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

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

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

Youden Two-Sample Method

Julia Martín, Nieves Velázquez and
Agustin G. Asuero

Additional information is available at the end of the chapter

Abstract

The results obtained when testing materials, equipment and procedures are not generally identical. Factors that influence the magnitude of the results are not fully controllable. As such, the interpretation and analysis of results must take into account the variations caused by numerous and random unavoidable causes. Intercomparison exercises are considered of being of importance, as they do allow the examination of the analytical process and their generated results. Youden plot is particularly aimed at interlaboratory comparisons. The raw results provided by the participating laboratories are treated by a statistical method applied by the centre performing the trial. In order to materialize this, two similar materials with small differences in the concentration of the characteristics are required. The advantage of Youden analysis is its ability to separate the random errors with a minimum effort by participants in the design from the point of view of the analytical requirement. This book chapter illustrates the method that has been applied to elaborate on data covering a diverse scientific field: polyunsaturated fatty acids in fat and oils, total blood cholesterol and aspirin in pharmaceutical preparations. Finally, liquid chromatography with tandem mass spectrometry detector has been applied to the determination of an emerging contaminant, methylparaben (MeP), in surface waters.

Keywords: Youden plot, confidence ellipse, quality control

1. Introduction

The main objective of quality systems when implanted in analytical laboratories is to ensure that the results obtained confirm to quality standards, in addition to them showing a level of harmonization [1–4] between obtained results.

In order to achieve this goal, quality assessment systems are implemented so as to allow the examination of the analytical process as well as of their results generated.

Quality assessment is coined as the systematic examination carried out by an entity to verify [5–7] that it meets specified requirements (fitness for purpose). This is a generic concept that can be refined more by relating it to the specific set of activities planned and executed with the aim of ensuring that the activities involved in the quality control are done in a proper and efficient way.

The quality assessment involves the methodical and continuous contrast of the product, system or quality service. In the specific area of the laboratory, it refers to examination of systems and to analytical results generated both in terms of accuracy [8, 9] and representativeness.

Intercomparison exercises are framed in this context [10–13], establishing the procedure for design, organizing and gathering information from a set of laboratories working with the same samples that undergo an assessment of their results. An intercomparison exercise is based on acceptance by several laboratories to perform the same analysis. This analysis is carried out under the co-ordination of an organization. The purpose is to assess the quality of their work, to evaluate the method of measuring, to determine the property of a material (the content of an element or compound, etc.).

The main mission of the organization is to establish the objectives and the conditions regarding the participation of the laboratories [14–18] while ensuring the quality and stability of the sample under study. Those institutions are also responsible of dealing with the statistical treatment of obtained outcomes. The participating laboratories should, in turn, commit themselves to follow the conditions set by the organization, which may change depending on the type of executed exercise.

A very important aspect of intercomparison exercises is the selection of material to be used for the study. In that sense the type of matrix and the nature and range of values of the parameter (or parameters) under study must be defined.

It is worth mentioning at this point that not all materials are suitable to carry out a study of this type. It is essential that the material used is representative, homogeneous and stable. The organization is responsible for ensuring that the criteria mentioned above are met.

The preparation of the material must follow a series of stages, after which the material is packaged. It should, hence, be homogeneous and stable. The submitted sample must be properly identified and packaged in order to prevent breakage.

The sample is tagged accordingly so as to show the state in which the sample has been sent to the participating laboratory. Guidelines for sample preservation and handling, a description of the analytical methods to be applied as well as the report methods must also be included in the shipment.

Youden two-sample diagram, two-sample collaborative testing, two-sample plan, Youden plot or Youden analysis is particularly aimed at interlaboratory comparisons [19, 20]. The raw results provided by the participating laboratories are treated by a statistical method applied by the organizing centre of the trial. The Youden approach or z-score marks are some of the tools used for the treatment of the results. Finally, based on these statistical treatments, a statement is sent to the laboratories that have presented inaccurate results, as well as appropriate suggestions in order to improve their work.

A literature search has been carried out in order to validate current status with regards to the use of the Youden approach. The information gathered is presented in **Table 1** [1–129].

Performed literature review reveals that the Youden chart has been successfully used in agriculture, environmental chemistry, geochemistry, industry and medicine. The invention of the Youden diagram may be regarded as setting a landmark in quality control in clinical chemistry [21].

The performance and evaluation of interlaboratory programmes by Youden's method are suited to laboratory monitoring and allow to obtain information concerning both precision and systematic errors from analytical results without much effort.

Content	Reference
Key feature in analytical proficiency testing	[5]
Comparative studies of Shewhart, Thompson, Howarth and Youden representations: advantages and disadvantages	[21]
Implementation of two new graphical methods recommended by the ISO standard, Mandel's h statistic and the Youden plot, to evaluate the consistency between laboratories and within laboratories for radon and thoron exposures	[22]
A proficiency testing scheme (CNAS T0419) is described involving 217 laboratories in China as participants using their regular analytical methods for the determination of lead and arsenic in foundation cream cosmetics	[20]
An optimized Youden chart was developed and compared with the traditional and trimmed traditional Youden charts	[23]
A robust Youden plot is constructed based on robust statistical parameters since these are scarcely affected by non-normally distributed data, and this approach is applied in an external quality assessment (EQA) programme.	[24]
Youden representation for mycotoxins (deoxynivalenol and ochratoxin A) and toxins (T-2 and HT-2) in wheat and corn	[25]
Metrology statistical manual: the Youden approach with standardized variables	[26]
Interlaboratory studies: statistical organization protocol and evaluation	[27]
Control blood for an external quality assessment scheme (EQAS) for international normalized ratio (INR) point-of-care testing (POCT) in the Netherlands and to assess the performance of the participants	[28]
Application of ISO 13528 robust statistical methods for external quality assessment of blood glucose measurements in China	[29]
A study under what conditions of measurement to assess bias and from the results of a six-round blind-duplicated interlaboratory proficiency programme for creatinine in urine shows that bias is present in each individual run with components from that batch and from the laboratory over the rounds of the programme	[30]
Brazilian interlaboratory programme study on anion measurement in synthetic water. The programme described is promoted regularly since 2007 and recommended the use of ion chromatography as analytical technique for all participant laboratories	[31]
Robust determination of the correlation coefficient, analytically validated using two types of statistical models and computational simulations	[32]
Comparison of the statistical Youden method (by Hotelling T2 test and bivariate normal distribution) in interlaboratory studies by ISO and National Association of Testing Authorities (NATA) standards	[33]
Collaborative study procedures	[14]
A method validation study was conducted according to the IUPAC harmonized protocol for the determination of ochratoxin A in *Capsicum* spp. (paprika and chilli). The study involved 21 participants representing a cross section of research, private and official control laboratories from 14 EU member states and Singapore	[34]

Content	Reference
Collaborative test on the count of *Escherichia coli*	[35]
Proficiency test for the determination of heavy metals in mineral feed. The importance of correctly selecting the certified reference materials during method validation	[36]
Evaluation tools to understand statistical methods related to the *z*-score for use in proficiency testing by interlaboratory comparisons	[37]
Evaluation of learning outcomes in quantitative analysis lab using Youden plots	[38]
State of the art with respect to the selection and use of proficiency testing schemes and the interpretation of results and evaluations given in proficiency testing schemes	[39]
Performances of analytical methods for atmospheric deposition and soil analysis assessed through intercomparison exercises	[40]
HIV external quality assessment (EQA) results by the KCDC from the 17 HIV testing laboratories that also performed HIV-1 western blot testing of the 585 laboratories	[41]
Second interlaboratory exercise on non-steroidal anti-inflammatory drug analysis in environmental aqueous samples	[42]
Statics and chemometrics for analytical chemistry	[7]
A proficiency testing scheme was developed for a limited number of analytical laboratories participating in the analysis of natural water in Israel	[43]
Proficiency test for heavy metals in feed and food in Europe	[44]
A multilaboratory proficiency testing programme was conducted by the National Accreditation Board for Testing and Calibration Laboratories (India) and coordinated by the Institute of Pesticide Formulation Technology. This programme was conducted to compare the performance of individual laboratories in the area of pesticide formulation (Chlorpyrifos 20 EC) analysis. A total of 24 laboratories in India participated	[45]
Proficiency testing for the determination of pesticides in mango pulp: a view of the employed chromatographic techniques and the evaluation of laboratories' performance	[46]
Investigations for the improvement of the measurement of volatile organic compounds from floor coverings within the health-related evaluation of construction products: application of the Youden method	[47]
Characterization of candidate reference materials for bone lead via interlaboratory study and double isotope dilution mass spectrometry	[48]
Implementation and methodology of an interlaboratory system that ensures the quality of glassware calibration and use in a large laboratory	[49]
An updated liquid chromatographic assay for the determination of glyphosate in technical material and formulations: application of the Youden method	[50]
Collaborative studies for quantitative chemical analytical methods	[51]
Description and results of the 2005 interlaboratory comparison exercise for trace elements in marine mammals. Two quality control materials derived from fresh-frozen marine mammal livers were produced and characterized at the NIST and were then distributed to over 30 laboratories	[52]
Youden method applied to the external quality control of semen analysis in Germany	[53]
Quality assurance in analytical chemistry application in the environmental, food and materials analysis, biotechnology and medical engineering	[1]
Brief note on the Youden method	[54]
Interlaboratory comparison by means of method performance precision and bias studies and proficiency testing schemes are described. The set-up of the experiments and the evaluation of the data by means of graphical and statistical methods are considered	[11]

Content	Reference
Practical advice on the Youden plot	[55]
Repeatability and reproducibility of determination of the nitrogen content of fishmeal by the combustion (Dumas) method and comparison with the Kjeldahl method: interlaboratory study	[56]
Application of the Youden method in clinical chemistry: cortisol determination	[57]
An investigation of the capability of the medium resolution imaging spectrometer validation teams to determine chlorophyll a, using the latest measuring protocols and advanced high-performance liquid chromatography and spectrophotometric and fluorometric method has been performed	[58]
Standardization of calibration and quality control surface-enhanced laser desorption/ionization time of flight mass spectrometry	[59]
A chlorophyll-a interlaboratory comparison was carried out to compare three different analytical chlorophyll-a determination methods: a German standard DIN 38412-16, a method of the HELCOM-Combine-Manual and the different "in-house" methods of participating laboratories	[60]
Results for total chloride content in four different types of Portland cement provided by testing laboratories participating in an interlaboratory comparison are presented. The data sets were evaluated by using different statistical methods	[61]
A proficiency test on the quantification of trace elements in serum was carried out to verify the performance of about 30 regional laboratories of the network of Italian laboratories. The exercise consisted of four runs in which the laboratories were free in choosing analytical methods to determine trace elements in freeze-dried animal serum. Laboratory performances were evaluated by the study of statistical functions as coefficients of variation (CV), Youden plot and z-score value	[62]
Collaborative studies for cereal analysis	[10]
Practical digest for evaluating the uncertainty of analytical assays from validation data according to the LGC/VAM protocol	[63]
Statistical methods for use in proficiency testing by interlaboratory studies, International Organization for Standardization	[13]
Youden analysis of Karl Fisher titration data from an interlaboratory study determining water in animal feed, grain and forage	[64]
Establishing measurement traceability in clinical chemistry: cholesterol, progesterone and aldosterone in serum	[65]
Worldwide and regional intercomparison for the determination of organochlorine compounds and petroleum hydrocarbons in mussel tissue IAEA-432	[66]
Interlaboratory study on the determination of ascorbic acid in serum	[67]
An intercomparison of in vitro chlorophyll-a determination	[68]
Guide about collaborative studies to validate characteristics of an analytical method	[69]
Youden method application to result in total and dissolved organic carbon in surface waters	[70]
Interlaboratory exercise conducted within the framework of a hydrological project on underground water	[71]
Interlaboratory study on the determination of trace elements in sea water	[72]
State of the art with respect to the selection and use of proficiency testing schemes and the interpretation of results and evaluations given in proficiency testing schemes	[15]
Interlaboratory studies in analytical chemistry: method performance studies (collaborative trials), laboratory-performance studies (proficiency tests), collaborative bias evaluation, interlaboratory evaluation of to-be standard methods as well as certification studies for reference materials	[12]

