Review article
Central Eur J Paed 2018;14(2):224-235
DOI 10.5457/p2005-114.221

Understanding the Basics of Quality Control in the Medical Laboratory – the Impact on Result Interpretation

Vesna Kušec[1], Nataša Rojnić Putarek[2], Nevena Krnić[2], Anita Špehar Uroić[2], Katja Dumić[2]

[1]Clinical Institute of Laboratory Diagnosis, University Hospital Centre Zagreb, [2]Department of Paediatrics, University of Zagreb School of Medicine, University Hospital Centre Zagreb, Croatia

Correspondence:
vkusec@kbc-zagreb.hr
Tel.: + 385 1 236 7 289
Fax.: + 385 1 236 7 395

Received: November 29, 2017
Accepted: June 18, 2018

Key words: Hormone measurement ▪ Measurement methods ▪ Immunoassay ▪ Quality management ▪ Six sigma.

The aim of this paper was to provide information on the limitations of current analytical methods and the basics in quality control in laboratory endocrinology which affect the results and their interpretation. For the practicing clinical endocrinologist it is important to understand the essentials of the quality of hormone measuring methods and their limitations. Awareness of the quality of methods will allow caution and/or ensure the reliability of result interpretation. A measuring method cannot be performed with equal precision and trueness for the entire measuring range, and may differ between parameters and manufacturers. According to the level of medical decision, the medical laboratory should determine the quality of the method and communicate that information to the clinician. The basic approach to quality control of measurement methods is useful knowledge for doctors working in a hospital setting or private practice. The advantages and limitations are presented of immunoassay-based methods and tandem mass spectrometry (LC-MS/MS) used for measurement in biochemistry laboratory. **Conclusion** – Hormone measurement is widely available but the methods may be affected by limitations. Understanding a method's limitations and options for improvement enables the practicing clinician to interpret the test results rationally.

Introduction

Hormone measurement remains an essential aspect of diagnosis and management in paediatric endocrinology, in conjunction with imaging and molecular diagnostic tests. Laboratory tests have advanced since the introduction of commercial radio-immunoassays for hormone measurement that have ensured the availability of laboratory tests at reasonable cost. Both methodological improvements and modern technology have been important milestones in laboratory diagnostics, allowing for the current state in which results are obtained in a shorter time and more frequently. Tests are in general more specific and precise, costs are lower and the development of automated instruments or platforms have helped make laboratory work less demanding for personnel. Thus, every hospital biochemistry laboratory is able to offer some hormone tests.

Moreover, awareness of good clinical and laboratory practice and safety is now an important part of routine health care. In laboratory medicine, measurement methods may be standardized, through use of certified reference materials for assay calibration, or, in the absence of higher-order reference materials, external inter-laboratory quality assess-

ment schemes can be used. The main reason behind national or international standards for medical laboratories is to standardise care and prevent harm to the patient/client caused by laboratory results (1, 2, 3).

Various approaches to quality control in biochemistry laboratories have long been in use. However, one basic approach that is commonly used is analysing samples of control material of known concentration along with patient samples, on a daily basis (4). A special area of laboratory medicine is devoted to the study and application of scientific knowledge in practice for the purposes of quality assurance. Standards such as the ISO standard for medical laboratories 15189, Clinical Laboratory Improvement Amendment (CLIA), College of American Pathologists, Rili-BAEK and others, require the use of control material in routine work, the assessment of the quality of measurement methods, and participation in proficiency testing, i.e. comparison of results among peer laboratories according to method or manufacturer, organized by a third party (5). These standards encompass all processes in a medical laboratory, including: documentation on quality policy, personnel competencies, instruments, environment issues, resolving of complaints, etc. The ISO standard is widely accepted and used for accreditation of medical laboratories (6). The CLIA and the German national quality standard Rili-BAEK clearly defined approaches to quality control that are easy to follow and implement. Rili-BAEK is very similar to the ISO standard, but comprehensive, and covers all areas of laboratory medicine, i.e. "polyvalent laboratory medicine, which comprises microbiology and transfusion medicine (7).

A very attractive approach to quality control in the biochemistry laboratory is advocated by J. O. Westgard and his son through his website (westgard.com), which is open access and very informative. The basics of quality control are described very clearly, and real examples from published data are critically presented. The main message is that method quality can be graded according to its imprecision and bias, with rules for controlling the stability of the methods and detecting system malfunctions causing erroneous results (8).

