Human errors in an analytical chemistry laboratory - Implementation of ordinal analysis of variation for risk assessment

Talanta Open 13 (2026) 100603

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open

Human errors in an analytical chemistry laboratory - Implementation of ordinal analysis of variation for risk assessment

Francesca R. Pennecchi [a], Tamar Gadrich [b], Ilya Kuselman [c,*], D. Brynn Hibbert [d], Angelique Botha [e], Anastasia A. Semenova [f]

[a] Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Turin, Italy
[b] Braude College of Engineering, Department of Industrial Engineering and Management, P.O. Box 78, 51 Snunit St., 2161002 Karmiel, Israel
[c] Independent Consultant on Metrology, 4/6 Yarehim St., 7176419 Modiin, Israel
[d] School of Chemistry, UNSW Sydney, Sydney NSW 2052, Australia
[e] National Metrology Institute of South Africa (NMISA), Private Bag X34, Lynnwood Ridge, 0040, Pretoria, South Africa
[f] V.M. Gorbatov Federal Research Center for Food Systems, 26 Talalikhina St., 109316 Moscow, Russia

ARTICLE INFO

Keywords:
Risk assessment
Analytical chemistry
Testing laboratory
Human errors
Expert responses
Ordinal data analysis

ABSTRACT

Decision-making risks caused by human errors in performing a chemical analysis are assessed using laboratory expert judgments (responses) on a specified ordinal scale. In the present paper, a new approach to assessment of risk is described based on implementation of the recently developed two-way ordinal analysis of variation – ORDANOVA. This approach calculates the number of expert responses related to the same category for each ordinal characteristic and then analyzes their relative frequencies as fractions of the total number of responses (of all categories) obtained for this characteristic. It does not violate the properties of ordinal data and allows for the correct interpretation of expert responses. Previously published expert responses on the risks in pH measurements of groundwater, in gas chromatography–mass spectrometry multi-residue pesticide analysis of fruits and vegetables, and in inductively coupled plasma–mass spectrometry analysis of geological samples, are analysed as examples. The datasets prepared for ORDANOVA calculations with the freely available software tool are provided in supplementary materials to the paper. The reduction of risk by different components of the laboratory quality system (QS) are estimated under several error scenarios. New multinomial scores characterizing risk reduction by the laboratory QS as a whole are proposed.

1. Introduction

Assessment of decision-making risks in an analytical chemistry (testing) laboratory, directly influencing the quality of measurement/test results, is an important part of the laboratory's technical competence [1]. Such risks in performing a chemical analysis are caused by human errors [2,3].

Human error in a routine analytical laboratory may lead to out-of-specification test results [4] in the pharmaceutical industry, and results that do not comply with regulatory, legislation or specification limits in other industries and fields, e.g., environmental and food analysis.

The joint Guide on the risks from human errors in a laboratory performing chemical analysis, developed by the International Union of Pure and Applied Chemistry (IUPAC) and Cooperation on International Traceability in Analytical Chemistry (CITAC), was published in 2016 [5]. Today it is still actual [6–8].

The reduction of risk by different components of the laboratory quality system (QS) are estimated under several error scenarios. Residual risk of human error, not prevented or blocked by the laboratory QS, decreases the quality of analytical results and may be interpreted as a source of measurement uncertainty [9].

A laboratory expert in a specific chemical analysis has the necessary information to estimate the risks. Estimates (responses or judgements) of an expert are ordinal data on a specified scale, which can have empirical relations only; they can be equal or unequal to each other, greater than or less than; algebraic operations among ordinal data are not applicable [10–12]. Ordinal data should not be treated as continuous quantities since such treatment may lead to their misinterpretation and to inadequate laboratory management actions.

* Corresponding author.
E-mail address: ilya.kuselman@bezeqint.net (I. Kuselman).

https://doi.org/10.1016/j.talo.2025.100603
Received 10 October 2025; Received in revised form 3 December 2025; Accepted 4 December 2025
Available online 5 December 2025

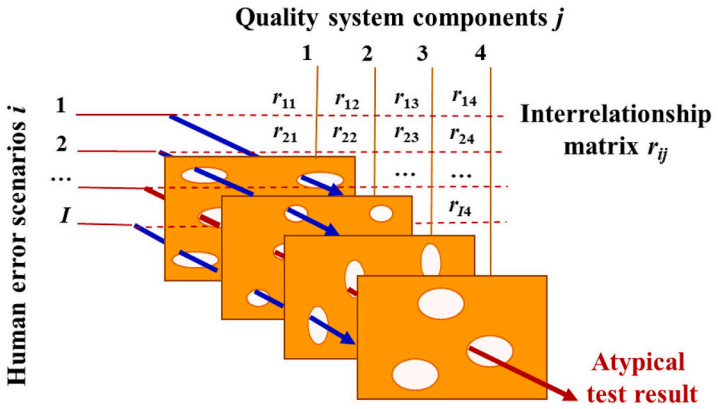

Fig. 1. A laboratory QS against human errors. Human error scenarios are indicated by rows $i = 1$ to I. QS components/layers are shown as the Swiss cheese slices (columns) $j = 1$ to 4. Bad outcomes blocked by the QS layers are presented by the blue pointers. Appearance of an atypical test result is depicted by the longest red pointer. Estimates of risk reduction r_{ij} of error scenario i, as the result of interaction between the error and layer j, form the interrelationship matrix. Figure modified from [5].

In the present paper, an approach to assessment of risks caused by human errors is described, based on the implementation of the two-way ordinal analysis of variation ORDANOVA [13] and the corresponding software tool [14]. This approach does not violate the properties of ordinal data, allowing for the correct interpretation of expert responses in an analytical chemistry (testing) laboratory [15] as it was supposed in the review [2].

2. Methods

The applied methods are briefly described below in the form allowing for their direct implementation in the discussed examples.

