The Assignment of Values to Dried Blood Spot Quality Control Samples

J.Gerard Loeber, Joanne V. Mei, Victor R. De Jesús

APHL-8th ISNS meeting, Atlanta, 070513

An anecdote

- Once upon a time, a few centuries ago, there was a small village, with a church, a pub, and a few shops including a watchmaker's shop.
- Every morning at 7 am the church's sexton rang the church bells to wake up the villagers. The watchmaker heard the chuch bells and adjusted the large clock in his shop window.

- The sexton on his daily morning round through the village walked along the watchmaker's shop and compared the church clock with the watchmaker's clock, went back to the church and adjusted the church clock.

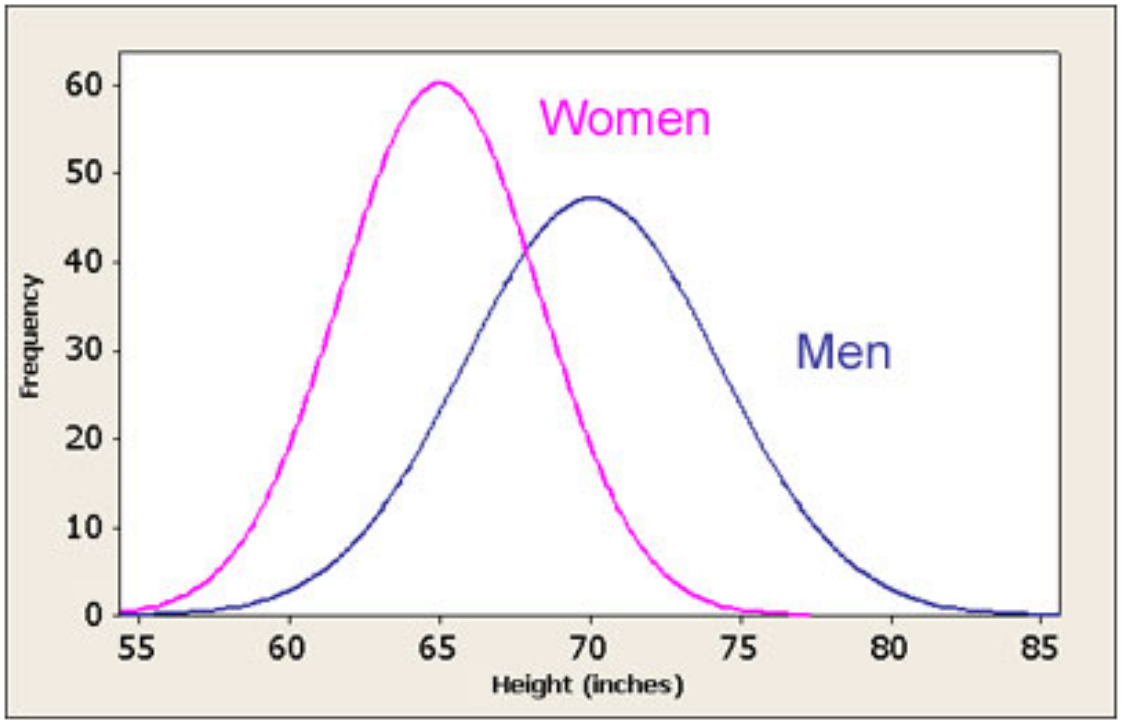
- After a few months a traveler came along and wanted to have lunch in the local pub, which was open between 12 noon and 1 pm.
- When he knocked at the door the owner explained that it was already 1.15 and that the pub was closed.
- The traveler replied that that was wrong because it was in fact only 12.05 and showed the other man his Swiss watch.

- An argument followed and along came the sexton and the watchmaker to help solve the issue.
- After a while the traveler was able to help the villagers understand what they did wrong.