From the perspective of the clinician, the aspects of quality management of greatest interest are the quality of the measuring method, i.e. evidence-based data on its limitations and reliability. Standard measurement methods in the biochemistry laboratory are recommended by international and national professional societies. Gold standard methods (e.g. mass spectrometry) may not be routinely available, are complex and costly, and the technology requires appropriately educated personnel (9, 10). Routine methods are typically automated platforms or commercial kits based on immunoassay technology. Both methodologies have advantages and limitations (11).

The method characteristics that are checked by the laboratory upon the introduction of a new method or change of manufacturer, include trueness or bias, imprecision, linearity (or reportable range) and detection limits. However, this can be reduced to two main traits, bias (i.e. the deviance of a result from its target value) and imprecision (i.e. the variability between multiple measurements from the same sample) (3, 8). In order for a method to achieve its clinical purpose (i.e. to reliably reflect the analyte in its respective matrix and not to be a result of a random effect), bias and imprecision have to conform to the desirable specifications recommended by professional societies or expert groups (e.g. 12). The estimated bias and imprecision of a method have to be less than those stated as desirable specifications, and these are parameter- and sometimes body fluid-specific (4).

Body fluid constituents, that is analytes, are subject to fluctuations and variations as a consequence of physiological conditions (sex, age, menopause, pregnancy, growth) and

pathophysiological processes (13, 14). This includes variations dependent on time of day or season, and variations related to specific traits or an individual set point for a test (13, 14). Biological variations can be described as dependent on time/condition of life, on different cyclical changes (e.g. monthly) or as of random character. Random biological variations are the traits of most tests, and vary around the individual set point for a test. Subject variations include biological variations and differences between set points of different individuals (examples for some hormones in Table 1).

A biochemistry laboratory should consider these biological variations when reporting results and defining quality specifications for analytical performances. Within-subject and between-subject biological variations for each test are implemented in the assessment of analytical specifications. This is also important for evaluation of patient monitoring with regard to significant changes between two tests in the same patient, the use of a test for distinction of health and disease, transfer of population-based reference ranges, decisions on the most likely pathognomonic tests, and various aspects of studies on new laboratory parameters (13, 14). Data from clinical studies are selected and compiled to form databases, which are updated on a regular basis with information on within- and between-subject biological variations used for assessment of analytical specifications and setting goals in quality management inf a medical laboratory. Within-subject variations in diseases are within the same order of magnitude as in state of health for most tests, which enables the use of data collected from healthy individuals to assess significant differences between consecutive tests in patients (15).

The aim of this paper was to provide information on the limitations of current analytical methods, and the basics of quality control in laboratory endocrinology which affect the results and their interpretation.

Quality Assessment of Laboratory Methods

The quality of products and services can be expressed in terms of the number of flawed occurrences per million (Table 2). This method was introduced in the early 1980s by the manufacturer Motorola. The lowest number of faulty products, e.g. 3.4 per million, is

Table 1. Data on Biological Variation and TEa for Selected Hormones in Serum from the Ricos Database (10) Presented as Rounded Numbers

Hormone	Biological variation		TEa (%)
	Intra-individual (%)	Between individuals (%)	
17-hydroxyprogesterone	20	50	30
Aldosterone	29	40	37
Androstendione	16	39	24
Cortisol	15	38	23
Estradiol	23	24	27
Follicle-stimulating hormone	11	47	21
Insulin	21	85	33
Luteinising hormone	23	27	28
Parathyroid hormone	26	24	30
Prolactin	23	35	29
Testosterone	9	22	14
Thyroid-stimulating hormone	19	25	24
Thyroxine	5	11	7

TEa=Total allowable error.

termed as six sigma, and is considered to be the highest quality production. An increase in number of defective products yields a lower sigma. In everyday life, the quality of commercial airline travel has a sigma higher than six, but the quality of luggage transport by air is sigma three. In health care, anaesthesia during surgery has a sigma higher than six, but drug prescription is three, and mammography outcome less than three. An average company and also medical laboratory function is at sigma three (16, 17).

Sigma can be also calculated for measurement methods, for qualitative methods, e.g. urine analysis or molecular diagnostics, as the number of erroneous results per one million tests. For quantitative methods, the formula takes into account bias and imprecision in relationship to the desirable specifications. For this purpose, a different metric of desirable specification is used, that is, total allowable error (TEa) (e.g. 7, 12). TEa describes the greatest percentage of measurement error that can be allowed for the method to have clinical utility. There is another practical use for information obtained by TEa. It represents a significant difference between the results of two samples from one individual. For example, if TEa is 10%, it means that the difference between two separate results of the same patient has to be ≥10% to be a biochemically significant difference. If the difference is <10%, it is within the measurement error and there is no significant change between the two results. This is very practical for doctors who are not yet experienced in clinical work, and also allows for an evidence-based approach to the assessment of laboratory tests.