2.1. Classification, modelling, and evaluation of human errors

There are nine kinds of human errors including seven kinds of commission errors of a sampling inspector and/or an analyst/operator (knowledge-, rule- and skill-based mistakes; as well as routine, reasoned, reckless, and malicious violations) and two kinds of omission errors – lapses and slips. These errors may happen at any step of the chemical analytical measurement/testing process – location of the error. The kind of human error and the step of the analysis, in which the error may occur, form the event scenario $i = 1$ to I, where the number of such scenarios I is the product of the number of kinds of human errors (9 in this study) and the number of the steps of the analysis taken into account [5].

A Swiss cheese model [16] is used for characterizing the interaction of errors with a laboratory QS. This model applied here considers the following typical four QS components $j = 1$ to $J = 4$ as protective layers against human errors: 1) validation of the measurement/analytical method and formulation of the standard operating procedure – SOP; 2) training of analysts and proficiency testing; 3) quality control (QC) using statistical charts and/or other means; and 4) supervision.

A technique for quantifying human errors in chemical analysis using expert judgments was formulated based on the Swiss cheese model and the house-of-security approach [17]. According to this approach, a laboratory expert may estimate the possible reduction r_{ij} of the risk of human error scenario i because of the error blocking by QS layer j: a negligible reduction is estimated as $r_{ij} = 0$, weak reduction – as $r_{ij} = 1$, medium – as $r_{ij} = 3$, and strong (maximal) reduction – as $r_{ij} = 9$. The interrelationship matrix of r_{ij} has I rows and four columns ($4I$ entries), as shown in Fig. 1.

Blocking human error according to scenario i by a QS component j can be more effective in the presence of another component j' ($j' \neq j$) because of the synergy $\Delta_{jj'}^{(i)}$ between the two components. The synergy between two components is considered equal to 0 or 1 whenever the effect is absent or present, respectively, according to the expert judgements. For all QS components, the synergy factor is estimated as $s_{ij} = 1 + \sum_{j' \neq j}^{4} \Delta_{jj'}^{(i)}/3$, $1 <\rightarrow s_{ij} <\rightarrow 2$. Then, the score characterizing the risk reduction by the laboratory QS, and several other scores are calculated also applying basic algebraic operations for averaging expert responses. Some results of these calculations are compared further in the paper with the scores obtained based on ORDANOVA.

2.2. Analysis of variation of ordinal data

The implemented approach of ORDANOVA is to calculate for each characteristic the number of expert responses related to the same category, and then to analyze their relative frequency as a fraction of the total number of responses (of all categories) obtained for this characteristic. This approach is independent of the scale of the responses, represented in words, by numeric codes, or other means [15].

A vector of expert responses for a given characteristic of human error as a random variable $\boldsymbol{Y} = (Y_1,\ Y_2,\ Y_3,\ Y_4)$ on the ordinal scale (0, 1, 3, 9) is related to the four categories k: category $k = 1$ of responses 0; $k = 2$ of responses 1; $k = 3$ of responses 3; and $k = 4$ of responses 9. The vector of theoretical probabilities p_k, $k = 1$ to 4, of the events Y_k is $\boldsymbol{p} = (p_1,\ p_2,\ p_3,\ p_4)$, where $\sum_{k=1}^{4} p_k = 1$. Let F_k denote the cumulative theoretical probability up to the k-th category, $F_k = \sum_{q=1}^{k} p_k$, $q = 1, \ldots,\ k$, and $F_4 = 1$. The probability P of receiving a set of responses $\boldsymbol{n} = (n_1, n_2, n_3, n_4)$, where n_k is the number of responses related to the k-th category $\left(\sum_{k=1}^{4} n_k = N\right)$, can be calculated based on the multinomial distribution with parameters (N, $\boldsymbol{p}$) as the probability mass function [18]:

$$P(\boldsymbol{Y} = \boldsymbol{n}) = \frac{N!}{n_1!n_2!n_3!n_4!} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} \quad (1)$$

Variability in $\boldsymbol{Y}$ is explained in the present study by two independent factors: human error scenarios $X1$ and components of a laboratory QS $X2$ as fixed factors. An expert response of category k on reduction r_{ij} of a risk of human error according to scenario i by a component j of the QS, is a realization Y_{ijk} of variable $\boldsymbol{Y}$, when the first factor $X1$ has I levels ($i = 1$ to I error scenarios) and the second factor $X2$ has J levels ($j = 1$ to 4 components of the QS). There are N expert responses in total for a given chemical analysis in the laboratory, each response of one of categories $k = 1$ to 4 of variable $\boldsymbol{Y}$. Thus, it is assumed that there is no

replication of the responses, and each of $4I = N$ cells (i, j) contain one only response of the chosen category k, i.e. $n_{ijk} = 1$ and zero for all other categories in the same cell.

Treating N responses as a statistical sample, and Y_{ijk} as a random variable, then for the chosen k probability of the event is $\widehat{p}_{ijk} = 1$ and zero for all other categories in the same cell. Thus, $\widehat{F}_{ijk} = \sum_{q=1}^{k} \widehat{p}_{ijq}$ denotes the sample cumulative frequency of responses of categories $q = 1$ to k, i.e. up to the k-th category in cell (i, j); $\widehat{F}_{i.k}$ – the sample cumulative relative frequency of responses up to the k-th category at level i of factor $X1$; $\widehat{F}_{.jk}$ – the cumulative relative frequency at level j of factor $X2$; and $\widehat{F}_{..k}$ – the cumulative relative frequency of all responses up to the k-th category:

$$\widehat{F}_{i.k} = \frac{1}{4}\sum_{j=1}^{4}\widehat{F}_{ijk}, \quad \widehat{F}_{.jk} = \frac{1}{I}\sum_{i=1}^{I}\widehat{F}_{ijk}, \text{ and } \widehat{F}_{..k} = \frac{1}{4I}\sum_{i=1}^{I}\sum_{j=1}^{4}\widehat{F}_{ijk}\ . \tag{2}$$

