Performance Monitoring in Newborn Screening - a Co-ordinated National Approach

Kate John, Jim Bonham, Christine Cavanagh and Rachel Carling

Introduction

In January 2015, the NHS newborn blood spot screening programme was expanded to include four additional disorders. Six conditions are now screened for by MSMS; PKU, MCADD, GA1, MSUD, IVA and HCU. Nationally agreed screening protocols were adopted with specified analytical and clinical cut-off values (COV) so harmonisation between labs was important. To assist with performance monitoring and provide assurance of the efficacy of the screening programme, the national co-ordinating centre commissioned a team to collect and analyse population data and results from common IQC material from 14 screening labs in England and Wales.

Population Data

Labs returned data monthly by instrument on 8 analytes. Results above the analytical cut off and from babies not aged between day 5 and 8 were excluded. The 10th, 50th, 90th and 99th centile was calculated for each analyte and plotted monthly and cumulatively. The graphs below show 1 years worth of data (>600,000 babies) by laboratory except tyrosine (>410,000) as not all labs submit tyrosine data.

Results

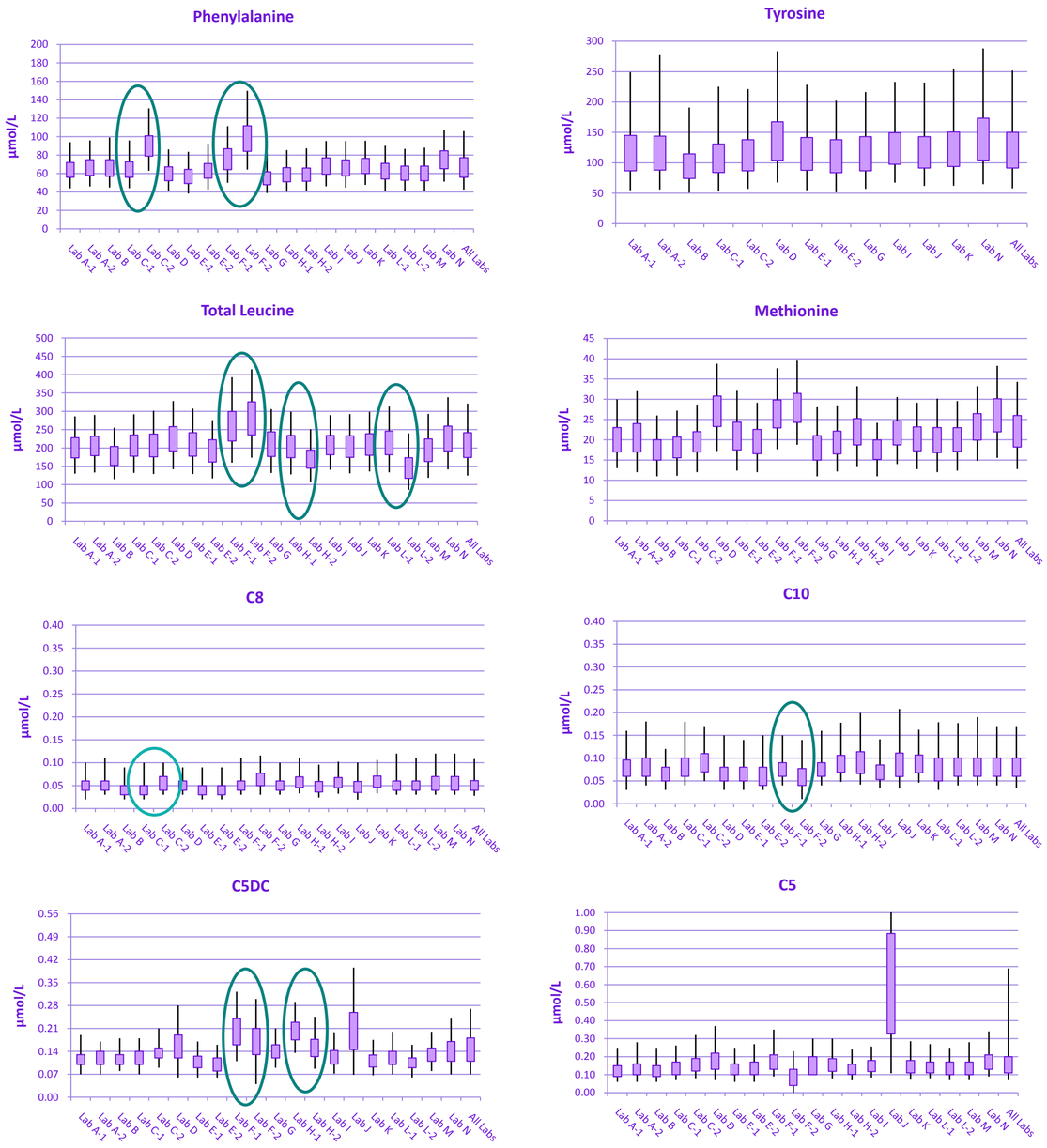

Figure 1: Population Data
(highlighted results show instruments within same lab with different population data)

Generally most analytes are acceptable with the 90th centile well removed from the COV, however the data does highlight a number of issues.

Phenylalanine and Leucine: Would expect these analytes to show less variation considering the concentrations being analysed, however, the data indicates some differences between the labs and also between instruments within the same lab.

Methionine: Considerable variation and concentrations approaching COV, however, 2nd tier testing for homocysteine is part of the protocol

Tyrosine: The difference between the 90th and 99th centile likely to reflect biological variation rather than analytical.

