Birmingham Quality

Birmingham Quality

Previously known as the *Wolfson EQA Laboratory*,
Birmingham Quality provides primarily
UK NEQAS External Quality Assessment
Services in Clinical Chemistry

UK NEQAS

Birmingham Quality
PO Box 3909
Birmingham
B15 2UE

tel: +44 (0)121 414 7300
fax: +44 (0)121 414 1179
birminghamquality@uhb.nhs.uk
www.birminghamquality.org.uk

Reference Intervals

Finlay MacKenzie

Organiser of the UK NEQAS for Thyroid Hormones
Organiser of the UK NEQAS for Steroid Hormones
Deputy Director Birmingham Quality *(formerly Wolfson EQA Laboratory)*
University Hospitals Birmingham NHS Foundation Trust

Reference Intervals ~ Introduction

Normal Ranges, Reference Intervals or Expected Values. Call them what you will, comparing an individual result against a predetermined estimate of an expected number or range is at the heart of Laboratory Medicine.

Gone are the days when a physician sat the patient down, did the history, made the examination and took some blood to rule-in or rule-out some specific disease state. In the UK, you are more likely to get a rushed consultation of seconds rather than minutes or hours and a blanket battery of tests requested and when the result comes back, the poor pressurised doctor has to decide whether the results are normal or abnormal.

Reference Intervals

We all know that within-individual variation is usually small when compared to the ‘Normal range’;

our individual set points don’t tend to shift, but nevertheless in what many consider to be dumbed-down, 21st Century Britain it can often mean that “9.9” and you’re fine, but if it’s “10.1” then you need to get your affairs in order and be ready to make your peace with your maker!

Reference Intervals

The empirical findings, or ‘real world’ cut-offs, for Newborn Screening were contrasted with the published, recommended, cut-offs. This meant that analytical performance was actually involved and it wasn’t just both sides reading a common set of guidelines and pretending that we agreed.

In summary, EQA is perfectly positioned to add a pragmatic dimension to the numbers generated. Until all assays agree numerically, assay bias must be taken into account when looking at cut-offs. HbA1c results which have a negative bias currently gain UK Primary Care Physicians a cash windfall. How perverse an incentive is that to use a method just on the right side of traceability?

Reference Intervals

EQA can ensure that the numbers agree. It can ensure that the numbers are accurate. But perhaps more importantly, it can try to ensure that the numbers are used correctly and not generated then just interpreted on a whim.

Reference Intervals

I will show how we have asked for, and coped with, these method principle-specific ranges and also for Trimester-specific TFT ranges. The dilemma for all of these systems is in the provenance of the values used. 'Lost in the mists of time' is often the way that Labs tell us where their ranges have come from.

Reference Intervals

We can assist in the collation and dissemination of reference intervals by means of Audit/Questionnaires – usually of the on-line variety. We can collate and produce our findings in a simple to understand graphical way, but it is up to others to help in their uptake and adoption. The Pathology Harmony Group worked most noticeably with the Tumour Marker and Haematology UK NEQAS Schemes. In Chemistry, there was much more of a pragmatic approach to the Type I analytes, with a reluctance to take on contentious assays. We, at UK NEQAS Birmingham, have tried to assist with some Enzyme Reference Intervals and hope to feed back into the Pathology Harmony process.

The Trouble with Guidelines

Rant Zero - MacKenzie's Maxim

When interpreting a laboratory's result for any given analyte, all of those components contributing to the 'uncertainty budget' must be taken into consideration.

This will not only be the background 'imprecision' in the laboratory, but will also take into account the method bias and whether or not the analyte can be truly 'calibrated' and 'measured'.

Again, an individual 'set point' and biological variation have their role in such considerations, too.

MacKenzie's Maxim

What if the lab has a bad day and all the results are particularly low?

Do you mis-treat a greater proportion of patients?

I know that in the real world clinicians might consider a TSH of '9' to be about the same as a TSH of '10'.

But where do you draw the line? At 8, 7 or where?

Round numbers

How do we know there is a God?

Simple, the cut-offs used in most guidelines are nice round numbers, usually involving a ‘10’ somewhere.

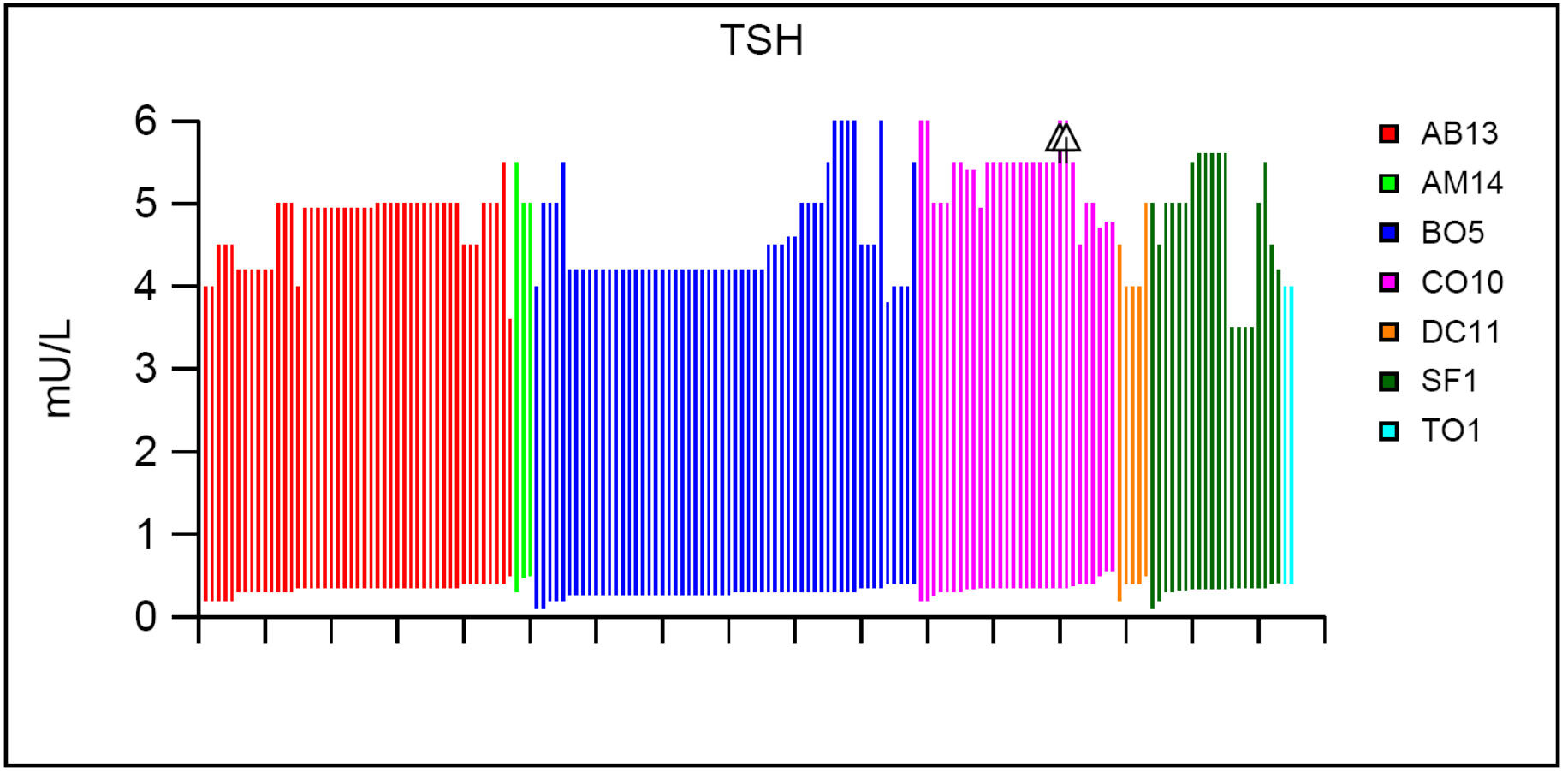
10 mU/L has been suggested as the cut-off for hypothyroidism. Does this mean that for TSH, God uses mU/L 3rd IS as his/her units and is bang-up-to-date?

But, because the cut-off for raised cholesterol is a nice round 200 mg/dL *(masquerading as a scientific-looking 5.19 mmol/L),* then God must also use American units and also simultaneously be, in a theological contradiction, only as modern as the Old Testament?