The total sample variation $\widehat{V}_{\mathrm{T}}$ of the response variable $\boldsymbol{Y}$, normalized to the [0, 1] interval, is defined in the two-way ORDANOVA model [13] for categories $k = 1$ to 4 as

$$\widehat{V}_{\mathrm{T}} = \frac{1}{3/4}\sum_{k=1}^{3}\widehat{F}_{..k}(1-\widehat{F}_{..k}). \tag{3}$$

In the model without replication of the responses, $\widehat{V}_{\mathrm{T}}$ is partitioned into the between covariation component $\widehat{C}_{\mathrm{B}}$ and the within residual variation $\widehat{V}_{\mathrm{W}}$. The individual effects of factors $X1$ and $X2$ (scenarios and QS components, respectively) can be evaluated using the $\widehat{C}_{\mathrm{B}}$ decomposition into corresponding components $\widehat{C}^{\mathrm{B}}_{X1}$ and $\widehat{C}^{\mathrm{B}}_{X2}$ [13]. In addition, $\widehat{C}_{\mathrm{B}}$ decomposition by response categories was also discussed in refs [19,20]. Such decomposition may include a component related to the possible interaction between the two factors. However, no interaction between the two factors can be analyzed, when only one expert response at the specified levels of the factors is obtained.

The null hypothesis of homogeneity of the responses states that the probability of classifying the responses as belonging to the k-th category does not depend on the levels of the first factor (levels i) nor on those of the second factor (levels j), i.e., $p_{ijk} = p_k$ for all $i = 1$ to I and $j = 1$ to 4. Under this hypothesis, the following relations are applicable for both factors:

$$\frac{E(\widehat{V}_{\mathrm{T}})}{df_{\mathrm{T}}} = \frac{E(\widehat{C}^{\mathrm{B}}_{X1})}{df_{X1}} = \frac{E(\widehat{C}^{\mathrm{B}}_{X2})}{df_{X2}} = \frac{V_{\mathrm{T}}}{N}, \tag{4}$$

where E denotes the mathematical expectation of the variation (in parentheses); $df_{X1} = I - 1$, $df_{X2} = 4 - 1 = 3$, and $df_{\mathrm{T}} = N - 1$ are degrees of freedom. The numerator of the last term in Eq. (4) is the population total ordinal variation corresponding to the probability vector $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$.

To check the statistical significance of effects of both the factors the following significance indices (test statistics) have been applied:

$$\widehat{SI}_{X1} = \frac{\widehat{C}^{\mathrm{B}}_{X1}/df_{X1}}{\widehat{V}_{\mathrm{T}}/df_{\mathrm{T}}} \text{ and } \widehat{SI}_{X2} = \frac{\widehat{C}^{\mathrm{B}}_{X2}/df_{X2}}{\widehat{V}_{\mathrm{T}}/df_{\mathrm{T}}} \tag{5}$$

Testing the null hypothesis H_0 on the effect significance requires knowledge of at least the asymptotical distribution of the index $\widehat{SI}$ for calculation of the critical values of the indices at a given level of confidence $(1-\alpha)\cdot 100$ %, where α is the probability of a Type I error to reject the correct null hypothesis H_0.

The sample vector of relative frequencies $\widehat{\boldsymbol{p}} = (\widehat{p}_{..1}, \widehat{p}_{..2}, \widehat{p}_{..3}, \widehat{p}_{..4})$, as well as the variations $(\widehat{C}^{\mathrm{B}}_{Xl}, \widehat{V}_{\mathrm{W}}, \widehat{V}_{\mathrm{T}})$ and the empirical significance indices $\widehat{SI}_{Xl}$, where $l = 1, 2$, are calculated from the expert responses. The critical values SI^{crit}_{Xl} for the indices in Eq. (5) are recovered through Monte Carlo simulations based on the multinomial distribution. The null hypothesis H_0 is rejected when the significance index $\widehat{SI}_{Xl}$ exceeds the critical value SI^{crit}_{Xl} at the $(1-\alpha)\cdot 100$ % level of confidence, concluding that a statistically significant effect on the response variable $\boldsymbol{Y}$ is detected. Calculation of the power values of the test $P_l = 1-\beta_l$, where β is the probability of a Type II error of not rejecting the H_0 when it is incorrect, is described for different effects w of the sample size in refs [14,15] also. The software tool is freely available [21].

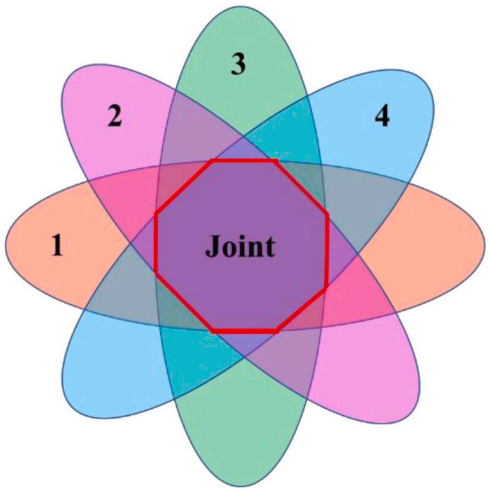

Fig. 2. Venn diagram of the events. The event $\boldsymbol{Y}_1 = \boldsymbol{n}_1$, when the probabilities of the expert responses to the risk reduction by QS component 1 are as in the specific vector $\boldsymbol{p}_1$, is shown with a semi-transparent brown ellipse 1; similar color ellipses indicate: 2 (violet) - the event $\boldsymbol{Y}_2 = \boldsymbol{n}_2$ for component 2; 3 (green) - the event $\boldsymbol{Y}_3 = \boldsymbol{n}_3$ for component 3; and 4 (blue) - the event $\boldsymbol{Y}_4 = \boldsymbol{n}_4$ for component 4. The joint (intersection) event, consisting of the corresponding responses to the four QS components simultaneously, is highlighted by the central red octagon shape. Figure modified from [22].