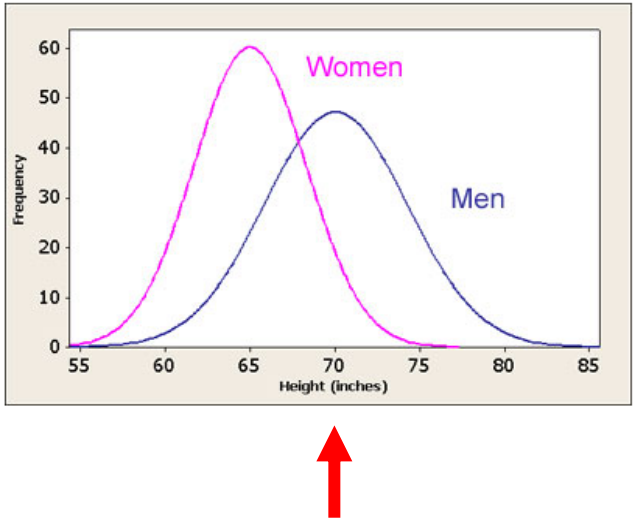
Objective 1

- To compare machines, and to calibrate them, we need an independent calibrator

Screening for gender via body length



- Not an ideal screening method, however, that's NOT the issue
- It's about the unit height, i.e. inches



- This approach is suitable for within one jurisdiction (e.g. USA) where everyone automatically knows that height is measured in *inches*
- However, if an American talks about the mean male length of "70" then most Europeans will not understand this number, because they use *centimeters*

Objective 2

- To compare numerical methods we need to use the same units.
- SI units have been available since the ’60s …

The Use of QC-samples

- In neonatal screening we use QC-samples for internal QC, and for external QC.

- For internal QC, bias is not essential. Whatever the numerical results, the screening system functions if we can discriminate between the healthy and the potentially ill populations. Yet, as scientists we want to have minimum bias.

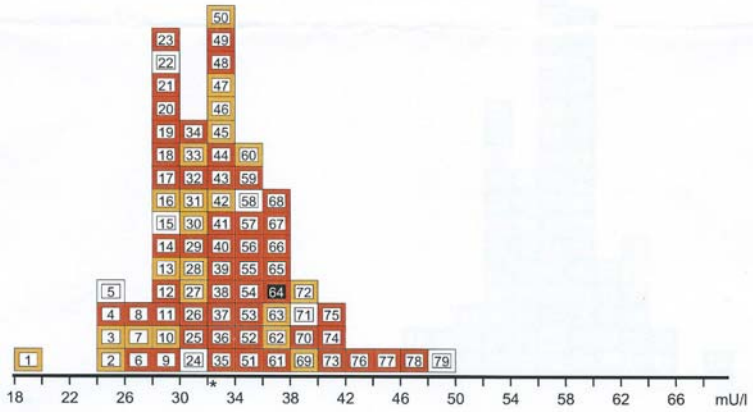
- However, for external QA (important in view of e.g. regulation (CLIA)) a minimum bias is essential. Results of EQA surveys are scrutinised and we are unhappy if we are not in the "middle" of the participants

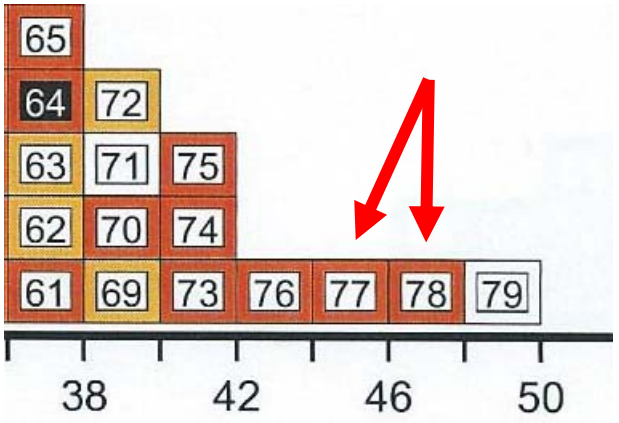
Results RfB (Germany), Survey October 2012 TSH

Findings TS4/12 -October 2012-
TSH - Screening, Sample 1
Participant No 9900639

RfB Referenzinstitut für Bioanalytik

Legend:
N = number of results
XM = mean value
SD = standard deviation
CV = coefficient of variation
M = method number
rn = result box/result code no.
rn rn = normal / C.H. possible
rn rn = C.H. evident / other
= 16. bis 84. percentile with median
★ = marks target value
C.H. = congenital hypothyroidism
rn = marks your result box





