We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

An Analytical Hierarchy Process to Decision Making in Human Blood, Cells, and Tissue Banks

Paulo Pereira and Sandra Xavier

Additional information is available at the end of the chapter

Abstract

The decisions are critical in transfusion and transplantation principally because they could affect directly and indirectly the post-transfusion and post-transplantation safety. However, the blood, cells, and tissues banks should not be uniquely centered in the satisfaction of receptor (patient) requirements. All the decisions affecting the sustainability of the organization should also be considered. Despite not to be systematically used in these banks, the analytical hierarchy process could be useful for the improvement in the reliability of the decisions. This chapter reviews the basics of the analytical hierarchy process applied to the production of blood, cells, and tissues, and presents a case study that could be interpreted as applicable to other situations in these organizations.

Keywords: AHP, Clinical decision, Comparisons, Decision making, Priorities, Ratings, Sustainability

1. Introduction

The decisions of blood, cells, and tissues banks, referred to as banks, are usually complex, considering the risk of failure during the collection, production, storage, and distribution of the blood, cells, and tissues. Differently from hospitals, where most of the clinical decisions are performed by physicians (an important and noncontrolled source of uncertainty), the clinical decisions in banks are taken according to what is required by the law or algorithms (a minor and controlled source of uncertainty). Decisions related to the specifications of products, such as based on laboratory quality control results, should happen using reliable methodologies to assure an acceptable risk associated with the post-transfusion or post-transplantation

safety [1, 2]. Clearly, critical decisions could not consider uniquely the safety of the receptor of blood, cells, or tissues. Some other variables are also significant to the selection of an individual product or service. Examples are the impact in the budget, which could be decisive in choosing of products or services satisfying the technical specifications. Unreliable decisions are happening from technical levels to the top management. Failures in decisions could contribute seriously to unsuccessful goals, such as the example of improper validation of an individual product is the cause of a receptor's infection.

The role of the decision maker is critical for the successful implementation of a methodology to the acceptance of new products or services. The decision must be carefully achieved. Erroneously, it could find that all the available information could be useful to contribute to accurate decisions. However, this is inappropriate since a larger quantity of information does not signify more substantial impact in the decision making (most of the information is nonsignificant). The decision maker identifies what is the critical information to a certain decision. This was already defined by Deming in the Pareto principles or "80-20 rule." This policy considers that approximately 80% of the effects come from 20% of the causes [3]. It is closely related to the cause-effect diagram, allowing an easier identification of the primary sources of a particular effect. The effect in the context of this chapter is the decision goal.

A methodology that is recognized in the industry and services is the analytical hierarchy process (AHP). However, it is not systematically applied in banks. Following this model, the goal is identified together with the criterion and subcriterion, stakeholders, and alternative actions. The options per criterion should be selected according to the considered priorities, usually ranked in a matrix table involving criteria and subcriteria [4, 5].

Frequently, banks are focused exclusively on the customer (donor and receptor). However, current management systems [6] consider it must not be uniquely centered in customer but also in all relevant interested parties (stakeholders). ISO 9000 defines stakeholder as the "person or organization that can affect, be affected by, or perceive itself to be affected by a decision or activity" (entry 3.2.3 of [7]). In the European banks, stakeholders are also the State (usually in the name of a regulatory agency) and the European Commission (the European banks must fulfill a set of directives transposed to the national law), manufacturers and suppliers including certification and accreditation providers

Despite ISO 9001 model is applied systematically to European banks, the use of decision-making models is not uniform and even rarely discussed. ISO 9001 is focused on the continuous satisfaction of the stakeholders. However, the management procedures are not standardized. The (AHP) is a recognized procedure that could be valuable to assure successful decision making. The use of this approach could also be helpful when setting the objective per process, such as when defining the target's criteria. Alternatives to the goal are considered. The AHP requires that the information related to the decision methodology is organized in a hierarchy tree using judgments to determine the ranking of the applied criteria. Banks staff easily recognize the major decisions critically related to the sustainability of the organization, such as:

1. Selection of a screening test to validate blood and tissues donations (e.g., an antihepatitis C virus immunoassay);

2. Formulating a successful marketing strategy to blood donation (e.g., growth in the number of donors from 1700 to 2000 during one year);

3. Selection of a metrology policy (e.g., to assure testing and calibration of all the monitoring and measurement equipment);

4. Choice of a statistical process control method to the monitoring of blood components production (e.g., a method based on variables and attributes charts, and capability indexes);

5. Selection of a strategy to decrease the cost of production (e.g., reduce the cost of red blood cells concentrates from €30.00 to €25.00 per component); and

6. Choosing a leader of an establishment or department (e.g., selection of a director of the bank).

This chapter briefly discusses the AHP theoretical approach applied to the banks. As an example, it uses an AHP methodology to the selection of a clinical chemistry laboratory test. Currently, the recommend methods to select a lab test are not related to AHP. Commonly, they are primarily focused on technical requirements, and in a second phase to the cost of reagent per result. However, there are other critical variables, such as reliability of the results, and the cost also has some components that are not considered. This example verifies if the application of the AHP approach could significantly improve the decision-making correctness. Although this is a unique case, the principles considered are generic.

2. Research methods

2.1. The AHP in the blood and transplantation scope

Hypothetically, the application of the AHP decreases significantly the risk of failure related to decisions. The judgment plays a crucial role in an AHP approach when considered in trans-fusion and transplantation. The judgments are required principally when standard practices are not required, i.e., when the law or standards are not considered. Some critical requirements are standardized in this scope, to assure that the products have normalized specifications to guarantee necessarily successfully transfusions or transplantations. This is also common to others industries where the production must fulfill a set of specifications usually coming from the law. Nonetheless, not all essential requirements are standardized, such as some regarding the management or even technical requirements closely related to the post-transfusion and post-transplantation safety. For example, testing and calibration methods, and quality control of laboratory tests. The staff must be skilled in judging the differentiated factors accurately to prevent biased conclusions in these examples.

The AHP methodology identifies the decision separated into phases according to the priorities. Four steps are considered [5]:

Stage 1 - Goal and knowledge: Determination of the goal to satisfy a questioned problem and selection of the knowledge needed to accomplish the objective.

Stage 2 - Decision hierarchy: Ranking the decision hierarchy from the decision goal (top), through intermediate stages, to a set of alternatives (bottom).

Stage 3 - Pairwise comparison: Conception of a set of pairwise comparison matrices where each variable is compared to upper level with the variables of the next lower level, this comparison allows the classification of criterion.

Stage 4 - Ranking the priorities: Use of the previous comparison to determine the degree of priority of each of the variables in the level below. The determination should be repeated to next level(s). An overall rating orders the classification of variables identifying the highest ranked variable.

2.2. Goal and knowledge

The banks define goals from top to bottom levels. The bottom level is related to the production processes, measured using a key performance indicator (KPI) such as the fulfillment of blood components specifications. Intermediate levels are related to the major goals such as the number of tissues produced per year. The top level is linked principally to goals regarding costumer's satisfaction.

