Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Analytical validation based on total error measurement and cut-off interpretation of a neonatal screening TSH-immunoassay☆

François Boemer [a,*], Vincent Bours [a], Roland Schoos [a], Philippe Hubert [b], Eric Rozet [b]

[a] Laboratoire de Biochimie Génétique, Centre de Génétique Humaine, CHU, University of Liège, Liège, Belgium
[b] Analytical Chemistry Laboratory, CIRM, Department of Pharmacy, University of Liège, Liège, Belgium

ARTICLE INFO

Article history:
Received 22 July 2008
Accepted 5 November 2008
Available online 13 November 2008

Keywords:
Neonatal screening
Congenital hypothyroidism
Validation
Total error
Accuracy profile

ABSTRACT

To prevent the severe developmental and physical morbidities associated with congenital hypothyroidism, we developed a home-made Enzyme-Linked Immunosorbent Assay (ELISA) method to quantify Thyroid Stimulating Hormone (TSH) levels on newborn dried blood spots. In order to agree with actual clinical laboratory quality referential (ISO 15189), we desired to update our analytical validation protocol. For this purpose, an approach using accuracy profiles based on tolerance intervals for the total error measurement was for first time applied to an immunological assay. According to acceptance limits fixed at ±30%, the method was found accurate over a concentration range from 17.48 to 250 mIU/L. Based on 99.5 percentile of a 16,459 newborn population, cut-off was fixed at 20.1 mIU/L and validated against normal and pathologic neonatal populations. Additionally, uncertainty regions around this value were obtained applying four different approaches. Finally, we demonstrated here our in-house immunological technique fulfils criterions of a neonatal screening policy.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Congenital hypothyroidism is the principal cause of preventable intellectual deficit, with a prevalence of 1 in 4000 neonates. Approximately, 85% of cases of congenital hypothyroidism are sporadic and 15% are hereditary. The most common cause of congenital hypothyroidism is some form of thyroid dysgenesis (i.e. agenesis, hypoplasia or ectopy). Thyroid ectopy accounts for two-thirds of the cases worldwide.

Newborn screening is the most effective method to prevent the severe developmental and physical morbidities associated with congenital hypothyroidism. Most newborn babies with congenital hypothyroidism have few or no clinical manifestations of thyroid deficiency. As a result, it is not possible to predict which infants are likely to be affected. However, since initiation of replacement therapy within the first weeks of life should significantly reduce the number of patients with severe mental retardation, newborn screening programs were developed. Since the mid-1970s, either Thyroxin (T4) or Thyroid-Stimulating Hormone (TSH) were measured in heel-stick blood specimens to detect affected neonates as early as possible [1–3].

Our laboratory began the screening for congenital hypothyroidism in 1978, quantifying both T4 and TSH on dried blood spots. Later, according to local legislation and economical aspects, screening policy was limited to the only TSH quantification and we finally chose to develop a home-made analytical method to measure TSH on blood spots.

In order to agree with actual clinical laboratory quality referential (ISO 15189 [4]), we desired to update the validation of our analytical protocol. For this purpose, a strategy based on uncertainty of measurements and total error assessments has been for first time implemented on an immunological test [5–9].

2. Experimental

2.1. Chemicals

Reference TSH (WHO 81/565) was purchased from the National Institute for Biological Standards and Control (London, UK).

Bovine serum albumin (BSA), activated horseradish peroxidase and 3,3′,5,5′-tetramethylbenzidine (TMB) substrate were acquired from Roche (Basel, Switzerland). Sodium borohydride was from Acros Organics (Geel, Belgium) and Tween 20 from Sigma–Aldrich (St. Louis, USA). Sodium dihydrogenophosphate, dipotassium hydrogenophosphate, sodium chloride and hydrogen

☆ This paper is part of a special issue entitled "Method Validation, Comparison and Transfer", guest edited by Serge Rudaz and Philippe Hubert.