Descriptive Statistics

N = 74
XM = 33,0 mU/l
SD = 5,2 mU/l
CV = 15,8 %

Statistics of classifications

normal	
C.H. possible	22
C.H. evident	50
other	2

Statistics of Implications, related to Classifications

	none	new card	plasma sample	other
normal		18	2	1
C.H. possible		2	42	4
C.H. evident		2		
other				

Kit Evaluation for Sample 1 (Target value = 32,4 mU/l)

M	Kit	N	Min	16.P	50.P	84.P	Max	Kit Classifications			
All	Kits	74	18.1	28.5	**32.4**	37.8	48.7	0	22	50	2
2	2	2	25.0		**33.5**		41.9	0	0	2	0
3	91	55	18.1	28.4	**32.4**	37.0	41.4	0	18	40	0
3	99	2	35.2		**36.9**		38.6	0	0	0	0
4	77	10	25.5	28.1	**31.2**	35.3	39.7	0	4	3	2
4	111	2	43.8		**45.3**		46.9	0	0	2	0

Other Kits: 2 99(1),3 72(1),4 48(1),

The deviation of your result from the total median M and from the median of the corresponding sub-collective (kit) Mu is: M 14 %, Mu 14 %

2	2	2	25.0		**33.5**		41.9
3	91	55	18.1	28.4	**32.4**	37.0	41.4
3	99	2	35.2		**36.9**		38.6
4	77	10	25.5	28.1	**31.2**	35.3	39.7
4	111	2	43.8		**45.3**		46.9

International Society for Neonatal Screening

- Participants 77 and 78 may want to speak to their vendor and ask why his kit measures higher than the mean of the rest.

- The vendor, afraid of losing his customers, may change the calibration of his kit to get better in line with the rest of the participants.

- If the majority of the participants use one vendor, this means that the calibration of that kit determines the measuring level of all participants. That may lead to an unwanted deviation from the “true” level thus increasing the bias!

Objective 3

- EQA schemes should not use the overall mean as the “target”, but use materials that have been calibrated against an independent reference preparation to the extent possible.

The use of QC-samples with “known” values

- CDC provides QC-samples, not for daily routine but for regular external QC
- The samples have been enriched with known amount of analyte and are offered with a certificate

- Let’s focus on TSH

Newborn Screening
Quality Assurance Program

Quality Control Hormones Specimen Certification
Set 1 – January 7, 2013

ENRICHMENT LEVELS (endogenous levels not included)

Analyte	*Lot*	*Low*	*Lot*	*Intermediate*	*Lot*	*High*
Thyroxine (T_4 µg/dL serum)	101	2	102	7	103	11
Thyroid-Stimulating Hormone (TSH µIU/mL serum)	211	25	212	40	213	80
17 α-Hydroxyprogesterone (17-OHP ng/mL serum)	251	25	252	50	253	100

ANALYTICAL INFORMATION

Analyte	*Lot*	*Mean/ 95% CL*	*Lot*	*Mean/ 95% CL*	*Lot*	*Mean/ 95% CL*
T_4	101	$\bar{x}$ = 1.8 CL = 1.1 – 2.6	102	$\bar{x}$ = 7.1 CL = 5.5 – 8.7	103	$\bar{x}$ = 11.2 CL = 8.7 – 13.6
TSH	211	$\bar{x}$ = 24.3 CL = 19.2 - 29.5	212	$\bar{x}$ = 39.2 CL = 31.0 - 47.3	213	$\bar{x}$ = 69.3 CL = 56.6 - 81.9
17-OHP	251	$\bar{x}$ = 25.1 CL = 17.0 - 33.2	252	$\bar{x}$ = 48.7 CL = 35.1 – 62.3	253	$\bar{x}$ = 109.2 CL = 79.9 – 138.5

***Analysis Method: MSMS Derivatized - MS/MS non-kit**

Note: The values provided in the above tables are for reference use only. The mean value and confidence limits (CL) are determined by CDC for each Quality Control (QC) lot. Each participating laboratory must establish its own mean values and CL for its test method with these QC materials. Temporary estimates of mean values and CL can be determined after 10 successive, independent measurements. *Sunyk WE, Hinson WH. Quality Assurance in the newborn screening laboratory. In: Therrell BL Jr, editor. Laboratory methods for neonatal screening. Washington (DC): American Public Health Association, 1993:23-46*

