Opinion Paper

Ferruccio Ceriotti*, Pilar Fernandez-Calle, George G. Klee, Gunnar Nordin, Sverre Sandberg, Thomas Streichert, Joan-Lluis Vives-Corrons and Mauro Panteghini, on behalf of the EFLM Task and Finish Group on Allocation of laboratory tests to different models for performance specifications (TFG-DM)

Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference

DOI 10.1515/cclm-2016-0091
Received February 2, 2016; accepted July 12, 2016; previously published online August 9, 2016

Abstract: This paper, prepared by the EFLM Task and Finish Group on Allocation of laboratory tests to different models for performance specifications (TFG-DM), is dealing with criteria for allocating measurands to the different models for analytical performance specifications (APS) recognized in the 1st EFLM Strategic Conference Consensus Statement. Model 1, based on the effect of APS on clinical outcome, is the model of choice for measurands that have a central role in the decision-making of a specific disease or clinical situation and where cut-off/decision limits are established for either diagnosing, screening or monitoring. Total cholesterol, glucose, HbA_{1c}, serum albumin and cardiac troponins represent practical examples. Model 2 is based on components of biological variation and should be applied to measurands that do not have a central role in a specific disease or clinical situation, but where the concentration of the measurand is in a steady state. This is best achieved for measurands under strict homeostatic control in order to preserve their concentrations in the body fluid of interest, but it can also be applied to other measurands that are in a steady state in biological fluids. In this case, it is expected that the "noise" produced by the measurement procedure will not significantly alter the signal provided by the concentration of the measurand. This model especially applies to electrolytes and minerals in blood plasma (sodium, potassium, chloride, bicarbonate, calcium, magnesium, inorganic phosphate) and to creatinine, cystatin C, uric acid and total protein in plasma. Model 3, based on state-of-the-art of the measurement, should be used for all the measurands that cannot be included in models 1 or 2.

Keywords: analytical performance specifications; analytical quality; biological variation.

***Corresponding author: Ferruccio Ceriotti,** Clinical Laboratory Service, Ospedale San Raffaele Hospital, Via Olgettina 60, 20132 Milan, Italy, Fax: +390226432640, E-mail: ceriotti.ferruccio@hsr.it
Pilar Fernandez-Calle: Department of Laboratory Medicine, Hospital Universitario La Paz, Madrid, Spain
George G. Klee: Mayo Clinic College of Medicine, Mayo Clinic, Rochester, MN, USA
Gunnar Nordin: External Quality Assessment for Clinical Laboratory Investigations in Sweden (Equalis), Uppsala, Sweden
Sverre Sandberg: Norwegian Quality Improvement of Primary Care Laboratories (Noklus), Haraldsplass Diaconess Hospital, Bergen, Norway; and Department of Global Health and Primary Health Care, University of Bergen, Bergen, Norway
Thomas Streichert: Institute for Clinical Chemistry, University of Cologne, Germany
Joan-Lluis Vives-Corrons: Red Cell Pathology Unit, Hospital Clinic-IDIBAPS, University of Barcelona, Spain
Mauro Panteghini: Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Italy

Introduction

The 1st EFLM Strategic Conference, held in 2014 in Milan, defined three models to be used to derive analytical performance specifications (APS) [1]:

- Model 1: based on the effect of analytical performance on the clinical outcome;
- Model 2: based on components of biological variation (BV) of the measurands;
- Model 3: based on the state of the art of the measurement, defined as the highest level of analytical performance technically achievable.

As indicated in [1], models 1 and 2 should be preferred and a rationale for assigning measurands to one of the models should be clearly stated. The APS, which can be derived

by applying the different models to the same measurand, can be different and the decision on which model to apply should have a clear scientifically sound reason and should be based on the patient needs.

In this paper, we propose a theoretical rationale for selecting the best model that should be applied to a specific measurand, the workflow being summarized in Figure 1. As shown in the figure, the preference should be given to the first two models, while the state-of-the-art model should temporarily be used for measurands still waiting for studies on outcome-based APS or for BV data, and for measurands for which the two previous models cannot be applied.