* Corresponding author at: Biochemical Genetic Laboratory, CHU Sart-Tilman, University of Liège, B35, 4000 Liege, Belgium. Tel.: +32 4 366 7696; fax: +32 4 366 8474.
E-mail address: F.Boemer@chu.ulg.ac.be (F. Boemer).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.11.004

peroxide were purchased from Merck (Darmstadt, Germany). Erythrocytes were obtained from EDTA-anticoagulated whole blood by centrifugation at 1000 × g for 5 min. All aqueous solutions were prepared with high purity water produced by a laboratory MilliQ system (Millipore, Billerica, USA).

Monoclonal antibodies for coating (subclass 1) and peroxidase-coupling (subclass 2a) were purchased from Medix Biochemica (Kauniainen, Finland).

2.2. Samples

All samples (whole blood dried on filter paper, Whatman 903 card) were obtained from 5-day-old newborns. These samples were collected as part of the mandatory neonatal screening program in Belgium.

2.3. Standard solutions

Standard solutions were prepared in a reconstituted whole blood matrix. Quantified amounts of a reference TSH stock solution were added into a synthetic serum matrix prepared by dissolving 60 g/L BSA into 0.05 M phosphate buffer at pH 7.5. These serums were then mixed with equal volumes of isolated red blood cells to the following final TSH concentrations: 250, 125, 62.5, 31.3, 15.6, 7.8 and 3.9 mIU/L.

Two types of those spiked samples were prepared, namely calibration standards and validation standards, both having the same concentrations. Each calibration standard was analyzed twice whereas each validation standard was analyzed at least eight times. Calibration and validation standards were prepared for four different days.

2.4. Immunoassay procedure

We have developed an in-house sandwich Enzyme-Linked Immunosorbent Assay (ELISA) to screen for congenital hypothyroidism, quantifying TSH.

Microplates were bounded by adding monoclonal mouse IgG_1 anti-TSH antibody to each well (50 ng in 0.05 M phosphate buffer at pH 7.5). Overnight incubation was carried out at 4 °C.

Monoclonal mouse IgG_{2a} anti-TSH antibody was conjugated to the activated horseradish peroxidase according to the periodate-mediated protocol [10–12]. Briefly, commercial monoclonal antibody solution was dialyzed against carbonate buffer to remove conservative agents (i.e. NaN_3). Conjugation with the periodate-activated peroxidase was then initiated, followed by a reduction with sodium borohydride of the Schiff's bases that have formed. Triethanolamine was added and finally, peroxidase-conjugated antibodies were stabilized in a PBS-Glycine buffer containing BSA (10 g/L).

After coating, microplates were saturated with 0.05 M phosphate buffer at pH 7.5 containing BSA (20 g/L). 3 mm dried blood spots were then eluted in each well under agitation and an overnight incubation at 4 °C was undertaken before blood spots discarding. After that, plates were washed several times with a rinsing buffer and a second 1-h incubation at room temperature using the peroxidase-conjugated monoclonal antibody (0.3 mg/L) was carried out. The microplate wells are thoroughly washed to remove unbound conjugate and peroxidase activity was revealed by addition of a TMB substrate solution (100 mg TMB and 100 μL H_2O_2 in 100 mL citrate buffer). After incubation for 15 min with gentle shaking at room temperature, the enzymatic reaction is stopped by addition of sulfuric acid and absorbance at 450 nm is read using an ELISA microtiter plate reader.

2.5. Method validation

An approach using accuracy profiles based on tolerance intervals for the total error measurement, including both bias and standard deviation for intermediate precision, was applied to demonstrate the method capability [5–9].

Method is considered as valid within the range for which the accuracy profile is fully included inside the accuracy acceptance limits set at ±30% [13–15]. The β-expectation tolerance interval, which describes a region where, on average, a proportion β of future measurements will fall, was fixed at 80% [16]. This proceeding gives the guarantee that each further measurement of unknown samples will be included within the tolerance limits at the fixed level and thus within the acceptance limits.

Seelva software v1.0 (Arlenda, Liège, Belgium) was used to compute accuracy profiles and validation results.