The goals should be well-defined. It is a risky practice to establish a goal based on subjectivity. Therefore, it is determined based on empirical data (evidence-based using objective evidence) (entry 3.8.3 of [7]). Some nontechnical goals are typically not considered in banks, for example, the financial viability.

For a successful implementation of the goals, the selection and adequate training of the decision makers are required. The staff must show proper skills acquired through academic and practice. This also does not systematically happens in the bank's management scope. Considering a worst case where the managers at different levels of organization regularly do not demonstrate adequate skills, the use of AHP is useless (this is applied to any other management tool). Accordingly, the banks managers require a long-term training monitored by advisors.

2.3. Decision hierarchy and pairwise comparison matrices

Sometimes, the decision hierarchy is complex in these banks, impacting negatively the decision-making part. The decision hierarchy must be reviewed in this case. It is considered a structure, commonly designed in a hierarchy tree diagram. The criteria to reach the goal are divided into classes, for example, training, cost, and reliability. These classes are selected following the Pareto principle. If the classes do not have a significant contribution to the "correct" judgment, the final decision will be critically affected. The alternatives to the goal are represented in lower classes and are classified using the preceding criteria. The hierarchy tree complexity depends on the nature of the problem. For example, the selection of a new laboratory test requires a simpler tree when compared to the selection of a new production

method for a certain blood component. The bank applies the criteria (and subcriteria when applied) unequally, grading according to the contribution to the decision. This ranking is based on several inputs such as data, experience, expertise, consulting, and judgments. Again, the use of subjectivity is erroneous and even nonconforming (entry 2.3.6 of [7]).The construction of the tree requires leadership and the engagement of people to permit an adequate design. One more time, the staff, including the leader or leaders, must have an appropriate matrix of competencies. It is a challenge to identify the leaders of decision making based on AHP. The criteria should be correctly defined. However, the team can revise when it is demonstrated that they are not useful for the decision making [8, 9].

The pairwise comparison matrix is suited to identify the relative importance of a criterion over another to rank the priorities. For example, the reliability and cost are classified differently according to their importance to the decision. This chapter considers the numerical scale in **Table 1**. This table is adapted from (**Table 1** of [5]). Further information about matrix operations can be found elsewhere [10].

Scale	Level of importance	Description
1	Equal	Equal contribution to the goal
3	Moderate	Somewhat advantage to one variable over another according to the knowledge and judgment applied
5	Strong	Strong advantage to one variable over another according to the knowledge and judgment applied
7	Very strong	Very strong advantage to one variable over another according to the knowledge and judgment applied with evidenced supremacy based in empirical data
9	Extreme	The advantage to one variable over another is related at the highest confident level of supremacy based in empirical data

Table 1. Numerical scale to be used in pairwise comparison matrixes.

AHP contemplate priorities as the contribution of the hierarchy variables to the goal using relative metrics, for example, test A and test B in decision making to a new laboratory test. The sum of the relative results is equal to one (synonymous of goal property). The metrics should be analyzed considering the weight per variable compared to the weight of another paired variable. For example, if test A = 0.426 and test B = 0.143, test A has around three times the test B weight, therefore, test A is a primary candidate to be selected. Accordingly, "weight" is similar to priority level. Simple hierarchy scenarios could be easily computed in a spreadsheet software. However, complex problems require a dedicated software since the correct estimation of priorities is an arduous task in this scenario [9]. Additional information about decision hierarchy and pairwise comparison could be found elsewhere [11-13].

3. Case study

3.1. Goal

In blood, cells, and tissues banks, the intended use of the reported results is primarily the validation of blood components, cells, and tissues. Test validation methodology is related to the residual risk (entry 2.29 of [14]) of the reaction/infection post-transfusion or post-transplantation. Consequently, the error measured must be lower or equal to the maximum allowable error, mostly recognized as the allowable total analytical error (TEa) [15]. This allowable error is the error that could be associated with the true result (*in vivo* result) without affecting the clinical decision.
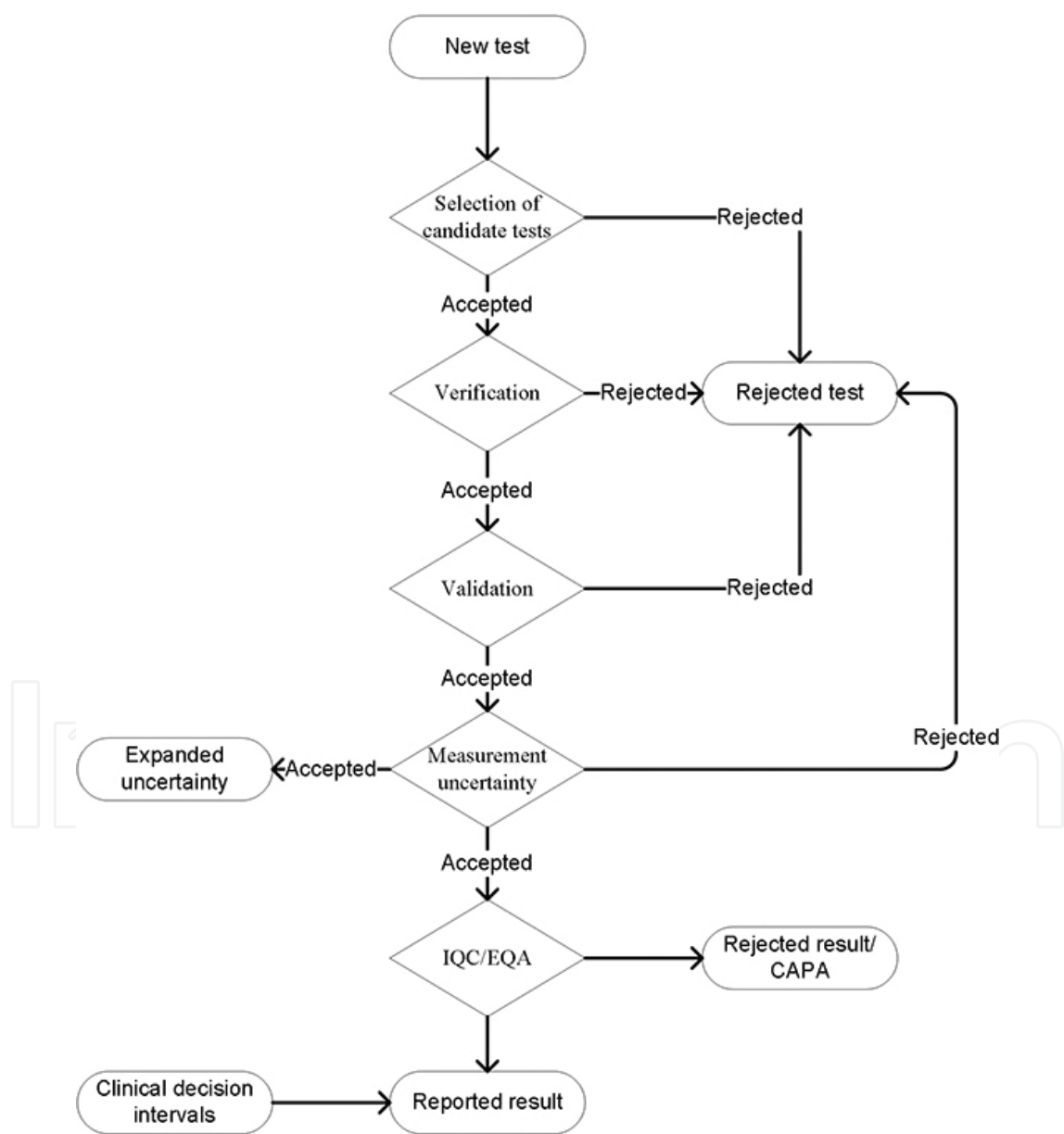