Content	Reference
Intercomparison exercise on the determination of organochlorine compounds and petroleum hydrocarbons in algae	[73]
Statistical model assumptions upon which the procedure is based. Provides validity tests for several of these assumptions, explains conditions under which Youden is not consistent with precision estimate and indicates when precision estimates based on the procedure should be interpreted with caution or should not be used	[74]
Intralaboratory testing of method accuracy from recovery assays	[8]
Application of the Youden method to the mass fraction Youden protein fodder	[75]
Succinct description of the two-sample Youden method	[19]
Performances of analytical methods for freshwater analysis assessed through intercomparison exercises	[76]
Proposed guidelines for the internal quality control of analytical results inthe medical laboratories	[77]
Application and improvement of the Youden analysis in the intercomparison between flowmeter calibration facilities	[78]
Basic of interlaboratory studies: the trends in the new ISO 5725 standard edition	[79]
Round-robin study of performance evaluation soils vapor-fortified with volatile organic compounds	[80]
Protocol for the design, conducting and interpretation of collaborative studies	[2]
Application of the Youden method to acid rain analites	[81]
A bivariate control chart for paired measurements	[82]
Polystyrene film as a standard for testing FT-IR spectrometers	[83]
Graphical diagnosis of interlaboratory quality control data for surface water samples	[84]
Nomenclature of interlaboratory analytical studies	[85]
Basic method for the determination of repeatability and reproducibility of a standard measurement method	[86]
Reviews on the life and work of Youden	[87]
Quality control in analytical chemistry	[6]
Assessment of overall accuracy of lead isotope ratios determined by inductively coupled plasma mass spectrometry using batch quality control and the Youden two-sample method	[9]
World Health Organization international intercalibration study on dioxins and furans in human milk and blood	[88]
Analytical quality assurance. A review	[89]
Multiway analysis of variance for the interpretation of interlaboratory studies	[90]
External quality control study on the reliability of current histamine determinations in European laboratories	[91]
Guidelines for the development of standard methods of collaborative study: organization of interlaboratory studies and a simplified approach to the statistical analysis of collaborative study results	[16]
Classic paper reprint. The collaborative test of Youden	[92]
Robust statistic and functional relationship estimation for comparing the bias of analytical procedures over extended concentration ranges	[93]
Bias-free adjustment of analytical methods to laboratory samples in routine analytical procedures	[94]
Protocol for the design, conducting and interpretation of method-performance studies	[3]
Exchange of comments on a new technique in chemical assay calculations	[95]

Content	Reference
Measurement, statistics and computation, analytical chemistry by open learning. Application to aspirin preparations	[96]
Quality assurance of chemical measurements	[4]
The use of statistics to develop and evaluate analytical methods	[17]
Interlaboratory evaluation of high-performance liquid chromatographic. Determination of nitroorganics in munition plant wastewater	[97]
Interlaboratory variability in trace element analysis	[98]
The limitations of models and measurements as revealed through chemometrics intercomparison	[99]
Considerations about the graphical representation	[100]
Reverse-phase HPLC method for analysis of TNT, RDX, HMX and 2,4-DNT in munitions wastewater	[101]
Determination of heavy metals in reference marine sediments. Application of the Youden method	[102]
Organization and evaluation of interlaboratory comparison studies amongst southern African water analysis laboratories	[103]
The use of the Youden plot for internal quality control in the immunoassay laboratory	[104]
An annotation on the Youden method: recognition of the systematic and random errors	[105]
Testing laboratory performance: evaluation and accreditation	[106]
Qualification of estimates for total trace elements in food stuffs using measurement by atomic-absorption spectrophotometry	[107]
A collaborative study for measuring polyunsaturated fatty acids in fats and oils	[108]
Application of interlaboratory studies on the quality of effluent wastewaters	[109]
Statistical techniques for collaborative tests. Planning and analysis of results of collaborative tests	[18]
Interpretation and generalization of Youden's two-sample method	[110]
Collaborative analysis and the standardization of analytical methods	[111]
Graphical diagnosis of interlaboratory test results (reprinted from industrial quality control)	[112]
Systematic versus random error laboratory surveys	[113]
Precision measurement and calibration. Statistical concepts and procedure	[114]
A graphic display of interlaboratory test results	[115]
Determination of systematic an accidental errors of analytical procedure by the Youden method	[116]
Collaborative test	[117]
The sample, the procedure and the laboratory	[118]
Graphical diagnosis of interlaboratory test results	[119]
Statistical aspects of the cement testing programme	[120]
Evaluation of chemical analyses on two rocks. A simple graphical technique is proposed to aid in the comparisons between laboratories	[121]
A plan for studying the accuracy and precision of an analytical procedure	[122]
Design and interpretation of interlaboratory studies of test methods	[123]

Table 1. Some published papers dealing with the Youden approach.

2. Literature review: the Youden plot

W. J. Youden (1900–1973) was a physical chemist during the first third of this life, who turned into a statistician later, employed by the National Bureau of Standards (NBS) (now National Institute of Standards and Technology, NIST) from 1948 until his death in 1971. One of his more memorable sentences states [22, p. 12] that "The best way to find out about some of the difficulties in making measurements is to make measurements" [22].

He approached interlaboratory testing as a means of uncovering biases in measurement processes, and the so-called Youden plot has become an accepted design and analysis technique throughout the world for comparing precision and bias amongst laboratories. Youden suggested in 1959 a very simple graphical procedure for plotting results obtained by different laboratories [23–25]. Work in graphical methods, which began with the Youden plot, continues today, notably in recent works of NIST chemists.

The above is also referred to as two-sample collaborative testing, two-sample diagram, two-sample plan or Youden plot.

The method focuses on intercomparison exercises. The main characteristic is its ability to separate the systematic and random errors with minimal effort on the part of the participants.

The method is implemented as follows:

Two nearly identical samples are prepared, divided and sent to each of the participating laboratories, as recommended by Youden. A scatter plot is drawn in which the x-axis indicates one of the reported values and the y-axis the other. The scale units are the same along each axis. Each pair of results, corresponding to a given laboratory, is a point in the Youden plot (see **Figure 1**).

The points will cluster in a circular pattern whose centre is the mean values for the two samples.

Once the results are represented in the plot, they are divided into four quadrants, which are identified as (+, +), (−, +), (−, −) and (+, −). When any laboratory's result exceeds the mean achieved for all laboratories, a plus sign is used, a minus sign indicates a value smaller than the mean. If the variation in results is dominated by random errors, it would be expected that the points fall randomly distributed in all quadrants, with similar number of points in each quadrant. When systematic errors are significantly larger than random errors, then the points occur primarily in the (+, +) and the (−, −) quadrants, forming an elliptical pattern around a line bisecting these quadrants at a 45° angle.

The plot is an effective method to qualitatively evaluate the results and the capabilities of the proposed method. As can be seen in **Figure 2**, the length of a perpendicular line from any point to the 45° line is proportional to the contribution of random error on a given laboratory's results (red arrow). The distance from the intersection of the axes (mean values for samples X and Y) to the perpendicular projection of a point on the 45° line is proportional to the laboratory's systematic error (green arrow).
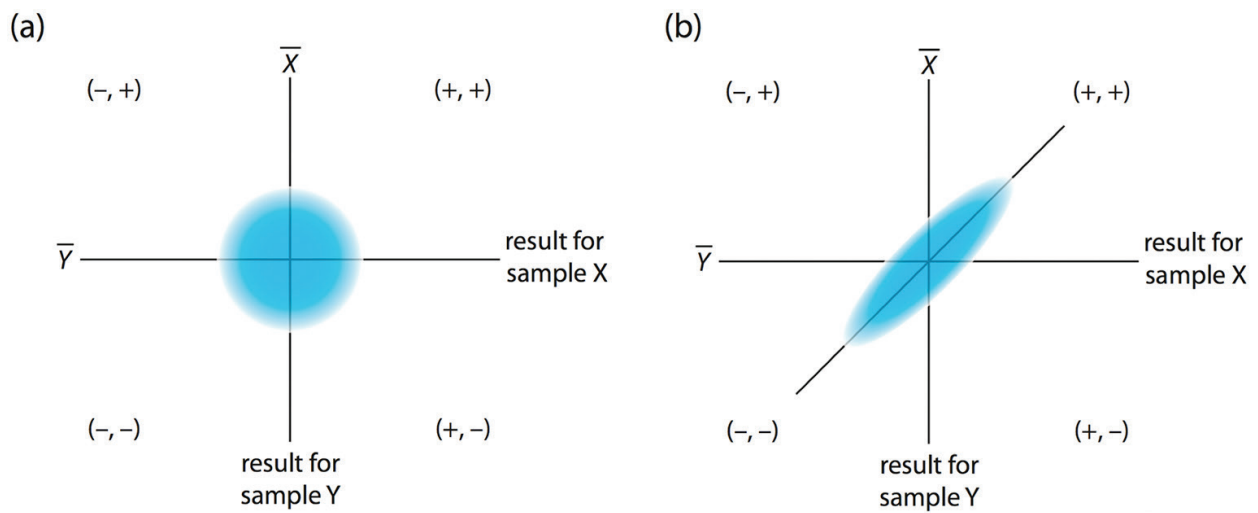
Figure 1. Typical Youden plots when (a) random errors are significantly larger than systematic errors due to the analysts and (b) when systematic errors due to the analysts are significantly larger than the random errors [126].
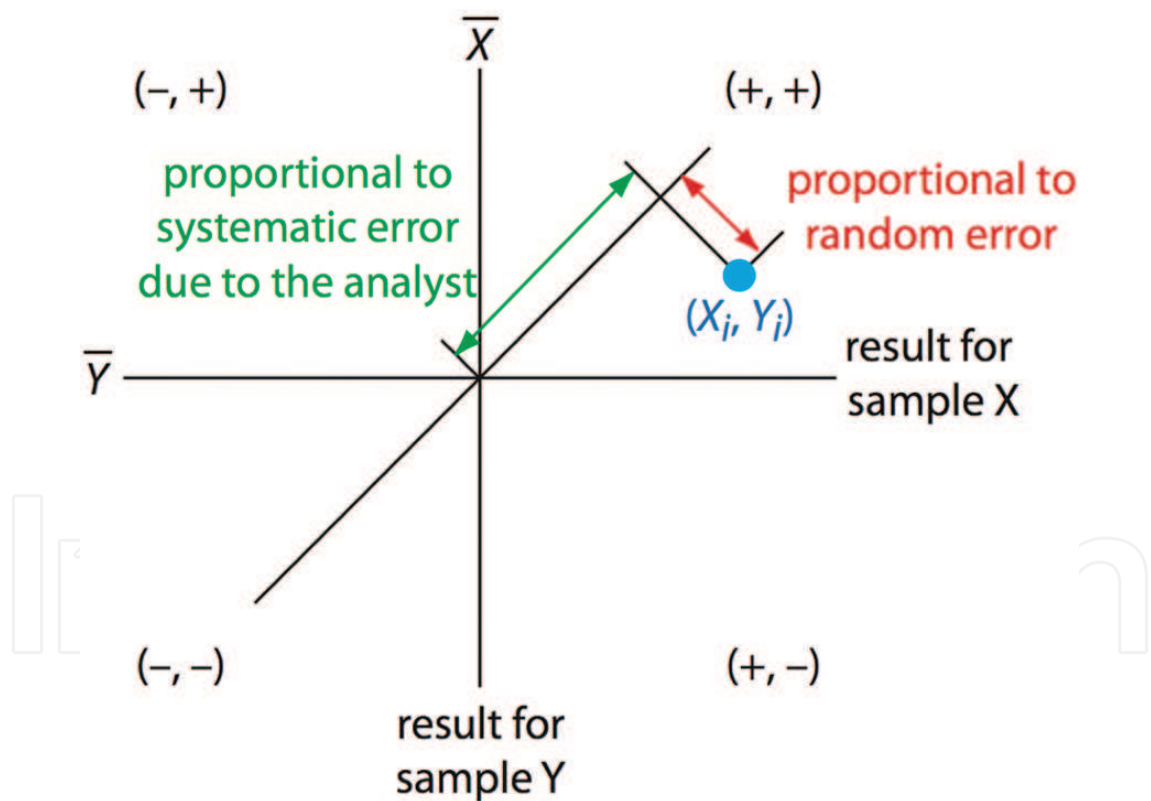

Figure 2. Relationship between the result for a single laboratory (in blue) and the contribution of random error (red arrow) and the contribution from the laboratory's systematic error (green arrow) [126].

An ideal standard method is linked with small random and systematic errors characterizing a circular compact cluster of points.

The Youden plot is a special case of the bivariate control chart to evaluate the performance of several laboratories, and the idea behind is the principal component analysis [26].

In 1974, the Youden plot was extended by Mandel and Lashof by using an ellipse instead of a circle [27]. In Youden's original method, the concentration of the analyte in the two materials was nearly the same, so that the repeatability as well as the laboratory biases would be the same for two materials [28].

Mandel and Lashof investigate the situation where two samples do not have a similar concentration so that random and systematic errors are no longer necessarily the same for both methods. They showed that in all cases, the points in the plot fall within an elongated ellipse. When Youden's original plot (two similar samples) is applied to, then the major axis forms a 45° angle. In retrospect, when samples are not similar to one another, different angles may be obtained. Their paper contains a procedure to decide whether lab bias occurs or not and also contains an estimate of all variance components.