3. Results and discussion

3.1. Response function

Response function is the relationship existing between the response (signal) and the analyte concentration. Several regression models were tested by the Seelva software, estimating various analytical parameters such as accuracy, concentration range, precision and trueness indexes (Table 1) [17]. Ten regression functions were tested in order to find the most for our purpose: quadratic, power four and five parameters logisitic (PL) function either weighted or not and linear regression with logarithmic transformation of both the *X* and *Y*-axis. From these function, algorithm convergence could not be achieved for the 2 five PL models and for the weighted

Table 1
Regression models tested for fitting calibration curve during the validation study. The concentration range, precision, trueness and accuracy indexes are defined in Ref. [17].

Model	Accuracy index	Concentration range index	Precision index	Trueness index
Unweighted quadratic regression	0.814	0.906	0.630	0.944
Weighted (POM) power regression[a]	0.804	0.945	0.630	0.874
Weighted (1/*X*) quadratic regression	0.801	0.912	0.632	0.891
Unweighted four parameter logistic regression	0.750	0.750	0.601	0.937
Weighted ($1/X^2$) quadratic regression	0.669	0.626	0.566	0.845
Unweighted log–log regression	0.360	0.168	0.500	0.555
Unweighted power regression	0.000	0.000	0.000	0.000
Weighted (POM) four parameter logistic regression[b]	N/A	N/A	N/A	N/A
Unweighted five parameter logistic regression[b]	N/A	N/A	N/A	N/A
Weighted (POM) five parameter logistic regression[b]	N/A	N/A	N/A	N/A

[a] Selected model.
[b] The fitting algorithm did not converge for the considered model.

Table 2
Evaluation of the trueness of the ELISA assay dedicated to TSH evaluation in neonate during the validation study. Trueness is expressed a absolute bias, relative bias and recovery at each validation standard.

TSH concentration level (mIU/L)	Mean back-calculated concentration (mIU/L)	Absolute bias (mIU/L)	Relative bias (%)	Recovery (%)
3.9	4.85	0.95	24.2	124.2
7.8	8.68	0.87	11.1	111.1
15.6	15.53	−0.10	−0.6	99.4
31.3	31.53	0.28	0.9	100.9
62.5	62.53	0.03	0.05	100.0
125.0	127.3	2.28	1.8	101.8
250.0	227.2	−22.81	−9.1	90.9

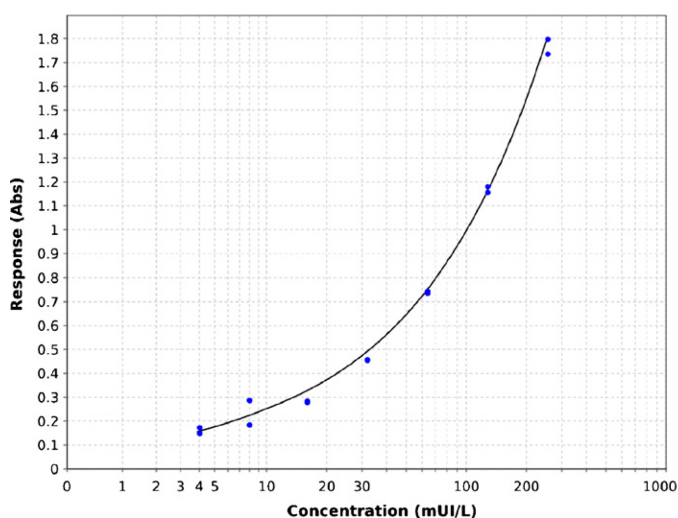

Fig. 1. Illustration of the selected calibration curve: weighted (POM) power regression ($Y = \beta_0 + \beta_1 X^{\beta_2}$).

four PL model. Nonetheless, considering our analytical procedure, weighted (using power of the mean – POM) power regression ($Y = \beta_0 + \beta_1 X^{\beta_2}$) was selected on account of its widest dosing range index [17]. Indeed with this model as standard curve the method is able to quantify accurately over the widest concentration range of TSH.

As presented in Fig. 1, calibrator concentrations only cover the lower part of the sigmoid curve, resulting in relative poor sensitivity for low TSH concentrations. Since, the purpose of a newborn screening policy is to detect babies with congenital hypothyroidism, prevention of "hook effect"-related false negative results linked to very elevated TSH values is fundamental. Thus, we focused our calibrators to fit under the inflection point of the sigmoid standard curve [18].