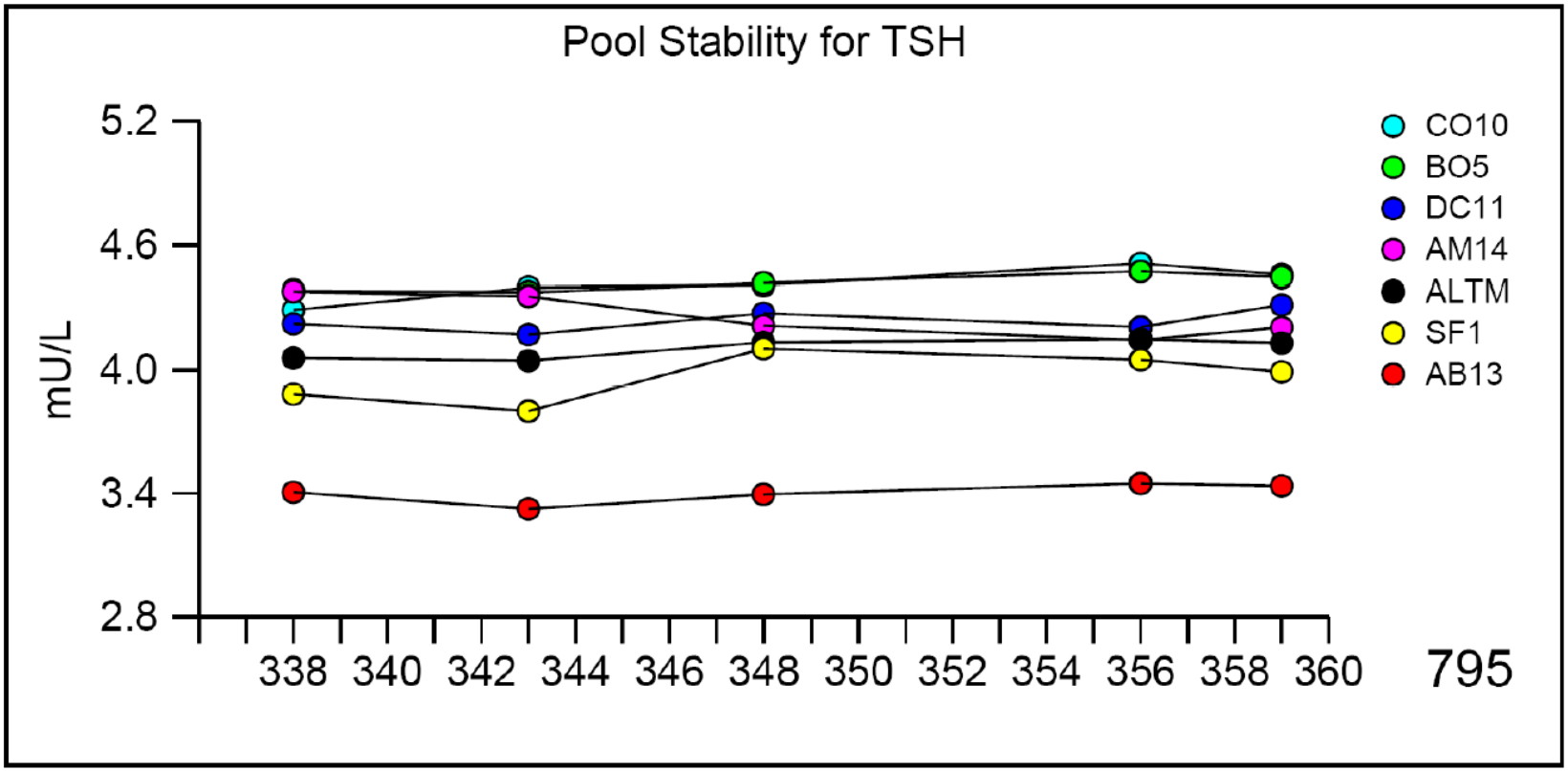
The first set of Guidelines set a precedent for an obsession with the number "10"

It was always a surprise to see the late Chuck Heston with a staff and not an assault rifle!