2.3. Multinomial scores

The probability mass function P by Eq. (1) of a multinomial random variable $\boldsymbol{Y}$, characterized by a vector $\boldsymbol{p}$ of response probabilities, is the probability that the event $\boldsymbol{Y} = \boldsymbol{n}$ occurs. Four such multinomial variables, each corresponding to the risk reduction r_{ij} by a QS component j applied against human errors, will be used further with corresponding subscripts j from 1 to 4 in the symbols of variables and parameters related to these QS components.

The probability P_{joint} of the joint (intersection) event, consisting of the expert responses to the four components simultaneously, can be formulated [15,22] as

$$P_{join} = p(\{\boldsymbol{Y}_1 = \boldsymbol{n}_1\} \cap \{\boldsymbol{Y}_2 = \boldsymbol{n}_2\} \cap \{\boldsymbol{Y}_3 = \boldsymbol{n}_3\} \cap \{\boldsymbol{Y}_4 = \boldsymbol{n}_4\}), \tag{6}$$

where $\cap$ is the symbol of intersection of events. The Venn diagram of the events is shown in Fig. 2.

When responses to these four QS components are independent, the probability of the joint event (joint probability) is the following product:

$$P_{\mathrm{joint}} = P_1 \cdot P_2 \cdot P_3 \cdot P_4. \tag{7}$$

Treating N responses to each component as a separate statistical sample, and corresponding frequencies n_k as random variables, the probability vector $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$ is estimated for the components as a vector of relative frequencies $\widehat{\boldsymbol{p}} = (\widehat{p}_1, \widehat{p}_2, p_3, \widehat{p}_4)$, where $\widehat{p}_k = n_k/N$. Then, an estimate of P_1 by Eq. (1) is $\widehat{P}_1$, similarly for the other probabilities, while the estimate of the joint probability by Eq. (7) is $\widehat{P}_{\mathrm{joint}} = \widehat{P}_1 \cdot \widehat{P}_2 \cdot \widehat{P}_3 \cdot \widehat{P}_4$. The estimate $\widehat{P}_{\mathrm{joint}}$ of the probability P_{joint} is a characteristic of the risk reduction by the laboratory QS as a whole, expressed initially as the expert responses r_{ij} on the ordinal scale.

Table 1
Results of two-way ORDANOVA of expert responses.

Example	$\widehat{V}_T$	$\widehat{C}_B$	$\widehat{V}_w$	l	df_{Xl}	$\widehat{C}^B_{Xl}$	$\widehat{SI}_{Xl}$	SI^{crit}_{Xl}	P_l
	pH *measurements of groundwater*								
1	0.766	0.460	0.306	1	35	0.192	1.023	1.270	0.198
				2	3	0.268	16.673	2.157	0.484
	GC–MS analysis of pesticide residues in fruits and vegetables								
2	0.647	0.368	0.279	1	53	0.225	1.408	1.217	0.244
				2	3	0.143	15.872	2.187	0.638
	ICP-MS analysis of geological samples								
3	0.480	0.220	0.260	1	35	0.179	1.525	1.284	0.192
				2	3	0.040	4.008	2.265	0.447

The responses to two or more QS components might not be independent, e.g., when a synergy between them is observed. In such cases the probability of the joint (intersection) event P_{joint} can be represented numerically by a Gaussian copula-based procedure [23,24]. This procedure is used for generating samples from a discrete multivariate random variable with prescribed experimental marginal cumulative distributions for the four QS components $\widehat{F}_{k\_1}$, $\widehat{F}_{k\_2}$, $\widehat{F}_{k\_3}$, $\widehat{F}_{k\_4}$ and an empirical correlation matrix of those quality properties. A large number (typically at least 10^6) of those multivariate samples, each of size N, are generated to estimate $\widehat{P}_{joint}$ of P_{joint} as the relative frequency of realization of the intersection event $(\{\boldsymbol{Y}_1 = \boldsymbol{n}_1\} \cap \{\boldsymbol{Y}_2 = \boldsymbol{n}_2\} \cap \{\boldsymbol{Y}_3 = \boldsymbol{n}_3\} \cap \{\boldsymbol{Y}_4 = \boldsymbol{n}_4\})$.

The multinomial score of the QS (a kind of quality index [15,22]) is the negative common logarithm $Q = -\lg(\widehat{P}_{joint})$. When $\widehat{P}_{joint}$ tends to 1, the score achieves its minimum value $Q = 0$. The greater Q, the higher the chance that the abilities of the QS against the human errors will differ from the estimated ones. The calculation code in the R programming environment is presented in the paper [22].

Other multinomial scores, e.g., for evaluation of an error likelihood and severity, effectiveness of a separate component of the laboratory QS against human errors and/or effectiveness of the QS at different steps of the chemical analysis, may also be calculated in the same way.

3. Experimental

The dataset for assessment of the risks in pH measurements of groundwater available in the Guide [5], Example 1, includes the expert judgments/responses on the nine kinds of human errors (listed above in Section 2.1) at the four steps of the measurement process: 1) choice of the method, corresponding equipment, and SOP; 2) sampling; 3) proper pH measurement; 4) calculation and reporting of the test results. The number of scenarios of human errors was $I = 9 \times 4 = 36$. There are responses on the reduction of the error risk by the four components of the QS ($J = 4$). Thus, the interrelationship matrix has $N = 4I = 144$ cells (i,j). This matrix, transformed into a comma-separated input text file for calculations with the ORDANOVA software tool [21], named "Supplementary dataset 1", is provided in supplementary materials to the present paper. The parameters (I, J, K, Y) of the two-way model without replication are represented in the file in four numerical columns: the first column contains numbers of the error scenarios $i = 1$ to I, the second – numbers of the QS components $j = 1$ to 4, the third – categories $k = 1$ to 4 of the ordinal scale (0, 1, 3, 9), and the fourth – expert responses Y_{ijk} equal to 1 for the chosen category k on the scale and 0 for others. Since each of $N = 144$ responses was chosen from 4 categories, there are $4N = 576$ rows in the file.