The confidence ellipse has been proposed in ISO 13528:2005 to indicate anomalies in the between- and the within-laboratory errors in qualitative terms. For laboratory monitoring, interlaboratory tests performed according to ISO 5725-2 require much effort, especially because a large volume of samples must be provided by the organizer for $K = 4$ repeated analyses per laboratory. It is worth noting at this point that this is only suited for process standardization.

As already mentioned (see pp. 3–4), the performance and evaluation of interlaboratory programmes by Youden's method are recommended above all for laboratory monitoring. It allows to obtain information concerning both precision and systematic errors from analytical results without much effort. In addition to the above, Youden's method requires less effort for organizers and participants alike. It equally showcases a simple evaluation meaning that a potential manipulation is less likely.

Youden's method has been recommended in modern statistical manuals, procedures and protocols [29, 30] as well as recent papers, reports and government agencies, as depicted in **Table 1** [31–34].

This study discusses the Youden method and elaborates its applications in a number of diverse areas.

The experimental systems selected first for gaining experience and training in the application of the method have been the determination of polyunsaturated fatty acids in fats and oils [35], total blood cholesterol [36] and aspirin in pharmaceutical preparations [37], i.e. food and clinical and pharmaceutical applications, respectively.

Finally, a detailed procedure for the determination of methylparaben in surface waters, of special relevance nowadays in the environmental field, has been developed by using liquid chromatography-mass spectrometry.

At last, the confidence ellipse as proposed by ISO 13528:2005 will be described, and a practical case of concentrations of antibodies for two similar allergens is used as an aid to interpret the plot.

2.1. Youden plot development

The Youden plot is developed as follows:

1. Draw on the graph the points (X, Y) with the results submitted by the laboratories and reject any obvious anomalous points.

2. Calculate the centroid (X-mean, Y-mean) and draw up the lines. The vertical line is the average value for sample X, and the average value for sample Y is shown by the horizontal line.

3. Draw up the line X = Y passing through the centroid.

4. The difference, D, between the results $D_i = X_i - Y_i$ is referred to as random error. To estimate the total contribution from random error, the standard deviation of these differences, S_D, for all laboratories is used as follows:

$$S_D{}^2 = \frac{\Sigma \, (Di - \text{Dmean})^2}{2(l-1)} \tag{1}$$

where l is the number of laboratories and the factor of 2 is the result of using two values to determine D_i.

5. In the same way, the total, T, of each laboratory's results ($T_i = X_i + Y_i$) contains contributions from both random error and twice the laboratory's systematic error:

$$\sigma_{tot}^2 = \sigma_{rand}^2 + 2\,\sigma_{syst}^2 \tag{2}$$

6. The standard deviation of the totals, S_T, provides an estimate for σ_{tot}:

$$S_T{}^2 = \frac{\Sigma(Ti - \text{Tmean})^2}{2(l-1)} \tag{3}$$

Again, the factor of 2 in the denominator is the result of using two values to determine T_i.

7. If the systematic errors are significantly larger than the random errors, then S_T is larger than S_D, a hypothesis that can be evaluated using a one-tailed F-test, where the degrees of freedom for both the numerator and the denominator are n−1.

8. If S_T is significantly larger than S_D, σ_{tot}^2 may be split into components representing random error and systematic error:

$$\sigma_{SIS}{}^2 = \frac{S_T{}^2 - S_D{}^2}{2} \tag{4}$$

9. Calculating the radius of the confidence circle:

$$R = SD \cdot b \tag{5}$$

$$b = -2 \ln\left(1 - \frac{P}{100}\right) \tag{6}$$

$$P = 100\left(1 - \exp\left(\frac{-b^2}{2}\right)\right) \tag{7}$$

where % P is the percentage of selected confidence level (usually 95%).

10. Draw a circle with radius R and centroid (X-mean, Y-mean). The laboratories falling outside the 95% circle are said to provide biased results. The radius of the circle is based on a multiple of SD, depending on the desired percentage of observations anticipated to fall within a bivariate normal distribution. A circle whose radius is a multiple of SD = (2.5; 3) represents the smallest circle that can be contained in almost every point, in the absence of bias.

2.2. An alternative approximation: the Z-score

An alternative to Youden Plot is the punctuation Z-Score. This value is used to "score" a parameter in a particular round of a laboratory's participation. This is done by means of the following calculations:

1. Calculate the median of X and Y.

2. Calculate the total error (ε_T) for each laboratory:

$$\varepsilon_T = \sqrt{(x_i - x_{\text{Me}})^2 + (y_i - y_{\text{Me}})^2} \tag{8}$$

3. Calculate the systematic error component (C_s) as

$$C_s = \frac{(x_i - x_{Med}) + (y_i - y_{\text{Me}})}{\sqrt{2}} \tag{9}$$

4. Calculate the random error component (C_R) as

$$C_R = \sqrt{\left(x_1 - \left(\frac{C_S}{\sqrt{2}}\right) + x_{\text{Me}}\right)^2 + \left(y_1 - \left(\frac{C_S}{\sqrt{2}}\right) + y_{\text{Me}}\right)^2} \tag{10}$$

5. Systematic and random components calculation required so that their sum is equal to the magnitude of the total error:

Systematic error:

$$\varepsilon_S = \frac{C_s}{|C_s| + C_R} \varepsilon_T \tag{11}$$

Random error:

$$\varepsilon_R = \frac{C_R}{|C_s| + C_R} \varepsilon_T \tag{12}$$

Thus,

$$E_T = |\varepsilon_S| + \varepsilon_R \tag{13}$$

6. Calculate the typical deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_{Ri}^{2}}{n-1}} \tag{14}$$

where n is the number of laboratories.

7. Finally, the z-score is calculated as

$$z = \frac{x_i - x_{Me}}{\sigma} \tag{15}$$

The results are classified as

$|z| \leq 2$	"satisfactory"
$2 < |z| \leq 3$	"questionable"
$|z| > 3$	"unsatisfactory"

3. Application to experimental systems

3.1. Determination of polyunsaturated fatty acids in fats and oils

To illustrate the procedure, data from the interlaboratory study of a method for determining polyunsaturated fatty acids (PUFA) in fats and oils is used. The procedure consisted of saponifying a sample, treating it with an enzyme and measuring the absorbing product at 234 nm. Palm oil, corn oil and three hydrogenated blends were used in the study. One blend of hydrogenated oil was separated into two parts and designated as samples X and Y, respectively. Random subsamples from the two samples were analysed in each of the 17 laboratories as blind duplicates. The test incorporated an official Food and Drug Administration (FDA) method and a slightly modified one using boron trifluoride-methanol (BF). The aim in this case was to identify the laboratories that have higher quality results.

The results for FDA method are shown in **Table 2**. The data from laboratory 1 were rejected because inaccurate results were proportioned; those from laboratory 11 are listed but were not used (because their results were beyond the ones shown by others (8.2 g/100 g for sample X and 26.3 g/100 g for sample Y) in the calculations.

The vertical line at 28.6 g/100 g is the average value for sample X, and the average value for sample Y is shown by the horizontal line at 28.2 g/100 g. To estimate σ_{rand} and σ_{syst}, the values for D_i and T_i are calculated first. Next, the standard deviations for the differences, S_D, and the totals, S_T, are computed using Eqs. (1) and (3), yielding S_D = 1.53 and S_T = 3.11. To determine if the systematic errors between the laboratories are significant, the F-test is applied so as to compare S_T and S_D.

Because the F-ratio (4.141) is larger than F(0.05,14,14), which is 2.484, it is concluded that the systematic errors between the analysts are significant at the 95% confidence level, which is estimated using Eq. (4) giving 3.67.

Laboratory	FDA method		Laboratory	BF method	
	Sample X	Sample Y		Sample X	Sample Y
2	26.1	28.5	2	24.9	29.4
3	29.6	28.6	3	30.3	29.4
4	29.2	26.8	4	29.1	31.7
5	29.5	26.9	5	31.4	29.3
6	30.3	30.8	6	29.1	30.4
7	27.5	25.9	7	26.6	26.6
8	25.8	26.9	8	30.0	30.7
9	30.0	28.0	9	29.5	29.7
10	29.0	25.0	10	28.3	29.3
11*	8.20	26.3	11*	10.3	29.6
12	31.3	32.0	12*	10.5	10.1
13	24.7	24.8	13	25.3	24.8
14	24.3	25.9	14	26.3	28.6
15	31.0	31.3	15	31.4	30.3
16	28.2	32.3	16	28.0	28.0
17	31.8	29.9	17	29.6	27.0

*Not included in mean.

Table 2. Determination of cis, cis-PUFA in blind duplicate samples by two methods (g trilinolein/100 g sample).

The results are plotted in **Figure 3**. The latter reveals that laboratories 11 (aberrant result), 12, 13, 14, 15 and 16 are outside the 95% circle, indicating high systematic errors.

The results for the BF method are shown in **Table 2**.

Again, the data from laboratory 1 were rejected because of a mistake; those from laboratories 11 and 12 are listed but were not used (10.3 and 10.5 g/100 g, respectively, for sample X and 29.6 and 10.1 g/100 g, respectively, for sample Y), as they lie beyond the set limit [35] in the calculations.

Again, the F-ratio (3.268) is larger than F(0.05,13,13), which is 2.577, so it is determined that the systematic errors between the laboratories are significant at the 95% confidence level. All the results are plotted in **Figure 4**. By observing the latter, one may observe that the laboratories 11, 12 (aberrants) and 2 and 13 are outside the 95% circle and are displaced far from the cluster of the others.

Notice that in both methods, about half the points lie above, and about half lie below the horizontal lines through the two means. Likewise, the vertical lines also separate the laboratories into equal groups, as do the 45° lines. However, in neither plot are the results equally distributed amongst the four quadrants; there are more in the upper right and lower left quadrants than in the upper left and lower right. Dispersion along the 45° line indicates that laboratories are high or low on both samples, while dispersion at right angles to the 45° line indicates a lack of agreement between results from the same laboratory.
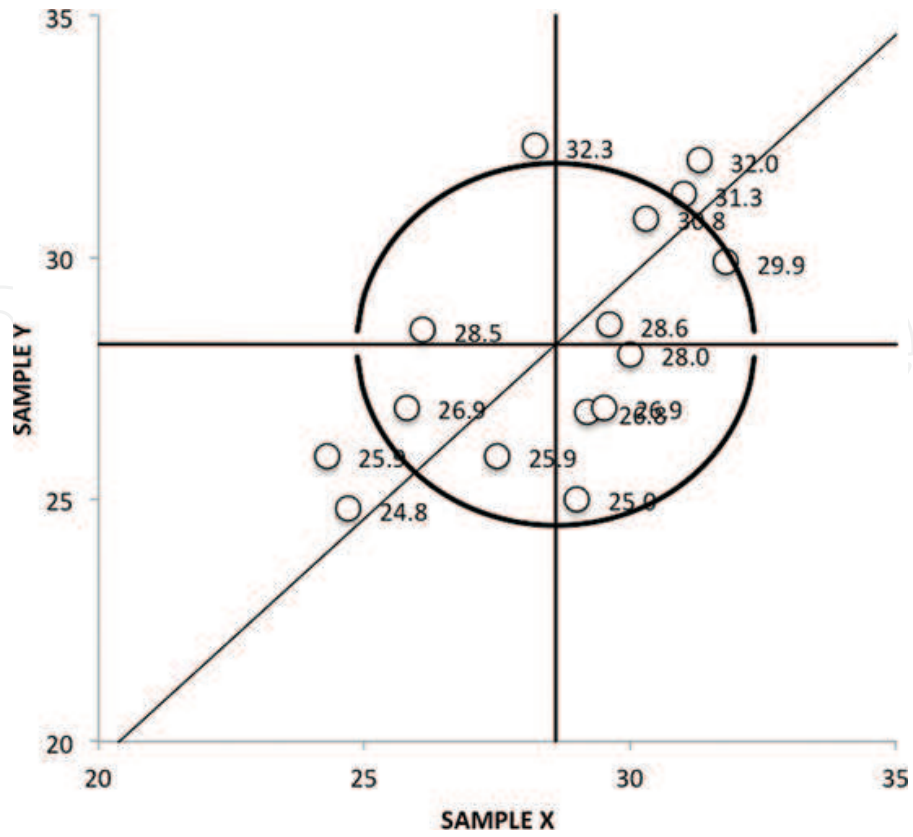
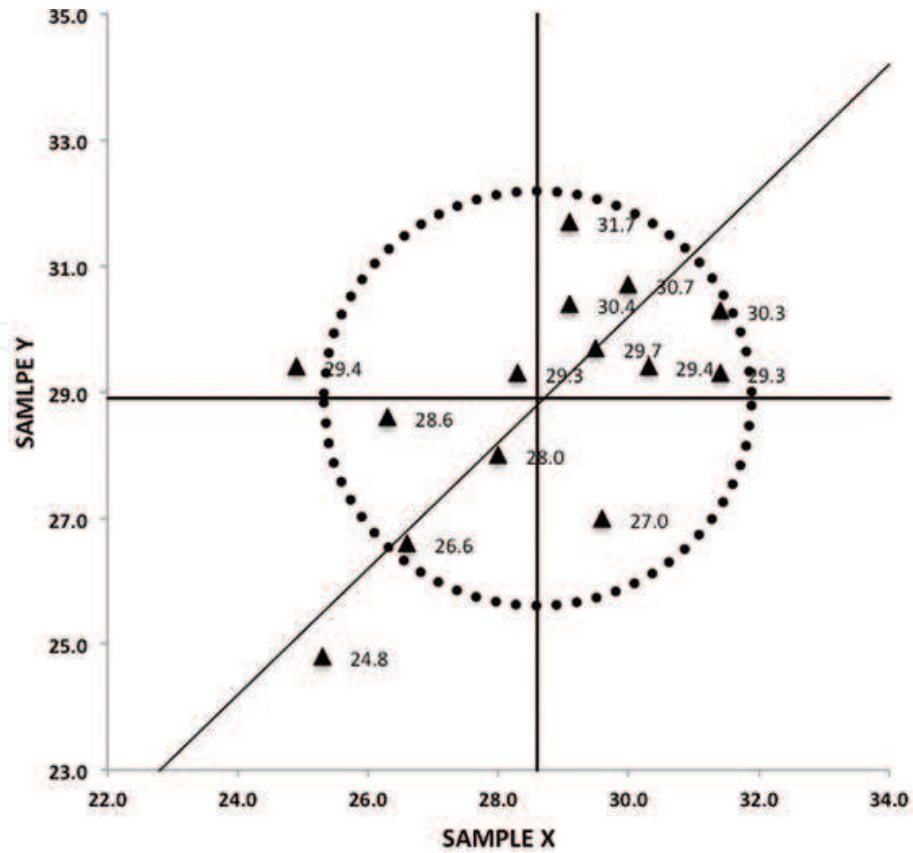**Figure 3.** Youden plot (determination of PUFA, FDA method).