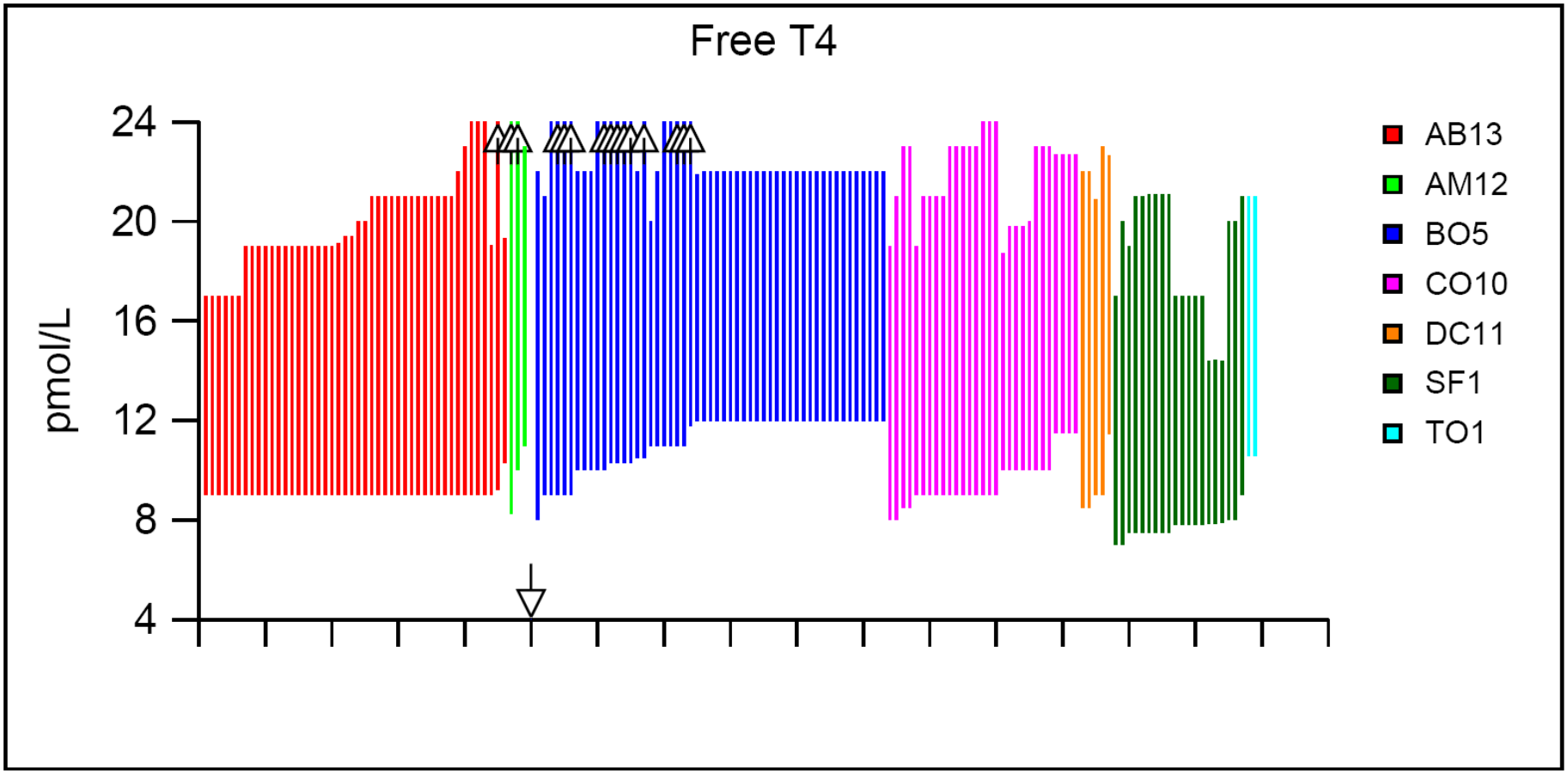
Reference Intervals



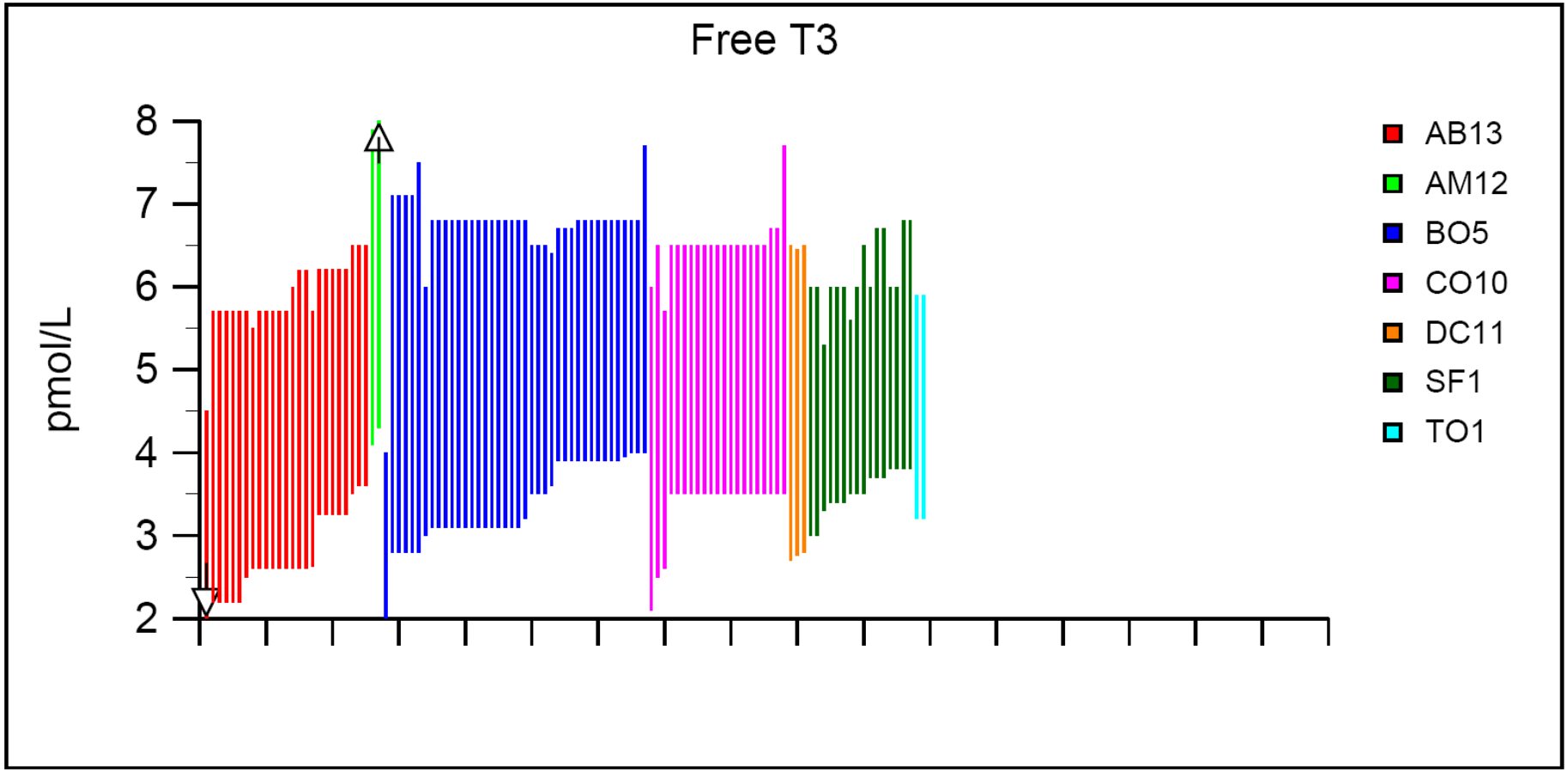
Stability of the ALTM and MLTMs



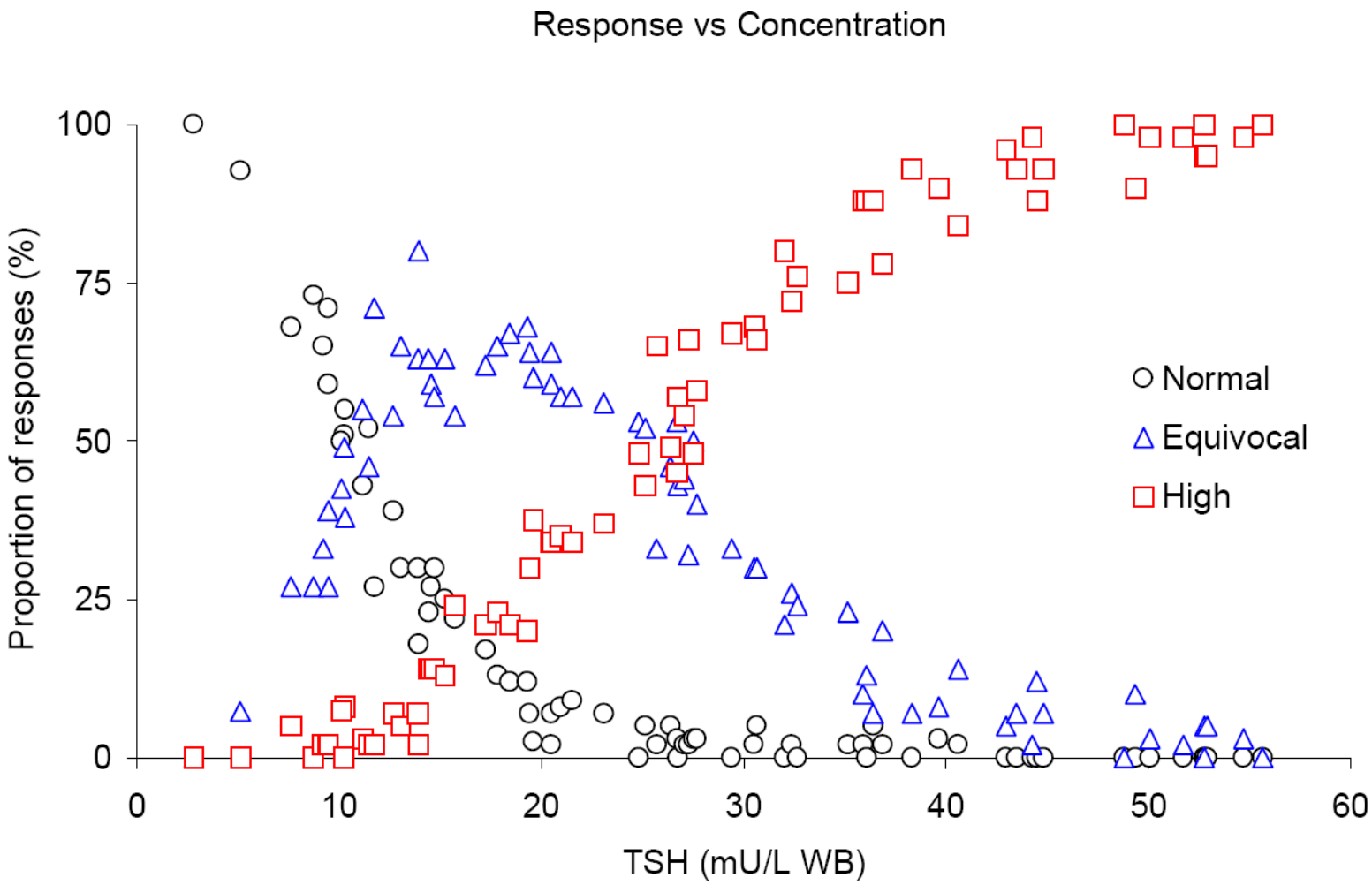
Reference Intervals