Figure 1. A flowchart to decide about of new test in a medical laboratory according to ISO 15189.

The case considers a hypothetical selection of a new laboratory test (goal). Let us consider a hematology test to the measurement of the hemoglobin (Hb) (the example could be applied similarly to any other medical laboratory test). It is chosen that three different laboratory tests are to be evaluated using the AHP. The candidate tests (alternatives) have the same photometric method.

The APH methodology differs from ISO 15189 requirements [16], intended to the accreditation of medical laboratory tests or methods. This standard aims to assure that reported results (*in vivo* results plus error) are nonsignificantly different from the *in vivo* result (true result). **Figure 1** illustrates the ISO 15189 technical requirements related to the control of the error in the results. The first step in the flowchart (entry 5.5.1.1 of [16]) is intended to decide what test or tests can be selected. The second step (entry 5.5.1.2 of [16]) is to verify if the claimed error is lower or equal to the allowable error. Tests with an error higher than the allowable limit in the bank's laboratory are rejected, despite being approved to be used in the European Union (note: manufacturers and medical laboratory specifications could not be the same). The thirds step is the validation (entry 5.5.1.3 of [16]). The lab measures the error in a representative sampling of the donors. It is compared to the allowable error. This phase is intended to verify in the laboratory conditions and with representative samples if the validation condition is acceptable. So, only tests with an estimated error lower or equal than the allowable error are accepted [20]. The fourth step is the determination of the expanded measurement uncertainty U (entry 5.5.1.4 of [16]). The approach used is commonly empirical and is expressed as follows: $U = k(u_c)$, where k is a constant related to the degrees of freedom v_{eff} (equal to the number of measurements minus one in this evaluation), and u_c is the combined standard uncertainty. In this case, the u_c is computed using an empirical mathematical model: $\sqrt{\left(u_b^2 + s_{Rw}^2\right)}$, whre u_b is the standard uncertainty of bias and s_{Rw} is the within-laboratory reproducibility standard deviation. ISO 15189 requires the U to be verified against an acceptable uncertainty interval. From the clinical point of view, the measurement uncertainty and the total analytical error concepts could be viewed as similar. The uncertainty approach (entry Annex A of [17]) is not systematically used in medical laboratories [18], different from the error approach. So, this step is like a second validation stage using the same data but with a different metrology concept. Since the purpose of the error validation and the measurement uncertainty validation are close, the AHP example considers uniquely the test validation. Further details about the estimation of measurement uncertainty are available elsewhere [17, 19]. The fifth and last step is the internal quality control (IQC) and external quality assessment (EQA) (also known as proficiency testing). Different from the validation and the measurement uncertainty models where the outcomes are primary verified before the test is used in routine, the IQC verifies the allowable error in daily runs (number of samples tested in 1 day) and it is used not to accept or reject a test but to accept or reject test results. This is a source to classify the test's reliability, i.e., the capacity to produce conforming results. The EQA has a different role, and it is used to compare the results in a group of laboratories using data arising from the same control sample. Since the EQA results are affected by the heterogeneity of the results from different laboratories, they are not considered in the AHP case. Further details about quality control schemes are available elsewhere [20].
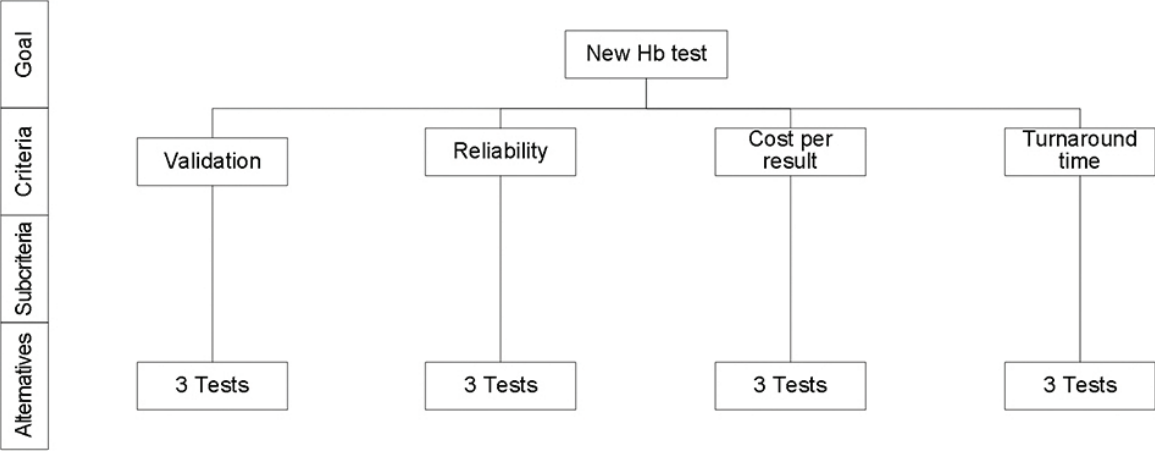
Figure 2. Hierarchical tree to the selection of a new Hb test.

The hierarchical tree of criterion is illustrated in **Figure 2**. It represents the hierarchy tree to select the Hb test. This example is a simple case. The criteria considered four variables: validation, reliability, cost per result, and turnaround time. It does not consider any subcriteria. There are three alternatives (candidates) to choose the new Hb test: test A, test B, and test C.

3.2. Pairwise comparisons matrices

A pairwise comparisons matrix is used to a paired comparison among the criteria. The fraction of a criterion over another is elaborated according to the scale in **Table 1**. It is based primarily on the relevance of each one to the post-transfusion and post-transplantation safety considering the impact of turnaround time in the production of blood components, cells, and tissues and the sustainability of the organization. This matrix is done by skilled personnel concerning the interest of parties, focused on the receptor (patient) satisfaction.