Table 2. Sigma Level According to the Number of Faulty Products or Services Per Million

Sigma level	Defects per million
6	3.4
5	233
4	6210
3	66,807
2	308,37
1	690,000

A method with a sigma six or higher performance is world class, five is considered excellent, four is good, and a method with sigma three is borderline (18). Sigma is also a term for the standard deviation for a large sample size, e.g. a population. For a method with six sigma this means that imprecision is so low that six standard deviations fit on both sides of the arithmetic mean within the limits of TEa. Conversely, for a method with three sigma, only three standard deviations can be placed on each side of the arithmetic mean of repeated measurement within the limits of TEa. Thus, an excellent method has small imprecision of repeated results, and a method of mediocre quality has wider imprecision (3, 8). The next characteristic is the difference between the result or arithmetic mean of repeated results and the target result. The formula for the calculation of sigma is (westgard.com, 3): sigma = (total allowable error – bias)/coefficient of variation.

All terms are expressed in percentages. Coefficient of variation (CV, i.e. imprecision) is calculated as the ratio of standard deviation and arithmetic mean multiplied by 100. Assessment of method quality facilitates several practical aspects of work in biochemistry laboratories. The most important is the appropriate use of control samples to detect errors in the measurement process. According to a specific sigma, rules exist for application of the number and frequency of control samples, and also for their interpretation. These are Westgard rules, devised by a scientist who devoted his career to understanding and implementing quality management systems in clinical laboratories. Westgard rules apply to the sigma of the method from six to three, and are based on a scientific approach ensuring the detection of the defective functioning of the method, but also preventing

the false rejection of good results (8, 19). The mathematical bases for these rules were developed in the nineteen-seventies, and complex graphs were used to read how many controls are required per run, and how they should be interpreted (20). With time, the rules have been simplified for use, and can be downloaded as easy to follow diagrams (https://www.westgard.com/westgard-sigma-rules.htm). The easiest approach is for a sigma six or higher method. In this case, one control sample is used in each run and the control result is allowed within 3 standard deviations (on each side of the arithmetic mean). This can be used for the manufacturer's data for controls, but data collected by the lab are preferred (19). For sigma five, two controls are measured in each run, and a malfunction is detected when the control results exceed the same side of 2 standard deviations on two consecutive occasions, or if one control result exceeds – 2 standard deviations, while the other exceeds + 2 standard deviation. For sigma of a smaller value, several rules or multi-rules are considered. Additional information is available at https://www.westgard.com/mltirule.htm. This enables detection in case of a failure of the measurement system which can then be stopped, the instrument recalibrated and the samples repeated (8, 19). False rejection is when the good functioning of the measurement system is not recognized by the operator, but it is assumed to be faulty due to erroneous interpretation of the quality control results. This may occur if the method has six sigma, but the control results are between 2nd and 3rd standard deviations and considered outside the acceptable limits. This causes unnecessary cost and delay in reporting the results (8).

Assessment of measurement methods may reveal an unsatisfactory sigma of three or less for a specific method, manufacturer or type of instrument (21, 22, 23, 24). The quality of the method is not improved by running it on an automated instrument, as this will mostly only result in imprecision. The problem usually lies in the method itself and its bias. The gold standard methods, e.g. tandem mass spectrometry and high pressure liquid chromatography, are precise methods, but not widely available for routine laboratory tests. Even then, the problem may be in the preparation of the sample i.e. it leads to imprecision. Methods with poor sigma can be controlled and kept stable by applying the Westgard rules for the number and frequency of controls and also interpretation of the control results. A method with poor sigma should be replaced in the long term, and this should be included in future plans for instrument purchase and is an opportunity for improvement of the quality of laboratory tests.

Control samples are usually purchased from the manufacturer of the instrument i.e. a method for one, two or three different levels covering the expected range of the parameters. For example, for glucose or cortisol measurement, low, middle and high concentrations may be of equal clinical interest. Control samples can also be purchased from independent manufacturers. The manufacturer provides the target value for the control sample with a range of 2 (or 3) standard deviations to fit its instruments working in various conditions. However, this is only for the labs' orientation. Each laboratory should establish its own data on controls run over 20 consecutive working days (7, 18). The mean and standard deviation thus obtained are used in interpreting the results of the controls and are specific to the use of a particular instrument in specific working conditions, personnel and clinical working load. The data collected in this way on the results of control samples are used to calculate imprecision (i.e. coefficient of variation). When the lot number of the reagents is changed, ten measurements of the controls are performed with the new lot of reagents and the results

are compared by statistical tests with those of the previous lot (19). In case of significant differences, the statistical data on control results are changed accordingly for further evaluation of the control sample results.