Reasons for choosing model 1 (outcome model)

In principle, to select the outcome model for defining APS, the measurand should have a central and well-defined role in the decision making of a specific disease or clinical situation and measurement results should be applied when agreed cut-off/decision limits are established. The results should be able to directly influence the management of the patient and, consequently, his/her outcome. Before outcome-based APS are taken in use, they must have been evaluated and proven useful in studies. A disadvantage with outcome-based APS is that they often are influenced by the current measurement quality [1]. If outcome-based studies do not exist, researchers should be encouraged to perform them, either direct outcome studies based on double-blind randomized controlled trials or indirect outcome studies based on the impact of analytical performance of the test on clinical classifications or medical decisions and thereby on probability of outcomes (e.g. by simulation or decision analysis). As direct outcome studies are very challenging, indirect outcome studies will be more feasible [2]. However, simulation studies, for example, of classification of false positive and false negative, should usually be accompanied by clinical (and economical) judgment to put weight to the different outcomes, which then will decide the APS. This principle applies better to measurands for which a complete reference measurement system exists, so it is possible to define a maximum acceptable bias in absolute terms, but the same principle can be applied to less standardized measurands in terms of relative bias; also in this case, in fact, it is possible to calculate the effects in terms of misclassification. It should be noted that the same measurand could have performance specifications set by another model when used in another clinical situation.

Examples of measurands for which model 1 should be used are the following:

– Total, HDL and LDL cholesterol. These measurands are central in the definition of the cardiovascular risk

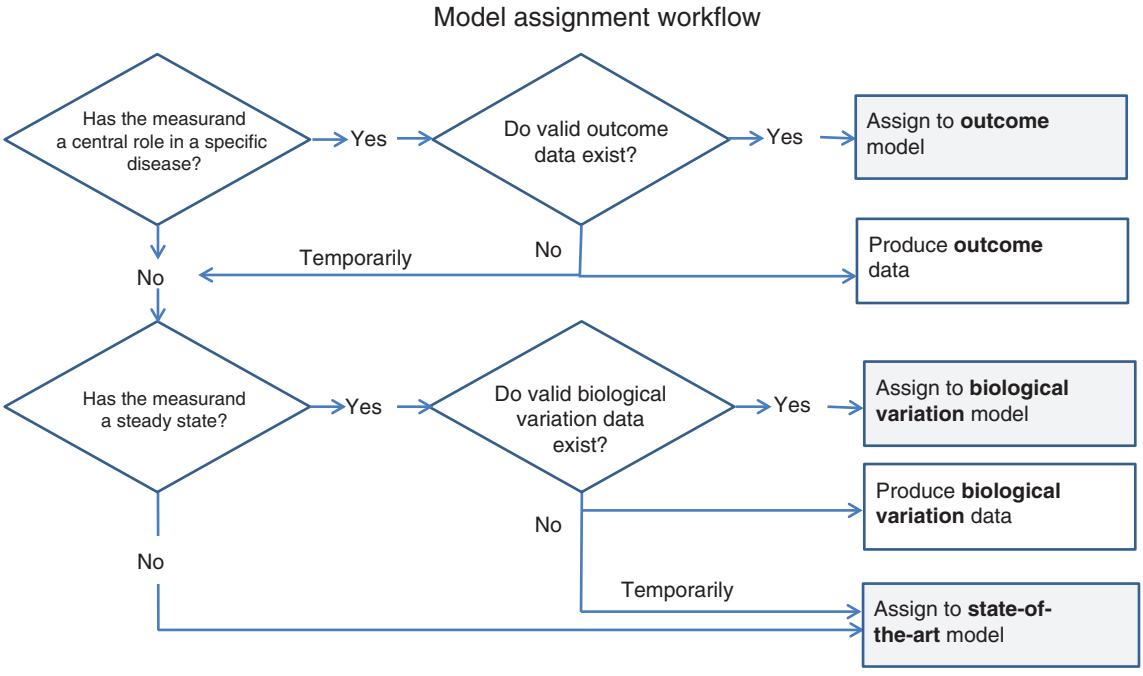

Figure 1: Workflow for assignment of a measurand to a defined analytical quality specification model.

and there are clearly defined decision thresholds and related treatment indications [3]. Even if more recent guidelines [4] are discouraging the use of multiple decision thresholds for making clinical decisions, the analytical quality of the measurement of plasma lipids and lipoproteins is crucial for avoiding misclassification of subjects with reference to their cardiovascular risk [5].