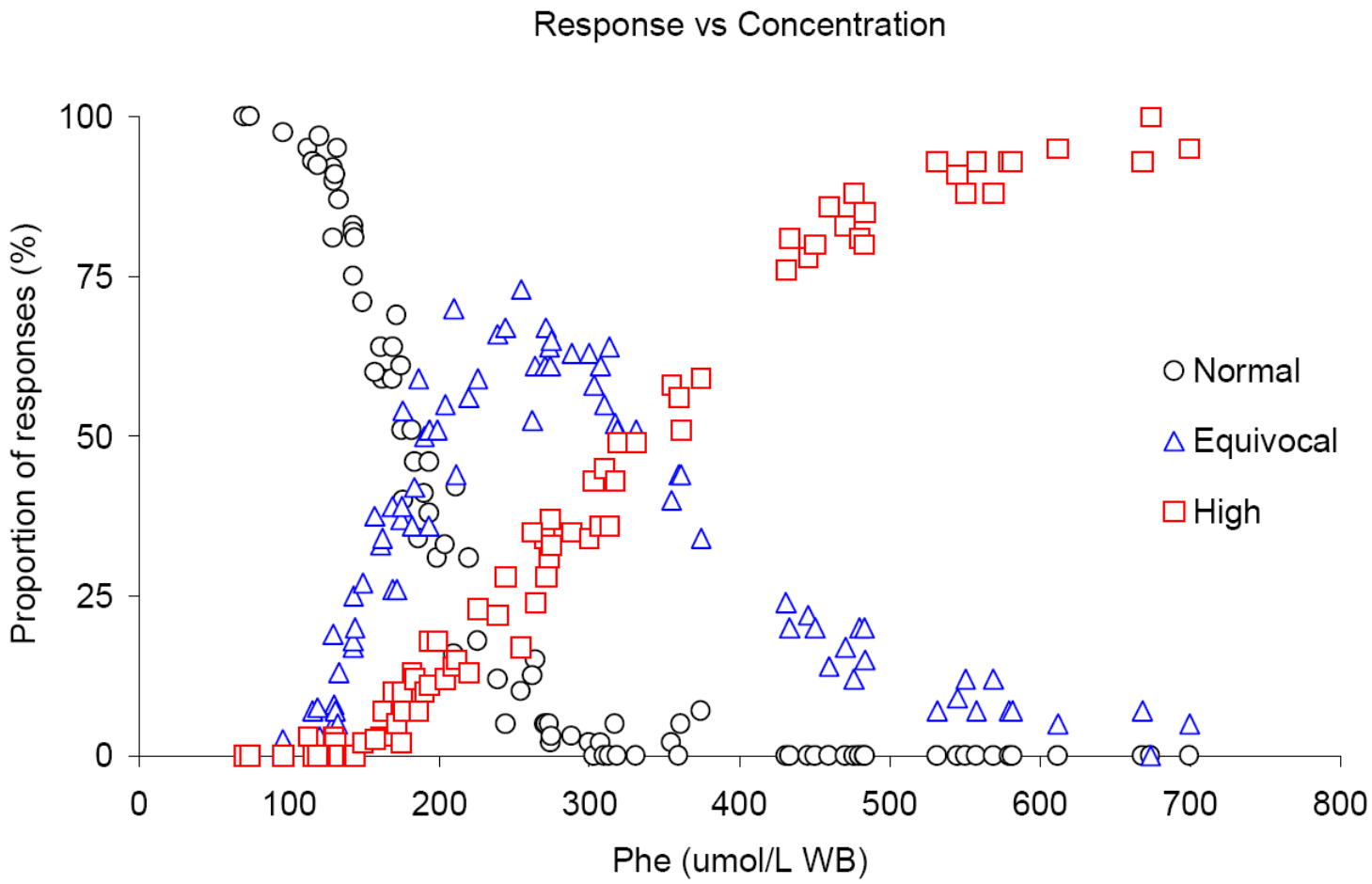
Reference Intervals



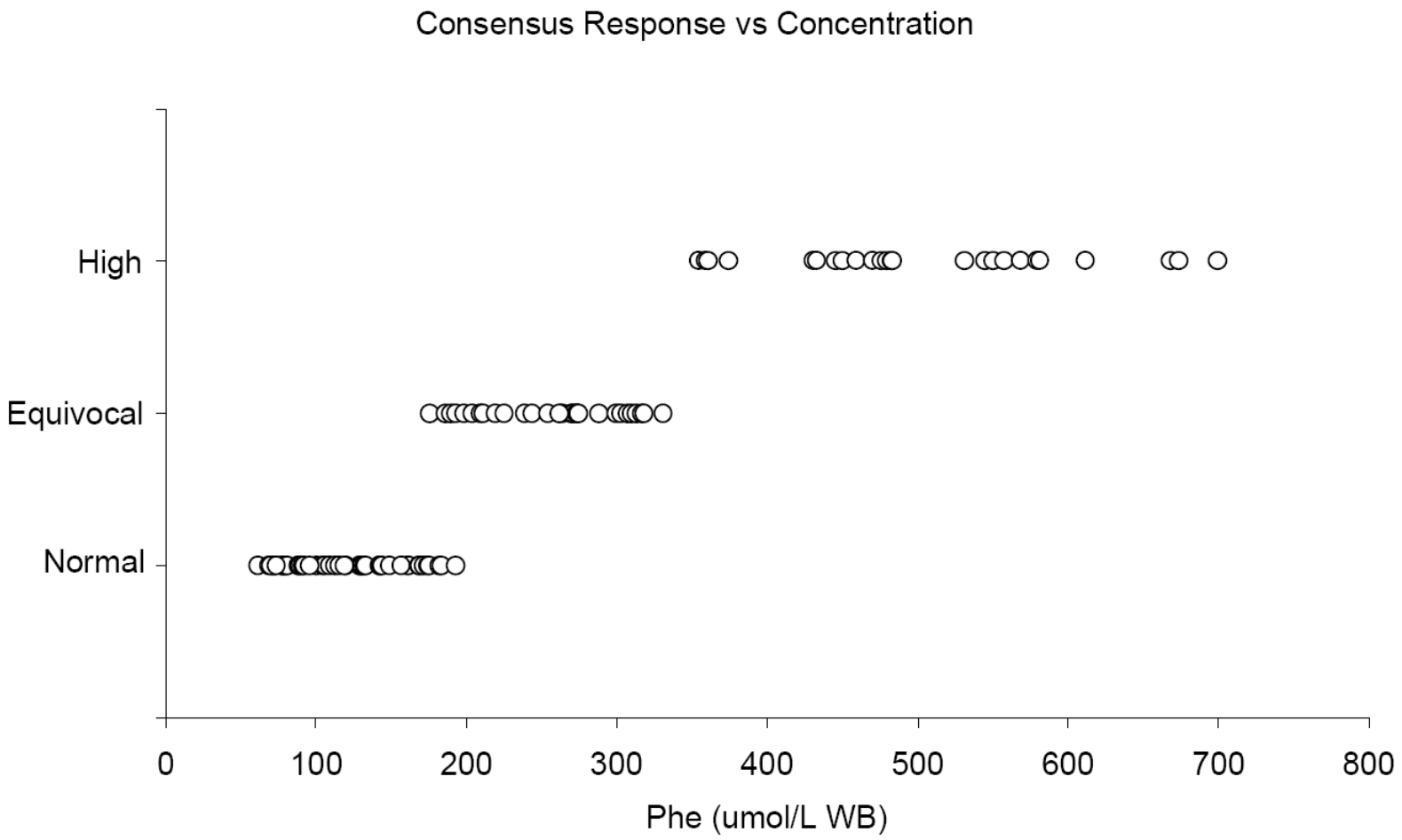
TSH Graph 1


Response vs Concentration
Proportion of responses (%)
100
75
50
25
0
0
10
20
30
40
50
60
TSH (mU/L WB)
Normal
Equivocal
High

