

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



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



The Basic Concepts of Quality Control Reference: Interval Studies, Diagnostic Efficiency, and Method Evaluation in Quality Control

Ayed Dera

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.76848>

Abstract

Laboratory data are very important in making majority of the patient's decisions. Before introducing a new test to the patients, it is very important that the acceptable performance of the test is carried out. Hence, "method evaluation" should be carried out to find out and verify the accuracy of a new test before it is used in patients. Once the method has been approved, it is the job of the laboratory personnel to utilize "quality control" techniques to maintain it. All these fall under the system of "quality management." Laboratorians use the concepts of "descriptive statistics" for comparing and analyzing different data. Descriptive statistics encompasses a variety of measures. Diagnosis in the medical field and initiation and management of various therapies depend upon the comparison of the patient's test result with a "reference interval." A specified percentage of the values for a population is used to set the lower and upper reference limits. Reference interval should be established and verified before it can be used in patients. After establishing the reference interval, the analytic and pre-analytic variables must also be standardized in order to verify and make validations of that particular reference interval. There are numerous requirement establishment of a reference interval. Establishment of reference interval requires data analysis. A number of parameters are used to find out how efficient a particular test is for predicting or nullifying a particular disease. These parameters fall under the broad heading of "diagnostic efficiency." Diagnostic efficiency encompasses "predictive values," "specificity," and "sensitivity." It is very important that accurate and reliable test results are provided by the clinical laboratory service. To enable this, a method undergoes the full process of "method evaluation." "Imprecision" and "inaccuracy" are the first estimates to be made in a method evaluation; then, they are compared with the maximum allowable medical criteria-based error. Then, the use of "quality control" and "quality control charts" follows.

Keywords: method evaluation, quality control, quality management, descriptive statistics, reference interval, diagnostic efficiency, predictive values, specificity, sensitivity, imprecision, inaccuracy

1. Introduction to method evaluation and quality management

The current nature of conducting medical transactions and procedures has revealed that most of the underlying medical decisions are arrived at utilizing laboratory data. As a result, there is the great significance that the outcomes emanating from the laboratory be of the high degree of accuracy. Determination and upholding of accuracy call for considerable cost and potential, involving the utilization of several approaches in accordance with the underlying test's complexity [1]. Invariably, commencing the entire decision-making process, one is entitled to acknowledge the necessary quality besides knowing how to measure the quality. In conjunction with that, there are several statistical techniques deployed to enable the medical practitioner to measure the resultant quality. Prior to enacting a modern test, there is the essence of determining whether the test can be pursued acceptably wherein method evaluation is deployed in verifying the acceptability that accrues to the new approaches before reporting the results to the patient. Immediately, an approach has been enacted, a necessity prevails regarding that the laboratory ensures its validity over time. Quality control is the process that facilitates the upholding of the validity accruing to the laboratory over time. All the two concepts—method evaluation and quality control—are effective constituents of quality management. Invariably, quality management entails that the aggregate testing process is directed to the chief goal of enhancing the accuracy that accrues to the laboratory results [2]. This chapter presents the basic statistical concepts besides providing a universal overview regarding the procedures crucial for enacting a new method to ensure its persistent accuracy.

2. Basic concepts of quality control

On a daily basis, too many clinical laboratories prove to generate a wide range of results. This pool of clinical lab data ought to be summarized with an aim of monitoring the test performance. The basis for tracking performance—the quality control—is descriptive statistics, which involves three key concepts: measures of spread, shape, and center.

2.1. Descriptive statistics: measures of spread, shape, and center

After close examination, a combination of nearly identical aspects typically exhibits at least some differences for a certain property like smoothness, color, potency, volume, weight, and size. Likewise, laboratory data will possess at least some measurement differences. An effective example entails that if the glucose present in a specimen is examined a hundred times in one row, then there would emerge a range of the resultant data wherein such differences in the lab values can affect outcomes of several sources. Despite the fact that measurements differ, their resultant values yield patterns whose visualization and analysis can prevail collectively.