Figure 4. Youden plot (determination of PUFA, BF method).

If there were no systematic variations amongst laboratories, the pattern of points would be expected to be circular. The greater the systematic variations, the more elliptical the pattern will become.

The results obtained by the laboratories may now be compared by using the two methods as follows:

- Three laboratories (3, 6 and 9) are within the 95% circle and near each other on both plots.

- Five laboratories (4, 5, 8, 10 and 17) are within the 95% circle on both plots but widely separated from each other.

- Three laboratories (2, 14, 16) are outside the limit on one plot but not on the other.

- Three laboratories (11, 12 and 15) are outside the limit on both plots.

- Laboratory 13 is systematically low using both methods.

- The two standard deviations are approximately equal.

3.2. Determination of cholesterol levels in the blood

As part of a collaborative study to assess a new method that allows the determination of the total amount of cholesterol in the blood, two samples and the instructions to analyse each sample are sent to ten laboratories [36].

Table 3 shows the results obtained in mg total cholesterol per 100 mL of serum.

Figure 5 provides a two-sample plot of the results. The clustering of points suggests that the systematic errors of the analysts are significant. Two laboratories (1, 5) are outside the limit on the plot and widely separated from each other.

The vertical line at 248.9 mg/100 mL is the mean value for sample X, whereas the horizontal line at 246.5 mg/100 mL corresponds to the mean value of sample Y. To estimate σ_{rand} and σ_{syst}, the values for D_i and T_i are calculated first, followed by the standard deviations.

Because the F-ratio (2.530) is lower than F(0.05,9,9), which is 3.179, it is concluded that the systematic errors between the analysts are insignificant at the 95% confidence level.

If the true values for both samples are known, it is possible to test the presence of systematic errors. When there are no systematic method errors, the sum of the true values, μ_{tot}, for samples X and Y is equal to

$$\mu tot = \mu X + \mu Y \qquad (16)$$

should fall within the confidence interval around T. A two-tailed t-test of the following null and alternate hypotheses is applied:

$$H0: T = \mu tot \quad HA: T \neq \mu tot$$

Laboratory	Sample X	Sample Y
1	245.0	229.4
2	247.4	249.7
3	246.0	240.4
4	244.9	235.5
5	255.7	261.7
6	248.0	239.4
7	249.2	255.5
8	255.1	224.3
9	255.0	246.3
10	243.1	253.1

Table 3. Determination of cholesterol in serum (mg cholesterol/100 mL of serum).
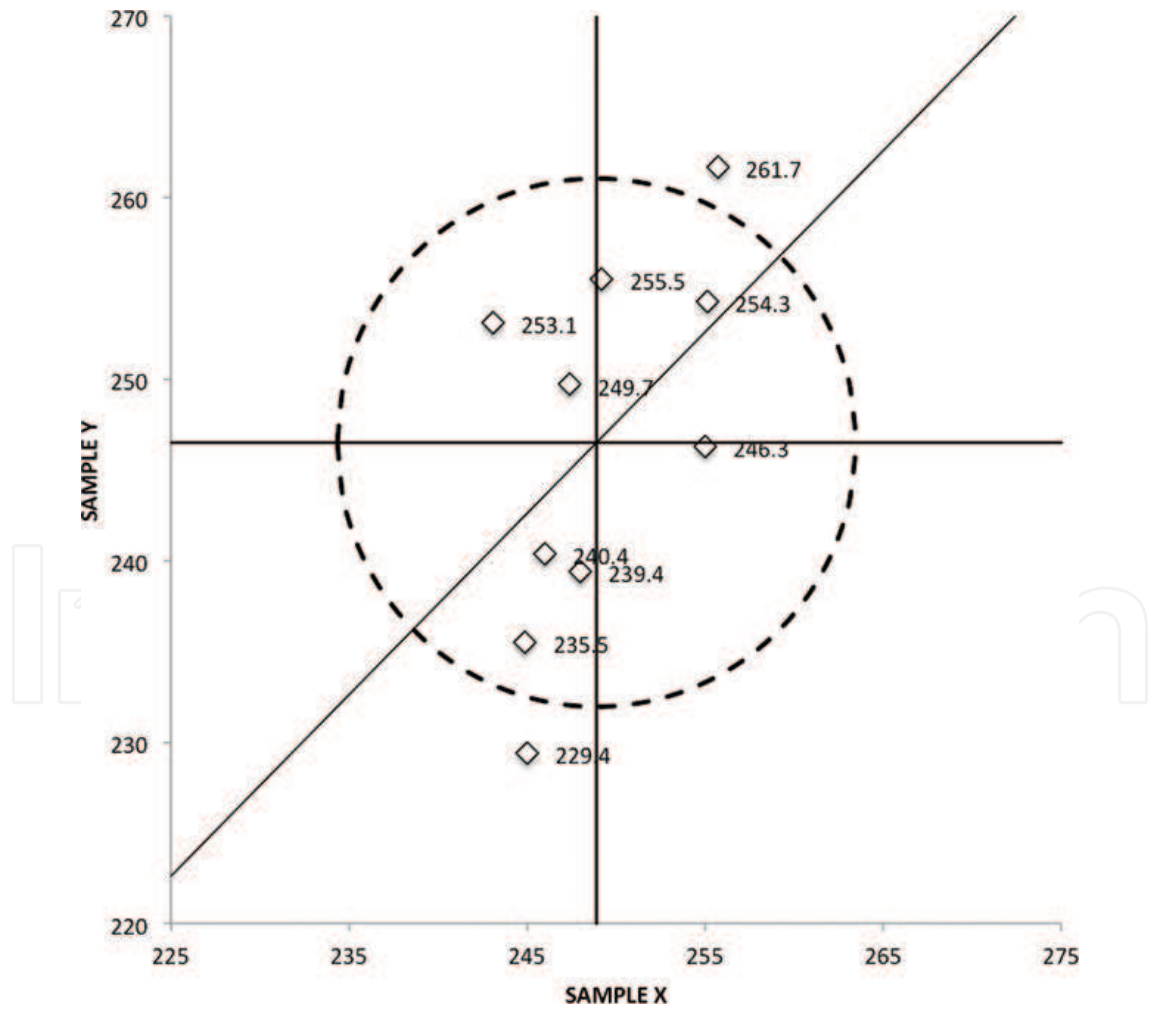

Figure 5. Youden plot (determination of cholesterol).

This occurs so as to determine if there is evidence for a systematic error in the method. The test statistic, $t$exp, is

$$t_{\text{exp}} = \frac{\left| \dot{T} - \mu_{\text{tot}} \right| \sqrt{n}}{S_T \sqrt{2}} \tag{17}$$

with $n-1$ degrees of freedom. The $\sqrt{2}$ in is included in the denominator because S_T underestimates the standard deviation when comparing T to μ_{tot}.

Because this value for $t$exp is smaller than the critical value of 2.26 for $t(0.05, 9)$, there is no evidence for a systematic error in the method at the 95% confidence level.

3.3. Determination of aspirin in pharmaceutical preparations

The results of determinations, made in ten laboratories, on two similar aspirin preparations are given in **Table 4** [37].

Laboratory	Sample X	Sample Y
1	50.45	52.55
2	49.89	52.00
3	49.60	51.70
4	50.26	52.11
5	49.78	51.79
6	49.92	51.81
7	50.22	52.35
8	50.40	52.26
9	50.17	52.24
10	49.85	51.87

Table 4. Weight content (%) of aspirin in pharmaceutical preparations [37].

The analysis of data concerning materials X and Y reveals the following:

- A high level of interlaboratory variation.

- For each of the laboratories, the observed difference between the aspirin content of the two materials is approximately the same.

The average values for the materials (50.054% for X and 52.068% for Y) are used for the centroid. The results are plotted in **Figure 6** where it can be deduced that the data pointed on the diagram fall in either the first or the third quadrant (line X = Y). This is a consequence of the fact that laboratories which obtained "high" values for material X also obtained "high" values for material Y. The opposite is also true; laboratories reporting "low" values for X also reported "low" values for Y.
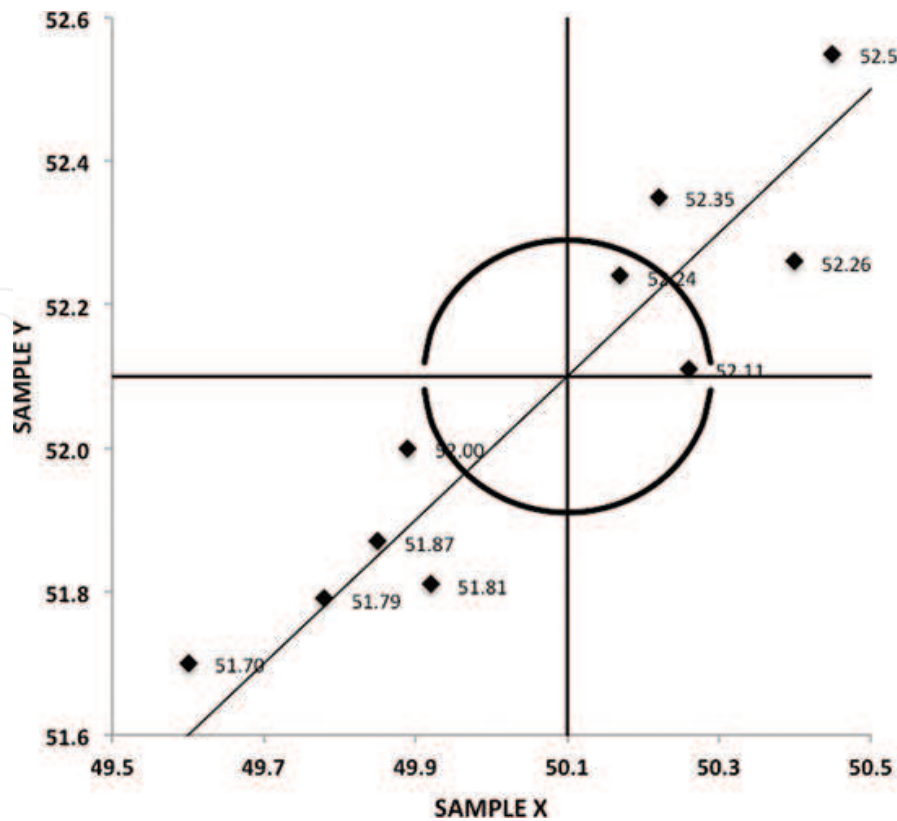
Figure 6. Youden plot (determination of aspirin).

Only two laboratories (2 and 9) are within the 95% circle. Because the F-ratio (25.834) is larger than F(0.05,9,9), which is 3.179, it is reasoned that the systematic errors between the analysts are significant at the 95% confidence level.

All this suggests that the variability in the difference between the two samples is evident. There are two factors which influence the variability of the differences. These are the random error of measurement and the heterogeneity of the materials. It is easy to infer that if the materials were relatively heterogeneous, then the reported differences would show considerable variability. That this is not so in this example is evidence that the laboratories were dealing with relatively homogeneous materials, i.e. the composition of the samples of materials X and Y received by each laboratory was essentially the same.

