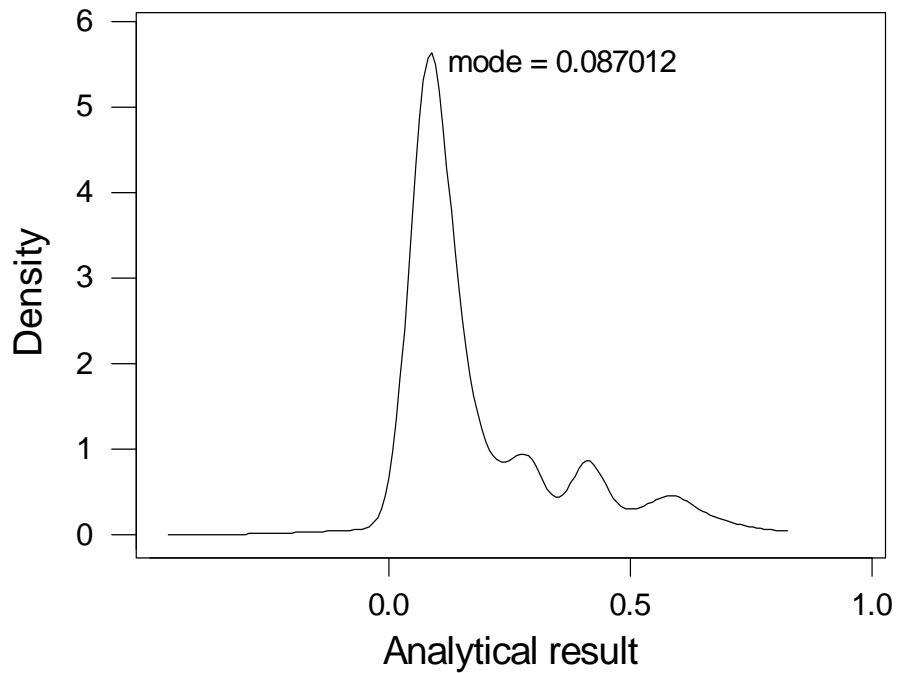
95% Confidence Interval for Median  
0.08998      0.17190





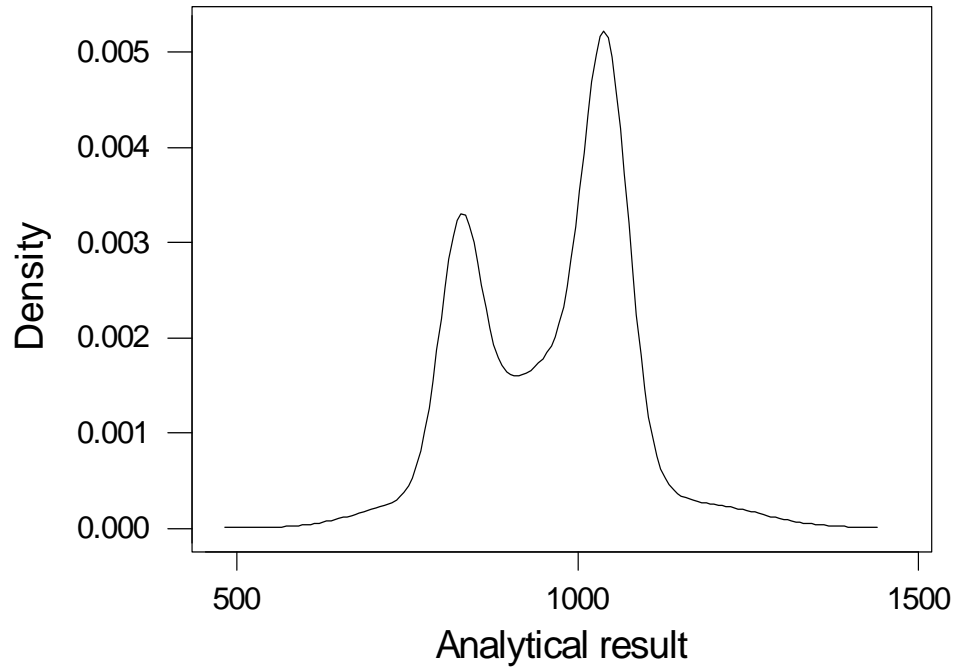
# ***‘Bump-hunting’ for the Mode(s)***

Adaptive kernel density plot – aflatoxin M<sub>1</sub>



# *Bimodal Results – Poor Methodology?*

Adaptive kernel density plot - chloride



## ***Influence of $\sigma$ -values on z-scores***

Total aflatoxins in peanut meal - mean 25.4 ppb

	<b>Acceptable range ppb</b>	<b>Numbers of participants</b>		<b>% Satisfactory</b>
		<b>Satisfactory</b>	<b>Unsatisfactory</b>	
Horwitz	11.3 - 39.6	102	19	84
95% $\pm 2$ z-score	2.4 - 48.4	114	7	95
Best practice	15.8 - 35.1	86	35	71