Phenylalanine Graph 1

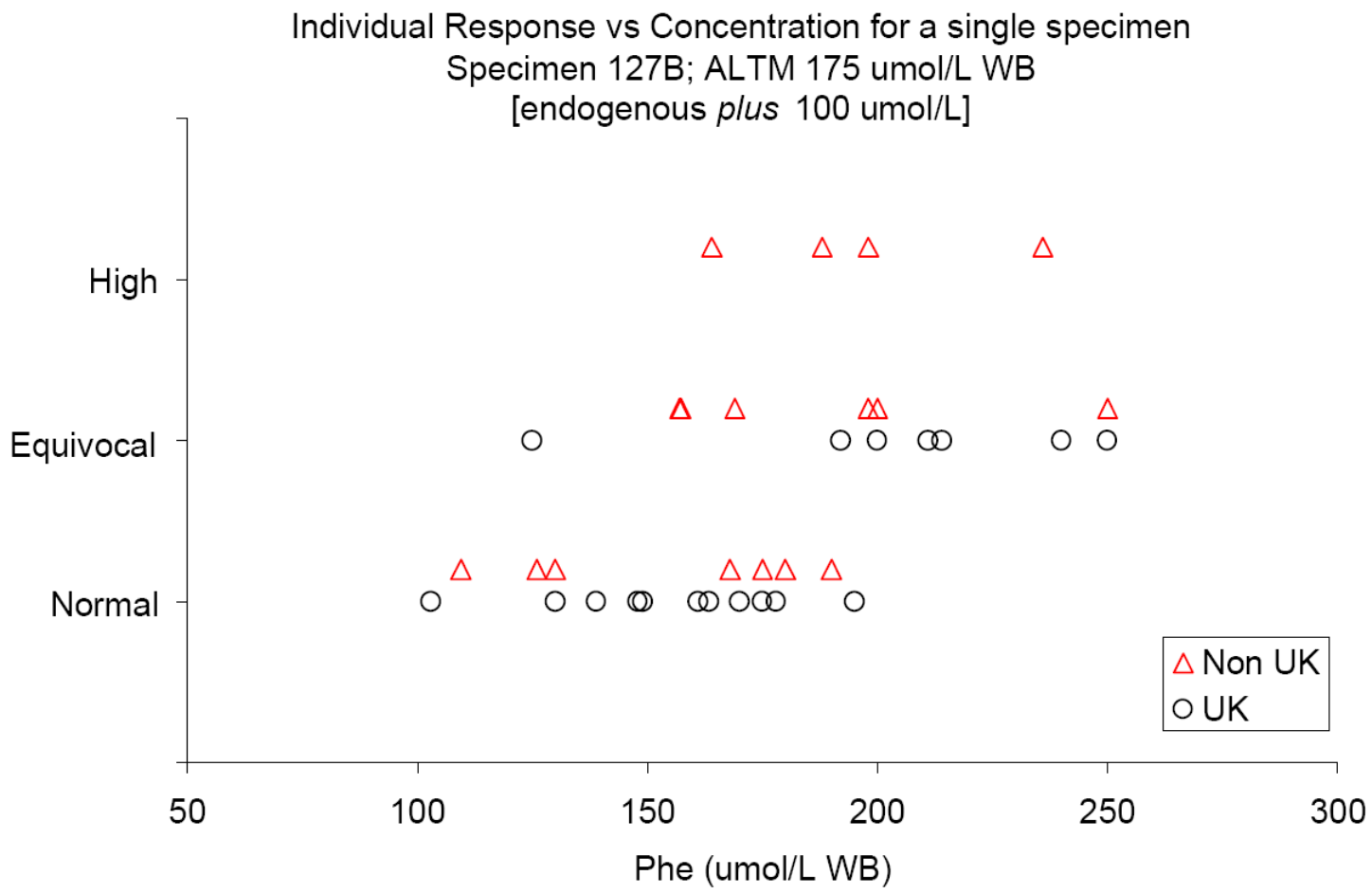


Phenylalanine Graph 2



UK NEQAS

Phenylalanine Graph 3



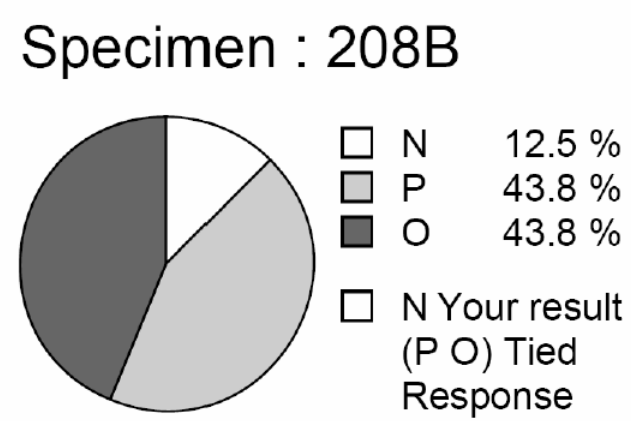
UK NEQAS

PKU Screening

I have chosen to highlight a specimen which was enriched with 240 umol/L Phenylalanine and 240 umol/L Tyrosine. Given the difficulty of measuring low levels, we cannot be sure as to what the unspiked base value was, but the ALTMs of 37 and 35 umol/L are probably not too far from the truth. The recovery of added Phenylalanine was around 95%, while closer to 90% for Tyrosine. Not quantitative, but certainly acceptable for a screening assay. Even if all labs were rather good at the analysis, we all know that if I were to dispatch a specimen containing exactly 240 umol/L Phenylalanine I would have half the labs getting above and half the labs getting below. No one is disputing this. The issue for an EQA Organiser is to try to quantify this effect in some way. For example, would it be acceptable nationally if 80% of labs got values above 200 umol/L on such a sample? Would it have to be 90%? I am trying to collect data to try to help make the judgement between practical and theoretical considerations of both analysis and interpretations. I am trying to provide data to assist in the debate, not trying to stifle debate.

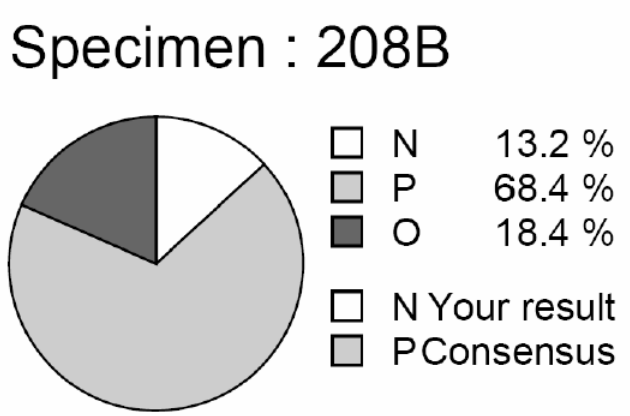
PKU Screening

British Isles subset



Your result	N
Total responses	16
N - Presumptive screen negative	2
O - Screen positive - other	7
P - Presumptive screen positive	7

Full Scheme data



Your result	N
Total responses	38
N - Presumptive screen negative	5
O - Screen positive - other	7
P - Presumptive screen positive	26

PKU Screening

Phe rank	Phe umol/L WB	Tyr umol/L WB	PKU I	Approriate response for results?	Right answer?	Approriate and Right?
1	195	182	N	Yes	No	No
2	232	230	N	Yes	No	No
3	242	243	O	Yes	Yes	Yes
4	247	234	P	Yes	No	No
5	248		O	No	Yes	No
6	254	230	O	No	Yes	No
7	255	283	P	No	No	No
8	256	231	O	No	Yes	No
9	264	214	P	Yes	No	No
10	268	226	P	Yes	No	No
11	270	239	P	Yes	No	No
12	272		P	Yes	No	No
13	274	283	P	No	No	No
14	277	285	no intepretation			
15	280	279	O	Yes	Yes	Yes
16	293	232	no intepretation			
17	320	258	O	Yes	Yes	Yes
18	328	326	O	Yes	Yes	Yes

Section 1 - Demographics

This has been sent to laboratory 10012

Please state any other Lab numbers to which the SAME answers apply.	

Unless otherwise stated, we will use the contact details below to clarify any anomalies about this questionnaire.

email(s) : philip.hyde@ulh.nhs.uk	
Name : Mr P A Hyde	
Phone : 0120 544 6339	

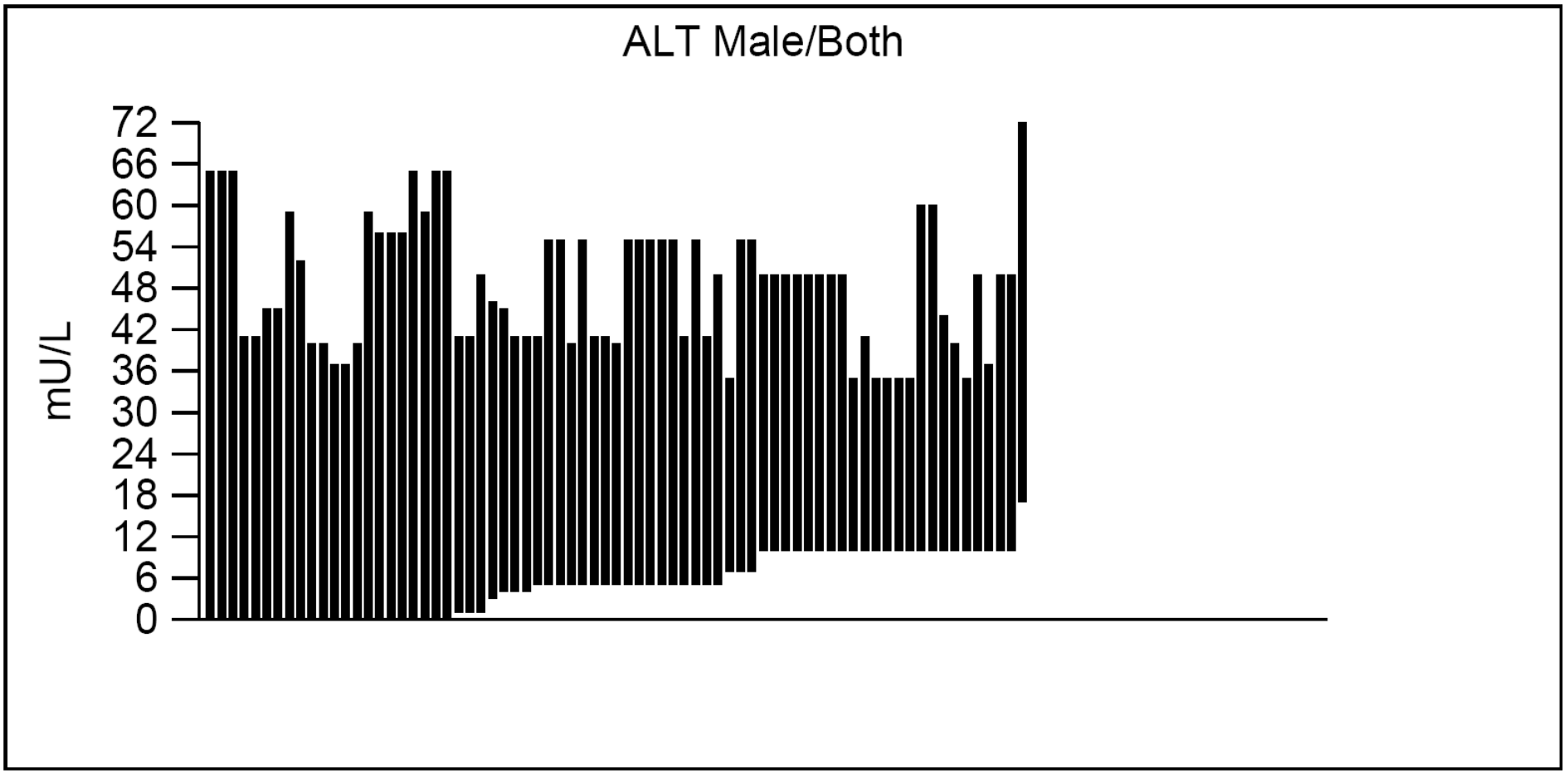
Section 2 - AST reference ranges

AST Male (or both) reference range - low		AST Male (or both) reference range - high	
AST Female (if different) reference range - low		AST Female (if different) reference range - high	
Source of range	○ In-house study ○ Kit insert ○ Literature ○ Lost in the mists of time! ◉ Reset		