- Plasma glucose and blood HbA_{1c}. As for lipids, for glycaemia and HbA_{1c}, there are clearly defined decision limits for diagnosis and treatment monitoring of diabetes mellitus [6]. For glucose monitoring, goals for treatment are not always established and then other models can be adopted [7].
- Plasma albumin. The 2015 KDOQI guidelines call for monitoring of albumin in blood plasma as a valid and clinically useful measure of protein-energy nutritional status in dialysis patients, identifying the maintenance of concentrations >40 g/L as treatment goal [8]. Values below this concentration are highly predictive of mortality risk when present at the time of initiation of chronic dialysis as well as during the course of maintenance dialysis therapy [8, 9]. In the USA, monitoring of serum albumin concentrations is recommended in the dialyzed population with a target value of 40 g/L as a quality indicator of the performance of dialysis centres in order to obtain the reimbursement from the healthcare system. The International Myeloma Working Group recommends an albumin concentration ≥35 g/L, associated with a concentration of β_2-microglobulin in plasma <3.5 mg/L, to classify individuals with multiple myeloma at stage 1 of disease [10]. Finally, plasma albumin measurement in patients undergoing replacement therapy with human albumin is recommended for calculation of the dose to be administered and for therapy monitoring [11]. For instance, in the case of cirrhotic patients with spontaneous bacterial peritonitis or with ascites refractory to diuretic therapy or in subjects with protein-losing enteropathy or undergoing major surgery, a concentration of plasma albumin <20 g/L represents the decision level for albumin infusion.
- C-reactive protein (CRP) in blood plasma, when the protein measurement is used for discriminating between viral and bacterial infections or to establish severity of acute pancreatitis [12].
- Cardiac troponins in blood plasma. Assuming unbiased results, an imprecision <10% CV at troponin decision limit permits physicians to keep diagnostic misclassification of evaluated patients lower than 1%, while a misclassification level of 0.5% can be reached with an analytical CV <6% [13]. Troponin at the decision limit measured with a CV of 16% gave a percentage misclassification between 1.8% and 3.8%. Note that this was probably a conservative estimate, given that the impact of imprecision was derived from the results of duplicate measurements in a single assay run.
- Hemoglobin in blood. Decision limits for anaemia diagnosis (130 g/L in men, 120 g/L in non-pregnant women, 110 g/L in children) [14], for performing blood transfusion (70–80 g/L, according to the clinical situation – if the patient is stable, transfusion may not be needed even at these concentration levels -) [15–17] or for increased hemoglobin (160 g/L in females and 180 g/L in males, respectively) are defined. However, when hemoglobin is used for monitoring, the BV model can be applied [18].
- Platelets in blood. Platelets transfusion is indicated when the number concentration of platelets are $<10\times10^9$/L in clinically stable patients, while the cut-off rises to 20×10^9/L for clinically instable patients, to 50×10^9/L for minor surgical interventions and to 100×10^9/L in case of major surgery [19].
- Neutrophil leukocytes in blood. When the number concentration of neutrophils decreases to severely neutropenic range ($\leq 0.5\times10^9$/L), there is a high risk of serious infections [20].
- Thyrotropin, thyroid stimulating hormone (TSH) in blood plasma. TSH is an important biomarker for the diagnosis and monitoring of therapy in both hypothyroidism and hyperthyroidism. Several clinical practice guidelines have been developed for TSH interpretation [21–24]. For instance, the European guideline defines a mild hypothyroidism with TSH concentrations between 4.0 and 10.0 mIU/L and severe hypothyroidism with TSH >10.0 mIU/L [21]. They recommend replacement therapy with thyroid hormone to therapeutic TSH limits of 0.4–2.5 mIU/L for adult patients and TSH limits of 1.0–5.0 mIU/L in older patients. The American guideline recommends treatment to a TSH goal of 0.4–4.0 mIU/L [22]. Another European guideline defines two grades of hyperthyroidism: grade 1, with TSH 0.10–0.39 mIU/L, and grade 2, with TSH <0.1 mUI/L [23]. They recommend treating both grades in patients >65 years and only grade 2 in younger subjects. Another American guideline defines overt hyperthyroidism when TSH is depressed to <0.01 mIU/L [24]. There are, however, no clinical trials evaluating the impact of the performance of TSH assays on the application of these recommendations.