## ***To satisfy PT Participants organisers MUST:-***

- ◆ Be confident in assigned value
- ◆ Be confident that there is no method dependence in dataset
  - bump-hunting
- ◆ Need to set target standard deviation at realistic value
- ◆ Be confident in homogeneity and in stability of test material



## ***Method performance – examples of insights from PT data***

- ◆ Tin in vegetable puree
- ◆ Chloramphenicol in prawns
- ◆ Acrylamide in cooked foods
- ◆ Deoxynivalenol in wheat



## ***FAPAS 0738 – Tin in Tomato Paste***

### ◆ Preparation

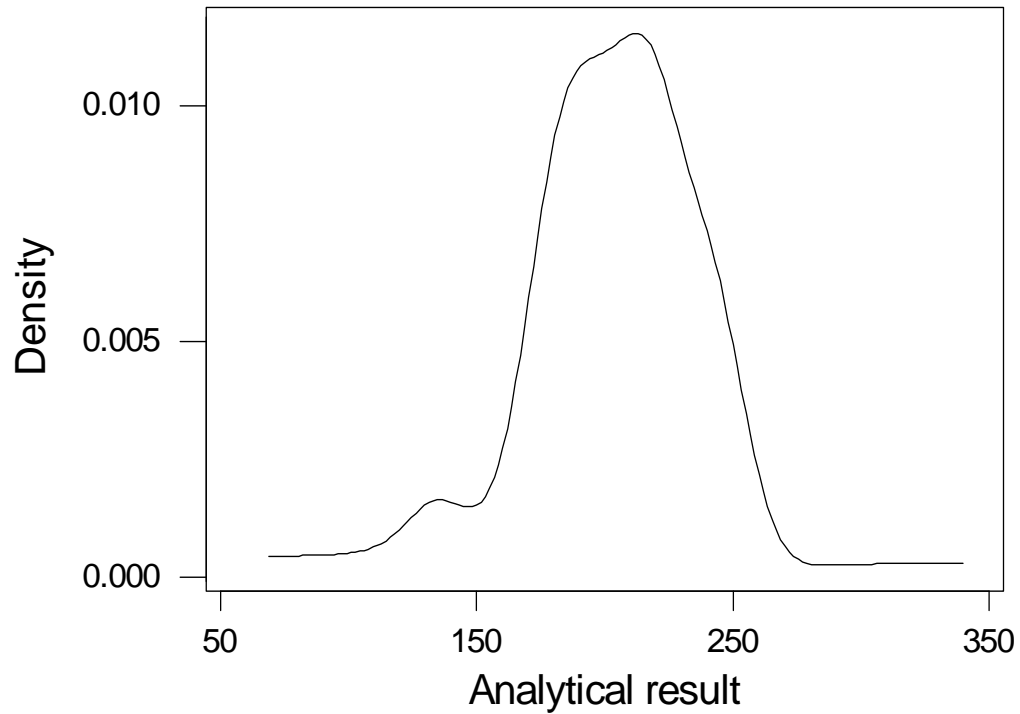
- Spiked at 250 mg/kg
- Homogeneity mean was 251 mg/kg

### ◆ Results from participants

- Consensus of results was 204 mg/kg
- Consensus was 20% lower than the spike
- Homogeneity mean would have resulted in a z-score of 3.2
- Using ICP-IDMS result was 247.8 mg/kg

# 0738 – Tin in Tomato Paste

Adaptive kernel density plot - all tin



## *Issues with PT for Tin Analysis*

- ◆ No correlation between methods used and the results received
- ◆ FAPAS protocol indicates that we should use the consensus value not the spike value
- ◆ Complaints from participants that obtained results that correlated with the spike value, but they received unsatisfactory z-scores