C5: Persistent gross error with one lab, yet still continuing to screen.

C5DC: Wide variability even within the same lab

C8 and C10: Variability much greater for C10 than C8 and the bias is not in sync which has implications for referral based on the C8/C10 ratio.

QC data

3 levels of commercial QC material was supplied to each lab for a period of one year and analysed once a day in singleton. Data was produced cumulatively by analyte. Results are shown for QC1 (nearest to the cut off values)

Results

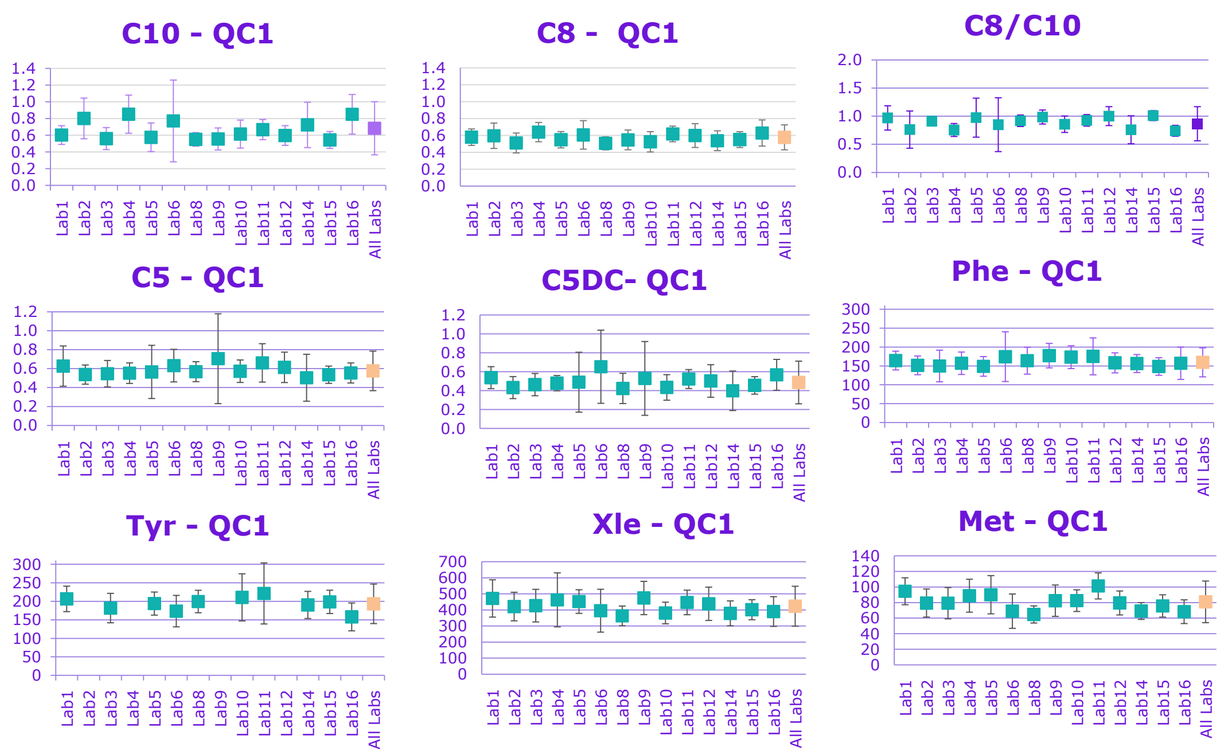

Figure 2: QC Data
(bars indicate +/- 2 SD)

Commercial QC would be expected to produce relatively precise results, however this is not apparent for a number of analytes. The QC was used to calculate the MU for the national programme as babies born in England and Wales are screened against a common cut off value.

Horwitz ratio indicates the between lab variation (MU) is greater than the predicted CV and the performance for phenylalanine and total leucine is unacceptable

Analyte	Clinical COV	Mean 90th Centile µmol/L	COV as Multiple of 90th centile	90th Centile Range (%diff)	MU	Horovitz Predicted between lab CV / Ratio
Phe	240	74	2.7	64-88 (27%)	158 ± 26 (24%)	8.8% / 1.23
Met	50	29	1.6	21-32 (34%)	80 ± 20 (25%)	10.9% / 1.0
Leu	600	244	2.05	203-279 (27%)	409 ± 94 (23%)	9.7% / 1.3
C5	2.0	0.19	8.4	0.16-0.24 (45%)	2.32 ± 0.56 (24%)	17.3% / 0.69
C8	0.5	0.07	5.7	0.06-0.07 (17%)	0.57 ± 0.10 (18%)	20.8% / 0.43
C5DC	0.7	0.15	3.7	0.12-0.22 (45%)	0.46 ± 0.14 (30%)	21.5% / 0.69
C10	-	0.12	-	0.09-0.15 (50%)	0.68 ± 0.3 (43%)	
Tyr	-	150	-	109-166 (38%)	193 ± 53 (27%)	

Discussion

Possible causes for the variation seen could be due to the differences in the instruments, setup parameters, mobile phases, internal standard. There is no certified reference material to ascertain what the actual concentrations should be.

Although the programme is not unsafe, there is the potential for unnecessary referrals due to the variability across the country, which has a detrimental effect on the parents. Considering the MU there is also the potential to miss a referral.

Is enough being done to troubleshoot the differences or to harmonize the results considering we are using a national cut off?

Should we consider setting population based COV within each lab?