Six steps of the measurement process of GC–MS analysis of pesticide residues in fruits and vegetables were highlighted in the Guide [5], Example 2: 1) sampling, 2) sample processing, 3) sample preparation, 4) identification and confirmation of pesticides, 5) measurement of their amount in the extract, and 6) calculation of the pesticide concentrations in the analyzed sample and reporting. Therefore, the number of scenarios of the nine kinds of human errors was here $I = 54$, and the number of cells of the interrelationship matrix at the same four components of the QS was $N = 216$. The corresponding file in the supplementary materials is named "Supplementary dataset 2." This file contains the above-mentioned 4 numerical columns and 864 rows.

The typical main steps of the ICP-MS analysis of geological samples described in the Guide [5], Example 3, are 1) sample preparation, 2) calibration of the ICP-MS measuring system, 3) measurement of analyte

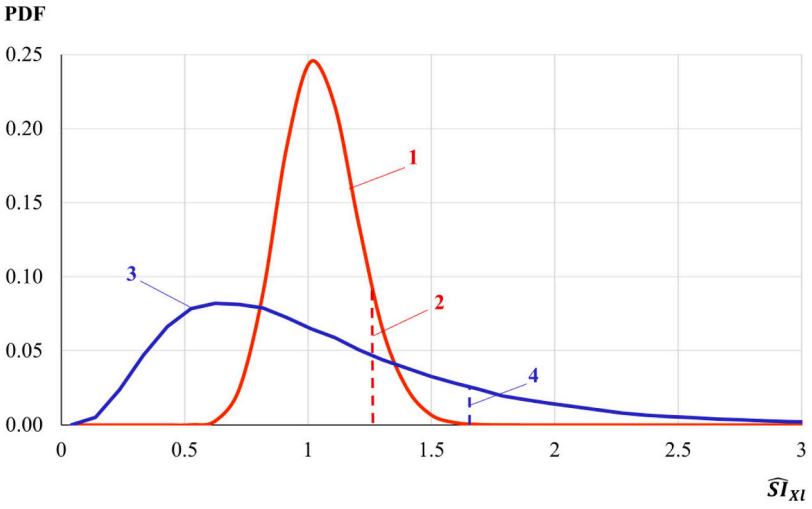

Fig. 3. Probability density functions (PDFs) of the significance index $\widehat{SI}_{Xl}$, $l = 1$ and 2. Red solid curve 1 is related to the scenarios of errors in the pH measurements of groundwater – factor $X1$; dash-dotted line 2 shows the critical value SI^{crit}_{X1} of the significance index $\widehat{SI}_{X1}$ at 95 % level of confidence. Blue solid curve 3 is for the QS components – factor $X2$; dash-dotted line 4 indicates critical value SI^{crit}_{X2} of the significance index $\widehat{SI}_{X2}$ at 95 % level of confidence.

concentrations in the prepared solutions, and 4) calculation of elemental mass fractions in analyzed samples and reporting. Given the nine kinds of human errors and the four components of the laboratory QS, there are $I = 36$ scenarios of human errors and $N = 144$ cells of the interrelationship matrix, as in Example 1 for pH measurements of groundwater. The file "Supplementary dataset 3", also provided in the supplementary materials, is of the same size as "Supplementary dataset 1."

4. Results and discussion

4.1. Analysis of ordinal values

The values of the total variation $\widehat{V}_T$ of the responses without replication by Eq. (3), partitioned into the between (inter) covariation component $\widehat{C}_B$ and the within (intra) residual variation $\widehat{V}_w$, are represented in Table 1.

There are also the individual effects $\widehat{C}^B_{Xl}$, $l = 1, 2$, of factors $X1$ and $X2$ (human error scenarios and components of the QS, respectively) evaluated using the $\widehat{C}_B$ decomposition. To check the statistical significance of effects of each factor by Eq. (4), significance indices $\widehat{SI}_{Xl}$ were calculated by Eq. (5) with degrees of freedom df_{Xl}. The critical index values SI^{crit}_{Xl} at 95 % level of confidence and the power of the test of hypothesis H_0 in Table 1 were also calculated using 10^5 iterations and the effect of sample size $w = 0.3$ in the input to the software tool [21]. The generated distributions of significance indices $\widehat{SI}_{Xl}$ and the critical values SI^{crit}_{Xl} are shown in Fig. 3 for pH measurements of groundwater as an example.

The vector of relative frequencies $\widehat{\boldsymbol{p}} = (0.174, 0.069, 0.201, 0.556)$ for categories $k = 1$ to 4 in the tool output for Example 1, pH measurements of groundwater, shows that expert responses of the highest category $k = 4$ dominated: $\widehat{p}_{.4} = 0.556$. There is no statistically significant difference in the influence of the scenarios on the risk reduction as $\widehat{SI}_{X1} < SI^{crit}_{X1}$ at 95 % level of confidence. In other words, the scenarios are homogeneous from the point of view of risk reduction. However, the differences between the QS components are significant: $\widehat{SI}_{X2}$ is considerably greater than the corresponding SI^{crit}_{X2}. The power P_2 of the applied criterion is more than twice the power P_1 in the case of scenarios.