## *FAPAS 0741 – Tin in Vegetable Puree*

- ◆ Preparation
  - Spiked at 200 mg/kg
  - Homogeneity mean was 204 mg/kg
  
- ◆ Results from participants
  - Consensus of results was 179 mg/kg
  - Consensus was 10% lower than the spike
  - Homogeneity mean would have resulted in a z-score of 1.9



## ***FAPAS 0741 – Tin in Vegetable Puree (continued)***

- ◆ Report production was delayed until discussed with Advisory Committee
- ◆ Comments from the Committee:
  - ‘Expert labs’ and LGC ICP-IDMS give good agreement with spike (0738 round)
  - No trends in method information seen
  - For tin analysis only, FAPAS would deviate from the protocol i.e. the assigned value would not be set as the consensus if this deviated from the spike value

# *Use of Reference Value rather than Consensus of Participants' Results 0745 – Tin in Tomato Puree*

## **Summary**

- Homogeneity mean was 282 mg/kg
- Consensus of results was 255.9 mg/kg
- Consensus was 10% lower than the spike
- Thus, assigned value was set at ICP-IDMS result provided by LGC (286.5 mg/kg)
- Homogeneity mean would have resulted in a z-score of  $-0.2$



# ***Consensus and Reference Values 0745 – Tin in Tomato Puree***

- ◆ Reduction in % satisfactory z-scores when reference value was used:
  - Using reference value – 60% satisfactory
  - Using consensus – 72% satisfactory
- ◆ Contacted participants to inform them of the problem
- ◆ On request, provided information on method used by the homogeneity laboratory



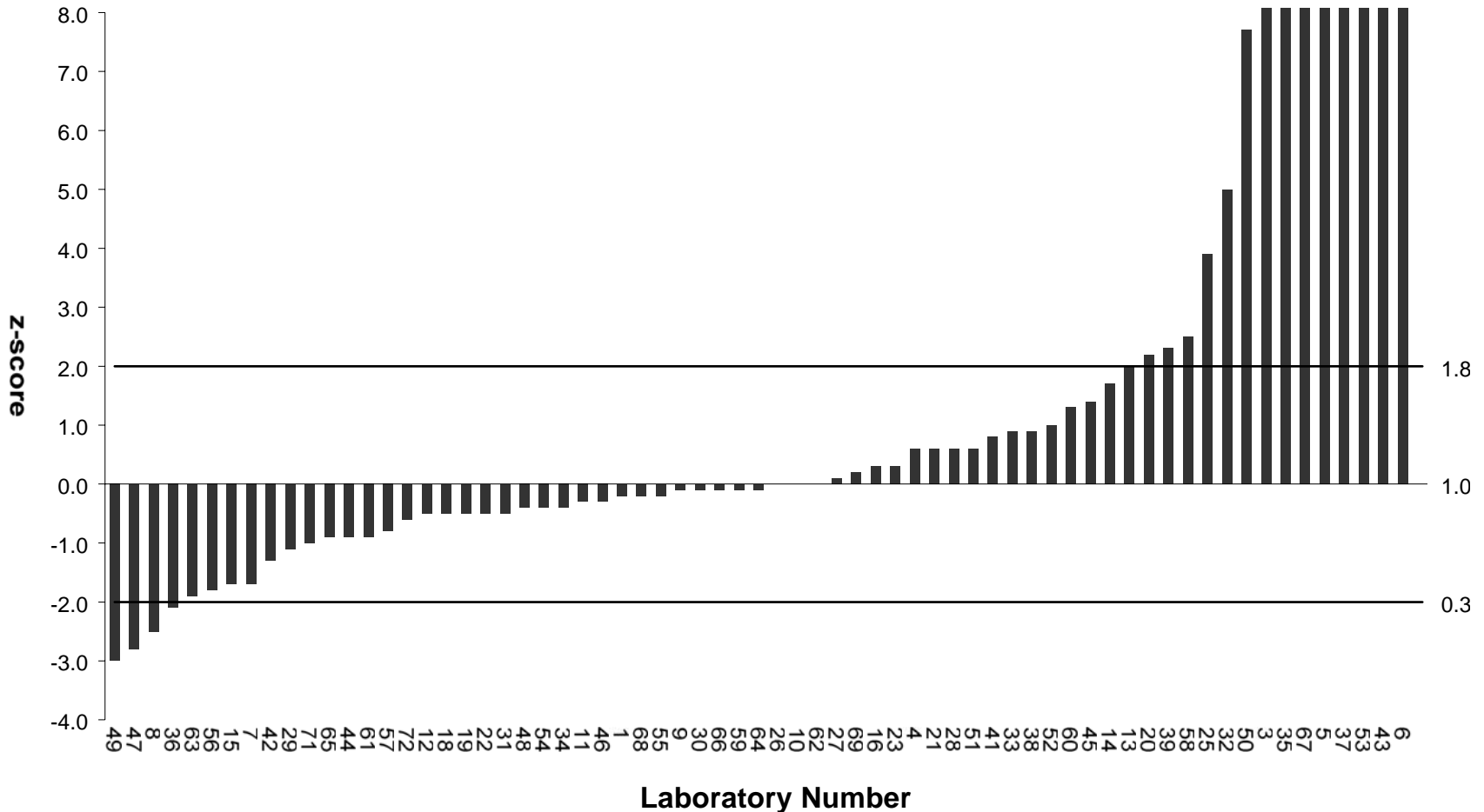
## *Most Recent Round of Tin in Vegetable Puree*