3.2. Trueness

Trueness refers to the closeness of agreement between a conventionally accepted value and a mean experimental one [4,19,20]. It gives information on systematic error.

Table 3
Evaluation of the precision of the ELISA assay dedicated to TSH evaluation in neonate during the validation study. Precision is expressed as relative standard deviation (R.S.D.) for intermediate precision and repeatability at each validation standard concentration.

TSH concentration level (mIU/L)	Repeatability (R.S.D.%)	Intermediate precision (R.S.D.%)
3.9	34.1	57.3
7.8	16.5	25.3
15.6	13.0	19.9
31.3	15.2	17.9
62.5	16.6	21.0
125.0	10.5	13.7
250.0	11.4	12.3

Table 2 reports trueness expressed as absolute bias, relative bias and recovery for the different level of validation standards. Recovery was close to 100% for concentration values higher than 15.6 mIU/L. The highest level showed a slight reduced value (90.9%).

3.3. Precision

Precision is the closeness of agreement among measurements from multiple sampling of a homogeneous sample under the recommended conditions [4,19,20]. It gives some information on random errors and it can be evaluated at two levels: repeatability and intermediate precision. Results are presented in Table 3.

Relative standard deviation for repeatability and intermediate precision seemed relatively high, with results varying between 10.5% and 21.0% for the five highest curve points. Purification of the peroxidase-conjugated monoclonal antibody to remove excess of activated peroxidase that does not have reacted should be a way of improvement to reduce variability. Consequently, such operation could generate better precision results.

3.4. Accuracy

Accuracy refers to the closeness of agreement between the test result and the accepted reference value, namely the conventionally true value [4,19,20]. The accuracy takes into account the total error, i.e. systematic and random errors, related to the test result (Table 4). It is assessed from the accuracy profile illustrated in Fig. 2. The

Table 4
Validation study: accuracy of the results obtained by the ELISA assay for TSH determination in neonates obtained by considering weighted (POM) power regression as standard curve.

TSH concentration level (mIU/L)	β-Expectation tolerance limit (mIU/L)	Relative β-expectation tolerance limit (%)	Total error (%)
3.9	[1.38, 8.33]	[−64.8, 113.2]	81.6
7.8	[5.68, 11.68]	[−27.3, 49.4]	36.4
15.6	[10.81, 20.25]	[−30.8, 29.6]	20.6
31.3	[23.75, 39.31]	[−24.0, 25.8]	18.8
62.5	[43.87, 81.19]	[−29.8, 29.9]	21.0
125.0	[102.7, 151.9]	[−17.9, 21.5]	15.5
250.0	[185.5, 268.9]	[−25.8, 7.5]	21.4

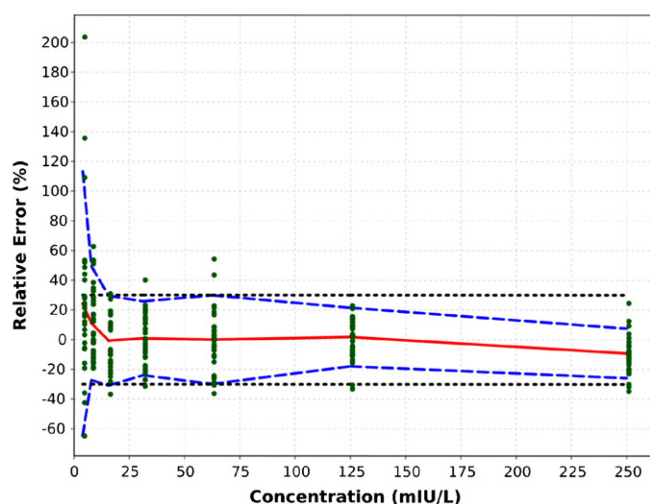

Fig. 2. Accuracy profile obtained by considering weighted (POM) power regression. Plain line is the relative bias, dashed lines are the β-expectation tolerance limit (β = 80%) and dotted curves represent the acceptance limit (±30%). The dots represent the relative back-calculated concentrations of the validation standards and are plotted according to their targeted concentration.

method is considered as valid within the range for which the accuracy profile is within the accuracy acceptance limits set at ±30%. This approach gives the guarantee that each further measurement of unknown samples will be included within the tolerance limits at the 20% level.