3.2.1. Criteria x criteria matrix

The matrix of criteria x matrix pondered is denoted by:

$$
\begin{array}{c}
 \\
\text{Validation} \\
\text{Reliability} \\
\text{Cost per test} \\
\text{Turnaround time}
\end{array}
\begin{bmatrix}
1/1 & 5/1 & 3/1 & 5/1 \\
1/5 & 1/1 & 1/3 & 1/3 \\
1/3 & 3/1 & 1/1 & 3/1 \\
1/5 & 3/1 & 1/3 & 1/1
\end{bmatrix}
=
\begin{array}{c}
 \\
\text{Validation} \\
\text{Reliability} \\
\text{Cost per test} \\
\text{Turnaround time}
\end{array}
\begin{bmatrix}
1 & 5 & 3 & 5 \\
0.200 & 1 & 0.333 & 0.333 \\
0.333 & 3 & 1 & 3 \\
0.200 & 3 & 0.333 & 1
\end{bmatrix}
$$

To normalize the matrix, each entry is calculated dividing each one by the sum to the associated column (e.g., the component of the first column and second row of the preceding matrix, 0.200/(1 + 0.200 + 0.333 + 0.200) = 0.115), as follows:

$$\begin{array}{c} \\ \text{Validation} \\ \text{Reliability} \\ \text{Cost per test} \\ \text{Turnaround time} \end{array} \begin{bmatrix} 0.577 & 0.417 & 0.643 & 0.536 \\ 0.115 & 0.083 & 0.071 & 0.036 \\ 0.192 & 0.250 & 0.214 & 0.321 \\ 0.115 & 0.250 & 0.071 & 0.107 \end{bmatrix}$$

The weight of criteria could be determined using the elements per row of the previous matrix. It uses the geometric mean of each matrix row to calculate the weight of a particular component. The geometric mean of a set of numbers $\{x_i\}_{i=1}^N$ is $\left(\prod_{i=1}^N x_i\right)^{1/N}$. It is calculated in Microsoft® Excel® 2016 (Microsoft Corporation, Microsoft Redmond Campus, Redmond, WA, USA) using the function =GEOMEAN(number1,[number2],…). The weight for validation, reliability, cost per test, and turnaround time is 0.536, 0.070, 0.240, and 0.122, respectively.

3.2.2. Criteria

3.2.2.1. Validation

3.2.2.1.1. Experimental procedures

Most of the statistical tools require a parametric distribution. Therefore, the laboratorian confirms the normality of the data distribution, such as the existence of outliers. The D'Agostino-Pearson K^2 normality test 21 verifies the normality of the samplings distribution. Since the p-value is equal or higher than 0.05, the normality distribution is not rejected: $p_{testA} = 0.4085$, $p_{testB} = 0.4090$, and $p_{testC} = 0.4678$. The outliers are tested using the p-value which is also equal or higher than 0.05 to the Grubbs double-sided [22]. Complementary, the Generalized Extreme Studentized Deviate (ESD) tests [23] for the area and Tukey's test [24] also does not identify outliers. The p-value is interpreted as the chance of a sample in this sampling to be an outlier and is not significant at the 0.05 significance level.

The validation follows a suitable protocol divided into two phases: preliminary phase and final phase [25]. The preliminary phase is intended to verify if random and systematic error components are conforming using simple approaches. These methods require the use of results in repeatability conditions (entry 2.20 of [26]), i.e., the within-run/nonsignificant variance of uncertainty sources, allowing the laboratorian to verify if there are significant causes of error previously identified using this tools. For example, during the transport of equipment or installation, incorrect use of mathematical models, power supply failure, wrong room temperature, and lack of personnel skills due to short experience. When detected as noncon-

forming types of error, the laboratorian avoids wasting time and resources/cost with more complex approaches (final phase).

The carry-over, interference, and recovery should be uniquely determined in the primary phase, differently from the standard deviation measurement. Their determination in the final stage is erroneous implying a waste of cost. For example, the cause of the estimation of a nonconforming reproducibility standard deviation or bias could be the carry-over of reagent. In this case, a correction is mandatory, and carry-over and reproducibility standard deviation must be reevaluated.

Carry-over determines the standard deviation increase from one sample to another sample. Carry-over could be caused by hardware failure, such as pipette malfunction. A clinically nonsignificant carry-over is the one where three times the standard deviation of the low concentration samples tested after a low concentration sample is less than the difference of the results' average of high concentration samples tested after low concentration samples, and the results' average of low concentration samples tested after low concentration samples. A low concentration sample is commonly a saline solution. Carry-over is computed using the model $\bar{x}_{high-low} - \bar{x}_{low-low} < s_{low-low}$, where $\bar{x}_{high-low}$ is the average of the difference between two consecutive samples, first with higher concentration and the second with low concentration, $\bar{x}_{low-low}$ is the average of the difference between two consecutive samples, both with low concentration, and $s_{low-low}$ is the standard deviation of the results of samples with low concentration [27]. If the difference for the three tests is lower than the $s_{low-low}$, the carry-over is classified as clinically irrelevant, as a result of test A: 0.08 < 0.12 g/dL, test B: 0.11 < 0.13 g/dL, and test C: 0.09 < 0.12 g/dL.

Interference verifies bias caused by an external component or property of the sample. It is tested a human sample added with a small volume of a pure solvent such as a saline solution. It is prepared a second solution of the same human sample but added with the same volume of the suspected interference material. These solutions are prepared to three different human samples. Each sample is tested in triplicate. For each sample the bias is computed using the difference between the result's average of sample added with interferer and result's average of sample combined with a solvent. The interference is equal to the average of the samples' bias. The interference to not be clinically significant must be equal or lower than the allowable total error. Consequently, $\bar{x}_{int} - \bar{x}_{dil} \leq TE_a$, where $\bar{x}_{int}$ is the average of the samples with interferent, and $\bar{x}_{dil}$ is the average of the samples without interferent [28]. The differences of the candidate outcomes are test A = 2%, test B = 2%, and test C = 3%, all equal or lower than 7%. Note that the TEa is available in tables for most of the medical laboratory tests, available elsewhere [29, 30]. Therefore, the error caused by interference is clinically nonsignificant. Simply, the interference test is a bias estimate in an unusual condition. Further information about carry-over and interference can be found elsewhere [31].

Recovery study confirms if any analytical bias occurs using a simple and low-cost approach. It validates the bias using two different concentration samples considering a linear regression of outcomes. Bias magnitude increases as the concentration of analyte rises. The preparation of the samples is identical to the interference model. However, the samples must have different

concentrations. Usually, it is used as supplied in the test kit and a dilution of the same standard. The dilution should not be higher than 10% to assure a lower dilution of the original sample matrix, to ensure no other significant sources of analytical bias. The recovery is calculated for each sample by dividing the difference between the sample with addition and the diluted sample by the additional volume. Then, the average of the recoveries of two samples is determined. The proportional error is calculated as follows: $\bar{x}_{high-conc} - \bar{x}_{\text{low-conc}}/\Delta * 100 \le \text{TE}_a$, where $\bar{x}_{high-conc}$ is the average of the results of the high concentration samples, $\bar{x}_{\text{low-conc}}$ is the average of the results of the low concentration samples, and Δ is the expected difference [21]. The differences of the candidates are: test A = 3%, test B = 2%, and test C = 3%, all equal or lower than 7%. Further information about recovery study can be found elsewhere [25].