- ◆ Under reporting not a problem (consensus agreed with the spike)
- ◆ Consensus used as the assigned value
- ◆ Using a reference laboratory pushed participants towards addressing problems with their methodology



# Results from recent FAPAS® Rounds

Round 41 (November 2002) chloramphenicol in prawns  
(assigned value 0.9 ppb)



***Choice of methods and FAPAS<sup>®</sup> scores  
(chloramphenicol in prawns)***

<b>Method</b>	<b>No. of Labs</b>	<b>No. Satisfactory</b>	<b>% Satisfactory</b>
GC & GC/MS	13	13	100
LC/MS	10	10	100
LC	27	16	60
ELISA	21	15	71
Overall	67	49	73

## *Indications of problem areas*

- ◆ Lack of specificity in HPLC detection
- ◆ High proportion of labs with unsatisfactory z-scores (70% - 7 out of 10) used UV detection in HPLC
- ◆ ELISA more appropriate for screening than quantification



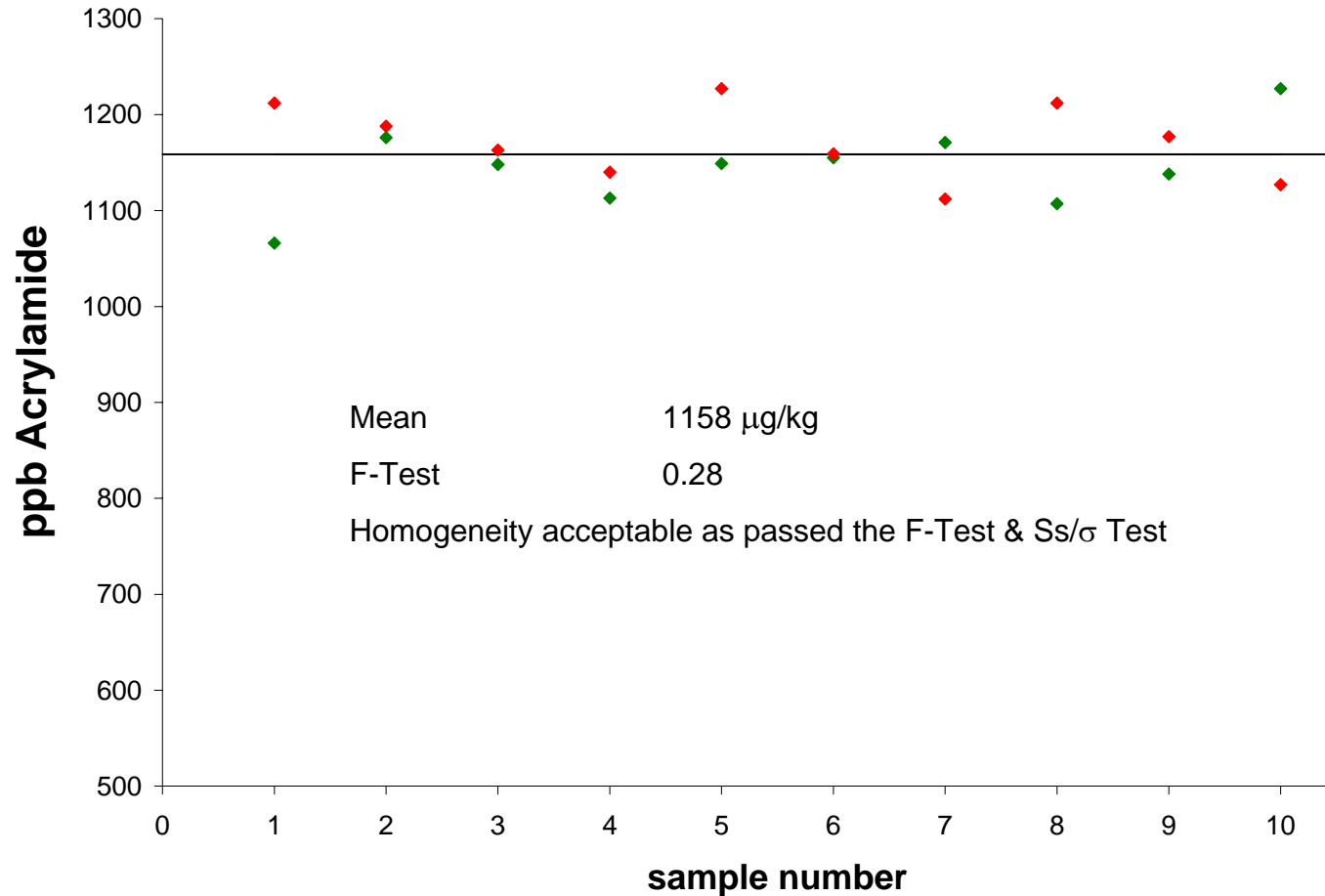


# **Acrylamide**

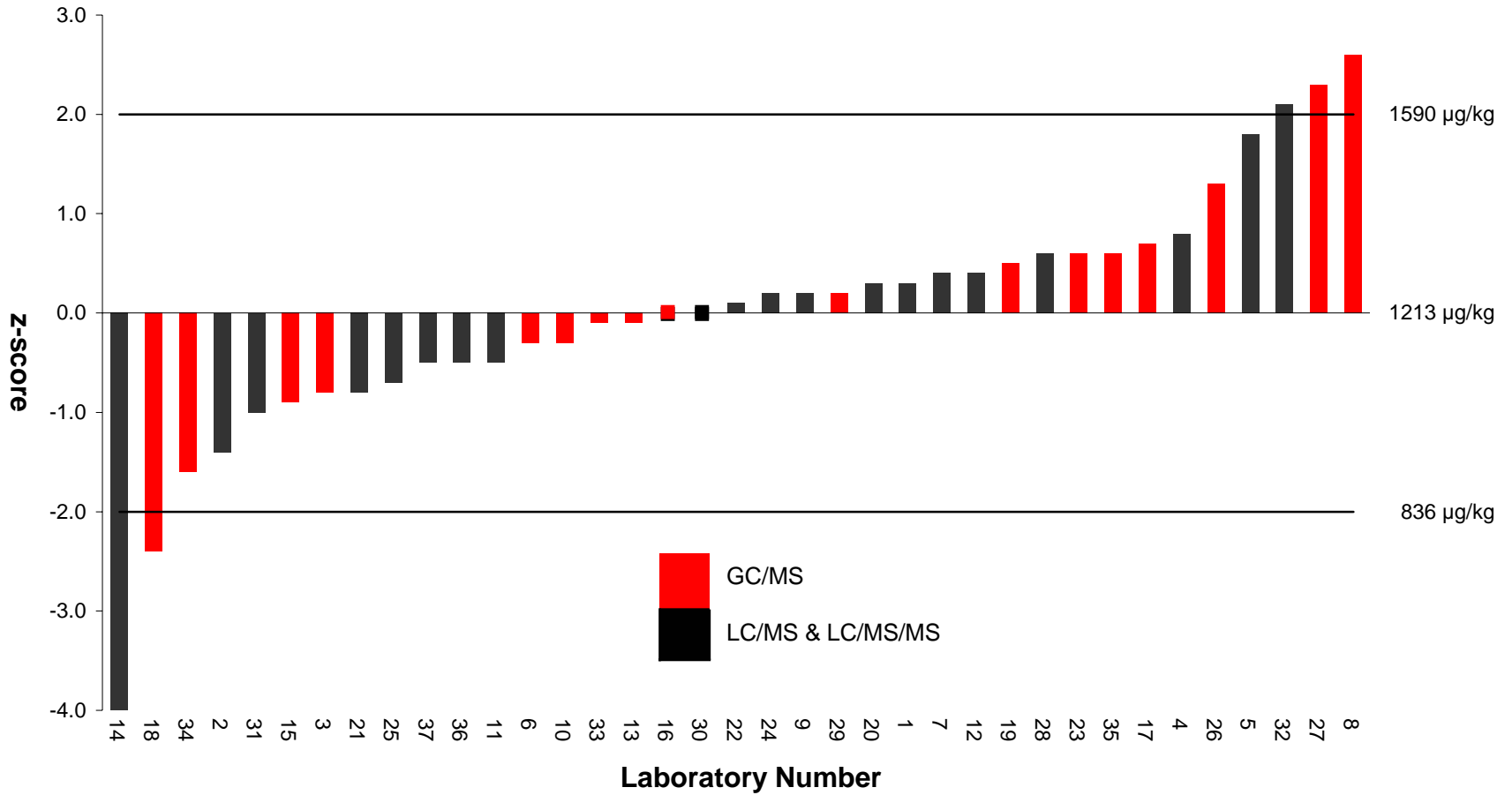
## **FAPAS<sup>®</sup> Rounds 2002-2003**