The larger the random error of measurement associated with an analytical procedure, the more varied the results of replicate measurements are. A relatively large random error of measurement will also cause the differences between them to vary considerably. The example presented herein indicates that the observed differences are approximately the same. This suggests that the random error experienced in each laboratory is relatively small. Perpendicular dispersions to the bisector, for homogeneous materials, are reflections of the within-laboratory variability. It is worth mentioning at this point that the example introduced here reveals that the within-laboratory variabilities are small, compared with the systematic between-laboratory variabilities.

Finally, it is sometimes believed that the centroid in a Youden diagram gives a good estimate of the true values of the two materials. It is the authors' view that this may often not

be the case. The scatter of the results obtained by several laboratories around the mean value is a result of both random within-laboratory and systematic between-laboratory variability. Averaging may result in the mean value being close to the true value. It may also result in a mean value which is, for example, much higher than the true value. Suppose, for example, that all the laboratories had a similar positive bias in one of the steps of the procedure. The result will be a scatter about a mean value which maybe higher than the true value.

3.4. Determination of methylparaben in surface water by liquid chromatography-negative electrospray ionization tandem mass spectrometry

Since parabens were discovered due to their antimicrobial activity, they have been widely used as bactericides, fungicides and preservative agents in many cosmetics, pharmaceuticals, personal care products and food, amongst other consumer products.

Although the toxicity of these compounds is very low, they present a weak estrogenic activity and are considered as endocrine disruptors. That is why they have been classified as emerging contaminants attracting scientific attention on a global scale. These compounds, after consumption, reach wastewater treatment plants, where they are not efficiently removed; thus, they end up in the environment. These chemical compounds consist of detergents, soaps and/or other products.

An analytical method for the determination of methylparaben (MeP) in surface water samples is applied during a period of 15 days.

The method is based on solid-phase extraction (SPE) and subsequent analysis by high-performance liquid chromatography-triple quadrupole mass spectrometry (HPLC-QqQ-MS) [38]. The Youden plot has been applied to the results. A detailed description of the completed experimental procedure is shown below:

3.4.1. Experimental part

3.4.1.1. Materials and reagents

HPLC-grade water, acetone and methanol were purchased by a company in Spain. Analytical-grade formic acid (98%), sulphuric acid (97%) and hydrochloric acid (37%) were acquired from another specialist industry in Spain. Ammonium acetate, MeP (≥99%), was bought from a firm in the USA.

Three millilitres SPE cartridges, packed with 60 mg of Oasis HLB, were purchased from Waters (Milford, MA, USA).

Stock solution, at a concentration of 1000 mg L^{-1}, was prepared in methanol and stored at 4°C. Working solutions were prepared by diluting the stock standard solutions in methanol.

3.4.1.2. Sample collection

Surface water samples were collected in May 2016 and were taken from Guadalquivir River (Seville, Spain). These samples were collected in amber glass bottles precleaned with acetone and methanol. In order to stabilize them, acetonitrile was immediately added after sampling

to achieve a final concentration of 0.5% v/v. Stabilized samples were stored at 4°C until further analysis, which was carried out within 48 h after sample collection. Prior to extraction, samples were filtered through a 1.2 µm glass-fibre membrane filter supplied by a British manufacturer.

3.4.1.3. Solid-phase extraction

Oasis HLB cartridges were conditioned using 3 mL of methanol followed by 3 mL of 0.5N hydrochloric acid and 3 mL of de-ionized water. Prior to extraction, the pH of the sample was adjusted to 2 by the addition of sulphuric acid 40% (v/v). The acidified sample (250 mL) was percolated through the cartridge at a flow rate of approximately 10 mL/min⁻¹. Then, the volumetric flask containing the sample was rinsed with 5 mL of de-ionized water, and the extract was added to the cartridge.

After loading the cartridges, they were washed with 5 mL of de-ionized water and dried for 10 min. The elution of the analytes was carried out with four successive aliquots of 1 mL of methanol at a flow rate of about 1 mL/min. The eluates were collected in 10-mL collection tubes and evaporated to dryness at room temperature by a gentle nitrogen stream. Finally, the extracts were reconstituted in 1 mL of methanol, filtered through a 0.45 µm nylon filter, and a 20-µL aliquot was injected into the HPLC instrument.

3.4.1.4. High-performance liquid chromatography-mass spectrometry

Separation was carried out using an Agilent 1200 series HPLC chromatography system equipped with a vacuum degasser, a binary pump, an autosampler and a thermostated column compartment. MeP was isolated with a Zorbax Eclipse XDB-C18 Rapid Resolution HT (4.6 mm × 50 mm i.d.; 1.8-µm particle size) column, using an isocratic elution with methanol (30%) and aqueous 5 mM ammonium acetate solution (70%) as mobile phase. Flow rate was 0.6 mL/min. The injection volume was 20 µL. The column temperature was maintained at 25°C.

The HPLC system was coupled to a 6410 triple quadrupole (QqQ) mass spectrometer (MS) equipped with an electrospray ionization source operating in negative mode. Two transitions were used for its identification (92.1m/z) and confirmation (136.1m/z).

The ionization of analytes was carried out using the following settings:

– MS capillary voltage 3000 V.

– Drying-gas flow rate 9 L/min⁻¹.

– Drying-gas temperature 350°C.

– Fragmentor 70 V.

– Collision energy 16 V.

– Nebulizer pressure 40 psi. Instrument control and data acquisition were carried out with MassHunter software.

3.4.2. Results and discussion

The results for the different days are shown in **Table 5**.

Because the F-ratio (3.000) is larger than F(0.05,14,14), which is 2.484, it is concluded that the systematic errors between the analysts are significant at the 95% confidence level.

Figure 7 depicts the results. By observing it, one may see that most of points fall in either the first or the third quadrant. Two days (5 and 11) are outside the 95% circle. The following comments may be drawn hereto:

– Days 4, 14 and 15 are within the 95% circle, near each other in the first quadrant.

– Days 2, 3 , 8, 10 and 13 are within the 95% circle, close together, in the third quadrant, although the latter is the very edge of the circle.

– Days 1, 9 and 12 are within the 95% circle, close together but in different quadrants.

– Day 7 is also within the circle but farther away from the other days.

– Day 6 is within the circle but at the very edge of it.

Finally, the z-score approximation is applied (pls. refer to Section 2.2) [39]. The results are shown in **Table 6**. A comparison of the results obtained by the laboratories using the Youden plot and Z-Score is done as follows:

Day	Area MeP (HPLC-MS/MS)		MeP concentration (ng/L)	
	X	Y	X	Y
1	3329	3574	16.1	17.8
2	3255	3388	15.6	16.5
3	3224	3302	15.3	15.9
4	3518	3886	17.4	20.1
5	3621	4095	18.2	21.6
6	3738	3435	19.0	16.8
7	3108	3862	14.5	19.9
8	3145	3200	14.8	15.2
9	3205	3531	15.2	17.5
10	3266	3133	15.6	14.7
11	3056	3076	14.1	14.3
12	3468	3537	17.1	17.6
13	3065	3162	14.2	14.9
14	3357	3758	16.3	19.1
15	3417	3877	16.7	20.0

Table 5. Determination of MeP in two similar surface water samples in different days.
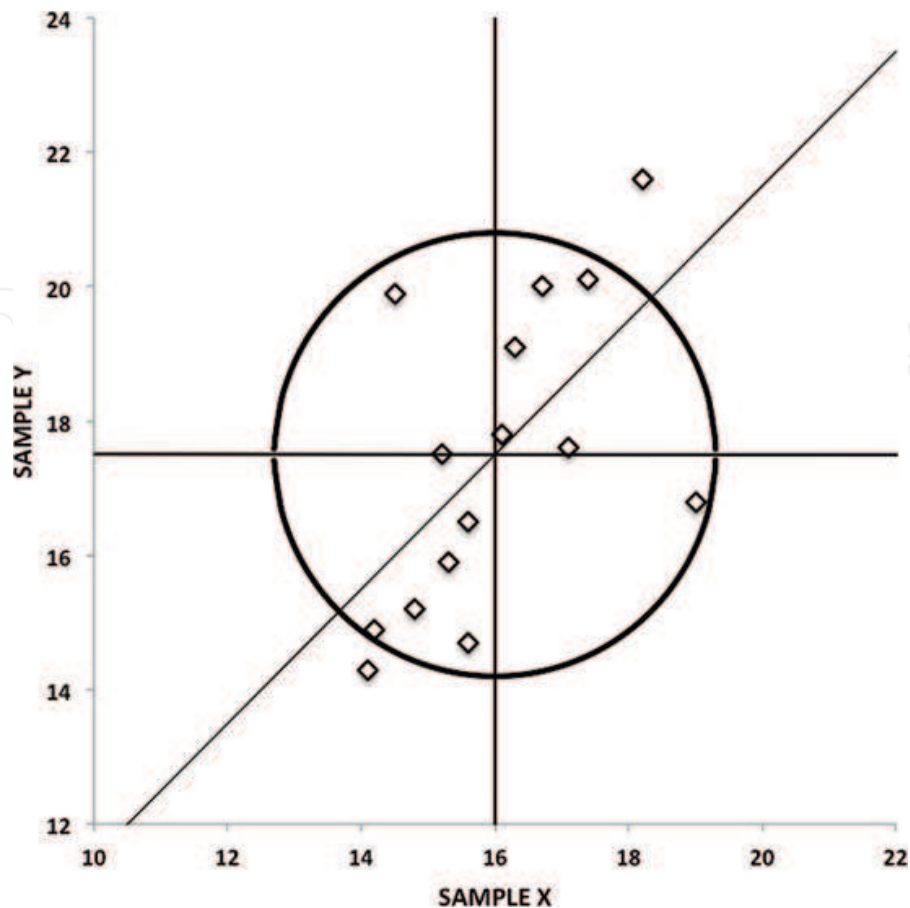
Figure 7. Youden plot (determination of MeP).

– The results obtained with the z-score are in line with those observed in the Youden plot.

– In both methods, days 5 and 11 whose systematic errors are relatively high compared to random errors showed unsatisfactory results.

– Day 6 is discarded using z-score but not on the Youden plot, although it is found very close to the boundary of the 95% circle.

– Days 1, 2, 3, 9, 12 and 14 showed satisfactory results with the z-score method applied.

– Days 4, 7, 8, 10, 13 and 15 showed questionable results. Day 7 is the furthest from the other days in the Youden Plot, and day 13 is on the edge of the 95% circle.

3.5. Antibody concentrations: an ISO-13528:2005(E) example

Finally, a confidence ellipse, calculated as described in ISO 13528 [40], has been used as an aid to interpret the plot so as to deal with those situations, in which the two samples differ in magnitude of the property measured. A Youden Plot for the original data may be derived from the z-scores (as explained below). It is constructed by plotting the z-scores obtained on one of the materials against the z-scores obtained on the other material.

Day	X	Y	ε_T	C_S	C_R	ε_S	ε_R	E_T	ε_R^2	Z-score	
1	16.1	17.8	0.54	0.53	0.11	0.45	0.09	0.54	0.01	0.52	Satisfactory
2	15.6	16.5	1.03	−0.78	0.67	−0.55	0.48	1.03	0.23	0.99	Satisfactory
3	15.3	15.9	1.67	−1.37	0.95	−0.99	0.68	1.67	0.46	1.60	Satisfactory
4	17.4	20.1	3.11	3.07	0.52	2.66	0.45	3.11	0.2	2.98	Questionable
5	18.2	21.6	4.77	4.65	1.06	3.88	0.89	4.77	0.78	4.57	Unsatisfactory
6	19.0	16.8	3.45	1.9	2.88	1.37	2.08	3.45	4.31	3.31	Unsatisfactory
7	14.5	19.9	2.62	0.87	2.47	0.68	1.94	2.62	3.76	2.52	Questionable
8	14.8	15.2	2.52	−2.29	1.07	−1.72	0.80	2.52	0.64	2.42	Questionable
9	15.2	17.5	0.44	−0.31	0.31	−0.22	0.22	0.44	0.05	0.42	Satisfactory
10	15.6	14.7	2.85	−2.02	2.02	−1.43	1.43	2.85	2.03	2.73	Questionable
11	14.1	14.3	3.59	−3.36	1.24	−2.62	0.97	3.59	0.93	3.44	Unsatisfactory
12	17.1	17.6	1.44	1.05	0.99	0.74	0.7	1.44	0.49	1.38	Satisfactory
13	14.2	14.9	3.01	−2.88	0.85	−2.32	0.69	3.01	0.47	2.88	Questionable
14	16.3	19.1	1.75	1.61	0.68	1.23	0.52	1.75	0.27	1.68	Satisfactory
15	16.7	20.0	2.`70	2.51	0.98	1.94	0.76	2.70	0.58	2.59	Questionable

Table 6. Application of z-score to the determination of MeP in two similar surface water samples.