Table 1 presents data on TEa for some hormones from the comprehensive Ricos database (12). It should be noted that for hormones, but also vitamins and tumour markers (the latter two groups are not included in Table 1), TEa is in the range of 20-30%, which is larger than for the parameters of routine clinical chemistry tests. Very few hormones have low TEa, e.g. free T4 is 7% (12). This indicates that the methodology for routine hormone tests is not optimal, and would benefit from improvements by the manufacturers. In addition, this is important information because repeated sampling in a single patient has to yield a difference of 20-30% to be considered a significant change.

Method performance is assessed by calculating sigma for all control sample materials, which are typically at low, medium and high concentrations. This process reveals that the method will not have similar performance for the entire measurement range. This poses the problem of which sigma to choose to decide on the method quality. The recommended approach is to decide on the level which is of clinical significance or is clinically decisive (8). For example, the upper limit of the reference range for most hormones is important for distinguishing increased secretion. Neonatal screening for hypothyroidism and detection of hypothyroidism in adults have different levels of clinical decision that require good sigma performance or awareness of poor method performance, but these levels are not low or at the limit of detection. If the method has poor performance at the level of clinical decision, the results should be interpreted with caution in relation to other clinical and diagnostic data. The clinician should be aware of this through communication with the medical laboratory. However, sigma six or five at the level of decision will support the reliability of clinical diagnosis regarding the test.

The manufacturers of large instruments or platforms for many analytes claim superior performance for the entire range of parameters. But this is never the case, due to the limitations of the methods. Routine chemistry analysers do not perform with sigma 6 or 5 or identical sigma for all tests, or for the entire measuring range. This makes decision-making when acquiring instruments for the medical laboratory difficult.

Limitations in Hormone Measurement

Besides understanding the quality of the measuring method, there are other limitations which are important for both the clinician and the laboratory doctor. These limitations may be related to biological variation or to the measurement method, e.g. cross reactivity or problems in separation of hormones from their binding protein.

17-hydroxyprogesterone

Routine immunoassay based methods for measurement of 17-hydroxy progesterone in newborns and infants suffer from interference from steroid metabolites from the temporarily active foetal adrenal zone, with the consequence of higher results (25, 26, 27). This can be solved by extraction of the serum with organic solvents (e.g. ether) for removal of the interfering steroids (28, 29). The screening method is compromised in newborns by interfering substances, but it is good for children beyond the first year of life and in adults because the perinatal activity of the adrenals resolves by itself with time (25, 26, 27). Newborn screening for congenital adrenal hyperplasia has several problems that need to be considered to ensure reliable de-

tection of the disease (specificity of the assay) and prevent false positive diagnosis. Some of the automated methods for newborn screening are immunoassay-based, and have limitations because the method has poor specificity for this analyte. Levels of 17-hydroxy progesterone are increased in newborns, in premature infants and in those with an illness or suffering stress, and thus they are difficult to distinguish from high levels in congenital adrenal hyperplasia (27). One approach is to set cut-off values by gestational age and birth weight, and include secondary measurement of several androgens (e.g. testosterone and androstenedione) after a screen-positive result (27, 30).

Cortisol

Physiological conditions may affect immunoassay-based cortisol measurements. In males and non-pregnant females, the performance of the measurement method may be similar. However, cortisol measurement in pregnant females is complicated due to the increased levels of binding protein that cannot be adequately separated in the assay, causing under-recovery. Conversely, in patients treated with metyrapone, which blocks cortisol synthesis but not the formation of precursor metabolites, over-recovery may be observed secondary to antibody cross-reactivity (31). Misinterpretation of cortisol results in pregnant females and patients on metyrapone might compromise patient diagnosis and/or management. Widespread use of various exogenous corticosteroids for the treatment of many diseases may also impact the assessment of a patient's adrenal reserve due to poor antibody specificity and therefore cross-reactivity (which is acknowledged by the manufacturers), resulting in over-recovery (31). Only dexamethasone does not cross-react in immunoassays, and can therefore be used safely to assess hypercortisolism.