The lower limit of quantitation (LLOQ) is the smallest quantity of the targeted substance in the sample that can be assayed under experimental conditions with well-defined accuracy [4,19,20]. The definition can also be applicable to the upper limit of quantitation, which is the highest quantity of the targeted substance in the sample that can be assayed under experimental conditions with well-defined accuracy. The limits of quantitation are obtained by calculating the smallest and highest concentration beyond which the accuracy limits or β-expectation limits go outside the acceptance limits. Experimentally, the LLOQ was evaluated at 17.48 mIU/L and the upper limit of quantitation at 250 mIU/L. The concentration range extends then to the interval between the lower and upper limits where the procedure achieves adequate accuracy.

Traditionally, LLOQ is defined as 2 S.D. above the mean of the lowest calibrator measurement [21,22]. Statistical comparison between the first point of curve and a blank sample represents an alternative for establishing the minimal quantitable concentration. Regarding to these proceedings, the new approach referring to accuracy profiles looks more stringent. Indeed, compared to our previous LLOQ set at 3.9 mIU/L, the new calculated LLOQ appeared to be high, with a concentration far above the normal population. Moreover, Table 4 shows from the β-expectation tolerance interval at 3.9 mIU/L that 80% of the future results that will be obtained when measuring a sample with true concentration of 3.9 mIU/L will be between 1.375 mIU/L and 8.329 mIU/L corresponding to a total error of 81.6% which is not acceptable for routine application of the method. However, at the LLOQ, it can be guaranteed that at least 80% of future results will have a total error of at most 30%, thus illustrating the reliability of this estimation of the LLOQ.

Referring to previous publications, a direct relationship has been defined between iodine deficiency and mild elevated TSH values on dried blood spot [23–26]. Moreover, the World Health Organization (WHO) [27] states that the increase in the number of neonates with moderately elevated TSH concentrations (above 5 mIU/L whole blood) is proportional to the degree of iodine deficiency during pregnancy. Assays that utilize monoclonal antibodies, which can detect TSH as low as 5 mIU/L in whole blood spots, are useful for recognizing iodine deficiency [27]. On that basis, our home-made technique does not seem suitable for identifying neonates with iodine deficiency since the WHO threshold value is below our LLOQ. As mentioned above, some method improvements could however reduce our LLOQ, although gaining a quantitation limit below 5 mIU/L seems unrealistic. On the other hand, it would nevertheless be interesting to estimate the LLOQ of the methods mentioned in those manuscripts assaying the same validation strategy, in order to ensure that the 5 mIU/L concentration would be fixed above the LLOQ.

3.5. *Uncertainty of measurement*

The uncertainty characterizes the dispersion of the values that could reasonably be attributed to the measurand, i.e., the concentration of TSH in our study [28–30]. Several uncertainty results (see Appendix A for formulas) were generated and are presented in Table 5 [28–32].

- The uncertainty of bias of the method at each concentration level of the validation standard.
- The composed uncertainty which combines the uncertainty of the bias together with the uncertainty of the method obtained during the validation step, i.e. the intermediate precision standard deviation.
- The expanded uncertainty which equals to the composed uncertainty multiplied by a coverage factor k = 2 representing an interval around the results where the unknown true value of the result can be observed with a confidence level of 95%.

In addition, the relative expanded uncertainties (%) for TSH obtained by dividing the corresponding expanded uncertainty with the corresponding introduced concentrations are given in Table 5. This means for instance for the 62.5 mUI/L concentration level validation standard that, with a confidence level of 95%, the unknown true value of TSH concentration in the sample is located at a maximum of ±44.2% around the measured result. For the smallest concentration level of TSH investigated the relative expended uncertainty is extremely high, going up to about 125%.

Table 5
Validation study: estimates of measurement uncertainties related to TSH measurement in neonates at each concentration level investigated.