The next preliminary phase test is the repeatability standard deviation s_r, which measures the random error in repeatability conditions. Westgard et al. [15] recommend a minimum of 20 replicate tests. The sample used to determine standard deviation and bias should be equal or close to the clinical decision value. High- or low-value samples should not be considered since the estimates are clinically not significant. The laboratorian should be considered using the risk of outliers (entry 4 of [32]) to control the chance of unrealistic measurements due to the use of erroneous data. A rule of thumb in statistics is to assume a normal distribution when sampling is equal or higher than 30. Note, when the distribution is unknown, and the data is less than 30, a nonparametric test is used. The criteria for acceptable performance requires it should be equal to a quarter or less of the TE$_a$ (see **Table 1**). It is not used as a final estimator of random error as its result is unrealistic since it is not affected by the long-term sources of imprecision. The clinical decision value of Hb tests is assumed to be 12 g/dL and the sample tested close to the value. It is computed using the following model $s_r \le 0.25 * \text{TE}_a$ [21]. The result for the three tests is $s_{r(\text{testA})}$ = 0.12 g/dL, $s_{r(\text{testB})}$ = 0.12 g/dL, and $s_{r(\text{testC})}$ = 0.14 g/dL. Considering the TE$_a$ = 7%, all the results should be equal or lower than 0.25*(12*0.07) = 0.8 g/dL to be accepted. Thus, the repeatability is classified as clinically nonsignificant in all the candidates.

After reporting that the claimed requirements in primary phase are fulfilled, the final stage of validation follows. In this phase, the detection limit is determined to check if the manufacturer's estimated lowest concentration of the analyte is acceptable, i.e., to verify if the standard deviation of the functional sensitivity is equal or lower than the test's standard deviation under reproducibility conditions. It uses the standard kit with a concentration equal to the claimed functional sensitivity. If the functional sensitivity is higher, it means the standard deviation is clinically significant, for which a higher concentration should be evaluated. In this case, the linearity (reportable range) is not verified, for which the reported manufacturer studies are accepted. The functional sensitivity is measured using the model $s_{fs} \le s_{\text{Rw}}$. The candidates results are $s_{(fs(\text{testA}))}$ = 0.11 g/dL, $s_{fs(\text{testB})}$ = 0.14 g/dL, and $s_{fs(\text{testC})}$ = 0.15 g/dL. All the results are equal or lower than the s_{Rw} (see the next paragraph). Further details on the evaluation of detection limit could be found in [33].

The reproducibility standard deviation s_{Rw} is the next final phase determination. All the other measurements and graphics in this stage are determined with MedCalc® software (Medcalc Software bvba, Ostend, Belgium). It expresses more realistic results due to use long-term data.

CLSI EP15-A3 requires using, at least, two samples with different concentrations. The samples are tested for at least 5 days with five replicates per daily run. If 5 days does not cover the primary imprecision sources, the period should be extended. The standard deviation is calculated by pooling the repeatability standard deviation s_r and the intermediate standard deviation s_I from between-runs. s_r and s_I are combined using the square root of the sum of the squares. The candidate tests results are $s_{Rw(testA)} = 0.19$ g/dL, $s_{Rw(testB)} = 0.21$ g/dL, and $s_{Rw(testC)} = 0.24$ g/dL, or $CV_{Rw(testA)} = 1.6\%$, $CV_{Rw(testB)} = 1.8\%$, and $CV_{Rw(testC)} = 2.0\%$. All the results are equal or lower to the TEa (7%). Further information about reproducibility standard deviation can be found elsewhere [34, 35].

Bias is determined using a comparison of paired results, where a number of samples n with different concentrations is measured in the candidate and the comparative or reference test obtaining n pairs of results (x, y). The comparison permits principally to calculate the regression equation (the case considers a linear regression) for two tests. This equation allows to correct the results of the evaluation test, transforming y in x. After the correction, the bias is determined on the clinical decision value(s). The sample is tested and replicated (minimum duplicate) during, at least, 5 days. The pair of results considers x (independent or explanatory variable) as the outcome of the reference test and y the result of the evaluation test (dependent variable). x and y are determined using the average of the replicates. The data should be displayed in the comparison plot such as in the difference plot:

a) The comparison plot reveals the interceptions of x and y results. This could indicate a simple linear regression or a constant or proportional biased regression. For easy understanding, let us consider the linear regression. It requires the calculation of the slope b and y-intercept a of the line of best fit, being $y = a + b \bullet x$. Accordingly, $x = (y - a)/b$, and bias equal to $(y - a)/b - x$. Note that in this equation x is assumed as the closest to the trueness, and $(y - x)/b$ as the corrected y, statistically identified as x, but is equivalent to x+residual systematic error. This chapter uses the Deming regression due to its robustness to outliers and to consider the random error also in the reference test [36]. **Figure 3** displays the regression between the candidate tests and a reference test. The results are significantly close due to the previous correction using the regression equation. It is shown that principally the tests A and C have nonsignificantly biased results. Test B apparently seems to have some significantly biased results. However, the regression equations in all the tests suggest significantly nonbiased estimates, since the slope is nearly one and the y-intercept is close to zero.