Testosterone

In blood, testosterone is bound to SHBG (sex hormone binding protein) and weakly bound to albumin. Only a small portion of testosterone is free and has biological activity together with the albumin-bound testosterone, which can easily dissociate. Testosterone is another test ordered frequently in paediatric and adult endocrinology for assessment of androgen status. This can be performed by LC-MS or immunoassay. A problem exists with measurement of low concentrations of testosterone, as found in women and small children due to the unsatisfactory sensitivity and specificity of the immunoassay methods used routinely. This is not encountered for the much higher levels found in adult men (32, 33). Sensitivity and specificity of the immunoassay can be improved by sample preparation, in which hormones are separated from proteins by extraction by gas or liquid chromatography. The extraction process requires expertise and is time consuming. In contrast, direct immunoassay techniques for which no sample preparation is necessary, are a more rapid alternative to the serum extraction approach (32, 34). Direct assays for total testosterone exist as commercial assays and on platforms. The recommended standard method, LC-MS, has high specificity and sensitivity, is appropriate for low levels, and is already in use in some laboratories. Advances in technology should enable more widespread use of this methodology in future for routine hormone measurement in clinical laboratories (33, 34). Comparison of testosterone results measured by direct immunoassay, with results measured by LC-MS demonstrated considerable differences, which were greater for low concentrations, as encountered in small children and women (32, 34, 35). Nevertheless, high quality LC-MS performance also requires regular maintenance and calibration by educated personnel to ensure reliable results, which is costly

and labour intensive, and consequently not always easy to implement due to costs (34). The availability and easy use of immunoassays for total testosterone, established in routine and clinical biochemistry laboratory assessment have highlighted the problem of variability between the results and methods used by different laboratories. Consequently, experts in the field have been engaged in order to recommend improvements in the measurement of testosterone (34). The conclusion in 2007 was that the current situation for testosterone measurement (both total and free testosterone) was unsatisfactory, but that the technology for accurate, precise and reproducible results already exists, and should become the standard option in future. In the meantime, the shortcomings of direct assay can be improved by validation with LC-MS. Another conclusion of the expert group was that proficiency testing assessment should be according to testosterone concentrations measured by the recommended methods and not compared to results obtained by peer methods, e.g. the result of a direct method compared to a group of results from laboratories also using direct immunoassay. Recommendations for clinicians ordering testosterone tests indicated the importance of information on the method and its quality, reference data associated to the method and laboratory, and advised avoiding direct methods for testosterone tests in children, females and hypogonadal men, due to their unsatisfactory performance. It was also suggested that serum should be extracted for immunoassay, and ideally LC-MS used when feasible. For free testosterone, the calculation method (using total testosterone and SHBG result) was preferred over immunoassays for measurement (34).

An attempt to harmonize testosterone measurement and other steroid hormones was undertaken by the Center for Disease Control, also setting the criteria of quality for certification in proficiency testing and publishing a list of certified laboratories according to the methods used (36). Among the certified laboratories using LC-MS methods, there is also a laboratory using the immunoassay method for testosterone, suggesting that implementing certain measures of quality management can improve the performance of a method which is not optimal.

Three methods exist for measurement of free testosterone, and two of them are reference methods (equilibrium dialysis and ultrafiltration). These two reference methods require expertise and are not widely available (37). In equilibrium dialysis the free testosterone from the serum passes through a semipermeable membrane, but not the protein bound testosterone (SHBG or albumin bound), and the free testosterone is measured by immunoassay in the dialysate. In ultrafiltration, the free testosterone is forced through a selective membrane by ultracentrifugation, and subsequently measured in the dialysate. Radioimmunoassay method (direct or "analogue", due to the competition of the testosterone in the reagent with the testosterone in the serum) has poor accuracy and underestimates the free testosterone, in comparison to the reference methods (38). The use of radioimmunoassay as a method for free testosterone is in decline in laboratories due to its poor quality. The third method calculates free testosterone from total testosterone and SHBG results, sometimes also including albumin results. The calculated free testosterone is comparable to the results obtained by equilibrium dialysis (39). Another calculated parameter is the free androgen index, the ratio of total testosterone and SHBG, which is considered unreliable and not clinically relevant (36). Immunoassay for free testosterone and assessment of free testosterone by calculation should not be considered as reliable as the results obtained by a reference method. The biochemistry laboratory should

advise the clinician of the limitations of the direct approach of testosterone measurement and formula-approach, which should only be used in cases of abnormal total testosterone result (37).

Vitamin D

Though not a hormone, but a precursor of the hormone 1,25-dihydroxy vitamin D, the measurement of the main circulating form 25-OH D is routinely undertaken by immunoassay-based methods (40, 41). Assessment of vitamin D status is of great clinical importance due to the epidemic proportions of hypovitaminosis D and the involvement of vitamin D in various biological processes. 25-OH D is a candidate for mass spectrometry as its measurement by immunoassay is burdened by several technical problems, e.g. the presence of metabolites that cross-react and the strong affinity to its binding protein. Commercial assays do not always have optimal clinical performance and some have been replaced (41, 42, 43, 44). In particular, at low concentrations, hypovitaminosis D may be over or under-diagnosed. Due to method-related problems, hypovitaminosis should not be diagnosed based only on the levels of immunoassay measured 25-OHD (43, 45, 46, 47).