Reasons for choosing model 2 (biological variation)

The main challenge when using this model is to minimize the analytical variation to the BV. In this case, there is no direct link to the clinical use of the test, but a low ratio of the analytical noise compared to the intrinsic variability of the biological signal will ease the clinical interpretation of results. This model is possible to use for measurands that do not have a central role in a specific disease or clinical condition. The advantage is that it can be applied to most measurands for which population-based or subject-specific BV data can be established. When using population data like reference values as source of BV, it has to be considered that these data include analytical variability and so this is a combination of model 2 and 3. There are limitations to this approach, including the need to carefully assess the relevance and validity of the BV data, e.g. the presence of 'steady state', the appropriate time intervals, effect of undercurrent illness and effect of measurand concentrations. Basically, we can recognize two different situations:

1. the situation where a measurand has to be kept at a certain concentration level in the serum/plasma otherwise the body will suffer and we will get symptoms (i.e. the measurands is under strict homeostatic control);
2. the situation where a measurand de facto has a stable concentration, but deviations from this concentration will not in itself cause symptoms.

Both within- and between-subject BVs are important to set APS, taking into account variability components related to both bias and imprecision. It has been underlined that the methodology to obtain BV data should be scientifically sound [25–27]; biological CVs >33% clearly suggest a non-Gaussian distribution of the data and consequently this information may be not appropriate to calculate APS [28]. Several measurands in the database developed by Ricos' group [29] present such characteristics and should temporarily be placed to model 3 while waiting for better studies or for different mathematical approaches, e.g. as proposed in [30] for D-dimer.

As an example, the BV model should be used for the following measurands:

– Electrolytes and minerals in blood plasma (i.e. sodium, potassium, chloride, bicarbonate, calcium, magnesium, inorganic phosphate). The concentrations of these measurands are strictly controlled by hormones (e.g. aldosterone and vasopressin for sodium and potassium, parathyroid hormone for calcium and inorganic phosphate, etc.) and other mechanisms, such as renal function.
– Creatinine, urea and cystatin C in plasma. Kidney function finely controls these biomarker concentrations.
– Urate in plasma. Kidney function compensates for differences in endogenous production and dietary supplementation of urate.
– Total proteins in plasma. The relatively long half-life of the most representative proteins (in terms of plasma concentrations) and the fine hormonal control of the body water content makes the total protein concentration in plasma quite stable.
– Hema constituents [erythrocytes number concentration, erythrocyte volume fraction (hematocrit), mean corpuscular volume of erythrocytes].
– Hemoglobin in blood (when used for patient monitoring).
– Some basic coagulation test with well-defined clinical application (prothrombin time for monitoring dicumarinic therapy, activated partial thromboplastin time for monitoring therapy with heparin).

Currently, available BV data and derived APS are compiled by Ricos' group of the Spanish Society of Clinical Chemistry and Molecular Pathology and can be consulted at www.westgard.com [29]. This database is currently under revision by the EFLM Task and Finish Group "Biological variation database" (TFG-BVD).

Reasons for choosing model 3 (state of the art)

As defined in [1], "state-of-the-art" performance of measurement means the highest level of analytical performance technically achievable by field methods. This is the least preferred method because there may be no relationship between what is technically achievable and what is clinically needed. There is no official agreement on how to set APS based on this model, but a possible way to derive them is from external quality assessment programs or with some empirical method as proposed by Haeckel et al. [31]. This model should be used for measurands that cannot be included in models 1 or 2, as described above. For example, it can be temporarily used for those measurands still waiting for the definition of outcome-based APS, while waiting for BV data, or for measurands for which the two previous models do not

Table 1: Proposal for assignment of some commonly requested laboratory measurands to the three models for analytical performance specifications (APS) as defined in the Milan Consensus.[a]