b) The difference plot shows the bias per pair of results (bias plotted on the y-axis, and the comparator/reference results on the x-axis). In a best case, all the points are in the zero line (unbiased results). A large dispersion of points suggests that y results are significantly biased. **Figure 4** indicates the bias percentage. It is around ±1.5% in tests A and C and around ±6% in test D.
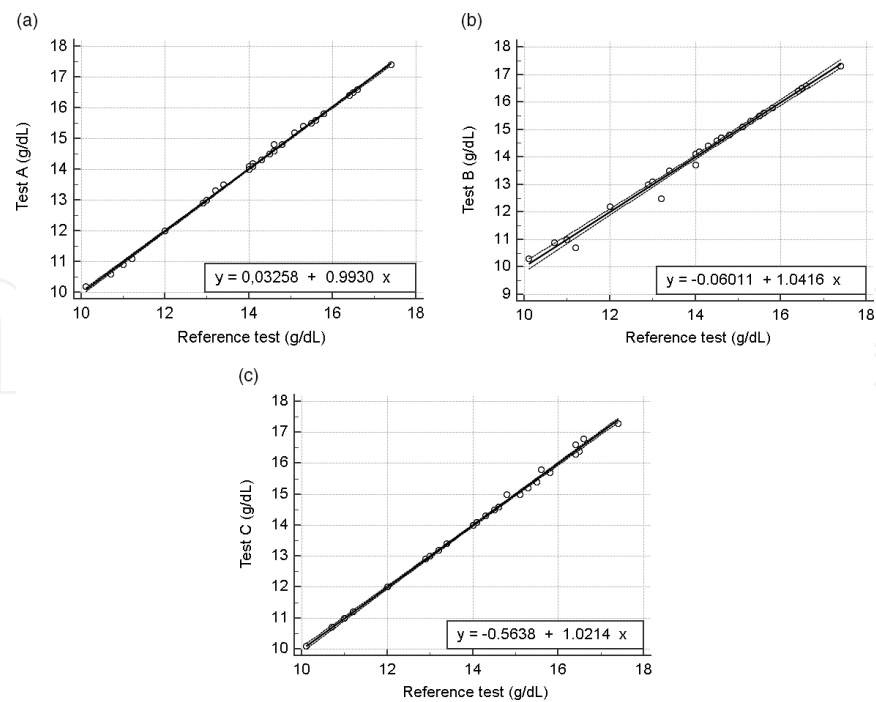
Figure 3. Scattergram, the regression line, and 95% confidence interval for the three candidate tests in MedCalc® software. The input data in the graphics is corrected according to the regression equations.
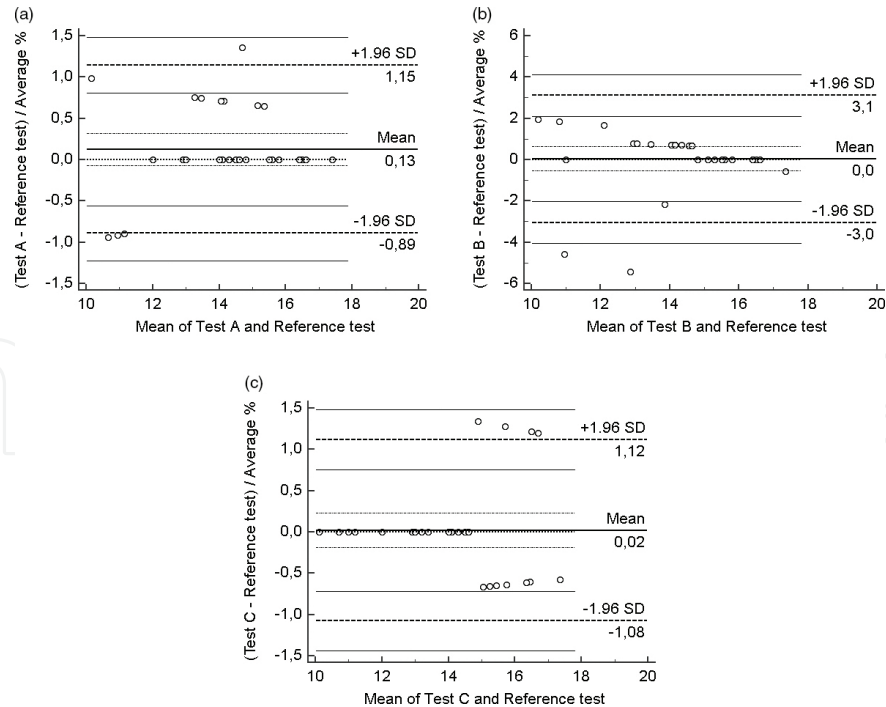

Figure 4. Bland-Altman plot for the three candidate tests (corrected input data) (MedCalc® software).

The correlation coefficient r is a measurement of the relationship between the two variables. It should be understood as the level of association between x and y. A perfect correlation result is one, in practice, if r is equal or higher than 0.99, the estimations of b and a are considered as reliable. In this case, it is recommended to determine simple linear regression statistics and calculate bias in the clinical decision value(s). If r is less than 0.99, the laboratorian should collect extra data increasing the concentration range. In this second case, bias should be estimated at the mean of the data from t-test statistics. The correlation coefficient should not be used as an outcome to validate a test, due to the bias effect to be omitted. For example, the laboratorian could watch in the graph a proportional bias but the r to be equal to 0.99. The coefficient of correlation of tests A, B, and C is $r_{testA} = 0.9993$, $r_{testB} = 0.9948$, and $r_{testC} = 0.9979$, significantly close to one [25].

The bias is determined to each of the tests in the decision value, 12 g/dL, i.e., the difference between the average of the test results and an accepted reference value. The sample chosen per test is the one which is the closest to the decision value. If there are more than one sample, it is preferred the worst case. The tests' results were previously corrected. Thus, b_{testA} = 12-12 = 0 g/dL (0%), b_{testB} = 12.2-12=0.2 g/dL (1.7%), and b_{testC} = 12-12 = 0 g/dL (0%).

Finally, the total analytical error (TAE) [37] is determined to verify the acceptability of the total error in novel test results. It is computed according to the model: TAE=b+$z \bullet s_{Rw}$, where b is the analytical bias, z is the coverage factor, which is related to the level of confidence chosen. For the approximate level of confidence of 95%, z is commonly established as z= 1.65 or z = 1.96, respectively, for a one-sided or a two-sided estimate. Alternatively, z could be determined easily in Excel® (Microsoft®, Redmond, Washington, USA) using the function =TINV(probability;deg_freedom). For a 95% confidence, the probability is equal to the difference between 1 and 0.95. Consequently, the criterion for acceptable performance is TAE ≤ TEa. The tests' TAE is as follows: TAE$_{testA}$: 0% + 1.95*1.61% = 3.1%; TAE$_{testB}$: 1.67% + 1.95*1.75% = 5.1%; and TAE$_{testC}$: 0% + 1.95*2.00% = 3.9%. All the tests satisfy the claimed TEa [14]. The sigma level is used to classify the TAE in an equation where it is related to the TEa: TEa% - b%/CV%, where CV is the coefficient of variation. Thus, Sigma$_{testA}$: (7%-0%)/1.61% = 4.4, Sigma$_{testB}$: (7%-1.67%)/1.75% = 3, and Sigma$_{testC}$: (7%-0%)/2.00% = 3.5. The test performance is classified as follows:

1. 6-sigma is classified as "world class quality";

2. 5-sigma as the goal related to quality improvement;

3. 4-sigma as the typical sigma level in industry; and

4. 3-sigma as the minimum acceptable quality level.

This case uses the TAE as the validation criterion. However, the sigma level could be employed alternatively. Detailed information about sigma level are available elsewhere [38].

These set of models differ according to different sources. CLSI harmonizes these models and is commonly the primary reference [39]. It is considered that tests with nonconforming TAE are rejected and are not included in the comparison matrices.