Biotin

A problem for some immunoassay-based methods is the supplement biotin, widely used for various conditions in high doses, e.g. as a skin and hair remedy. Biotin is also a common reagent in immunoassays and its presence in blood can cause interferences. Biotin in human serum can cause the displacement of antibodies in a reaction or can block binding sites, depending on the type of method. Consequently, the results can be too low or too high. This is important for the clinician as this type of immunoassay is used in hormone measurements, but also for other analytes, such as vitamins. In cases of a discordant test result with clinical presentation and other diagnostic procedures, information on supplements should be obtained from the patient (48).

Current Standpoint and the Future of Hormone Measurement

Hormone measurement has advanced since its introduction into clinical diagnostics, and mass spectrometry for the measurement of steroid hormones is superior to commercial immunoassays. Its clear advantage is in its superior specificity, but it is not yet widely available due to technical requirements and costs. Advances in manufacturing LS-MS technology has already enabled its application in some laboratories for hormone measurements and 25-OH D. The clear advantages of immunoassays are the simplicity of the process in various hospital settings and regarding personnel, the availability of the reagents, automation, and the very acceptable costs with high throughput (11). It should be also kept in mind that the contemporary diagnostic approach demands confirmation and exclusion of diseases within a short time period. With good internal quality control (individualized to the specific laboratory and methods) and proficiency testing, the test results should be reliable with respect to known limitations and problems. Additionally, immunoassay should have traceable calibration against mass spectrometry, which can contribute to its quality.

Manufacturers are obliged to provide all the necessary information about their products, which is also mandatory according to the ISO standard for manufacturers of medical equipment, but they often fail to do so (8,11). In the race to launch a product, a thorough investigation of the method is not

performed regarding the intended use of the reagents, which may partly fail to conform to requirements for certain patient populations. The manufacturer should provide useful data for the clinician end-user, e.g. whether the method is equally reliable throughout the measuring range, in particular with regard to the range of clinical decision.

Mass spectrometry is not a uniform solution to laboratory diagnostics in endocrinology, as assay performance is dictated by clinical requirements and not by technology. As is the case with all methods in biochemistry laboratory, it is also important for the LC MS to be validated and quality control measures used to reduce inter-laboratory variations (11). For the average hospital laboratory, immunoassay will remain the solution in laboratory diagnostics for some time. Calibration of an immunoassay-based method against a reference method and using reference material for control samples, appropriate quality management, method validation and quality control, including information on the specificity and sensitivity of the assay, should ensure good medical service to the clinical specialist.

Conclusion

The medical laboratory provides a service of quick and reliable result reporting for clinical use, based on improved methodology and regulated procedures of quality management, and also supports medical decisions by supplying the clinician with specific information on the limitations of tests. Cases of incongruence of clinical presentation and test result (s) may sometimes be resolved in the joint efforts of the specialist in the medical laboratory and the clinical setting.

Authors' contributions: Conception and design: VK, NRP, NK, AŠU, and KD; Acquisition, analysis and interpretation of data: VK, NRP, NK, AŠU, and KD; Drafting the article: VK, NRP, NK, AŠU, and KD; Revising the article critically for intellectual content: VK, NRP, NK, AŠU, and KD; Approved final version of the manuscript: VK, NRP, NK, AŠU, and KD; VK, NRP, NK, AŠU, and KD.

Conflict of interest: The authors declare that they have no conflict of interest.