For ease of reference, let A and B denote the two materials:

1. Calculate the averages and standard deviations of the two sets of data and the correlation coefficient ($\hat{\rho}$).

2. Calculate the z-scores for the two materials as follows:

$$Z_{A,i} = \frac{(X_{A,i} - \overline{X_A})}{S_A}; \quad and \quad Z_{B,i} = \frac{(X_{B,i} - \overline{X_B})}{S_B} \tag{18}$$

3. Calculate the combined scores for the two materials:

$$Z_{A,B,i} = \sqrt{Z_{A,i}^2 - 2\hat{\rho}\, Z_{A,i}\, Z_{B,i} + Z_{B,i}^2} \tag{19}$$

4. In terms of the standardized variables, the confidence ellipse may be written in terms of Hotelling's T^2:

$$Z_A^2 - 2\hat{\rho}\, Z_A\, Z_B + Z_B^2 = \left(1 - \hat{\rho}^2\right) T^2 \tag{20}$$

where

$$T^2 = 2\left\{\frac{(p-1)}{(p-2)}\right\} F_{(1-\alpha)}(2, p-1) \tag{21}$$

Here, $F_{(1-\alpha)}(2, p-1)$ is the tabulated $(1-\alpha)$-fractile of the F-distribution with 2 and (p-1) degrees of freedom.

As recommended by the International Organization for Standardization (ISO), the ellipse may be drawn on a graph with the z-scores Z_A and Z_B as the axes by plotting a series of points for $-T < Z_A < T$ with

$$Z_B = \hat{\rho} Z_A \pm \sqrt{\left(1 - \hat{\rho}^2\right)\left(T^2 - Z_A^2\right)} \tag{22}$$

To interpret the Youden Plot, the combined z-scores may be used. The highest combined z-score corresponds to the highest significance level of 100%.

Also, the combined z-scores aid to identify the outlying points.

When a Youden Plot is constructed, it may be interpreted as follows:

– If a point is well separated from the rest of the data, it means that the result is subject to bias because the laboratory did not follow the test method correctly. Points far away from the major axis could also represent laboratories showing a considerable variation and inadequate repeatability outcomes.

– A positive relationship between the results for the two materials indicates that there is a cause of between-laboratory variation that is common to many of the laboratories, suggesting that the methodology may not have been adequately specified. If the method is reproduced, it may lead to an overall improvement.

Table 7 shows data obtained by testing two similar samples for antibody concentrations and the calculations required to derive the confidence ellipse. With p = 29 laboratories and using

Row	Data		Z-score		Combined Z-score
	Allergen A (U)	Allergen B (U)	Z_A	Z_B	Z_{AB}
1	12.950	9.150	0.427	0.515	0.370
2	6.470	6.420	−1.540	−0.428	1.275
3	11.400	6.600	−0.043	−0.366	0.336
4	8.320	4.930	−0.978	−0.942	0.737
5	18.880	13.520	2.228	2.023	1.641
6	15.140	8.220	1.092	0.194	0.965
7	10.120	7.260	−0.432	−0.138	0.349
8	17.940	9.890	1.942	0.770	1.501
9	11.680	4.170	0.042	−1.204	1.234
10	12.440	7.390	0.272	−0.093	0.344
11	6.930	7.780	−1.400	0.042	1.430
12	9.570	5.800	−0.599	−0.642	0.477

Row	Data		Z-score		Combined Z-score
	Allergen A (U)	Allergen B (U)	Z_A	Z_B	Z_{AB}
13	11.730	5.770	0.057	−0.652	0.693
14	12.290	6.970	0.227	−0.238	0.429
15	10.950	6.230	−0.180	−0.493	0.388
16	10.950	5.900	−0.180	−0.607	0.497
17	11.170	7.740	−0.113	0.028	0.134
18	11.200	8.630	−0.104	0.335	0.415
19	7.640	3.740	−1.185	−1.353	0.985
20	12.170	7.330	0.190	−0.114	0.282
21	10.710	5.700	−0.253	−0.676	0.529
22	7.840	6.070	−1.124	−0.549	0.833
23	20.470	15.660	2.710	2.762	2.098
24	12.600	11.760	0.321	1.415	1.210
25	11.370	4.910	−0.052	−0.949	0.913
26	11.360	13.510	−0.055	2.019	2.059
27	10.750	5.480	−0.241	−0.752	0.607
28	12.210	9.770	0.203	0.729	0.603
29	7.490	5.820	−1.230	−0.635	0.902
Mean	11.543	7.659	0.000	0.000	
Standard deviation	3.294	2.897	1.000	1.000	

Units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material

Hotelling's T^2	6.927
T	2.632
$\hat{\rho}$	0.706
F (5%)	3.34

Table 7. Data and calculations on concentrations of antibodies for two similar allergens.

a significance level of 100% = 5%, $F_{(1-\alpha)}(2, p-1) = 3.34$. Hence, T = 2.632. The ellipse is shown, together with the points representing the z-scores, in **Figure 8**, in tandem with the ellipses pertaining to probability levels of 100% = 1% and 0.1%.

As depicted in **Figure 8**, laboratories 5 and 23, with combined z-scores of 1.641 and 2.099, respectively, are found in the top right-hand quadrant. Laboratory 26 has a high z-score on material B (2.019) compared to material A (−0.055) and a combined z-score of 2.059 followed by laboratory 8 with a combined z-score of 1.501. The points for laboratories 23 and 26 fall between the ellipses for the 5% and 1% probability levels. Thus, the results may be perceived as giving rise to warning signals.
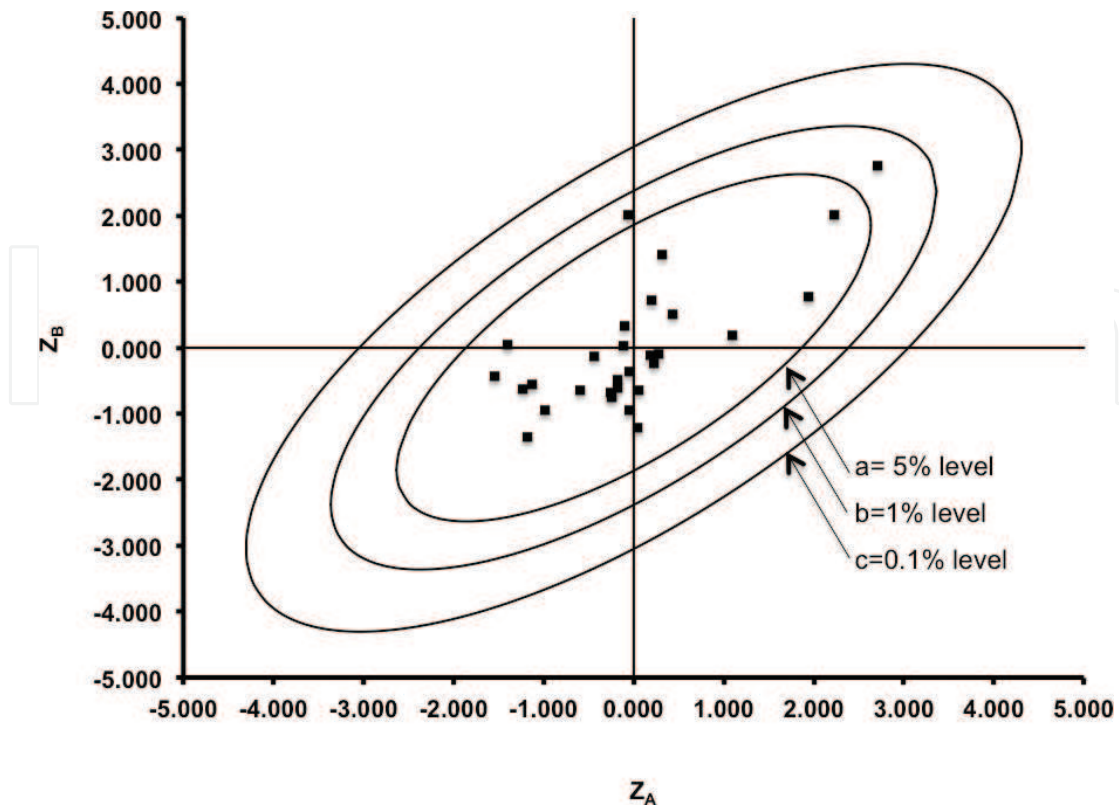
Figure 8. Youden plot of z-scores from **Table 7** (concentrations of antibodies for two similar allergens).

4. Conclusions

Intercomparison exercises are of great value in systems of quality assessment allowing the examination of the analytical process and the generated results. Youden plot or Youden analysis is particularly aimed at interlaboratory comparisons, obtaining accurate information without much effort. The main characteristic is its ability to separate the systematic and random errors with minimal effort on the part of the participants. To implement the method:

- Two similar materials (samples A and B) with small differences in the concentration of the characteristics (magnitude) are required with the purpose of determining their content [1993].

- A scatter plot is drawn in which the x-axis indicates one of the reported values and the y-axis the other, being the scale units the same along its axis. Each pair of results, corresponding to a given laboratory, is a point in the Youden plot 2.0.

- The points occur mainly in the (+ +) and (− −) quadrants, forming an elliptical pattern around a line bisecting these quadrants at a 45° angle, when systematic errors are larger than random errors. The circle centred at the intersection of the reported value medians (once outliers removed) affords a test on the randomness of results, its radius being a multiple of the within-laboratory standard deviation. The suitability and benefits of the Youden method have been applied to a number of results obtained from different fields such as food; clinical and pharmaceutical applications, in order to determine the concentration of

polyunsaturated fatty acids in fats and oils; total blood cholesterol; and aspirin in pharmaceutical preparations, respectively.

- In most cases, it is observed that systematic errors may be regarded as the main cause of variation with most of points in quadrants (+ +) and (− −). Finally, a detailed procedure for the determination of methylparaben in surface waters, of special relevance nowadays in the environmental field, has been developed by liquid chromatography-tandem mass spectrometry. In this experimental system, an alternative to the Youden method based on the z-score has also been assessed showing no discrepancy between both methods.

- A confidence ellipse is proposed in ISO 13528:2005 to deal with those situations where the two samples differ in magnitude of the property measured. The extension of the Youden method based on the confidence ellipse may be used as the building platform for further studies, incorporating amongst other three- and four-dimensional Youden Plots.

Abbreviation list

BF	Boron trifluoride-methanol
FDA	Food and Drug Administration
ISO	International Organization for Standardization
MeP	Methylparaben
MS	Mass spectrometer
PUFA	Polyunsaturated fatty acids
QqQ	Triple quadrupole
SPE	Solid-phase extraction

Author details

Julia Martín, Nieves Velázquez and Agustin G. Asuero*

*Address all correspondence to: asuero@us.es

Department of Analytical Chemistry, Faculty of Pharmacy, The University of Seville, Seville, Spain

References

[1] Funk W, Dammann V, Donnevert G. Quality assurance in analytical chemistry application in environmental, food and materials analysis, biotechnology and medical engineering. 2nd ed., Wiley-VCH: Weinheim, Germany, 2007; pp. 179–188.

[2]   Horwitz W. Protocol for the design, conduct and interpretation of collaborative studies. Pure Appl. Chem. 1995; **67**(2):331–343.

[3]   Horwitz W. Protocol for the design, conduct and interpretation of method-performance studies. Pure Appl. Chem. 1988; **60**(6):855–864.

[4]   Taylor JK. Quality assurance of chemical measurements. CRC: Boca Raton, FL, p. 19.

[5]   AMC Fitness for purpose: the key feature in analytical proficiency testing. Analytical Methods Committee, AMCTB No. 68, Anal. Methods. 2015; **7**:7404–7405.

[6]   Kateman G, Buydens L. Quality control in analytical chemistry. Wiley: New York, 1993; 3.4.3 Youden Plot, pp. 132–136.

[7]   Miller JN, Miller JC. Statics and chemometrics for analytical chemistry. Pearson: Harlow, England, 2010.

[8]   González AG, Herrador MA, Asuero AG. Intralaboratory testing of method accuracy from recovery assays. Talanta. 1998; **48**:729–736.

[9]   Ketterer ME. Assessment of overall accuracy of lead isotope ratios determined by inductively coupled plasma mass spectrometry using batch quality control and the Youden two- sample method. J. Anal. At. Spectrom.. 1992; **7**(7):1125–1129.

[10] Delwiche SR, Palmquist DE, Lynch JM. Collaborative studies for cereals analysis. Cereals Food World. 2005; **50**(1):9–17.

[11] Heyden YV, Smeyers-Verbeke J. Set-up and evaluation of interlaboratory studies. J. Chromatogr. A. 2007; **1158**(1–2):158–167.