International Society for Neonatal Screening

ENRICHMENT LEVELS (endogenous levels not included)

Analyte	*Lot*	*Low*	*Lot*	*Intermediate*	*Lot*	*High*
Thyroid-Stimulating Hormone (TSH, µIU/mL serum)	211	25	212	40	213	80

ANALYTICAL INFORMATION

Analyte	*Lot*	*Mean/ 95% CL*	*Lot*	*Mean/ 95% CL*	*Lot*	*Mean/ 95% CL*
TSH	211	$\bar{x}$ = 24.3 CL = 19.2 - 29.5	212	$\bar{x}$ = 39.2 CL = 31.0 - 47.3	213	$\bar{x}$ = 69.3 CL = 56.6 - 81.9

Note: The values provided in the above tables are for reference use only. The mean value and confidence limits (CL) are determined by CDC for each Quality Control (QC) lot. Each participating laboratory must establish its own mean values and CL for its test method with these QC materials. Temporary estimates of mean values and CL can be

- The CDC- analytical results have been obtained by using one method (PerkinElmer)

- Although CDC states that all users should establish their own mean values, many users believe that the CDC characterization values must be the “right” values.

- If they deviate too much they tend to believe that their assay system may be at fault.

- However, if the commercial kit used by CDC by chance had not been calibrated correctly then the user’s kit may well have been correct in the first place.

Objective 4

- CDC gives a mean value and the 95% upper and lower confidence limits (CL)
- Values are determined based on 20 independent runs, in duplicate
- The values provided are for **reference use only**
- Each participating laboratory must establish its own mean values and CL for its test method with these QC materials
- Values may differ by method

Performance of commercial kits

- By measuring QC-samples over longer periods of time information becomes available about the analytical performance of commercial kits
- CDC provides data for the analytes in its QC program for all reported methods
- Example: TSH method performance summary data for 2012

2012 Quality Control Data
Summaries of Statistical Analyses

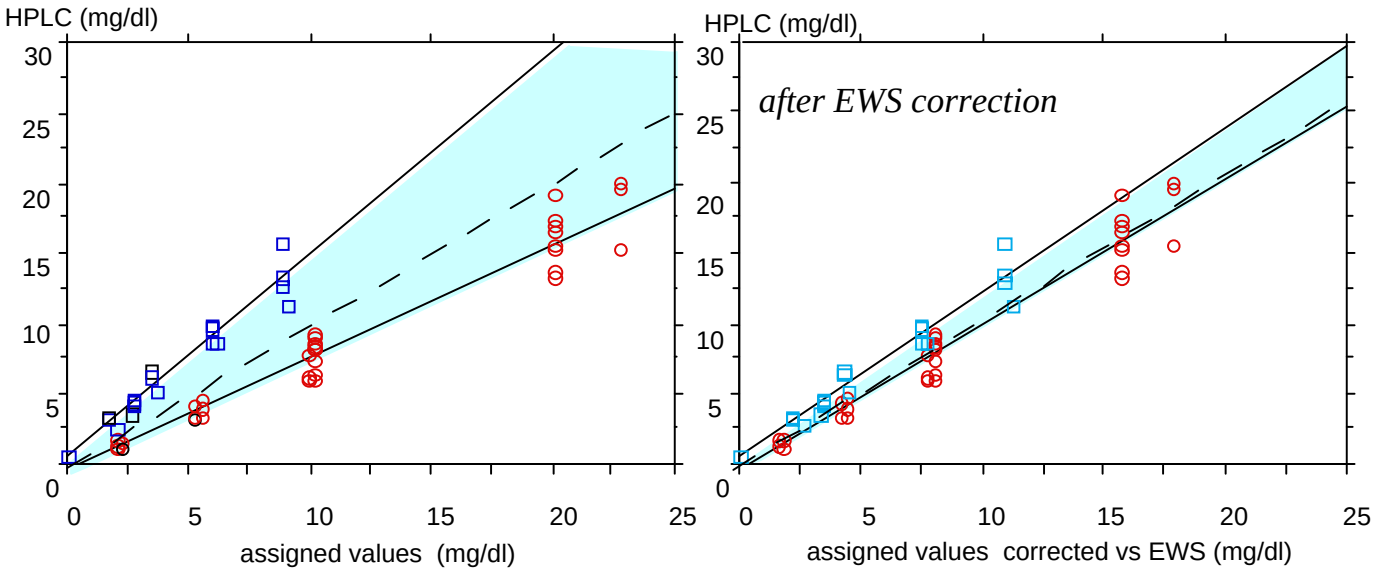
THYROID-STIMULATING HORMONE (µIU TSH/mL serum)