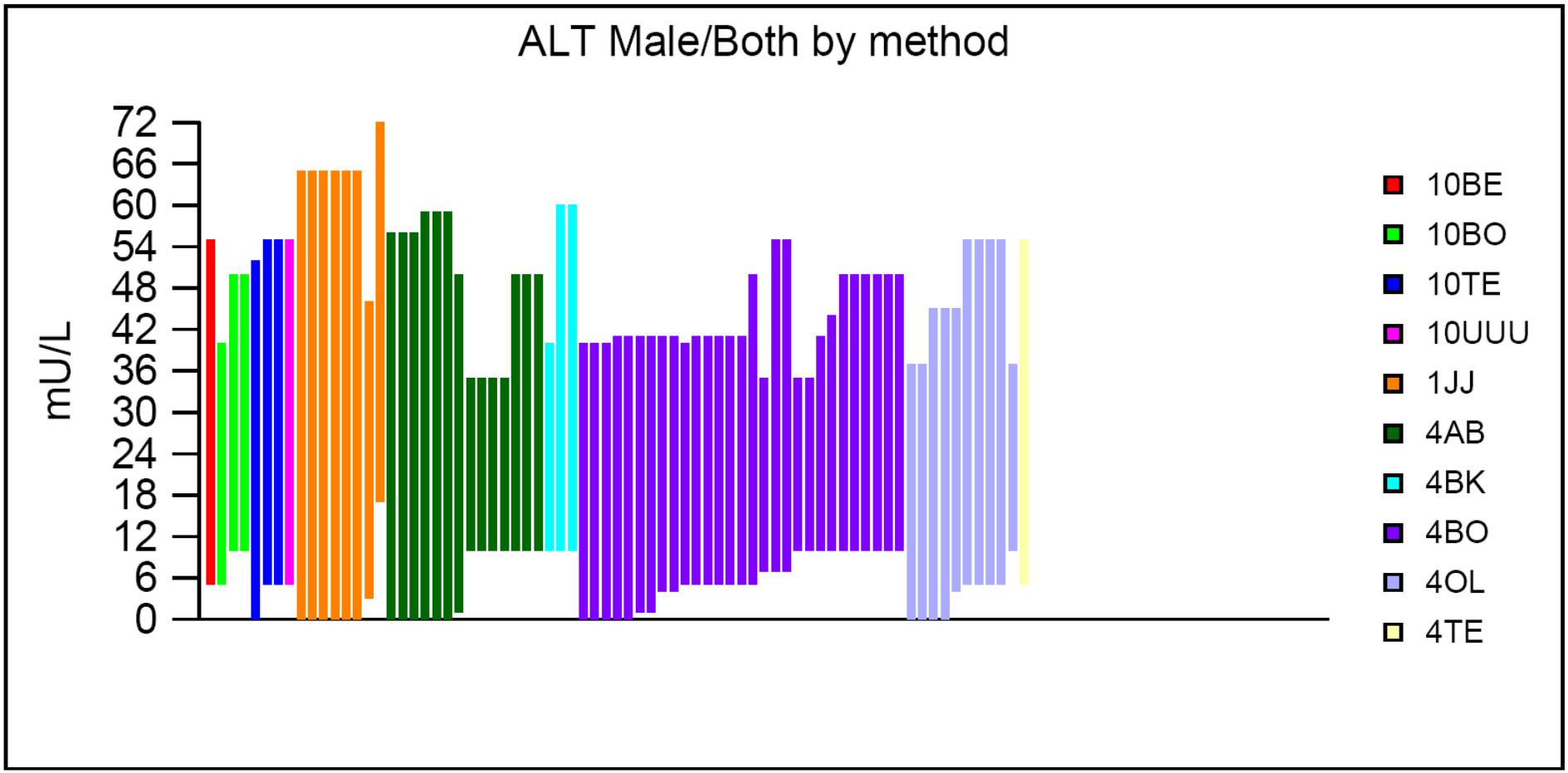
Section 3 - ALT reference ranges

ALT Male (or both) reference range - low		ALT Male (or both) reference range - high	
ALT Female (if different) reference range - low		ALT Female (if different) reference range - high	
	○ In-house study		

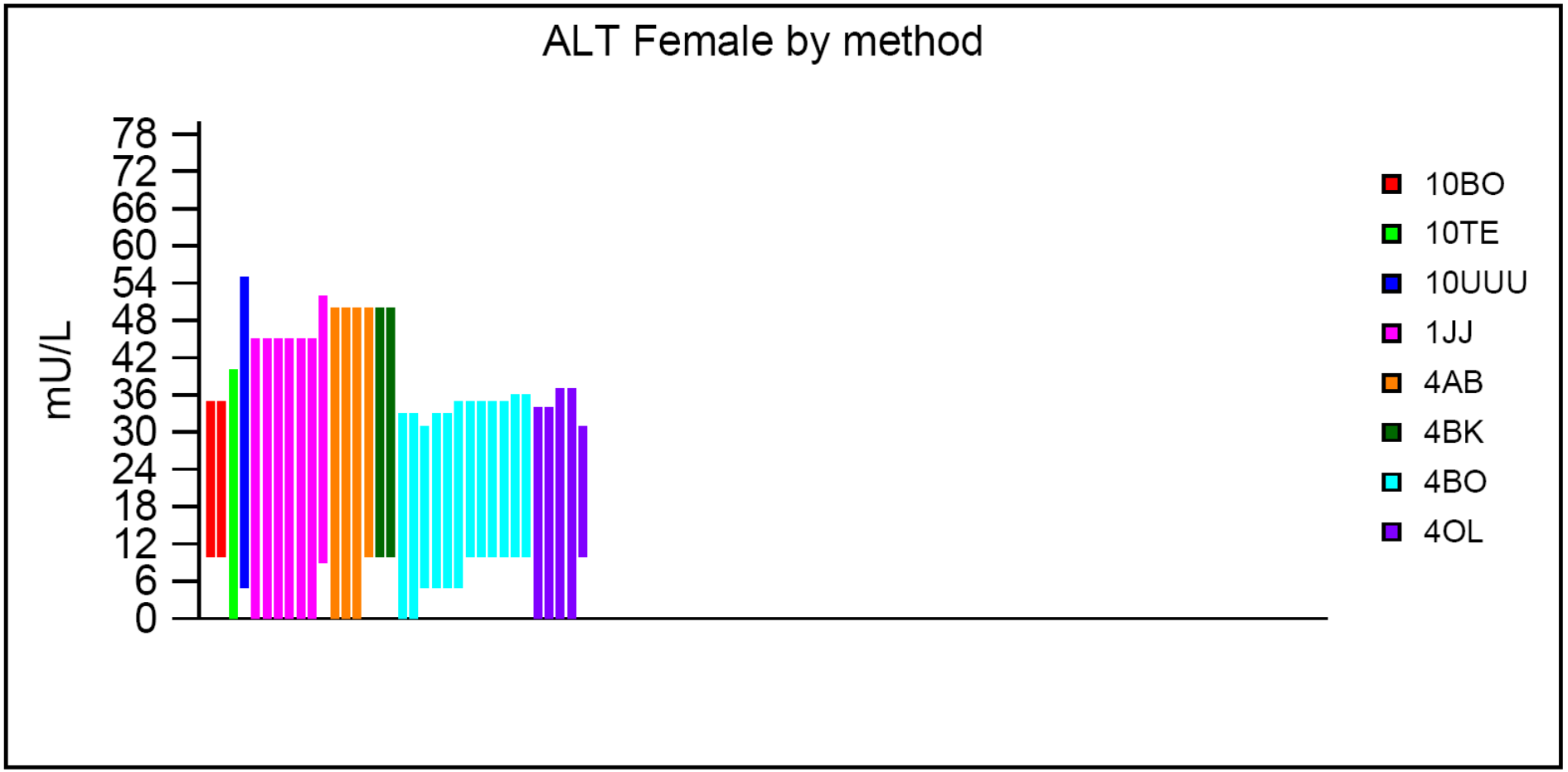
Enzymes



Enzymes


ALT Male/Both by method
mU/L
72
66
60
54
48
42
36
30
24
18
12
6
0
10BE
10BO
10TE
10UUU
1JJ
4AB
4BK
4BO
4OL
4TE

Enzymes


ALT Female by method
mU/L
78
72
66
60
54
48
42
36
30
24
18
12
6
0
10BO
10TE
10UUU
1JJ
4AB
4BK
4BO
4OL

Thyroid guidelines – are thyroid-stimulating hormone assays fit for purpose?