Round No.	Material	Fat	Carbohydrate	Protein	Acrylamide ppb
1	Crispbread	2%	63%	9%	1200
2	Potato crisps	34%	49%	6%	170
3	Breakfast cereal	3%	68%	11%	100-300
4	Ground roasted coffee	12%	35%	3%	400-600

# *FAPAS<sup>®</sup> Homogeneity Data for Acrylamide in Crispbread*



# Z-scores for Acrylamide ( $1213 \mu\text{g}/\text{kg}$ ) in Crispbread Test Material



## ***Information Available on Methodology from FAPAS®***

- ◆ Literature reference to method if available
- ◆ Sample weight taken
- ◆ Internal standard used
- ◆ Extraction solvent used



## ***Information Available on Methodology from FAPAS<sup>®</sup> (continued)***

- ◆ Treatment of extract – filtered, centrifuged etc.
- ◆ Clean-up – SPE etc.
- ◆ Determination step GC/MS; LC/MS; LC/MS/MS etc.
- ◆ GC/LC – column, length, temperature programme etc.
- ◆ MS – ions monitored, scan mode etc.



## ***Correlating Performance with Methodology***

- ◆ Sample size taken
- ◆ Use of internal standard
- ◆ Choice of extraction solvent
- ◆ Clean-up
- ◆ GC/MS or LC/MS



## ***Sample Size Taken for Analysis***

Nos of unsatisfactory Labs			
	1g – 2g	2g – 5g	5g – 35g
Round 1	0	3	2
Round 2	2	2	1
Round 3	3	5	3
No evidence of influence of sample size			

## *Use of Internal Standards*

Nos of unsatisfactory Labs				
	None	d <sub>3</sub> -acrylamide	<sup>13</sup> C-acrylamide	other
Round 1	1 (50%)	2 (11%)	-	2 (20%)
Round 2	1 (50%)	3 (13%)	1 (14%)	-
Round 3	1 (50%)	4 (20%)	2 (30%)	4 (50%)

High proportion of Labs not using an internal standard or not using a labelled internal standard were unsatisfactory



## ***Choice of Extraction Solvent***

Nos of unsatisfactory Labs			
	Water	Aqueous Organic	Organic
Round 1	1 (4.5%)	-	4 (40%)
Round 2	3 (13%)	-	2 (40%)
Round 3	8 (33%)	1 (20%)	2 (33%)

High proportion of Labs using organic solvents were unsatisfactory

## *Choice of Sample Clean-up*

Nos of unsatisfactory Labs				
	None	Liquid/Liquid	SPE	Other
Round 1	1	1	-	-
Round 2	1	1	1	-
Round 3	1	-	8	2

No trend or patterns between clean-up and satisfactory performance

## ***GC/MS or LC/MS or LC/MS/MS***

Nos of unsatisfactory Labs			
	GC/MS	LC/MS	LC/MS/MS
Round 1	3 (15%)	2 (14%)	-
Round 2	2 (11%)	2 (25%)	-
Round 3	3 (21%)	4 (44%)	4 (31%)