[12] Hund E, Massart DL, Smeyers-Verbeke J. Interlaboratory studies in analytical chemistry. Anal. Chim. Acta. 2000; **423**(2):145–465.

[13] ISO 13528:2005 (E) Statistical methods for use in proficiency testing by interlaboratory studies. International Standardization Organization: Genève, 2005.

[14] AOCS Procedure M4-86. Evaluation and Design of Test Methods. Collaborative Study Procedures. The American Oil Chemists' Society; 2012. pp. 1–9.

[15] Boley N. Selection, Use and Interpretation of Proficiency Testing (PT) Schemes by laboratories – 2000. The Eurachem Nederland, Task Group "Proficiency Testing Schemes" and the Laboratory of the Government Chemist (LGC), United Kingdom.

[16] Pocklington WD. Guidelines for the development of standard methods of collaborative study: organization of interlaboratory studies and a simplified approach to the statistical analysis of collaborative study results, 5th ed.. Laboratory of the Government Chemist: Teddington, England, 1990.

[17] Wernimont GT, Spendley W. Use of statistics to develop and evaluate analytical methods. Assoc. Off. Anal. Chem: Arlington, VA, USA, 1987.

[18]   Youden WJ, Steiner EH. Statistical manual of the association of official analytical chemists. Statistical techniques for collaborative tests. Planning and analysis of results of collaborative tests. AOAC: Washington DC, 1975.

[19]   Juran JM, Godfrey AB. (Eds.) Juran's quality handbook, 5th ed.. McGraw-Hill: New York, 1999; Youden Two-Sample Plan, 47.28.–47.29.

[20]   Zhong Z, Li G, Luo J, Chen W, Liu L, He P, Luo Z. Proficiency testing for determination of lead and arsenic in cosmetics: comparison of analytical procedures and evaluation of laboratory performances. Anal. Methods. 2015; **7**(7):3196–3177.

[21]   Juneja A, Anand A.Understanding statistical concepts in laboratory quality control measures in biomedical research. Int. J. Res. Med. Sci.. 2015; **3**(11):3443–3445.

[22]   Youden WJ. Experimentation and measurement. National Institute of Standards and Technology, NIST, Special Publication 672. U.S. Department of Commerce. Reprinted; 1977.

[23]   Youden WJ. Graphical diagnosis of interlaboratory test results. Ind. Qual. Control. 1959; **15**(11):24–28.

[24]   Youden WJ. Statistical aspects of the cement testing program. Proc. Am. Soc. Testing Mat. Philadelphia, PA.. 1959; **59**: 1120–1128.

[25]   Youden WJ. Evaluation of chemical analyses on two rocks. Technometrics. 1959; **1**(4):409–417.

[26]   Tracy ND, Young JC. A bivariate control chart for paired measurements. J. Qual. Technol.. 1995; **27**(4):370–376.

[27]   Mandel J, Lashof TW. Interpretation and generalization of Youden's two- sample. J. Qual. Technol.. 1974; **6**(1):22–36.

[28]   Cornell JA. The man and his methodology. AQQC Statistical Division Newsletter. 1993; **13**(2):9–18.

[29]   Lavetz R. Statistical Manual. Chemical Proficiency Testing – NMI north ryde. Chemical & biological metrology. Document No 3. Australian Government. National Measurement Institute; 2014. pp. 1–15.

[30]   Starink RJ and Visser RG. Interlaboratory studies: protocol for the organisation, statistics and evaluation. Institute for Interlaboratory Studies. Spijkenisse, The Netherlands; 2014. pp. 1–27.

[31]   De Girolamo A, Ciasca B, Stroka J, Bratinova S, Visconti A, La anzio VMT. Determination of DON, FB1, FB2, ZEA, T-2, HT-2, OTA, AFB1, AFG1, AFB2, AFG2 in maize and determination of DON, ZEA, T-2, HT-2, OTA in wheat. Report of the 2014 Proficiency Test for LC-MS(MS) multi-mycotoxin methods. National Research Council of Italy. Institute of Sciences of Food Production. Italy; 2014.

[32] Kunsagi Z, Stroka J. Report on the validation of a method for the determination of Ochratoxin A in Capsicum spp. (Paprika and chilli). European Commission, Joint Research Centre, Institute for Reference Materials and Measurements: Geel, Belgium, 2012.

[33] Marchetto A, Mosello R, Tartari G. Atmospheric Deposition and Soil Solution Working Ring Test 2011. Laboratory ring test for deposition and soil solution sample analyses for the laboratories participating in the EU/Life + FutMon Project LIFE07 ENV/D/000218. Deliverable QA-Rwater-11. FutMon "Further Development and Implementation of an EU-level Forest Monitoring System". Consiglio Nazionale delle Ricerche. Istituto per lo Studio degli Ecosistemi. Verbania Pallanza; 2011. pp. 1–59.

[34] Villeneuve JP, de Mora SJ, Cattini C. World-wide and regional intercomparison for the determination of organochlorine compounds and petroleum hydrocarbons in mussel tissue IAEA-432. International Atomic Energy Agency, Marine Environment Laboratory: Monaco, Report No 74, March 2004.

[35] Sheppard AJ, Walting AE, Zmachinski H, Jones ST. Two lipoxidase methods for measuring cis, cis-methylene interrupted polyunsaturated fatty acids in fats and oils: collaborative study. J. Assoc. Offic. Anal. Chem.. 1978; **6**:1419–1423.

[36] Analytical chemistry 2.0.An electronic textbook for introductory courses in analytical chemistry. Chapter 15 quality assurance.Available from: http://acad.depauw.edu/harvey_web/eText%20Project/AnalyticalChemistry2.0.html.

[37] McCornic D, Roach A. Measurement, statistics and computation, analytical chemistry by open learning. ACOL, Wiley: Chichester, 1987; 4.7.4. Youden plot, pp. 322–325.

[38] Martín J, Camacho-Muñoz D, Santos JL, Aparicio I, Alonso E. Determination of emerging and priority industrial pollutants in surface water and wastewater by liquid chromatography–negative electrospray ionization tandem mass spectrometry. Anal. Bioanal. Chem.. 2014; **406**:3709–3716.

[39] Pérez A. Evaluación Estadística (Comparaciones Interlaboratorios - Análisis de Youden). Sociedad Española de Bioquímica Clínica y Patología Molecular. https://es.scribd.com/document/317778285/Estudios-Interlaboratorios-09-07-1.

[40] Jackson JE. Quality control methods for two related variables. Ind. Qual. Control. 1956; **7**: p. 2–6.

[41] Garrett RG. A comparison of Shewhart, Thompson and Howarth, and Youden plots – advantages and disadvantages. Explore Newsl. Assoc. Appl. Geochem.. 2015; **167**:5–11.

[42] Janik M, Yonehara H. The most recent international intercomparisons of radon and thoron monitors with the NIRS radon and thoron chambers. Radiat. Prot. Dosimetry. 2015; **164**(4):595–600.

[43] Zhou Q, Hu J, Li X, Li S, Gao Z, Xie W, Xu J. Comparison of traditional, trimmed traditional and robust Youden charts. Clin. Chim. Acta. 2015; **446**:213–217.

[44] Zhou Q, Hu J, Li X, Li S, Gao Z, Xu J, Xei W. Construction and application of the robust Youden plot in a EQA program. Accred. Qual. Assur. 2015; **20**(3):195–201.

[45] Van der Bresselaar AMHP, Abdoel CF, Ardanary D, van de Kamp G, Versluijs FAC. Preparation and control blood for external quality assessment of point-of-care international normalized ratio testing in the Netherland. Am. J. Clin. Pathol.. 2014; **14**:879–883.

[46] Xiao YL, Zhang CB, Zhao HJ, Kang FF, Wang W, Zhong K, Yuan S, Wang ZG. Application of ISO 13528 robust statistical methods for external quality assessment of blood glucose measurements in China. Accred. Qual. Assur.. 2014; **19**(5):397–401.

[47] O'Donnel GE, Hibbert DB. A study of the conditions of measurement required to evaluate bias in analytical results illustrated by the use of data from a multi-round, blind-duplicated, proficiency test. Analyst. 2013; **138**:3673–3678.

[48] Monteiro LR, Grafitti D, Albano F, Porfírio D, Fernandes Jr LP, Cotrim MEB, Pires MAF. Evaluation of a Brazilian ion chromatography interlaboratory study.Accred. Qual. Assur.. 2013; **18**(3):207–215.

[49] Shirono K, Iwase K, Okazaki H, Yamazawa M, Shikakume K, Fukumoto N, Murakami M, Yanagisawa M, Tsugoshi T. A study on the utilization of the Youden plot to evaluate proficiency test results. Accred. Qual. Assur.. 2013; **18**(3):161–174.

[50] Yu W, Tan Q, Yu L, Cao H. Comparison of statistical model of Youden plot for proficiency testing in ISO13528 and NATA. Metall. Anal.. 2013; **33**(12):74–80.

[51] Bremser W, Lücke FK, Urmetzer C, Fuchs E, Leist U. An approach to integrated data assessment in a proficiency test on the enumeration of *Escherichia coli*. J. Appl. Microbiol.. 2011; **110**(1):128–138.

[52] de la Calle Guntiñas MB, Semeraro A, Wysocka I, Cordeiro F, Quétel C, Emteborg H, Charoud-Got J, Linsinger TPJ. Proficiency test for the determination of heavy metals in mineral feed. The importance of correctly selecting the certified reference materials during method validation. Food Addit. Cont. Part A Chem. Anal. Control Expo. Risk Assess.. 2011; **28**(11):1534–1546.

[53] Kanefuji K, Tsugoshi T, Iwase K. Evaluation between statistical methods relate to the z score for use in proficiency testing. Bunseki Kagaku. 2011; **60**(7):571–577.

[54] Lavine BK. Learning Outcome Assessments in Quantitative Analysis Lab Using Youden Plots. New Trends in the Teaching of Analytical Chemistry. FACSS Analytical Science and Innovation. 2011 . https://www.scixconference.org/program/archive?p=4424&yearSelect=2011

[55] Mann I, Brookman B. Selection, use and interpretation of proficiency testing (PT) Schemes. 2nd ed. EA-Eurolab-Eurachem; 2011.

[56] Wang JS, Kee MK, Choi BS, Kim CW, Kim SS. Evaluation of external quality assessment results for HIV testing laboratories in Korea using current analytical methods. Clin. Chim. Acta. 2011; **412**(11–12):1127–1132.

[57] Heath E, Kosjek T, Farre M, Quintana JB, De Alecastro LF, Castiglioni S, Gans O, Langford K, Loos R, Radjenovic J, Rocca LM, Budzinski H, Tsipi D, Petrovic M, Barcelo

D. Second interlaboratory exercise on non- steroidal anti-inflammatory drug analysis in environmental aqueous samples. Talanta. 2010; **81**:1189–1196.

[58] Pankratov I, Elhanany S, Hening S, Zaritsky S, Ostapenlo I, Kuselman I. Development of a proficiency testing scheme for a limited number of participants in the field of natural water analysis. Accred. Qual. Assur.. 2010; **15**(8):459–466.

[59] de la Calle Guntiñas MB, Mysocka I, Quétel C, Vassileva E, Robouch P, Emteborg H, Taylor P. Proficiency test for heavy metals in feed and food in Europe. Trends Anal. Chem.. 2009; **28**(4):454–465

[60] Sanyal D, Rani A. Proficiency test for chemical laboratories for the analysis of a pesticide in a formulated product: interlaboratory study. J. AOAC Int.. 2009; **92**(1):271–278.

[61] Violante FGM, Bastos LHP, Cardoso MHWM, Rodrigues JM, Gourêa AV, Borges CN, Da Santos PR, Da Santos D, De A Goes HC, Souza VA, de Sâo J, Bandeira RDCC, Cunha V, Nóbrega A. Proficiency testing for the determination of pesticides in mango pulp: a view of the employed chromatographic techniques and the evaluation of laboratories performance. J. Chromatogr. Sci.. 2009; **47**(9):833–839.

[62] Wilke O, Horn W, Wiegner K, Jann O, Bremser W, Brödner D, Kalus S, Juritsch R, Till C. Investigations for the improvement of the measurement of volatile organic compounds from floor coverings within the health-related evaluation of construction products. BAM, Research-number ZP 52-5-20.49.1-1251/07. By Fraunhofer IRB Verlag, ISBN: 3-8167-8253-1, 2009.

[63] Bellis DJ, Hetter KM, Verostek MF, Parsons PJ. Characterization of candidate reference materials for bone lead via interlaboratory study and double isotope dilution mass spectrometry. J. Anal. At. Spectrom.. 2008; **23**(3):289–308.

[64] Flores L, Santo C, Trías M. Interlaboratory system to ensure and improve the quality of glassware calibration and use in a large laboratory. J. AOAC Int.. 2008; **91**(1):247–251.