The laboratorians describe and perceive these patterns deploying graphical representations as well as descriptive statistics. Nevertheless, once comparing and also analyzing sets of lab data, the description of the patterns can occur focusing on their spread, shape, and center. Even though the comparison of the data's center is quite typical, comparison of the spread is fairly more powerful. Nonetheless, data dispersion enables the lab practitioners to evaluate the predictability, as well as the lack of, in the lab test or rather a measurement.

2.2. Measures of center

The three typically deployed descriptions regarding the center include the mode, the median, and the mean. The mean is sometimes termed as the average of various data values. The median encompasses the "middle" point accruing to the data and is frequently deployed with fairly skewed data. The mode encounters its use rarely in describing the center of data but is often utilized in describing the data that deems to have two centers or rather bimodal data. The mean of the lab data can be acquired by summing up the total data values and dividing by the total number of samples or objects (**Figure 1**). Computing the median necessitates arrangement of the data values as per their ranks—either in an ascending manner or descending manner. Two values dominate the middle of the data, and then the median is an average of the two middle values. On the other hand, the mode entailed the most frequently appearing data value in the underlying dataset. It is often deployed in conjunction with the data's shape, bimodal distributions.

2.3. Measures of spread

The spread of the data depicts the distribution of the various data values. The spread further denotes the correlation of the entire data points to the data's mean. The descriptions of spread include standard deviation (SD), range, and coefficient of variation (CV). The range simply refers to the largest value regarding the dataset minus the dataset's smallest value. It denotes the data's extreme that one may identify standard deviation is a frequently deployed approach, especially when measuring variation. The SD and the variance denote the "average" distance notably from the data's center (mean) to every other value in the underlying dataset.

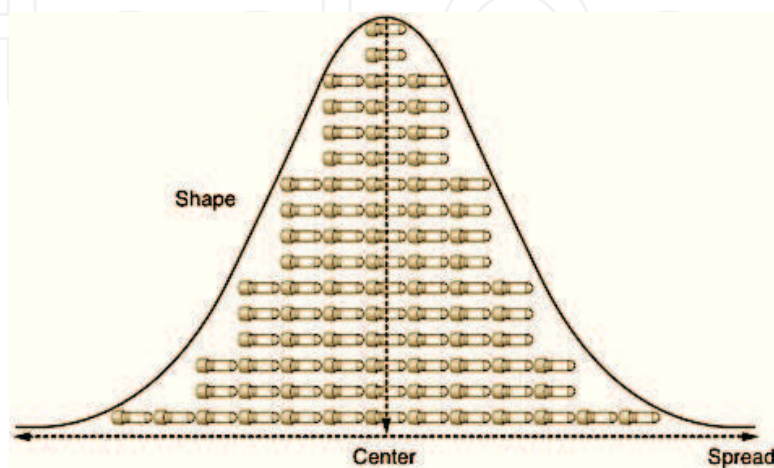


Figure 1. Basic measurements of data include the center, spread, and shape [1].

Furthermore, the CV enables the laboratorians to put up an effective comparison regarding the SDs with varying units. Computation of a dataset's SD necessitates prior computation of the dataset's variance (s^2). Variance precisely implies the average accruing to the squared distances of all the dataset's values from the set's mean. Variance, as a dispersion measure, denotes the difference dominant between each data value and the data's average. Afterward, the SD is simply the variance's square root. An additional approach of connoting SD is using the CV, which is computed via division of the SD by the mean of the data, and multiplying the quotient by 100 to represent it as a percentage (**Figure 1**). The CV proves to simplify the comparison of SDs accruing to test outcomes connoted in varying concentrations and units. The CV encounters extensive application in summarizing the underlying QC data, and it can be less than 1% for the highly precise analyzers.

2.4. Measures of shape

The most prevalent shape distributions accruing to datasets include the normal distribution (or the Gaussian distribution). This distribution proves to describe many lab variables that are continuous besides sharing various unique properties—the mode, median, and mean are identical. This distribution is further symmetric—since half of the values dominate the left side of the mean, whereas the other half is on the right side of the mean value. The symmetrical shape normally encounters the perception of being a “bell curve.” The aggregate area covered by the Gaussian curve totals to 1.0 or rather 100%. Precisely, selecting a value in a Gaussian distributed dataset reveals that there is a 68% probability of finding the value between ± 1 SD and the mean value. Likewise, there is 95% likelihood of finding the value between ± 2 SDs and the mean value. There is further 99% probability of finding the value between ± 3 SDs and the mean value of the dataset (**Figure 1**). Universally, plotting patient data in histograms makes it a simple approach to visualize the underlying distribution of the dataset. Nonetheless, one can as well perform other mathematical analyses like normality tests to affirm whether data fits into a certain distribution.