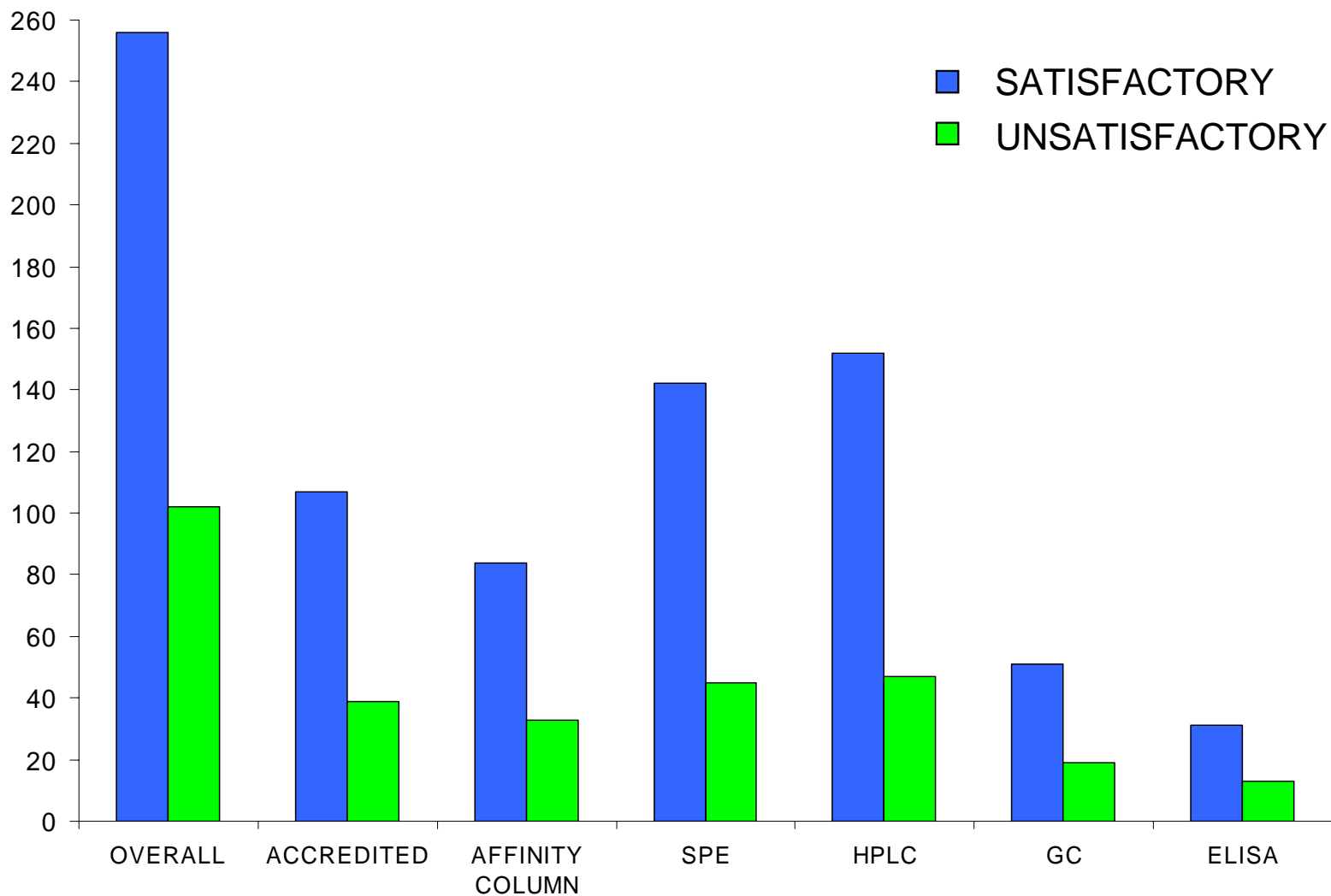
No evidence of GC/MS, LC/MS or LC/MS/MS influencing performance

## *Acrylamide – Proficiency Testing*

- ◆ Variety of methods used
- ◆ Methods using organic solvents and not using labelled internal standards gave poorer performance
- ◆ No evidence of better results by GC/MS or LC/MS or LC/MS/MS



# *Method Performance for Deoxynivalenol in Wheat - combined results for several PT rounds*



# *Conclusions*

- ◆ Participation in proficiency testing is leading to improved laboratory performance

Leads to sounder compliance decisions

- ◆ PT data can be used to identify problems with methodology

- ◆ PT data can provide insights into method validation

BUT limitation in extent to which methods are 'customised'