APS model 1: outcome-based	APS model 2: biological variation	APS model 3: state-of-the-art
P-Cholesterol+ester	P-Sodium ion	U-Sodium ion
P-Cholesterol+ester in LDL	P-Potassium ion	U-Potassium ion
P-Cholesterol+ester in HDL	P-Chloride	U-Chloride
P-Triglycerides	P-Bicarbonate	U-Calcium ion
P-Glucose	P-Calcium ion	U-Magnesium ion
B-Hemoglobin A_{1c}	P-Magnesium ion	U-Phosphate (inorganic)
P-Albumin	P-Phosphate (inorganic)	U-Creatinine
P-Troponin T and P-troponin I	P-Creatinine	U-Urate
P-Thyrotropin	P-Cystatin C	
B-Hemoglobin	P-Urate	
B-Platelets	P-Proteins	
B-Neutrophil leukocytes	B-Erythrocytes	
	B-Erythrocyte volume fraction	
	B-Erythrocyte volume	
	P-Prothrombin time	
	P-activated partial thromboplastin time	

[a]Some of the measurands can also have APS from other models depending on their clinical use. P and B denotes the system blood plasma or whole blood, respectively. Measurements might be performed in different types of sample matrices, such as serum, heparin plasma, citrate plasma, etc., as appropriate for the method.

apply (e.g. many urinary components). As an example, this model may be used for:

– Measurands in urine, such as sodium, potassium, chloride, calcium, magnesium, inorganic phosphate, creatinine, urea, urate, etc.

Conclusions

The present paper includes a preliminary general proposal for allocating laboratory measurands to different model for deriving APS proposed in the 1st EFLM Strategic Conference Consensus Statement [1]. It includes some examples related to commonly requested tests. However, in principle, the concepts here reported should be used to define the APS for all measurands used in the clinical setting. According to Fraser et al. [32], the use of BV in defining APS may easily permit to elaborate APS at different levels of quality (i.e. minimum, desirable and optimum). The same categorization should be applied to any type of model deriving APS. This will allow starting, for example, using minimum APS and, in the meantime, asking in vitro diagnostics manufacturers to work for improving the quality of assay performance for that specific measurand in order to fulfill desirable goals [33].

Table 1 displays a preliminary list of the measurands mentioned in the paper allocated according to the three Milan models for APS. In the present paper the authors have not evaluated studies on outcome or BV data for the proposed measurands. Therefore, the measurands in Table 1 relate to the first question in Figure 1: has the measurand a central role in a specific disease (model 1) and is the measurand in a steady state in the body fluids? (model 2).

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Sandberg S, Fraser FG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–5.
2. Horvath AR, Bossuyt PM, Sandberg S, St John A, Monaghan PJ, Verhagen-Kamerbeek WD, et al. Setting analytical performance specifications based on outcome studies – is it possible? Clin Chem Lab Med 2015;53:841–8.
3. Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults. Executive summary of the third