3. Descriptive statistics for groups of paired observations

COM (comparison of method) is common for laboratorians dealing with data for many patients per unit time. A COM examination entails evaluating the patient's specimens by a reference (existing) technique and a test (new) approach. The resultant data from such comparisons encompass two measurements accruing to each of the patient's specimen. Convention enables plotting of the values acquired via the reference approach on the x-axis, whereas the values yielded by the test approach dominate the y-axis. Nevertheless, linear regression is a statistical approach whose analysis offers objective measures accruing to the dispersion and location of the best fit line. A linear regression yields three aspects—the y-intercept, the correlation coefficient, and the slope. The sign of the correlation coefficient indicates the relationship between the two plotted variables, and a higher coefficient indicates the prevalence of a splendid agreement notably between the comparative methods and the test [3, 4].

The difference plot, also called the Bland–Altman plot, is an additional approach regarding visualization of paired data. This approach graphs the absolute bias or even the percent bias

(difference) prevalent between the test method and the reference approach values divided by the range of the dataset. The difference plot further enables simple comparison regarding the differences in order to previously set up maximum limits [1]. Invariably, the main difference between reference and test method depicts the underlying error. COM experiments have a correlation with prevalence of two types of errors—systematic errors and random errors. The random errors are dominant in nearly all measurements besides being either negative or positive. Random error can emanate from environmental variations, an instrument used, reagent, and operator variations. Computation of the random error calls for calculation of the dataset's SD regarding the regression line. This error implies the average distance notably between the regression line and the data. A larger random error implies a wider scattered data values. Nevertheless, if the data points were perfectly in the same alignment as the regression line, the dataset's random error or rather the standard error would be zero. On the other hand, the systematic error affects observations in a consistent manner and also in one direction. The measures of y-intercept and slope yield an estimate regarding the systematic error. Invariably, systematic error can encounter categorization into proportional and constant errors. The constant systematic errors prevail once a continual difference exists between the test approach and the underlying comparative technique values, irrespective of the dataset's concentration. A proportional error prevails once the differences accruing to the test approach and the comparative approach values are fairly proportional to the underlying analyte concentration. Whenever the slope is not equal to one, a proportional error is present in that dataset.

4. Inferential statistics

Inferential statistics is the subsequent degree of complexity past paired descriptive statistics. They are deployed in drawing conclusions or rather inferences convening the SDs or mean of two datasets. Nevertheless, inferential statistics acknowledges the relevance of data distribution regarding shape. The respective distribution is key in determining the type of inferential statistics to use in analyzing the underlying data. Data depicting Gaussian distribution is normally analyzed deploying "parametric" tests that encompass ANOVA (a Student's t-test or analysis of variance). "Nonparametric" analysis is used for the data that is not normally distributed. Reference interval studies mostly depict nonparametric tests, wherein population data frequently depict skewness [1]. A precaution entails that an inappropriate analysis regarding sound data can direct the practitioner toward drawing a wrong conclusion.

5. Reference interval studies

Lab examination data are deployed in making clinical diagnoses, managing therapy, and assessment of physiologic functionalities. Interpretation of lab data implies that the clinicians are comparing the evaluated test outcome from a certain patient with a certain reference interval. Nevertheless, reference intervals encompass all the data values defining the observations' range. All normal ranges are indeed referenced intervals, but not all reference intervals outstand to be normal ranges. The following example asserts the validity of this statement.