TSH concentration (mIU/L)	Uncertainty of the bias (mIU/L)	Uncertainty (mIU/L)	Expanded uncertainty (mIU/L)	Relative uncertainty (%)
3.9	0.93	2.42	4.85	124.2
7.8	0.78	2.13	4.268	54.5
15.6	1.23	3.35	6.70	42.9
31.3	1.69	5.84	11.68	37.4
62.5	4.38	13.82	27.65	44.2
125.0	5.95	18.10	36.21	29.0
250.0	7.54	31.70	63.40	25.4

Table 6
Newborn study: biological sensitivity and specificity have been calculated according to various cut-off values, based on a receiver operating characteristic (ROC) curve analysis. Results for concentrations enclosing the cut-off value (20.1 mIU/L) are presented.

Cut-off values	Sensitivity	Specificity
>15.6	100.00	98.20
>20.1	100.00	99.73
>22.0[a]	100.00	99.83
>22.3	97.67	99.83
>40.2	97.67	99.99
>44.4	97.67	100.00
>60.0	95.35	100.00

[a] Value corresponding with the highest average of sensitivity and specificity.

3.6. Cut-off interpretation

All analytical protocol for screening assays is basically a qualitative test providing general answers based simply on binary yes/no responses. These methods are characterized by two (or three) interpretation ranges delimited by concentration values named "cut-off". As a result, an unknown sample measured at a concentration belonging to the lower or the upper interval will be considered as a normal or a pathological specimen, according to the interpretation assigned to the cut-off value.

For our immunological assay, we established the pathological threshold on 16,459 newborns samples. A reasonable recall (positivity) rate was determined at 0.5%, by calculating the 99.5 percentile of our population. Cut-off was then fixed at 20.1 mIU/L and results above this value were considered as pathological. This concentration is superior to the LLOQ and agrees with usual cut-off reported for TSH neonatal screening [33,34]. Moreover, the two calibration levels enclosing this threshold present good recovery results (Table 2). Additionally, this cut-off has been validated against 43 confirmed congenital hypothyroidism samples and more than 10,000 normal neonates. Diagnostic sensitivity and specificity for critical concentrations were calculated using the MedCalc software (Mariakerke, Belgium). Results are reported in Table 6. According to those results, we can confirm the reliability of the selected threshold.

A cut-off value is of less value if the uncertainty region around it is not given simultaneously [35]. This uncertainty region defines a region of results where no conclusion about the status of patient can be made without making an excessive error. Results included in this region should be confirmed by another method. Several options are available to define this uncertainty region:

1. From validation results:
   - using the estimated uncertainty of measurement (see Appendix A for formula),
   - using the estimated total error (see Appendix A for formula),
   - using the prediction interval or β-expectation tolerance interval (see Appendix A for formula).
2. From the newborn study:
   - using the confidence interval of the 99.5 percentile [36].

Table 7 gives the uncertainty regions obtained using these four methodologies. Each of these approaches has pitfalls; the approaches based on the validation study results considers that the cut-off value is the true amount of TSH which is not the case, furthermore using uncertainty of measurement does not take into account the bias of the method whereas the other two do. Advantage of the methods using measurement uncertainty and tolerance interval is that they are predictive, whereas the total error one is not. The approach based on the confidence interval of the newborn study gives an uncertainty region that depends on the sample size and does not takes into account the random and systematic error of the analytical method, furthermore it is not a predictive uncertainty region and thus gives an optimistic uncertainty region. Nonetheless, as can be seen form Table 7, the uncertainty regions estimated using the approaches based on the validation phase of the method are rather close. As there is a negligible bias (Table 2) few differences exists between the measurement uncertainty and tolerance interval definition of the uncertainty region. The one defined by the total error approach is the smallest one, due to its lack of prediction ability. As expected the uncertainty region determined by the newborn study is extremely small. The selected uncertainty region for our method was the one estimated using the tolerance interval as it is a methodology that is predictive, considers the total error of the analytical method and is based on a statistically sound methodology. However, further researches should be made in order to find the most adequate way for defining the uncertainty region around cut-offs values due to their essential significance for diagnostics.

Table 7
Newborn study: uncertainty regions around the cut-off value (20.1 mIU/L) obtained according to four different approaches.