In Example 2, GC–MS analysis of pesticide residues in fruits and vegetables, the vector of relative frequencies $\widehat{\boldsymbol{p}} = (0.083, 0.116, 0.315, 0.486)$ is similar to that in Example 1, but the domination of the expert responses of the category $k = 4$ was not so expressed in comparison with the frequency $\widehat{p}_{.3} = 0.315$ of category $k = 3$. Both factors, the risk scenarios and the QS components, influence the risk reduction statistically significantly here as $\widehat{SI}_{X1} > SI^{crit}_{X1}$ and $\widehat{SI}_{X2} > SI^{crit}_{X2}$ at 95 % level of confidence. The power P_2 of the applied criterion for the QS components is greater again, more than twice the power P_1 for the scenarios. However, both the power values in this example are greater than those in Example 1 since the number of scenarios I and corresponding total number N of expert responses in Example 2 are also greater.

Example 3, ICP-MS of geological samples, has the same parameters I and J as in Example 1. However, the vector $\widehat{\boldsymbol{p}} = (0.028, 0.083, 0.514, 0.375)$ demonstrates a superiority of the moderate expert responses of category $k = 3$ with relative frequency $\widehat{p}_{.3} = 0.514$. The scenarios and the QS components are also statistically significant factors here for risk reduction. The criteria power values P_1 and P_2 are practically the same as in Example 1, since the statistical samples of the elicited expert responses were of the same size.

Note that the ability of ORDANOVA to test statistical significance of human error scenarios and QS components, as the factors influencing risk reduction, is a new tool in the risk assessment in comparison with the house-of-security approach [5] allowing the understanding of the important factors in a specific chemical analytical method.

Table 2
Spearman's rho correlation coefficients of the expert responses to risk reduction by QS components.

Example	Component, j	Validation 1	Training 2	QC 3	Supervision 4
	pH measurements of groundwater				
1	1	1.000	0.648**	0.651**	-
	2	0.648**	1.000	0.016	-
	3	0.651**	0.016	1.000	-
	4	-	-	-	-
	GC–MS analysis of pesticide residues in fruits and vegetables				
2	1	1.000	0.680**	0.357**	0.153
	2	0.680**	1.000	0.413**	0.314*
	3	0.357**	0.413**	1.000	0.606**
	4	0.153	0.314*	0.606**	1.000
	ICP-MS analysis of geological samples				
3	1	1.000	0.635**	0.366*	0.261
	2	0.635**	1.000	0.307	0.248
	3	0.366*	0.307	1.000	0.216
	4	0.261	0.248	0.216	1.000

* Significant at a level of confidence of 95 %.
** Significant at a level of confidence of 99 %.

4.2. Testing correlation of the expert responses

Spearman's rho correlation coefficients $\rho_{jj'}$, $j' \neq j$, calculated with the IBM SPSS software tool [25] were used for evaluating the synergy between the QS components for the risk reduction, according to the expert responses. The synergy is complete (the expert responses as variables are strongly correlated) when $\rho_{jj'} = \pm 1$, and the synergy is absent when $\rho_{jj'} = 0$. The matrices of $\rho_{jj'}$ of the responses to reduction of the risks by each pair of the four components j & j' of the QS, are presented in Table 2.

Correlation coefficients $\rho_{jj'}$, significant at a level of confidence of 95 % (probability of the two-tailed Type I error $\alpha = 0.05$) are marked in Table 2 by one asterisk, while those significant at a level of confidence of 99 % ($\alpha = 0.01$) are marked by two asterisks; other $\rho_{jj'}$ are considered as not significant. No negative correlation ($\rho_{jj'} < 0$) was observed.

In Example 1, pH measurement of groundwater, $\rho_{jj'}$ are statistically significant at a level of confidence of 99 % and $\alpha = 0.01$ for the pairs "validation & training" ($j = 1$ & $j' = 2$) and "validation & QC" ($j = 1$ & $j' = 3$). Supervision, as QS component $j = 4$, was not involved in the correlation testing here since all the expert responses to the error reduction by supervision were equally high (category $k = 4$ on the ordinal scale) for the 36 discussed risk scenarios: the response values are not a variable in such a case. Note that the synergy factor s_{i1} applied in the house-of-security approach (Section 2.1) for quantification of the risk reduction at different scenarios i of the errors by validation ($j = 1$) in Example 1 was 1 to 1.67. For training and QC, as QS components $j = 2$ and 3, s_{i2} and s_{i3} were 1 to 1.33. The value $s_{i4} = 1$ used for supervision at all scenarios [5] means that no correlation with other QS components was assumed.

The correlation matrix is more complex for Example 2, GC–MS analysis of pesticide residues in fruits and vegetables. While ρ_{12} of the pairs "validation & training" is evaluated to be close to that in Example 1 at a level of confidence of 99 % and Type I error $\alpha = 0.01$, ρ_{13} for the pairs "validation & QC" is about half as much, being statistically significant at the same level of confidence.

There is also a similar correlation between "training & QC" ($j = 2$ & $j' = 3$) and "QC & supervision" ($j = 3$ & $j' = 4$). With the larger probability of Type 1 error $\alpha = 0.05$, at a level of confidence of 95 %, the correlation between "training & supervision" ($j = 2$ & $j' = 4$) can also be considered statistically significant. The synergy factors s_{ij} applied in the house-of-security approach in Example 2 [5] were the same as in Example 1.

In Example 3, ICP analysis of geological samples, ρ_{12} of the pairs "validation & training" is also evaluated to be close to those in Examples 1 and 2 at a level of confidence of 99 % and Type I error $\alpha = 0.01$. Correlation of the pairs "validation & QC" is similar to the case of

Table 3
Probabilities of the responses related to QS components and multinomial score values.