Considering the reference interval that accrues to therapeutic drug levels, a “normal” individual will not have any drug dominating his/her system, while a patient undergoing therapy exhibits a certain target range. The theory of developing reference intervals involves standardization of collection approaches, application of statistical techniques in analyzing reference values, and selection regarding reference populations. There are two key forms of reference interval examinations—verification of reference interval and establishment of reference intervals. Establishment of a reference interval prevails once there lacks an existing analyte or rather methodology regarding the reference or clinical lab entitled to hold the comparative studies. This approach is labor intensive besides being costly since it entails lab resources at nearly all levels and may call for 120–700 study persons. Nonetheless, verification of a reference interval, or rather transference, is done with an aim of confirming the validity accruing to a prevalent reference interval provided that the analyte is utilizing identical analytic systems (methodology and/or instrumentation). This approach is fairly common regarding the operation of the clinical labs and can call for a few study individuals like 20. In addition to that, application of reference interval can be categorized into three primary classes—diagnosis of a condition or disease, monitoring a physiologic condition, and therapeutic management. The paradigm for verification or establishment of reference intervals can be damn overwhelming notably for the clinical lab that deals with multiple degrees of reference intervals-partitions. The personnel, resource, and cost requirements necessitate that the underlying reference interval examination ought to be well structured and defined to yield timely and accurate reference intervals for the productive clinical application.

5.1. Selection of reference interval study persons

This identification of people worth of inclusion in a certain reference interval experiment necessitates definition of detailed exclusion/inclusion criteria. The inclusion criteria state the factors crucial for use in the study, whereas the exclusion criteria specify the factors that make persons inappropriate for the experiment. Selection of the right individuals facilitates the acquisition of optimal specimens that exhibit acceptable degrees of confidence. Moreover, collecting the appropriate information regarding the exclusion and inclusion criteria, like donor health status, frequently necessitates a well-documented and confidential questionnaire as well as a consent form. An additional consideration regarding the selection of the individuals encompasses additional determinants that may necessitate partitioning persons into subgroups. Such subgroups may need separate reference interval experiments.

5.2. Pre-analytic and analytic considerations

After selection of individuals for a specific reference interval examination, a key consideration entails the pre-analytic and analytic variables capable of influencing certain lab tests. Control and standardization of both variables are crucial for the generation of valid reference intervals. Additionally, some approaches are damn sensitive to interferences. For instance, mass spectrometry is resistant to interferences, while chemical approaches are sometimes highly sensitive to the same. Additional consideration entails the specific reagents used since altering to a modern agent amidst a reference examination can widen the underlying reference interval or rather transform the data distribution, maybe from bimodal to normal. Universally, a valid

reference interval study necessitates extensive knowledge regarding the analyte, methodology, instrumentation, and analytic parameters.

Furthermore, plotting a reference approach versus a test approach and establishing a linear regression are key for determining whether to verify or establish a new reference interval. A correlation coefficient of one, the slope of one, and y-intercept of zero assert that the two approaches concur and hence a mere reference interval verification examination is necessary. Conversely, a considerable difference between the two approaches implies the necessity for establishing a modern reference interval. Nonetheless, analysis of reference values involves four key approaches—bias, confidence interval, parametric method, and nonparametric approach. The nonparametric approach is suitable for the majority of the reference range intervals involving analytes that are not normally distributed. A parametric approach is valid for the observed values that depict a Gaussian distribution. Confidence interval involves a range of values covering a specific probability and it serves to show the estimates' variability besides quantifying the variability. Bias implies the difference between the reference mean and the observed means wherein a negative bias implies that the reference value exceeds the test values, whereas a positive bias implies that the test values are higher [5]. Nonetheless, there is a current development regarding statistical software packages like MedCalc, JMP, SAS/STAT, Minitab, EP Evaluator, and GraphPad Prism [1]. This development has made a manual determination of reference intervals rare.

5.3. The statistical evaluation of reference values

It consists of [6]:

- i. Segregation of the reference values into suitable groups
- ii. Assessment of the dispersal of each group
- iii. Finding out the outliers
- iv. Establishment of the reference limits

5.3.1. Segregation of the reference values into suitable groups

The corresponding reference values and the reference individuals should be segregated into suitable groups according to age, sex, etc. It is done with the purpose of reducing biological “noise” and variations among the people. Various authors have developed various criteria for segregation and statistical methods for this purpose [7].

5.3.2. Assessment of the dispersal of each group

Graphical representation of the dispersal of each group should be done, and the data should then be assessed.

5.3.3. Finding out the outliers

An outlier means a person or thing situated away or detached from the main body or system or a person or thing differing from all other members of a particular group or set. In Ref. value setup, it means a value which is incorrect or inaccurate that drifts or digresses from the

established or accepted reference values. Too many methodical problems arise during the determination of the outliers; some methods developed in 2005 seems to be the solution for it [8].