Method	Uncertainty region	
	Lower limit (mIU/L)	Upper limit (mIU/L)
Uncertainty	14.26	25.94
Total error	15.44	24.76
Tolerance interval	14.52	27.42
Confidence interval	19.2	21.2

4. Conclusion

We presented here a methodology to validate a neonatal screening assay for congenital hypothyroidism taking into account the total error of the method through the use of tolerance interval. To our knowledge, we illustrated for first time the applicability of such approach to ELISA assays.

An accuracy profile was defined and the method was considered reliable within the range for which the accuracy profile is within the accuracy acceptance limits set at ±30%. We found our method presented a relatively elevated lower limit of quantification, far above the criteria defined by the WHO on TSH-based neonatal iodine deficiency screening. Method cut-off has been determined on a statistical basis and validated subsequently against normal and pathologic populations. Furthermore, uncertainty regions encompassing this cut-off have been determined applying four different approaches.

To conclude, our home-made method has been initially developed to screen for newborns with congenital hypothyroidism and so, the reported validation results tend to prove our technique suits with its intended use.

Acknowledgements

The authors thank M.A. Coeme for its excellent technical assistance. Thanks are also due to the Fonds Leon Fredericq and the University of Liège for a research grant provided to Eric Rozet.

Appendix A

A.1. Total error

Total error (TE) is the simultaneous combination of systematic and random error. Systematic error is measured by a bias ($\hat{\delta}$) and

random error by a variance $\hat{\sigma}^2$.

$$\mathrm{TE} = |\hat{\delta}| + (\hat{\sigma}^2_{I.P.})^{0.5} \tag{1}$$

A.2. β-Expectation tolerance interval

Tolerance intervals are intervals that contain a proportion β of the individual values, such as results, of the population (*e.g.* 0.95). These intervals allow to describe the entire population. The following formula describes a β-expectation tolerance interval:

$$\left[\hat{\delta} - Q_t\left(\nu; \frac{1+\beta}{2}\right)\hat{\sigma}_{Tol}; \hat{\delta} + Q_t\left(\nu; \frac{1+\beta}{2}\right)\hat{\sigma}_{Tol}\right] \tag{2}$$

where

- $Q_t(\nu;\theta)$ is the θth percentile of a Student $Q_t(\nu)$ distribution;
- $\hat{\sigma}_{Tol} = \sqrt{1 + \frac{n(\hat{\sigma}^2_\alpha/\hat{\sigma}^2_\varepsilon)+1}{pn((\hat{\sigma}^2_\alpha/\hat{\sigma}^2_\varepsilon)+1)}}\hat{\sigma}_{I.P.}$
- n: number of replicates per run (p runs);
- $\hat{\sigma}^2_\alpha$ is the run-to-run variance, and $\hat{\sigma}^2_\varepsilon$ is the within-run or repeatability variance. The overall variability of the analytical method is measured by the intermediate precision variance: $\hat{\sigma}^2_{I.P.} = \hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon$.

A.3. Measurement uncertainty

The measurement uncertainty $u(x)$ of a result x is estimated by:

$$u(x)^2 = \hat{\sigma}^2_{I.P.} + u(\hat{\delta})^2 \tag{3}$$

where $\hat{\sigma}^2_{I.P.}$ is the estimated intermediate precision standard deviation and $u(\hat{\delta})$ is the uncertainty associated with the estimator of the bias δ of the method (expressed in term of standard error). $u(\hat{\delta})$ can be estimated as

$$u(\hat{\delta}) = \sqrt{\frac{\hat{\sigma}^2_{I.P.}\left(1 - \gamma + (\gamma/n)\right)}{p}} \tag{4}$$

where $\gamma = \hat{\sigma}^2_\varepsilon/\hat{\sigma}^2_{I.P.}$ with $\hat{\sigma}^2_\varepsilon$ being an estimate of the repeatability variance.