Example	Validation	Training	QC	Supervision	$\widehat{P}_{joint}$		$Q \pm s_Q$
	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	Uncorrelated	Correlated	
	pH *measurements of groundwater*						
1	0.006	0.033	0.025	1.000	$4.8{\cdot}10^{-6}$	$6.5{\cdot}10^{-6}$	5.2 ± 0.1
	GC–MS analysis of pesticide residues in fruits and vegetables						
2	0.003	0.021	0.017	0.047	$4.7{\cdot}10^{-8}$	$1.2{\cdot}10^{-7}$	6.9 ± 0.2
	ICP-MS analysis of geological samples						
3	0.006	0.029	0.137	0.132	$3.0{\cdot}10^{-6}$	$5.1{\cdot}10^{-6}$	5.3 ± 0.1

Example 2, but the statistical significance of ρ_{13} is supported at a level of confidence of 95 % and Type I error $\alpha = 0.05$ as the size of the statistical sample of the responses in Example 3 is smaller than in Example 2. The synergy factors s_{ij} applied in the house-of-security approach in Example 3 [5] were the same as in Examples 1 and 2.

Thus, the expert estimates/judgements concerning the synergy of the QS components do not contradict the Spearman's correlation coefficients $\rho_{jj'}$, calculated independently from these judgements.

4.3. Calculation of the multinomial scores

The probability mass functions $\widehat{P}_j$, $j = 1$ to 4, calculated by Eq. (1) and interpreted as probabilities of the sets of the expert responses of different categories on risk reduction by QS components, are presented in Table 3. The most probable responses in Examples 1 and 2 are related to supervision ($j = 4$), while in Example 3 – to QC ($j = 3$). For comparison, according to the results of the data analysis based on the house-of-security approach in the report [5], the most effective QS component for risk reduction in Example 1 was training ($j = 2$), whereas in Example 2 – supervision, and in Example 3 – QC, as in the present study. The difference in the analysis results for Example 1 may be caused by violation of the properties of ordinal data in the house-of-security approach.

There are also $\widehat{P}_{joint}$ values for the QS as a whole, estimated both for the case of the uncorrelated system components, as a product of $\widehat{P}_j$ according to Eq. (7), and for the observed correlation structure represented in Table 2. The observed correlation (synergy) of the system components was considered by the Gaussian copula-based procedure [23,24] coded in the R programming environment [22]. The code exploits the package GenOrd [26] for generating samples from a multivariate discrete random variable with a pre-specified correlation matrix and marginal distributions. The procedure [26] was developed in two steps: the first step (function *ordcont*) sets up the Gaussian copula in order to achieve the desired correlation matrix on the target random discrete components; the second step (*ordsample*) generates samples from the target variables. The intermediate function *contord* computes the correlations of the multivariate discrete variable derived from correlated variables through discretization. Function *corrcheck* returns the lower and upper bounds of the correlation coefficient of each pair of discrete variables given their marginal distributions, i.e. returns the range of feasible bivariate correlations.

When running the *ordcont* and the *ordsample* functions in Examples 1 and 2, a problem has arisen that the corresponding correlation matrix in Table 2 was not coherent with the given marginal distributions of the expert responses. Substituting the original correlation matrix with the intermediate one provided by the function *contord* solved this issue. The number of generated datasets (each of the same size N as of the original one) was 10^7. For each Example, the simulation was performed 10 times to assess the reproducibility of the results. The multinomial score of the QS taken as the mean Q value obtained through the simulations, and its standard deviation s_Q from the mean, are presented in Table 3. Corresponding $\widehat{P}_{joint}$ for the correlated expert responses in Table 3 is recovered as the antilogarithm of the negative mean Q value.

While the calculation reproducibility is characterized by the s_Q values, the accuracy is assessed comparing the mean Q values for the diagonal correlation matrices (when correlations are absent) with the "theoretical" values $Q = -\lg(\widehat{P}_{joint})$, where $\widehat{P}_{joint}$ are estimated by Eq. (7): they practically coincided.

The probabilities $\widehat{P}_{joint}$ are naturally larger when the synergy (correlation) of the system components is considered for all Examples in Table 3. The Q value for Example 2 is the largest here as the corresponding probability $\widehat{P}_{joint}$ is the smallest. In other words, the expert judgements concerning risk reduction by the QS in the case of Example 2 were the most cautious.

5. Conclusions

A new approach to the assessment of decision-making risks caused by human errors in performing a chemical analysis is described based on implementation of the recently developed two-way ORDANOVA. This approach calculates the number of expert responses related to the same category for each ordinal characteristic and then analyzes their relative frequency as a fraction of the total number of responses (of all categories) obtained for this characteristic. It does not violate the properties of ordinal data, allowing for the correct interpretation of expert responses. Previously published datasets of the expert responses on the risks in pH measurements of groundwater, in GC–MS analysis of pesticide residues in fruits and vegetables, and in ICP-MS analysis of geological samples, were analysed as examples. Risk reduction by different components of the laboratory QS is estimated under several error scenarios. Unlike the house-of-security approach, calculation of a mean of the expert responses on the ordinal scale (or another algebraic operation on those responses) was neither required nor applied.

The multinomial scores characterizing risk reduction by the laboratory QS as a whole are proposed. This is also a new tool in risk assessment allowing for comparison of risk reduction between different methods and different laboratories.

CRediT authorship contribution statement

Francesca R. Pennecchi: Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Tamar Gadrich:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis. **Ilya Kuselman:** Writing – original draft, Validation, Investigation, Conceptualization. **D. Brynn Hibbert:** Writing – review & editing, Supervision. **Angelique Botha:** Writing – review & editing, Validation. **Anastasia A. Semenova:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge organizational support from the International Union of Pure and Applied Chemistry, IUPAC Project 2024–012–2–500.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.talo.2025.100603.

Data availability

Data are presented in the paper supplementary materials.

