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.](image-url)
...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 HbA1c. As for lipids, for glycaemia and HbA1c, 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 β2-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 <10×10⁹/L in clinically stable patients, while the cut-off rises to 20×10⁹/L for clinically instable patients, to 50×10⁹/L for minor surgical interventions and to 100×10⁹/L in case of major surgery [19].

- Neutrophil leukocytes in blood. When the number concentration of neutrophils decreases to severely neutropenic range (≤0.5×10⁹/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 mU/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.

- Heme 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
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