References

[1] J.H. Dussault, P. Coulombe, C. Laberge, J. Letarte, H. Guyda, K. Khoury, J. Pediatr. 86 (1975) 670.
[2] M. Irie, K. Enomoto, H. Naruse, Lancet 2 (1975) 1233.
[3] A.H. Klein, A.V. Agustin, T.P. Foley Jr., Lancet 2 (1974) 77.
[4] ISO 15189 Medical Laboratories – Particular requirements for quality and competence, International Organization for Standardization (ISO), 2007.
[5] A.G. Gonzales, M.A. Herrador, Talanta 70 (2006) 896.
[6] E. Rozet, R. Morello, F. Lecomte, G.B. Martin, P. Chiap, J. Crommen, K.S. Boos, P. Hubert, J. Chromatogr. B 844 (2006) 251.
[7] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 82.
[8] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 70.
[9] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579.
[10] E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988.
[11] P. Tijssen, E. Kurstak, Anal. Biochem. 136 (1984) 451.
[12] P.K. Nakane, A. Kawaoi, J. Histochem. Cytochem. 22 (1974) 1084.
[13] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, Pharm. Res. 20 (2003) 1885.
[14] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V.P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, Pharm. Res. 24 (2007) 1962.
[15] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M. Laparra, M. Laurentie, J. Nivet, Anal. Chim. Acta 391 (1999) 135.
[16] B. Boulanger, W. Dewé, A. Gilbert, B. Govaerts, M. Maumy-Bertrand, Chemometr. Intell. Lab. Syst. 896 (2007) 198.
[17] E. Rozet, V. Wascotte, N. Lecouturier, V. Preat, W. Dewe, B. Boulanger, P. Hubert, Anal. Chim. Acta 591 (2007) 239.
[18] D. Rodbard, Y. Feldman, M.L. Jaffe, L.E. Miles, Immunochemistry 15 (1978) 77.
[19] ISO 5725 Application of the statistics-accuracy (trueness and precision) of the results and methods of measurement. Part 1 to 6, International Organization for Standardization (ISO), 1994.
[20] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, J. Chromatogr. A 1158 (2007) 111.
[21] E.C. Neto, J. Schulte, Clin. Chem. 44 (1998) 2372.
[22] L.F. Hofman, T.P. Foley Jr., J.J. Henry, E.W. Naylor, J. Med. Screen. 10 (2003) 5.
[23] F. Delange, Thyroid 8 (1998) 1185.
[24] F. Delange, Acta Paediatr. Suppl. 88 (1999) 21.
[25] D.L. Copeland, K.M. Sullivan, R. Houston, W. May, I. Mendoza, Q. Salamatullah, N. Solomons, D. Nordenberg, G.F. Maberly, Public Health Nutr. 5 (2002) 81.
[26] A. McElduff, P. McElduff, J.E. Gunton, G. Hams, V. Wiley, B.M. Wilcken, Med. J. Aust. 176 (2002) 317.
[27] WHO, Assessment of Iodine Deficiency Disorders and Monitoring their Elimination: A Guide for Programme Managers, World Health Organization, Geneva, 2007.
[28] CB EA-4/16, EA Guidelines on the expression of uncertainty in quantitative testing, 2004, http://www.european-accreditation.org.
[29] Eurachem/Citac guide quantifying the uncertainty in analytical measurement, second ed., 2000.
[30] Analytical Methods Committee, Analyst 120 (1995) 2303.
[31] M. Feinberg, B. Boulanger, W. Dewé, P. Hubert, Anal. Bioanal. Chem. 380 (2004) 502.
[32] R. Marini, P. Chiap, B. Boulanger, S. Rudaz, E. Rozet, J. Crommen, P. Hubert, Talanta 68 (2006) 1166.
[33] S.R. Rose, R.S. Brown, T. Foley, P.B. Kaplowitz, C.I. Kaye, S. Sundararajan, S.K. Varma, Pediatrics 117 (2006) 2290.
[34] J.A. Lott, M. Sardovia-Iyer, K.S. Speakman, K.K. Lee, Clin. Biochem. 37 (2004) 791.
[35] E. Trullois, I. Ruisanchez, F. Ruis, J. Huguet, Trends Anal. Chem. 24 (2005) 516.
[36] G. Hahn, W. Meeker, Statistical Intervals: A Guide for Practitioners, John Wiley & Sons, Inc., New York, 1991.