5.3.4. Establishment of the reference limits

Many parametric and nonparametric methods are available for this [6–9].

6. Diagnostic efficiency

Universally, healthy patients depict entirely different lab values from the patients having epidemics. Nonetheless, lab values typically overlap, especially between various populations. Diagnostic efficiency is the key determinant regarding the appropriateness of a test at detecting and foretelling the prevalence of a disease. Diagnostic efficiency can encompass predictive values, specificity, and sensitivity. Diagnostic sensitivity entails the potential of a test regarding detection of a certain condition, whereas diagnostic specificity involves a test's potential to correctly detect the absence accruing to a certain condition or disease [10]. A positive predictive value depicts the probability of a person having a certain disease or condition once the test is not normal, whereas negative predictive value depicts a chance for an individual not having a certain condition or disease once the test is in the reference interval. The measures of diagnostic efficiency quantify the usefulness of a test regarding a certain condition or disease. Analytical sensitivity entails the lower extent of detection regarding a certain analyte, while clinical sensitivity encompasses proportion of people who test positive to show the presence of the underlying disease. True positives (TPs) are the patients confirmed by the test to have a certain disease, while those classified as not having the condition are false negatives (FNs). Contrary to specificity and sensitivity, predictive values rely on the condition's prevalence in the population under study. Measures of the diagnostic efficiency entirely rely on the distribution accruing to test outcomes for the TPs and FNs and the cutoff utilized in defining abnormal extents. Definition of effective cutoff necessitates laboratorians to frequently deploy a graphical tool—the ROC (receiver operator characteristic) [11].

7. Method evaluation

The value accruing to medical lab service depends on its potential to offer accurate and reliable test outcomes. Method evaluation targets at the production of outcomes within clinically acceptable error to assist physicians to optimally merit their patients. Regarding the regulatory issues of method evaluation, the Centers for Medicare and Medicaid Services (CMS) and the FDA outstand as the key government agencies influencing lab testing approaches in the USA. Invariably, the FDA controls lab reagents and instruments, while the CMS controls the Clinical Lab Improvement Amendments (CLIA) [12]. Nevertheless, method selection entails gathering the technical information linked to the test, its scientific literature, and presentations. Key reasons for selecting a new approach to entail a reduction of costs, improving efficiency and quality of outcomes besides amplifying client satisfaction. A method pre-evaluation follows

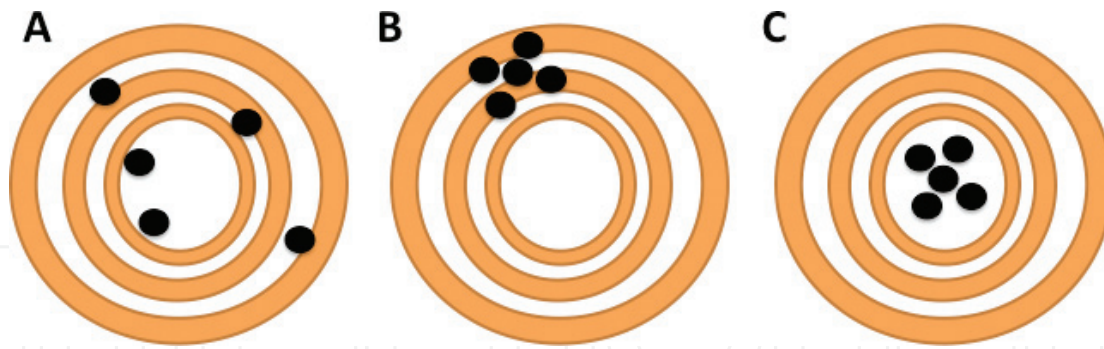


Figure 2. Graphic representation of (A) imprecision and inaccurate, (B) accurate but imprecision, and (C) precise and accurate [1].

which involves analysis of several standards with an aim of verifying the replicate analysis and linear range of two controls in order to acquire estimates regarding short-term imprecision. Inaccuracy and imprecision should be compared to the highest allowable error linked to medical criteria wherein acceptability prevails when the estimates are below the allowable highest error. After determination of imprecision, accuracy can be estimated via recovery, interference, and the patient-sample comparison. The key aspect regarding method evaluation entails determining whether the total error (systematic and random errors) does not exceed the allowable analytic error [13, 14]. The CLIA publishes the allowable analytic errors by the federally mandated proficiency examination (**Figure 2**).