3.2.2.1.2. Scale

Tests with the same TAE should be classified as equal (to one). When tests have different results, the test with the lowest TAE rank is classified with one, and the other tests are categorized using the rule of three (cross-multiplication), i.e., $a/b = c/x <=> x = b*c/a$. Note that the TAE is ranked according to its statistical significance, i.e., it is assumed that the lower the result is, closer to the *in vivo* will be the *in vitro* results. Then, it is calculated as the difference between x and the highest value of the scale. The result is the closest to the scale value. Next, the metrics are applied to the next test. **Figure 5(a)** displays the primary pairwise comparisons matrix of validation and **Figure 6(a)** shows the normalized matrix and scores. Thus, in test A and test B, $x = 3.1\%*9/5.1\% = 5.5$, and $9-x = 3.5$. Therefore, the value of scale is equal to three. So, in test B and test A, $x = 1/3$.

a. Validation

	Test A	Test B	Test C
Test A	1/1	3/1	3/1
Test B	1/3	1/1	1/3
Test C	1/3	3/1	1/1

b. Reliability

	Test A	Test B	Test C
Test A	1/1	1/1	3/1
Test B	1/1	1/1	3/1
Test C	1/3	1/3	1/1

c. Cost per test

	Test A	Test B	Test C
Test A	1/1	7/1	5/1
Test B	1/7	1/1	/3
Test C	1/5	3/1	1/1

d. Turnaround time

	Test A	Test B	Test C
Test A	1/1	1/1	1/1
Test B	1/1	1/1	1/1
Test C	1/1	1/1	1/1

Figure 5. Pairwise comparisons between tests and criterion.

a. Validation

	Test A	Test B	Test C	Scores
Test A	0.600	0.429	0.692	0.563
Test B	0.200	0.143	0.077	0.130
Test C	0.200	0.429	0.231	0.270

b. Reliability

	Test A	Test B	Test C	Scores
Test A	0.429	0.429	0.429	0.429
Test B	0.429	0.429	0.429	0.429
Test C	0.143	0.143	0.143	0.143

c. Cost per test

	Test A	Test B	Test C	Scores
Test A	0.745	0.636	0.790	0.721
Test B	0.106	0.091	0.053	0.080
Test C	0.149	0.273	0.158	0.186

d. Turnaround time

	Test A	Test B	Test C	Scores
Test A	0.333	0.333	0.333	0.333
Test B	0.333	0.333	0.333	0.333
Test C	0.333	0.333	0.333	0.333

Figure 6. Normalizes matrix and scores of pairwise comparisons between tests and criterion.

3.2.3. Reliability

A reliable test is the one that has a low risk to produce nonconforming results, i.e., internal quality control results that are rejected by a simple rule or multirules in a Levey-Jennings chart. Unreliable tests are a primary cause of waste in the budget. For example, if one analytical run is rejected, the repetition of that rule means that the cost per test in the new run will be twice. The reliability is determined considering the percentage of internal quality control results during five consecutive analytical runs. The result is the percentage of the division of the accepted runs by the total number of runs. Tests A and B have five approved runs, and test C has one rejected run. In the previous criterion, the rank is determined using the rule of three. The ratios to reliability are presented in **Figure 5(b)**, and **Figure 6(b)** shows that test A and test B are equally reliable, and tests A and B have a somewhat advantage to test C. Tests A and B have five accepted runs while test C has one rejected series. Accordingly, $x = (4/5)*9/(5/5) = 7.2$, and $9-x = 1.8$. Like 1.8 is closest to three, it is assumed as the value of the scale.

3.2.4. Cost per result

The cost of each reported result is 20c, 50c, and 80c to test A, test B, and test C, respectively. This cost is uniquely related to the price of the reagent. The cost is sometimes the only criterion in decision making, which is erroneous. For example, a less expensive cost per result in an unreliable test will critically increase the cost because the rejected run will require at least a new run/new series of samples tested. The value of scale is determined using the previous criterion. **Figure 5(c)** and **Figure 6(c)** show that there is a significant difference between the costs, principally between the price of tests A and B. Using the mathematical expression, $x = 20c*9/80c = 2.3$, and $9-x = 6.7$. So, seven is the closest value of scale to the fraction (test A, text B). For test A and test C, $x = 20c*9/50c = 3.6$, and $9-x = 5.4$. Therefore, five is selected as the closest value of scale. For test B and test C, $x = 50c*9/80c = 5.6$, and $9-x = 3.4$. Then, three is the closest value.

3.2.5. Turnaround time

Turnaround time is the period that a test requires to provide the results. This is critical for the efficiency principally related to blood components acceptance and in emergency cases, such as those happening in blood and transplantation fields. A bank defines the maximum turnaround time, i.e., the time that does not affect the production of components, cells, and tissues, or uniquely reported results. So, it is needed to know if the turnaround time is equal or lower than the maximum period allowed.

The turnaround time is equal or less than 90 minutes in this study. Test A and test C time are 60 minutes, and test B is 50 minutes, for what all the ratios are equal to one. If one of the tests has a turnaround time higher than 90 minutes, the rule of three should be applied. Shorter turnaround times could be understood as an opportunity to review to enable a faster production. The fraction and outcome are displayed in **Figure 5(d)** and **Figure 6(d)**, respectively.

3.3. Overall scores

The overall rating for each alternative requires the calculation of the matrix product of two arrays. The outcome is an array with the equal number of rows as array1 and the same number of columns as array2. It is determined in Excel® spreadsheet using the function =MMULT(array1,array2). In the case, array 1 is the elements of matrix table and array 2 is the weight of criteria (see Section 3.2.1). The overall score determines the test with the highest score take from a matrix using the scores presented in:

$$
\begin{array}{c}
\text{Test A} \\
\text{Test B} \\
\text{Test C}
\end{array}
\begin{bmatrix}
0.563 & 0.429 & 0.721 & 0.333 \\
0.130 & 0.429 & 0.080 & 0.333 \\
0.270 & 0.143 & 0.186 & 0.333
\end{bmatrix}
\quad
\begin{array}{c}
\text{Overall} \\
\text{scores} \\
0.545 \\
0.160 \\
0.240
\end{array}
$$

Test A is the one with the highest score, for what it is the selected test. The result is due to its highest scores in the validation and cost per test. Comparing the overall scores with the standard test validation, it permits to classify different tests with same or close (acceptable) TAE. Therefore, the use of the AHP in decision making to a new test (or another decision, namely, technical decisions) is a more complete approach, since it contemplates other important criteria related to the goal. If the decision making considers uniquely part of the significant criteria, for instance, the validation and cost per test in the example, there is the chance to select an unreliable test. The present case could even be more complex, for instance, considering the TAE and sigma level as subcriteria of criterion validation.

4. Conclusion

In summary, the AHP provides a logical framework to determine the benefits of each alternative. In this context, discussing costs together with benefits does not bring biased estimations/answers. The case study could be inferred to a wide range of decisions in the bank, principally when dealing with specifications. The AHP demonstrates to improve the accuracy of the decision making in the study. In the example if decision making does not use the AHP methodology, the decision is biased. The incorrect decision could be demonstrated in the long term (due to reliability), on the bank's budget (due to cost per test), or on the costumers' satisfaction (due to turnaround time) in general decision making to new lab tests. The AHP application could be viewed as a complement or an extension to part of the decision approaches currently used in the banks.