METHOD	N	Mean	Average Within Lab SD	Total SD	Y-Intercept*	Slope
Lot 111 - Enriched 25 µIU/mL serum						
Siemens Healthcare Diagnostics	28	29.4	3.1	3.4	-3.6	1.3
Neo-Genesis Accuwell	48	24.3	2.3	2.3	-4.0	1.1
Delfia	472	25.7	2.4	3.9	-2.1	1.1
AutoDelfia	830	26.4	1.9	2.9	-1.5	1.1
Ani Labsystems	74	25.7	3.3	3.6	-2.5	1.1
Bio-Rad Quantase	110	31.5	3.0	11.1	3.7	1.1
TecnoSuma UMELISA	20	28.6	3.4	6.0	-7.5	1.4
Bioclone ELISA	30	33.8	4.2	12.2	6.8	1.1
DiaSorin	98	27.6	1.9	3.0	1.2	1.0
PerkinElmer GSP Neonatal	133	24.7	2.6	2.9	-5.4	1.2
In House	70	25.3	2.3	4.3	-1.9	1.1
Lot 112 - Enriched 40 µIU/mL serum						
Siemens Healthcare Diagnostics	26	48.5	4.0	5.4	-3.6	1.3
Neo-Genesis Accuwell	50	41.1	3.7	4.7	-4.0	1.1
Delfia	479	40.3	4.1	6.3	-2.1	1.1
AutoDelfia	833	42.2	3.4	4.9	-1.5	1.1
Ani Labsystems	77	42.8	4.0	6.0	-2.5	1.1
Bio-Rad Quantase	109	48.8	3.9	8.9	3.7	1.1
TecnoSuma UMELISA	18	49.9	4.1	4.1	-7.5	1.4
Bioclone ELISA	30	52.2	6.1	23.2	6.8	1.1
DiaSorin	99	41.4	3.7	7.2	1.2	1.0
PerkinElmer GSP Neonatal	134	39.5	3.2	3.7	-5.4	1.2
In House	69	40.7	3.8	6.8	-1.9	1.1
Lot 113 - Enriched 80 µIU/mL serum						
Siemens Healthcare Diagnostics	29	101.3	3.6	14.0	-3.6	1.3
Neo-Genesis Accuwell	48	86.3	7.8	10.2	-4.0	1.1
Delfia	468	84.9	6.9	10.4	-2.1	1.1
AutoDelfia	837	86.9	6.0	8.7	-1.5	1.1
Ani Labsystems	74	87.9	10.2	18.9	-2.5	1.1
Bio-Rad Quantase	106	93.4	12.2	19.4	3.7	1.1
TecnoSuma UMELISA	20	107.7	16.5	22.0	-7.5	1.4
Bioclone ELISA	30	95.2	7.5	18.4	6.8	1.1
DiaSorin	96	83.8	6.2	8.8	1.2	1.0
PerkinElmer GSP Neonatal	133	87.7	6.2	8.0	-5.4	1.2
In House	70	84.3	6.6	13.5	-1.9	1.1