report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation and treatment of high blood cholesterol in Adults (Adult Treatment Panel III). J Am Med Assoc 2001;285:2486–97.
4. Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129(25 Suppl 2):S1–45.
5. Langlois MR, Descamps OS, van der Laarse A, Weykamp C, Baum H, Pulkki K, et al. Clinical impact of direct HDLc and LDLc method bias in hypertriglyceridemia. A simulation study of the EAS-EFLM Collaborative Project Group. Atherosclerosis 2014;233:83–90.
6. American Diabetes Association. 5. Glycemic targets. Diabetes Care 2016;39(Suppl 1):S39–46.
7. Thue G, Sandberg S. Analytical performance specifications based on how clinicians use laboratory tests. Experiences from a post-analytical external quality assessment programme. Clin Chem Lab Med 2015;53:857–62.
8. National Kidney Foundation. KDOQI Clinical Practice Guideline for hemodialysis adequacy: 2015 update. Am J Kidney Dis 2015;66:884–930.
9. Davison AM. European best practice guidelines for haemodialysis. Nephrol Dial Transplant 2002;17(Suppl 7):1–111.
10. Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Bladé J, et al. International staging system for multiple myeloma. J Clin Oncol 2005;23:1–9.
11. Liumbruno G, Bennardello F, Lattanzio A, Piccoli P, Rossetti G. Raccomandazioni SIMTI sul corretto utilizzo degli emocomponenti e dei plasma derivati. Raccomandazioni per l'uso dell'albumina. Milano: Società Italiana di Medicina Trasfusionale e Immunoematologia (SIMTI) ed., 2008:45–58. Available at: http://www.simti.it/linee_guida.aspx?ok=1. Accessed: 4 Dec 2015.
12. Panteghini M. Laboratory evaluation of the pancreas. In: Clarke W, editor. Contemporary practice in clinical chemistry, 2nd ed. Washington DC: AACC Press, 2011:333–41.
13. Sheehan P, Blennerhassett J, Vasikaran SD. Decision limit for troponin I and assay performance. Ann Clin Biochem 2002;39:231–6.
14. World Health Organization. Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia. Geneva, Switzerland: World Health Organization, 2008.
15. Vincent JL, Baron JF, Reinhart K, Gattinoni L, Thijs L, Webb A, et al. Anemia and blood transfusions in the critically ill: an epidemiological, observational study. J Am Med Assoc 2002;288:1499–507.
16. Corwin HL, Gettinger A, Pearl RG, Fink MP, Levy MM, Abraham E, et al. The CRIT Study: anemia and blood transfusion in the critically ill–current clinical practice in the United States. Crit Care Med 2004;32:39–52.
17. Carson JL, Grossman BJ, Kleinman S, Tinmouth AT, Marques MB, Fung MK, et al. Red blood cell transfusion: a clinical practice guideline from the AABB. Ann Intern Med 2012;157:49–58.
18. Thue G, Sandberg S, Fugelli P. Clinical assessment of hemoglobin values by general practitioners related to analytical and biological variation. A study based on case stories. Scand J Clin Lab Invest 1991;51:453–9.
19. Kaufman RM, Djulbegovic B, Gernsheimer T, Kleinman S, Tinmouth AT, Capocelli KE, et al. Platelet transfusion: a clinical practice guideline from the AABB. Ann Intern Med 2015;162:205–13.
20. Newburger PE, Dale DC. Evaluation and management of patients with isolated neutropenia. Semin Hematol 2013;50:198–206.
21. Pearce SH, Brabant G, Duntas LH, Monzani F, Peeters RP, Razvi S, et al. 2013 ETA Guideline: management of subclinical hypothyroidism. Eur Thyroid J 2013;2:215–28.
22. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association Task Force on Thyroid Hormone Replacement. Thyroid 2014;24:1670–751.
23. Biondi B, Bartalena L, Cooper DS, Hegedus L, Laurberg P, Kahaly GJ. The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. Eur Thyroid J 2015;4:149–63.
24. Bahn RS, Burch HB, Cooper DS, Garber JR, Greenlee MC, Klein I, et al. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Endocr Pract 2011;17:456–520.
25. Carobene A, Braga F, Roraas T, Sandberg S, Bartlett WA. A systematic review of data on biological variation for alanine aminotransferase, aspartate aminotransferase and γ-glutamyl transferase. Clin Chem Lab Med 2013;51:1997–2007.
26. Bartlett WA, Braga F, Carobene A, Coşkun A, Prusa R, Fernandez-Calle P, et al. Biological Variation Working Group, European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). A checklist for critical appraisal of studies of biological variation. Clin Chem Lab Med 2015;53:879–85.
27. Carobene A. Reliability of biological variation data available in an online database: need for improvement. Clin Chem Lab Med 2015;53:871–7.
28. Braga F, Panteghini M. Generation of data on within-subject biological variation in laboratory medicine: an update. Crit Rev Clin Lab Sci 2016;53:313–25.
29. Available at: http://www.westgard.com/biodatabase1.htm. Desirable Biological Variation Database specifications. Accessed: 30 Dec 2015.
30. Kristoffersen AH, Petersen PH, Sandberg S. A model for calculating the within-subject biological variation and likelihood ratios for analytes with a time-dependent change in concentrations; exemplified with the use of D-dimer in suspected venous thromboembolism in healthy pregnant women. Ann Clin Biochem 2012;49:561–9.
31. Haeckel R, Wosniok W, Streichert T. Optimizing the use of the "state-of-the-art" performance criteria. Clin Chem Lab Med 2015,53:887–91.
32. Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposals for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8–12.
33. Bais R, Armbruster D, Jansen RT, Klee G, Panteghini M, Passarelli J, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med 2013;51:973–9.