Author details

Paulo Pereira[1*] and Sandra Xavier[2]

*Address all correspondence to: paulo.pereira@ipst.min-saude.pt

1 Quality Management Department, Portuguese Institute of Blood and Transplantation, Lisbon, Portugal

2 School of Health, Polytechnical Institute of Beja, Beja, Portugal

References

[1] Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. Transfus Apher Sci 2015; 52(2): 245-251. DOI: 10.1016/j.transci. 2015.02.014

[2] Pereira P, Westgard J, Encarnação P, Seghatchian J, de Sousa G. The role of uncertainty in results of screening immunoassays in blood establishments. Transfus Apher Sci 2015; 52(2): 252-255. DOI: 10.1016/j.transci.2015.02.015

[3] Koch R. The 80/20 Principle. London: Nicholas Brealey Publishing; 1997.

[4] Saaty T. How to make a decision: the analytical hierarchy process. Eur J Oper Res 1990; 48(1):9-26. DOI: 10.1016/0377-2217(90)90057-I

[5] Saaty T. Decision making with the analytic hierarchy process. Int J Serv Sci 2008; 1(1): 83-98. DOI: 10.1504/IJSSCI.2008.017590

[6] International Organization for Standardization. ISO 9001 Quality Management Systems - Requirements. 5th ed. Geneva: ISO; 2015.

[7] International Organization for Standardization. ISO 9000 Quality Management Systems - Fundamentals and Vocabulary. 4th ed. Geneva: ISO; 2015.

[8] Saaty T, Ernest F. The Hierarchon: A Dictionary of Hierarchies. Pittsburgh (PA): RWS Publications; 1992.

[9] Saaty T. Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World. Pittsburgh (PA): RWS Publications; 2008.

[10] Brown W. Matrices and Vector Spaces. New York (NY): Marcel Dekker; 1991.

[11] Kwak N, McCarthy K, Parker G. A human resource planning model for hospital/medical technologists: an analytic hierarchy process approach. J Med Syst 1997; 21(3): 173-187. DOI: 10.1023/A:1022812322966

[12] Saaty T. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York (NY): McGraw-Hill; 1980.

[13] Figuera J, Greco S, Ehrgott M. Multiple Criteria Decision Analysis: State of the Art Surveys. Berlin: Springer; 2005.

[14] International Organization for Standardization. ISO 31000 Risk Management - Principles and Guidelines. Geneva: ISO; 2009.

[15] Westgard J, Neil Carey R, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974; 20(7):825-833.

[16] International Organization for Standardization. ISO 15189 Medical Laboratories-Requirements for Quality and Competence. 2nd ed. Geneva: ISO; 2012.

[17] Bureau International des Poids et Mesures. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Sèvres: BIPM; 2008. Retrieved from http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed: May 1, 2016.

[18] Westgard S. Practical value of TEa in laboratory quality management - Historical review of quality control, quality assessment and quality management from Dr. Harris at the Aspen Conference in 1976, through "Westgard Rules" into biology and beyond. Presented at the 17th TQM conference - Quality reflections in laboratory medicine, Quality in the Spotlight, March 14 and 15, Antwerp, Belgium.

[19] Magnusson B, Näykk T, Hovind H, Krysell M. NordTest NT TR 537 Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. 3.1th ed. Oslo: Nordic Innovation; 2011. Retrieved from http://www.nordtest.info/index.php/technical-reports/item/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-3.html. Accessed: May 1, 2016.

[20] Clinical and Laboratory Standards Institute. C24-A3 Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions. 3rd ed. Philadelphia (PA): CLSI; 2006.

[21] D'Agostino, R. Transformation to normality of the null distribution of g1. Biometrika 1970; 57(3):679-681. DOI:10.1093/biomet/57.3.679

[22] Grubbs, F. Procedures for detecting outlying observations in samples. Technometrics 1969; 11(1):1-21. DOI: 10.1080/00401706.1969.10490657

[23] Rosner, B. Percentage points for a generalized ESD many-outlier procedure. Technometrics 1983; 25(2):165-172. DOI: 10.1080/00401706.1983.10487848

[24] Tukey, J. Exploratory Data Analysis. Reading (MA): Addison-Wesley Publishing Company; 1977.

[25] Westgard J. Basic Method Validation. 3rd ed. Madison (WI): Westgard QC; 2008.

[26]   Bureau International des Poids et Mesures. JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms, 2008 version with minor corrections. 3rd ed. Sèvres: BIPM; 2012. Retrieved from http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf. Accessed: May 1, 2016.

[27]   Ashwood E, Burtis C, Aldrich J (editors). Tietz Fundamentals of Clinical Chemistry. 4th ed. Philadelphia (PA): WB Saunders Co; 1996.

[28]   Burtis C, Ashwood E. Tietz Textbook of Clinical Chemistry. 3rd ed. Philadelphia (PA): WB Saunders Co; 1999.

[29]   Westgard QC. Retrieved fromml: http://www.westgard.com/quality-require-ments.htm/ Accessed: May 1, 2016.

[30]   Data Innovations. Retrieved fromml: http://www.datainnovations.com/products/ep-evaluator/allowable-total-error-table. Accessed: May 1, 2016.

[31]   Clinical and Laboratory Standards Institute. EP7-A2 Interference Testing in Clinical Chemistry. 2nd ed. Philadelphia (PA): CLSI; 2005.

[32]   Clinical and Laboratory Standards Institute. EP9-A3 Measurement Procedure Comparison and Bias Estimation Using Patient Samples. 3rd ed. Philadelphia (PA): CLSI; 2013.

[33]   Clinical and Laboratory Standards Institute. EP17-A2 Valuation of Detection Capability for Clinical Laboratory Measurement Procedures. 2nd ed. Philadelphia (PA): CLSI; 2012.

[34]   Clinical and Laboratory Standards Institute. EP15-A3User Verification of Precision and Estimation of Bias. 3rd ed. Philadelphia (PA): CLSI; 2014.

[35]   Clinical and Laboratory Standards Institute. EP5-A3 Evaluation of Precision of Quantitative Measurement Procedures. 3rd ed. Philadelphia (PA): CLSI; 2014.

[36]   Cornbleet P, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 1979; 25(3):432-438.

[37]   Clinical and Laboratory Standards Institute. EP21-A Estimation of Total Analytical Error for Clinical Laboratory Methods. Philadelphia (PA): CLSI; 2003.

[38]   Westgard J. A method evaluation decision chart (MEDx chart) for judging method performance. Clin Lab Sci 1995; 8(5):277-283.

[39]   Clinical and Laboratory Standards Institute. Retrieved fromml: http://clsi.org/ Accessed: February 26, 2016.