References

1. Westgard JO. Six Sigma Risk Analysis. Designing Analytic QC Plans for the Medical Laboratory. Madison, WI:Westgard QC, Inc; 2011.
2. Westgard JO. Risk Analysis QC Plans: Principles and Priorities. aacc12.pdf; 2012. [cited 2017 November 9]. Available from: http:// www.westgard.com.
3. Westgard JO. Developing QC Plans in a Medical Laboratory; Webinar Series sponsored by Thermo Fisher Scientific; 2014. [cited 2017 November 9]. Available from: http//www.westgard.com.
4. Ricós C, Álvarez V, Minchinela J, Fernández-Calle P, Perich C, Boned B, et al. Biologic Variation Approach to Daily Laboratory. Clin Lab Med. 2017;37(1):47-56.
5. Graham RDJ, Albarede S, Kesseler D, MacKenzie F, Mammen J, Pedersen M, et al. for the EFLM Task Finish Group – Analytical Performance Specifications for EQAS (TFG-APSEQA) Analytical performance specifications for external quality assessment – definitions and descriptions. Clin Chem Lab Med. 2017;55(7):949-55.
6. Pereira P. ISO 15189:2012 Medical laboratories - Requirements for quality and competence. Febraury 2017. [cited 2017 October 19]. Available from: https://www.westgard.com/iso-15189-2012-requirements-1.htm.
7. Guideline of the German Medical Association on Quality Assurance in medical laboratory examinations. Revision of the "Guideline of the German mecial Association on Quality Assurance in Medical Laboratory Examinations – Rili-Baek" (unauthorized translation). J Lab Med. 2015;39(1):26-69.
8. Westgard JO, Westgard SA. Basic Quality Management Systems. Chapter 12. Designing SQC procedures. Madison WI: Westgard QC, Inc.; 2014.
9. Wartofsky L, Handelsman DJ. Standardization of hormonal assays for the 21st century. J Clin Endocrinol Metab. 2010;95 (12):5141-3.

10. Vesper HW, Thienpont LM. Traceability in laboratory medicine. Clin Chem. 2009;55(6):1067-75.

11. Taylor AE, Keevil B, Huhtaniemi IT. Mass spectrometry and immunoassay: how to measure steroid hormones today and tomorrow, Eur J Endocrinology. 2015;173(2), D1-D12.

12. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, et al. Current databases on biologic variation: pros, cons and progress. Scand J Clin Lab Invest. 1999;59(7):491-500. (This database was updated in 2014).

13. Ricós C, Perich C, MinchinelaJ, Álvarez V, Simón M, Biosca C, et al. Application of biological variation – a review. Biochemia Medica. 2009;19(3):250-9.

14. Fraser CG. Biological Variation Data for setting Quality Specifications. [cited 2017 October 16]. Available from: https://www.westgard.com/guest12.htm.

15. Ricos C, Iglesias N, Garcıa-Lario JV, Simon M, Cava F, Hernandez A, et al. Within-subject biological variation in disease: Collated data and clinical consequences. Ann Clin Biochem. 2007; 44(Pt 4):343-52.

16. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010;363(22):2124-34.

17. Llopis MA, Trujillo G, Llovet MI, Tarrés E, Ibarz M, Biosca C, et al. Quality indicators and specifications for key analytical-extranalytical processes in the clinical laboratory. Five years' experience using the Six Sigma concept. Clin Chem Lab Med. 2011;49(3):463-70.

18. Westgard JO. Six Sigma quality design & control, 2nd ed. Madison WI: Westgard QC, Inc.; 2006.

19. Loh TP, Lee LC, Sethi SK, Deepak DS. Clinical consequences of erroneous laboratory results that went unnoticed for 10 days, J Clin Pathol. 2013:66(3):260-1.

20. Westgard JO, Groth T. Power functions for statistical control rules. Clin Chem. 1979;25(6):863-9.

21. Westgard S. Thinking about Three Sigma. [cited 2017 October 12] Available at: https://www.westgard.com/lesson87.htm.

22. Lenters-Westra E, Slingerland RJ. Three of 7 Hemoglobin A1c Point-of-Care instruments do not meet generally accepted analytical performance criteria. Clin Chem. 2014;60(12):1062-72.

23. Fu H, Yan, S, Wang Y, Gu D, Li X, Yang G, Liu Y. Methods comparison and bias estimation of three distinct biochemistry analytical systems in one clinical laboratory using patient samples. Clin Lab. 2016;62(1-2):187-94.

24. Bachmann LM, Boyd JC, Bruns DE, Miller WG. State of Harmonization of 24 Serum Albumin Measurement Procedures and Implications for Medical Decisionste of Harmonization of 24 Serum Albumin Measurement Procedures and Implications for Medical Decisions Clin Chem. 2017;63(3):770-9.

25. Wudy SA, Hartmann M, Homoki J. Hormonal diagnosis of 21-hydroxylase deficiency in plasma and urine of neonates using benchtop gas chromatography-mass spectrometry. J Endocrinol. 2000;165(3):679-83.

26. al Saedi S, Dean H, Dent W, Stockl E, Cronin C. Screening for congenital adrenal hyperplasia: the Delfia Screening Test overestimates serum 17hydroxyprogesterone in preterm infants. Pediatrics. 1996;97(1):100-2.