Geoff Beckett[1] and Finlay MacKenzie[2]

Addresses
[1]University Department of Clinical Biochemistry, The Royal Infirmary of Edinburgh, 51 Little France Crescent, Little France, Edinburgh EH16 4SA;
[2]UK NEQAS for Thyroid Hormones, Wolfson EQA Laboratory, Institute of Research and Development, Birmingham Research Park, Vincent Drive, Edgbaston Birmingham B15 2SQ, UK

Correspondence
Dr Geoff Beckett
Email: g.j.beckett@ed.ac.uk

Ann Clin Biochem 2007;
44 203 - 208
and editorial

Abstract

Most thyroid-stimulating hormone (TSH) assays now have the sensitivity required by thyroid guidelines and allow the reliable identification of patients with both overt and subclinical hyperthyroidism. Clinical guidelines usually quote decision limits for TSH, but often ignore the issue of whether variability in bias between assays should be considered when such decision limits are implemented. Clinicians and laboratories should appreciate that these decision limits arise largely from historical data that used TSH assays with poorly defined bias. It is thus unlikely that laboratories will be able to apply an appropriate method-related bias adjustment to these TSH cut-offs. Clinicians should appreciate that TSH decision limits should thus be regarded as typical target figures rather than an absolute cut-off and thus can be applied with some degree of flexibility. There is currently insufficient evidence to justify a significant lowering of the upper reference limit for TSH, but fine-tuning of current reference ranges is required since there appears to be no association between the ranking of the assay bias in the UK National External Quality Assessment Service scheme and the manufacturers' quoted reference ranges. There is room for further improvement in TSH assays and this can best be achieved if manufacturers, laboratories and clinicians work together to produce TSH assays and reference ranges that show closer agreement between methods. Until this is achieved, future studies that examine the relationship of TSH with symptoms and treatment should ensure that sufficient information is included in the publication to allow the method related bias of the TSH assay to be clearly described.

Fit for Purpose?

Subclinical hyperthyroidism

Some suggest the use of TSH as an aid to guiding the treatment of endogenous subclinical hyperthyroidism. Thus a Consensus Development Conference on Subclinical Thyroid Disease held in 2002 in the USA and sponsored by the American Association of Clinical Endocrinologists, The American Thyroid Association and The Endocrine Society, suggested the following:

- *Patients with a TSH of 0.1–0. 45 mU/L* should have the measurement repeated for confirmation. If on repeat testing the TSH remains within 0.1–0.45 mU/L and the patient has no signs or symptoms of cardiac disease or arrhythmia, thyroid function tests should be repeated at 3–12 month intervals. In contrast, the group suggested that
- *individuals with a TSH <0.1 mU/L* confirmed by repeat should be considered for treatment if Graves' disease or multinodular goitre was diagnosed.[3,15,16]

TSH ~ differences in numerical values between methods

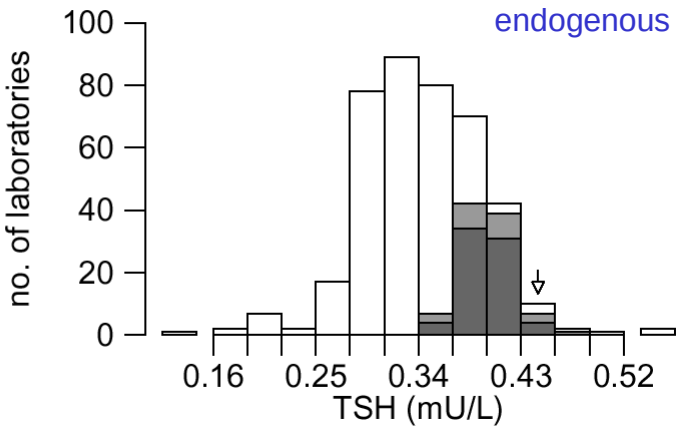
method	low	high	measured TSH on Specimen 292A	Decreasing method bias%
Roche Elecsys	0.27	4.20	*1.07*	
DPC Immulite 2000 Rapid L2KRT	0.40	4.00	*1.01*	
Tosoh AIA	0.34	3.80	*0.98*	
ALTM			***0.96***	
Bayer Advia:Centaur	0.35	5.50	*0.95*	
Beckman Access	0.35	3.50	*0.94*	
DPC Immulite 2000 3rd Gen L2KTS	0.40	4.00	*0.88*	
Abbott AxSym	0.49	4.67	*0.85*	
Abbott Architect	0.35	4.94	*0.79*	
Ortho Vitros ECi TSH-30*	0.30	3.05	*0.62*	

Fit for Purpose?

TSH in the 0.1 to 0.45 mU/L range [endogenous]

Specimen : 282B

	n	Mean	SD	CV(%)
All methods	394	0.350	0.050	14.2
Abbott Architect	57	0.292	0.017	5.8
Abbott AxSym	11	0.314	0.030	9.4
Bayer Advia:Centaur	108	0.329	0.025	7.6
Beckman Access	22	0.344	0.016	4.6
DPC Immulite 2000/2500	66	0.362	0.033	9.1
Ortho Vitros ECi TSH-30*	8	0.204	0.019	9.5
Roche Elecsys	96	0.40	0.02	5.2
Tosoh AIA	8	0.358	0.017	4.9



Again, the repeat interval quoted as 3 to 12 months

Free T4 ~ differences in numerical values between methods

method	low	high	measured Free T4 on Specimen 292A	decreasing method bias%
DPC Immulite 2000	10.3	24.5	17.1	
Roche Elecsys	12.0	22.0	16.1	
Tosoh AIA	9.0	21.9	15.8	
ALTM			15.0	
Bayer Advia:Centaur	11.5	22.7	14.6	
Ortho Vitros ECi			14.1	
Abbott Architect	9.0	19.1	13.6	
Abbott AxSym	9.1	23.8	12.6	
Beckman Access	7.7	14.2	11.9	

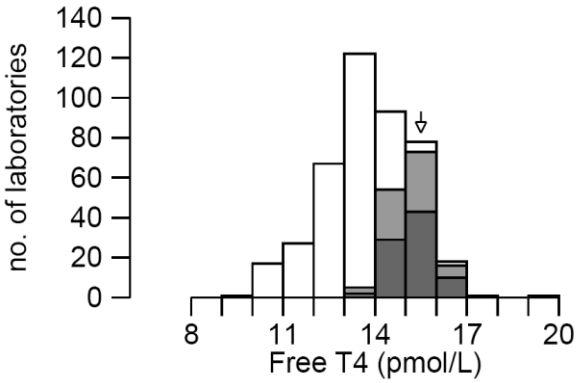
Table F4.4.1 Manufacturers' Quoted Reference Intervals for Free T4

The methods have been ranked in decreasing bias order for results on a single representative euthyroid specimen. The method giving the highest numerical result is listed at the top, the method giving the lowest numerical value is at the bottom.

Free T4

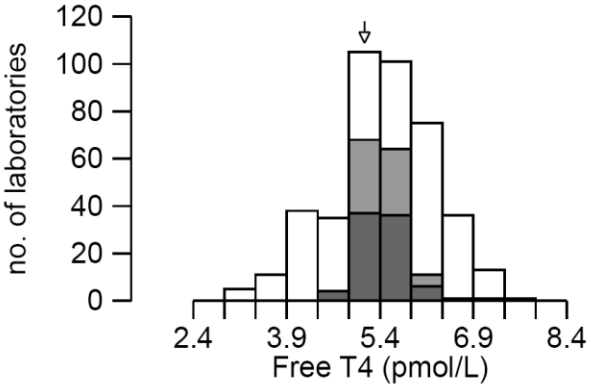
Free T4 Example of a typical Euthyroid Pool

Specimen : 359C	n	Mean	SD	CV(%)
All methods [ALTM]	425	13.91	1.43	10.3
Abbott Architect	116	13.40	0.58	4.3
Beckman Access/Dxi	35	11.10	0.78	7.0
OCD (J&J) VITROS	9	14.69	1.13	7.7
Roche Elecsys	149	15.25	0.62	4.1
Siemens Immulite 2000/2500	26	13.28	0.68	5.1
Siemens ADVIA Centaur	82	13.38	1.02	7.6