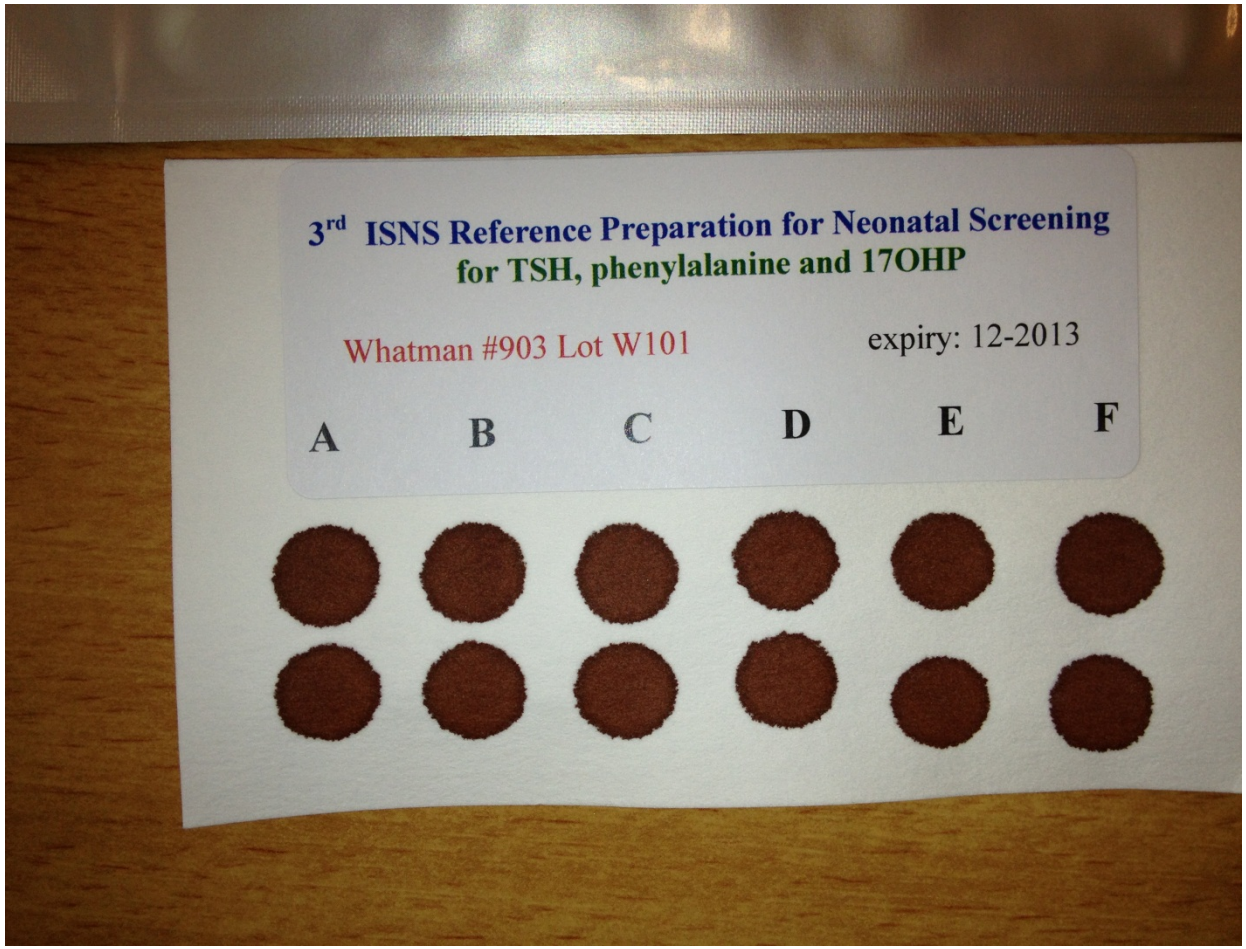
Objective 5

- Whenever possible, commercial kits should be calibrated against an independent reference preparation or using an independent method

- This can be done! See Dhondt et al (1998) Preparation of the first European Working Standard for Phenylalanine determination in dried blood spots J.Med.Screening 5, 63-66



ISNS reference material is available



peter.schielen@rivm.nl

Conclusions

- It is in a lab's best interest to be able compare their results with colleagues anywhere in the world
- Harmonization of results is an important goal for method comparisons
- Vendors want to provide the best data from their methods
- EQA providers want to certify their materials with methods that give accurate results
- The newborn screening community benefits from method harmonization
- Quality improvements in global neonatal screening can be accomplished with international cooperation