27. Matern D, Tortorelli S, Oglesbee D, Gavrilov D, Rinaldo P. Reduction of the false-positive rate in newborn screening by implementation of MS/MS-based second-tier tests: The Mayo Clinic experience (2004–2007). J Inherit Metab Dis. 2007;30(4):585-92.

28. Lee JE, Moon Y, Lee MH, Jun YH, Oh KI, Weon Choi JW. Corrected 17-Alpha-Hydroxyprogesterone Values Adjusted by a Scoring System for Screening Congenital Adrenal Hyperplasia in Premature Infants. Ann Clin Lab Sci. 2008;38(3):235-40.

29. Fingerhut R. False positive rate in newborn screening for congenital adrenal hyperplasia (CAH)–ether extraction reveals two distinct reasons for elevated 17α-hydroxyprogesterone (17-OHP) values. Steroids. 2009;74(8):662-5.

30. Janzen N, Peter M, Sander S, Steuerwald U, Terhardt M, Holtkamp U, et al. Newborn Screening for Congenital Adrenal Hyperplasia: Additional Steroid Profile using Liquid Chromatography-Tandem Mass Spectrometry J Clin Endocrinol Metab 2007;92(7): 2581-9.

31. Hawley JM, Owen LJ, Lockhart SJ, Monaghan PJ, Armston A, Chadwick CA, et al. Serum cortisol: An up-to-date assessment of routine assay performance. Clinical Chemistry 2016;62(9):1220-9.

32. Trost LW, Mulhall JP. Challenges in Testosterone Measurement, Data Interpretation, and Method-

ological Appraisal of Interventional Trials. J Sex Med. 2016;13(7):1029-46.

33. Matysik S, Liebisch G. Quantification of steroid hormones in human serum by liquid chromatography-high resolution tandem mass spectrometry.J Chromatogr A. 2017;1526:112-8.

34. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab. 2007;92(2):405-13.

35. Steinberger E, Ayala C, Hsi B, Smith KD, Rodriguez-Rigau LJ, Weidman ER, eta al. Utilization of commercial laboratory results in management of hyperandrogenism in women. Endocr Pract. 1998;4(1):1-10.

36. Centers for Disease Control and Prevention. Laboratory Quality Assurance and Standardization Programs. Available from: [updated 2017 July 25; cited 2017 September 9]. Available from: https://www.cdc.gov/labstandards/index.html.

37. Morales A, Collier CP, Clark AF. A critical appraisal of accuracy and cost of laboratory methodologies for the diagnosis of hypogonadism: the role of free testosterone assays. Can J Urol. 2012;19(3):6314-8.

38. Fritz KS, McKean AJ, Nelson JC, Wilcox RB. Analog-based free testosterone test results linked to total testosterone concentrations, not free testosterone concentrations. Clin Chem. 2008;54(3):512-6.

39. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al.; Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536-59.

40. DEQAS (Vitamin D External Quality Assessment Scheme). [cited 2017 September 9] Available from: http://deqas.org/.

41. Carter GD. 25-hydroxyvitamin D: A difficult analyte. Clin Chem. 2012;58(3):486-8.

42. Carter GD, Berry JL, Gunter E, Jones G, Jones JC, Makin HL, et at. Proficiency testing of 25-hydroxyvitamin D (25-OHD) assays. J Steroid Biochem Mol Biol. 2010;121(1-2):176-9.

43. Jones G. Interpreting vitamin D assay results: proceed with caution. Clin Am Soc Nephrol. 2015:10(2):331-4.

44. Freeman J, Wilson K, Spears R, Shalhoub V, Sibley P. Performance evaluation of four 25-hydroxyvitamin D assays to measure 25-hydroxyvitamin D2. Clin Biochem. 2015;48(16-17):1097-104.

45. Stőckl D, Sluss PM, Thienpont L M.Specification of trueness and precision of a reference measurement system for serum/plasma hydroxyvitamin D analysis. Clin Chim Acta. 2009;408(1-2):8-13.

46. Viljoen A, Singh DK, Farrington K, Twomey PJ. Analytical quality goals for 25-vitamin D based on biological variation. J Clin Lab Anal. 2011;25(2):130-3.

47. Ong L, Sawa S, Bte Sahabdeen N, Teng Tey K, Ho CS, Sethi SK. Current 25-hydroxyvitamin D assays: Do they pass the test? Clinica Chimica Acta. 2012;413(13-14):1127-34.

48. Samarasinghe S, Meah F, Singh V, Basit A, Emanuele N, Emanuele MA, et al. Biotin Interference With Routine Clinical Immunoassays Understand the Causes and Mitigate the Risks. Endocr Pract. 2017;23(8):989-98.