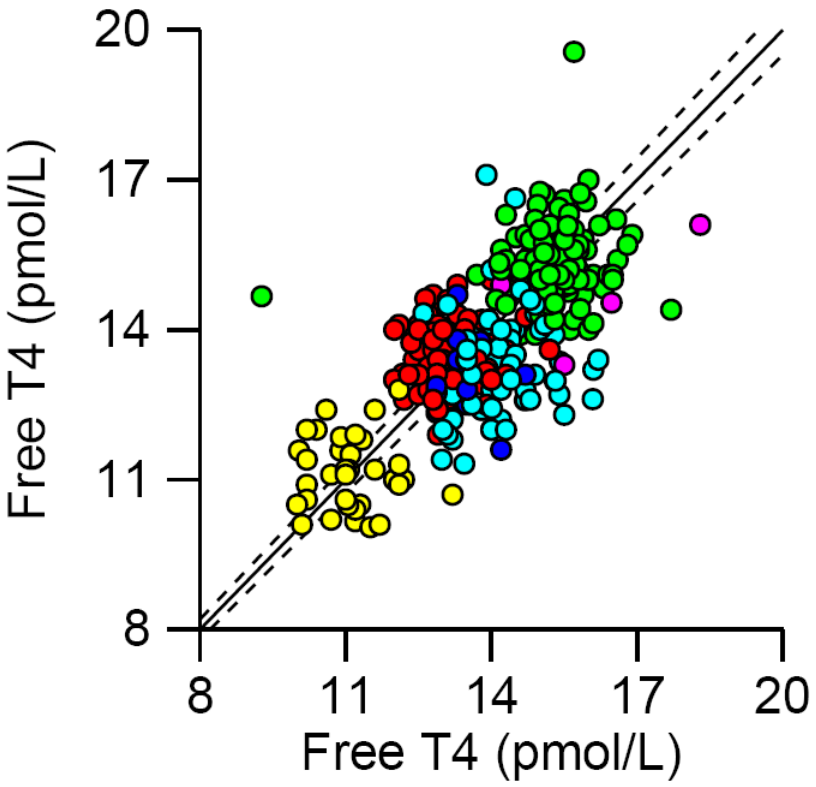


Free T4 Example of a low-level (manipulated) Pool

Specimen : 359E	n	Mean	SD	CV(%)
All methods [ALTM]	420	5.49	0.81	14.8
Abbott Architect	114	6.26	0.51	8.1
Beckman Access/Dxi	35	5.97	0.61	10.2
OCD (J&J) VITROS	9	4.50	0.23	5.2
Roche Elecsys	149	5.46	0.32	5.9
Siemens Immulite 2000/2500	26	4.87	0.51	10.4
Siemens ADVIA Centaur	80	4.53	0.69	15.2



Free T4 ~ 14 pmol/L



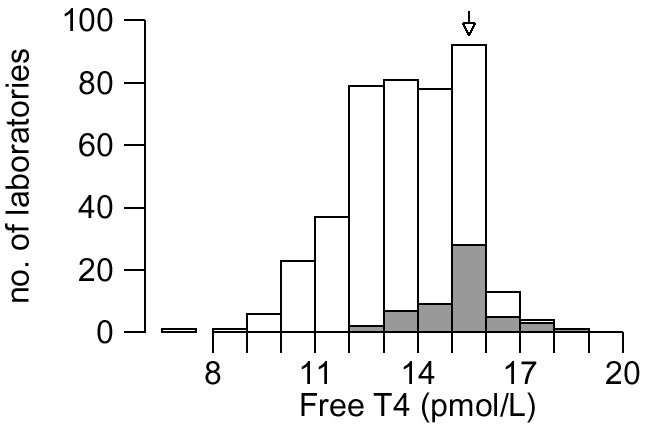
- Abbott Architect
- OCD (J&J) VITROS
- Roche Cobas/Modular
- Siemens ADVIA Centaur
- Siemens Immulite 2000/2500
- Beckman Access/Dxi

845

Free T4 results on a Euthyroid Pool

Specimen : 314D

	n	Mean	SD	CV(%)
All methods	416	13.75	1.76	12.8
Abbott Architect	97	12.54	0.81	6.4
Bayer Advia:Centaur	104	13.63	0.92	6.8
Beckman Access	28	10.38	0.62	6.0
DPC Immulite 2000/2500	55	15.24	0.99	6.5
Ortho Vitros ECi	7	11.67	0.67	5.7
Roche Elecsys	110	15.10	0.71	4.7
Tosoh AIA	6	14.52	1.84	12.7

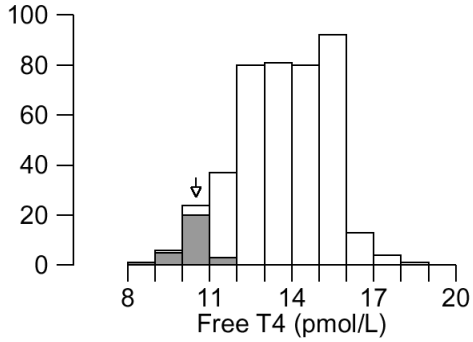


Note that the x-axis is close to some laboratories' own Reference Intervals

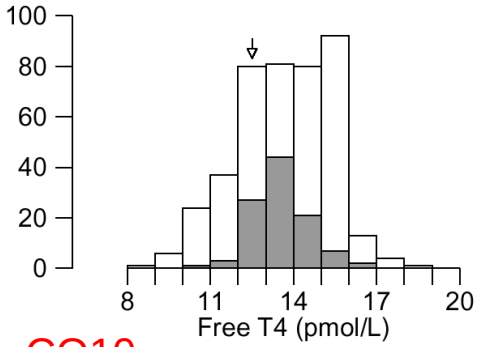
The shape of the overall distribution is a consequence of the overlap of a number of different 'normal' distributions; the relative size of each method data set gives rise to the final shape

Users of a single method can agree between themselves, but there are large between-method differences

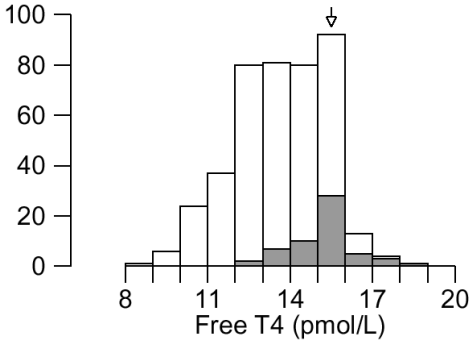
Free T4 results on a Euthyroid Pool



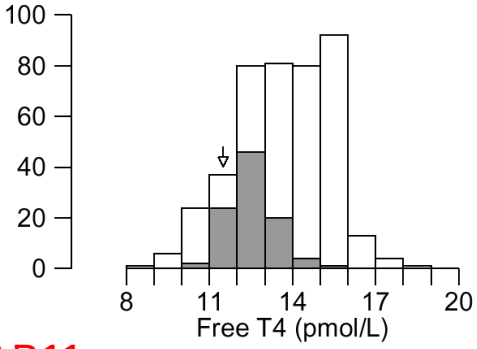
SF1



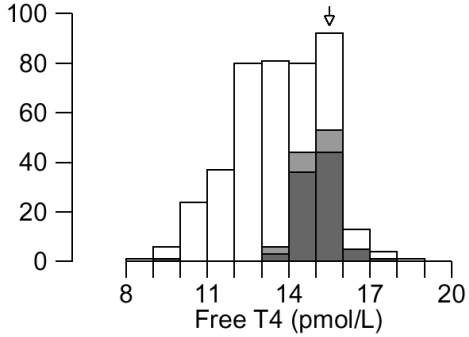
CO10



DC11



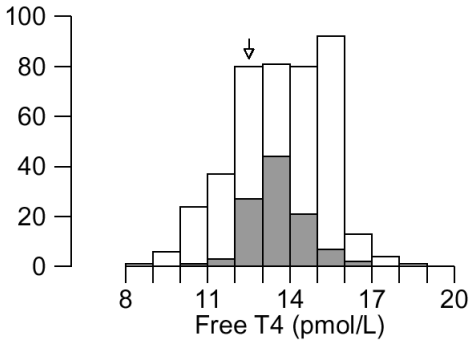
AB11



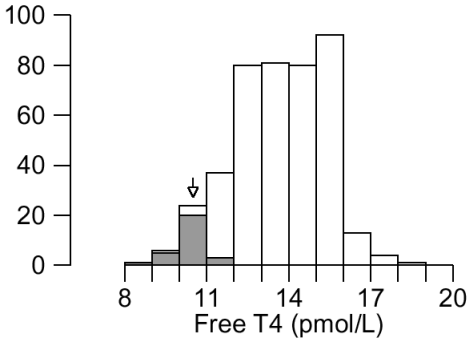
BO5

The shape of the overall distribution is a consequence of the overlap of a number of different 'normal' distributions; the relative size of each method data set gives rise to the final shape

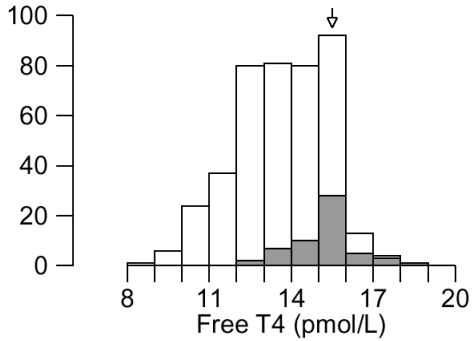
You don't have to be far from the target to get a Red Double Arrow



Your result	12.9
Target (MLTM)	13.62
Your specimen:	
%bias	-5.3
transformed bias	-106
Accuracy Index	106
Your method mean	13.62
Bayer Advia:Centaur	



Your result	10.7
Target (MLTM)	10.40
Your specimen:	
%bias	+2.9
transformed bias	+52
Accuracy Index	52
Your method mean	10.40
Beckman Access	



Your result	15.3
Target (MLTM)	15.24
Your specimen:	
%bias	+0.4
transformed bias	+10
Accuracy Index	10
Your method mean	15.24
DPC Immulite 2000/2500	