[65] Gavlick WK, Tomkins DF. An updated liquid chromatographic assay for the determination of glyphosate in technical material and formulations. J. AOAC Int.. 2008; **91**(1):1–4.

[66] Nelsen TC, Wehling P. Collaborative studies for quantitative chemical analytical methods. Cereals Food World. 2008; **53**(5):285–288.

[67] Christopher SJ, Pugh RS, Ellisor MB, Mackey EA, Spatz RO, Porter BJ, Bealer KJ, Kucklick JR, Rowles TK, Becker PR. Description and results of the NIST/NOA a 2005 interlaboratory comparison exercise for trace elements in marine mammals. Accred. Qual. Assur.. 2007; **12**(3):175–187.

[68] Cooper TG, Hellenkemper B, Nieschlag E. External quality control for Semen analysis in Germany. J. Reproduct. Med. Endocrin.. 2007; **4**(6):331–335.

[69] Hare LB. It's not always what you say, but how you say it. Qual. Progress.. 2007; **40**(8):64–66.

[70] Lau A. Practical advice on Youden plot. Qual. Progress. 2007; **40**(10):10.

[71] Miller EL, Bimbo AP, Barlow SM, Sheridau B. Repeatability and reproducibility of determination of the nitrogen content of fishmeal by the combustion (Dumas) method and comparison with the Kjeldahl method: interlaboratory study. J. AOAC Int.. 2007; **90**(1):6–20.

[72] Siekmann L, Breuer H. Determination of cortisol in human plasma by isotope dilution-mass spectrometry. Definitive methods in clinical chemistry, I. J. Clin. Chem. Clin. Biochem.. 1982; **20**(12):883–892.

[73] Sorensen K, Grung M, Röttgers R. An intercomparison of in vitro chlorophyll A determinations for MERIS level 2 data validation. Int. J. Remote Sensing-MERIS. 2007; **28**(3–4):537–544.

[74] Bons, JA, de Boer, D, van Dieijen-Visser, MP and Wodzig, WK. Standardization of calibration and quality control surface-enhanced laser desorption/ionization time of flight mass spectrometry. Clin. Chim. Acta. 2006; **366**:249–256.

[75] Schilling P, Powilleit M, Uhlig S. Chlorophyll-a determination: results of an interlaboratory comparison. Accred. Qual. Assur. 2006; **11**(8):462–469.

[76] Svegl F, Strupi JS, Svegl IG. Proficiency testing of chloride content in different types of Portland cement. Accred. Qual. Assur. 2006; **11**(8):414–421.

[77] Costantini S, Ciaralli L, Ciprotti M, D'Ilio S, Giordano R, Mosca M, Sepe A, Jenofonte O. The network of the Italian laboratories: a proficiency test on the quantification of trace elements in serum. Ann. Ist. Super. Sanita. 2005; **41**(1):171–179.

[78] González AG, Herrador MA, Asuero AG. Practical digest for evaluating the uncertainty of analytical assays from validation data according to the LGC/VAM protocol. Talanta. 2005; **65**(4):1022–1030.

[79] Jones FE. Youden analysis of Karl Fisher titration data from an interlaboratory study determining water in animal feed, grain and forage. J. AOAC Int.. 2005; **88**(6):1840–1841.

[80] Siekmann L. Establishing measurement traceability in clinical chemistry. Accred. Qual. Assur.. 2004; **9**(1):5–17.

[81] Margolis SA, Vangel M, Duewer DL. Certification of standard reference material 970, ascorbic acid in serum, and analysis of associated interlaboratory bias in the measurement process. Clin. Chem.. 2003; **49**(3):463–469.

[82] Sorensen K, Grung M, Röttgers R. An intercomparison of in vitro chlorophyll A determination – preliminary results, Proceedings of the ENVISAT Validation Workshop (ESA SP-531), 9–13 December, Frascati, Lacoste, H. (Ed.), Italy, 2003.

[83] Official Methods of Analysis (AOAC). Guidelines for collaborative study procedures to validate characteristics of a method of analysis; 2002 AOAC International; Appendix D, pp. 2–12.

[84] Gadmar TC, Vogt RD, Osterhus B. The merits of the high-temperature combustion method for determining the amount of natural organic carbon in surface freshwater samples. Inter. J. Environ. Anal. Chem. 2002; **82**(7):451–461.

[85] Grijsen JG. Findings of third inter-laboratory AQC exercise. Inter-Laboratory AQC Exercise. Technical Assistant. Hydrology Project. Government of India & Government of The Netherlands; 2002. pp. 1–24.

[86] Berman S. Seventh intercomparison exercise on trace metals in sea water. Institute for National Measurement Standards National Research Council. Marine Chemistry Working Group. International Council for the Exploration of the Sea. Denmark. ICES Cooperative Research Report No. 237; ISSN 1017-6195. 2000.

[87] Carvhalo FP, Villeneuve JP, Cattini C. The determination of organochlorine compounds and petroleum hydrocarbons in a seaweed sample: results of a world-wide intercomparison exercise. Trends Anal. Chem.. 1999; **18**(11):656–664.

[88] McClure FD. A statistical evaluation of the Youden matched-pairs procedure. J. AOAC Int.. 1999; **82**(2):375–381.

[89] ISO 5725-5:1998. Accuracy (trueness and precision) of measurement methods and results – Part 5: alternative methods for the determination of the precision of a standard measurement method, 1998. https://www.iso.org/

[90] Marchetto A, Bianchi M, Geiss H, Muntan H, Serrini G, Serrini-Lanza G, Tartari GA, Mosello R. Performances of analytical methods for freshwater analysis assessed through intercomparison exercises. Mem. Ist. Ital. Idrobiol. 1997; **56**:1–13.

[91] Petersen PH, Ricós C, Stöckl D, Libeer JC, Baadenhuijsen H, Fraser C, Thienpont L. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. Eur. J. Clin. Chem. Clin. Biochem. 1996; **34**(12):983–999.

[92] Wu GB, Meng H. Application and improvement of the Youden analysis in the intercomparison between flowmeter calibration facilities. Flow Measurement Inst. 1996; **7**(1):19–24.

[93] Feinberg M. Basic of interlaboratory studies: the trends in the new ISO 5725 standard edition. Trends Anal. Chem. 1995; **14**(9):450–457.

[94] Hewitt AD, Grant CL. Round-Robin study of performance evaluation soils vapor-fortified with volatile organic compounds. Environ. Sci. Technol. 1995; **29**:769–774.

[95] Mosello R, Bianchi M, Geiss H, Marche o A, Serrini G, Serrini-Lanza G, Tartari GA, Muntau H. AQUACON-MedBas Subproject No. 6. Acid rain analysis. Intercomparison 1/94. Join Research Centre European Commission, Rep. EUR 163332 EN, Istituto Italiano di Idrobiologia, National Research Council, Italy; 1995. p. 48.

[96] Barnes D, Dent G. Polystyrene films as a performance check for FTIR spectrometers. Spectrosc. Eur. 1994; **6**(2):8–14.

[97] Gaskin JE. Graphical diagnosis of interlaboratory quality control data for surface water samples. Analyst. 1994; **119**:1531–1535.

[98]   Horwitz W. Nomenclature of interlaboratory analytical studies. Pure Appl. Chem. 1994; **66**(9):1903–1911.

[99]   ISO 5725-2:1994. Accuracy (trueness and precision) of measurement methods and results – Part 2: basic method for the determination of repeatability and reproducibility of a standard measurement method. 1994. https://www.iso.org/

[100]   Stephens RD, Rappe C, Hayward DG, Nygren M, Startin J, EsbØll A, Carlé J, Yrjänheikki EJ. World health organization international intercalibration study on dioxins and furans in human milk and blood. Anal. Chem.. 1992; **64**(24):3109–3117.

[101]   Mesley RJ, Pocklington WD, Walker RF. Analytical quality assurance. A review. Analyst. 1991; **116**(10):975–990.

[102]   Jones NE. Multiway analysis of variance for the interpretation of interlaboratory studies. Anal. Chem.. 1990; **62**:1532–1536.

[103]   Oosting E, Neugebauer E, Keyzer JJ, Lorenz W. Determination of histamine in human plasma: the European external quality control study 1988. Clin. Exp. Allergy. 1990; **20**:349–357.

[104]   Youden WJ. Classic paper. The collaborative test. (Reprinted from J-Assoc-Off-Agric-Chem, Vol. 46, PG 55-62, 1963). J. Assoc. Off. Anal. Chem.. 1990; **73**(2):194–201.

[105]   Thompson M. Robust statistic and functional relationship estimation for comparing the bias of analytical procedures over extended concentration ranges. Anal. Chem.. 1989; **61**(17):1942–1945.

[106]   Ferrus R, Torrades F. Bias-free adjustment of analytical methods to laboratory samples in routine analytical procedures. Anal. Chem.. 1988; **69**(13):1281–1285.

[107]   Abern AM, Garrell RL. Exchange of comments on a new technique in chemical assay calculations. Anal. Chem.. 1987; **59**(23):2816–2818.

[108]   Bauer CF, Grant CL, Jenkins TF. Interlaboratory evaluation of high-performance liquid chromatographic. Determination of nitroorganics in munition plant wastewater. Anal. Chem.. 1986; **58**(1):176–182.

[109]   Boyer KW, Horwitz W, Albert R. Interlaboratory variability in trace element analysis. Anal. Chem.. 1985; **57**(2):454–459.

[110]   Currie LA. The limitations of models and measurements as revealed through chemometric intercomparison. J. Res. Natl. Bur. Stand. 1985; **90**(6):409–419.

[111]   Cleveland WS and McGrill R.The many faces of a scatterplot. J. Am. Stat. Assoc.. 1984; **79**(388):807–822.

[112]   Jenkins TF, Leggett DC, Grant CL, Bauer CF. Reversed-phase high-performance liquid chromatographic determination of nitroorganics in munitions wastewater. Anal. Chem.. 1986;**58**:170–175.

[113] Mcdonald RW and Nelson H. A laboratory performance check for the determination of metals (Hg, Zn, Cd, Cu,Pb) in reference marine sediments. Canadian Technical Report of Hydrography and Ocean Sciences, No. 33. Institute of Ocean Sciences Department of Fisheries and Oceans, Sidney; 1984.

[114] Smith R. Organization and evaluation of interlaboratory comparison studies among southern African water analysis laboratories. Talanta. 1984; **31**(7):537–545.

[115] Jeffcoate SL. Use of Youden plot for internal quality control in the immunoassay labora-tory.Ann. Clin. Biochem. 1982; **19**(6):435–437.

[116] Usui T. Quality control data evaluation: tolerance ellips expression of Youden plot. Japan. J. Clin. Chem.. 1982; **11**(2):98–103.

[117] Berman GA. Testing laboratory performance: evaluation and accreditation. NBS Publication 591, National Bureau of Standards: Gaithersburg, MD, U.S.A 1980.

[118] Evans WH. Qualification of estimates for total trace elements in food stuffs using mea-surement by atomic- absorption spectrophotometry. Analyst. 1978; **103**:452–468.

[119] Green A and Naegele R. Development of a system for conducting inter-labo-ratory tests for water quality and effluent measurements. EPA-600/4-77-031, Environmental Monitoring and Support Laboratory. Office of Research and Development. U.S. Environmental Protection Agent. Cincinnati, Ohio; 1977.

[120] Egan H. Collaborative analysis and the standardization of analytical methods. Proc. Society Anal. Chem.. 1972; **9**(11):245–249.

[121] Youden WJ. Graphical diagnosis of interlaboratory test results. (reprinted from Industrial Quality Control, XV, No 11, May 1959). J. Qual. Technol. 1972; **4**(1):29–33.

[122] Shendzel LP, Youden WJ. Systematic versus random error laboratory surveys. Am. J. Clin. Pathol.. 1970; **54**(3):448–450.

[123] Ku HH. Precision measurement and calibration. Statistical concepts and procedure. NBS Special Publication 300 –Volume 1, National Bureau of Standards: Washington, DC, 1969.

[124] Shendzel LP, Youden WJ. A graphic display of interlaboratory test results. Am. J. Clin. Pathol.. 1969; **51**(2):161–165.

[125] Schulz G. Determination of systematic and accidental errors of analytical procedure by Youden method. Zellstoff Papier. 1967; **16**(9):281.

[126] Youden WJ. Collaborative test. J. Assoc. Off. Agric. Chem.. 1963; **46**(1):55–62.

[127] Youden WJ. The sample, the procedure and the laboratory. Anal. Chem.. 1960; **32**(13):23A–27A.

[128] Linning FJ, Mandel J, Peterson JM. A plan for studying the accuracy and precision of an analytical procedure. Anal. Chem.. 1954; **26**(7):1102–1110.

[129] Wernimont GT. Design and interpretation of interlaboratory studies of test methods. Anal. Chem.. 1951; **23**(11):1572–1576.