8. Quality control

QC entails the systematic tracking of the analytic procedures in the lab to detect the analytic errors that prevail during analysis and finally curb reporting of incorrect test outcomes. An analytic approach is functioning optimally if the expected values lie within the underlying control limits. QC materials entail the specimens that are analyzed for QC functionality, and they ought to be of the similarity matrix as the tested specimens. Additionally, QC charts graphically denote the control material's observed values over time within the control limits. Multi-rule simplifies the various control rules to judge if an analytic approach is within the control or not. Proficiency testing is key to validating key measurement processes.

9. Quality management

Regarding quality improvement, Lean Six Sigma offers an infrastructure and methodology for quality enhancement. Additionally, define, measure, analyze, improve, and control (DMAIC) approach facilitates quality promotion. Regarding metrics, Lean Six Sigma targets at reducing cycle time, whereas Six Sigma targets at reducing error. Combining both ideologies yields a synergetic positive influence on the quality and process performance [15].

Author details

Ayed Dera

Address all correspondence to: ayedd@kku.edu.sa

Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia

References

- [1] Bishop ML, Fody EP, Schoeff LE. Clinical Chemistry: Principles, Techniques, and Correlations. 7th ed. Baltimore, MD and Philadelphia, PA: Lippincott Williams and Wilkins; Jan 13, 2013. ISBN-10: 1451118694, ISBN-13: 978-1451118698
- [2] Kaplan LA, Pesce AJ. Clinical Chemistry: Techniques, Principles, & Correlations. 5th ed. Mosby; July 2009
- [3] Westgard JO, de Vos DJ, Hunt MR, et al. Concepts and practices in the evaluation of clinical chemistry methods. V. Applications. The American Journal of Medical Technology. 1978;**44**:803-813
- [4] Wakkers PJ, Hellendoorn HB, Op de Weegh GJ, et al. Applications of statistics in clinical chemistry. A critical evaluation of regression lines. Clinica Chimica Acta. 1975;**64**:173-184
- [5] Villanova PA. C28-A2: How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline. 2nd ed. Clinical and Laboratory Standards Institute (CLSI); 2008
- [6] Burtis CA, Bruns DE. Tietz Fundamentals of Clinical Chemistry. Vol. 6. Elsevier and Saunders; 2008. pp. 231-235
- [7] Harris EK, Boyd JC. Statistical Bases of Reference Values in Laboratory Medicine. New York: Marcel Dekker; 1995
- [8] Solberg HE, Lahti A. Detection of outliers in reference distributions: Performance of Horn's algorithm. Clinical Chemistry. 2005;**51**:2326-2332
- [9] Solberg HE, Grasbeck R. Reference values. Advances in Clinical Chemistry. 1989;**27**:1-79
- [10] Galen RS, Gambino SR. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses. New York, N.Y: Wiley; 1975
- [11] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine [published erratum appears in Clin Chem 1993;39:1589]. 1993;**39**:561-577

- [12] Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid Services (CMS), Health and Human Services. Medicare, Medicaid, and CLIA programs; laboratory requirements relating to quality systems and certain personnel qualifications. Final Rule. Fed Reg 2003;**68**:3639-3714
- [13] Villanova, PA. Approved Guideline for Precision Performance of Clinical Chemistry Devices. National Committee for Clinical Laboratory Standards (NCCLS); 1999
- [14] Westgard JO, de Vos DJ, Hunt MR, et al. Concepts and practices in the evaluation of clinical chemistry methods: IV. Decisions of acceptability. The American Journal of Medicine. 1978;**44**:727-742
- [15] Ceccaroli B, Lohne O. Solar grade silicon feedstock. In: Luque A, Hegedus S, editors. Handbook of Photovoltaic Science and Engineering. 2nd ed. Chichester: Wiley; 2011. pp. 169-217. DOI: 10.1002/978047974704.ch5