References

[1] International Organization for Standardization, General requirements for the competence of calibration and testing laboratories, ISO/IEC 17025:2017, ISO/IEC, Geneva, 2017. https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html.
[2] I. Kuselman, F.R. Pennecchi, D.B. Hibbert, A. Botha, T. Gadrich, A.A. Semenova, Advanced methods for assessment of risks of false decisions in analytical chemistry (testing) laboratories – a review, Talanta 294 (2025) 128208, https://doi.org/10.1016/j.talanta.2025.128208.
[3] I. Kuselman, F. Pennecchi, W. Bich, D.B. Hibbert, Human being as a part of measuring system influencing measurement results, Accredit. Qual. Assur. 21 (2016) 421–424, https://doi.org/10.1007/s00769-016-1239-3.
[4] Center for Drug Evaluation and Research (CDER), Investigating Out-Of-Specification (OOS) Test Results For Pharmaceutical Production - Level 2 Revision: Guidance for industry, U.S. Department of Health and Human Services Food and Drug Administration, 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/investigating-out-specification-oos-test-results-pharmaceutical-production-level-2-revision.
[5] I. Kuselman, F. Pennecchi, IUPAC/CITAC Guide: classification, modeling and quantification of human errors in a chemical analytical laboratory (IUPAC technical report), Pure Appl. Chem. 88 (2016) 477–515, https://doi.org/10.1515/pac-2015-1101.
[6] Y. Zhao, A Bayesian approach to comparing human reliability analysis methods using human performance data, Reliab. Eng. Syst. Saf. 219 (2022) 108213, https://doi.org/10.1016/j.ress.2021.108213.
[7] L.-X. Hou, R. Liu, H.-C. Liu, S. Jiang, Two decades on human reliability analysis: a bibliometric analysis and literature review, Ann. Nucl. Energy 151 (2021) 107969, https://doi.org/10.1016/j.anucene.2020.107969.
[8] B. Kirwan, A Guide To Practical Human Reliability Assessment, 1st ed., CRC Press, London, 2017, p. 587, https://doi.org/10.1201/9781315136349.
[9] I. Kuselman, F. Pennecchi, Human errors and measurement uncertainty, Metrologia 52 (2015) 238, https://doi.org/10.1088/0026-1394/52/2/238.
[10] I.E.C. BIPM, I.L.A.C. IFCC, I.U.P.A.C. ISO, O.I.M.L. IUPAP, International vocabulary of metrology – basic and general concepts and associated terms (VIM), JCGM 200 (2012) 2012, https://doi.org/10.59161/JCGM200-2012.
[11] International Organization for Standardization, Statistics – Vocabulary and symbols – Part 2: Applied statistics, ISO 3534–2:2006, ISO, Geneva, Switzerland, 2006, pp. 133.
[12] D.B. Hibbert, Compendium of Terminology in Analytical Chemistry: IUPAC Orange Book, 4th edition, RSC, London, 2023, p. 666, https://doi.org/10.1039/9781788012881.
[13] T. Gadrich, Y.N. Marmor, Two-way ORDANOVA: analyzing ordinal variation in a cross-balanced design, J. Stat. Plan. Inference 215 (2021) 330–343, https://doi.org/10.1016/j.jspi.2021.04.005.
[14] T. Gadrich, Y.N. Marmor, F.R. Pennecchi, D.B. Hibbert, A.A. Semenova, I. Kuselman, Power of a test for assessing interlaboratory consensus of nominal and ordinal characteristics of a substance, material, or object, Metrologia 61 (2024) 045004, https://doi.org/10.1088/1681-7575/ad5846.
[15] I. Kuselman, T. Gadrich, F.R. Pennecchi, D.B. Hibbert, A.A. Semenova, A. Botha, IUPAC/CITAC Guide: interlaboratory comparison of categorical characteristics of a substance, material, or object (IUPAC technical report), Pure Appl. Chem. 97 (2025), https://doi.org/10.1515/pac-2025-0408.
[16] J. Reason, Human error: models and management, BMJ 320 (2000) 768–770, https://doi.org/10.1136/bmj.320.7237.768.
[17] S. Dror, E. Bashkansky, R. Ravid, House of security: a structured system design & analysis approach, Int. J. of Safety and Security Eng. 2 (2012) 317–329. https://www.witpress.com/elibrary/sse-volumes/2/4/668.
[18] National Institute of Standards and Technology, E-Handbook of Statistical Methods: Multinomial PDF, 2004. https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/multpdf.htm.
[19] T. Gadrich, I. Kuselman, F.R. Pennecchi, D.B. Hibbert, A.A. Semenova, P.S. Cheow, V.N. Naidenko, Interlaboratory comparison of the intensity of drinking water odor and taste by two-way ordinal analysis of variation without replication, J. Water. Health 20 (2022) 1005–1016, https://doi.org/10.2166/wh.2022.060.
[20] T. Gadrich, F.R. Pennecchi, I. Kuselman, D.B. Hibbert, A.A. Semenova, P.S. Cheow, Ordinal analysis of variation of sensory responses in combination with multinomial ordered logistic regression vs. chemical composition: a case study of the quality of a sausage from different producers, J. Food Qual. 2022 (2022) 4181460, https://doi.org/10.1155/2022/4181460.
[21] Y.N. Marmor, Research Areas - Factor Analysis Calculator Tool for Categorical Data, 5 May 2025. https://w3.braude.ac.il/lecturer/dr-yariv-n-marmor/.
[22] T. Gadrich, F.R. Pennecchi, I. Kuselman, D.B. Hibbert, A.A. Semenova, M. Salikova, A novel multisensory quality index of a food product: an analysis of a sausage properties, Chemometrics Intell. Lab. Syst. 237 (2023) 104815, https://doi.org/10.1016/j.chemolab.2023.104815.
[23] A. Panagiotelis, C. Claudia, H. Joe, Pair copula constructions for multivariate discrete data, J. Am. Stat. Assoc. 107 (2012) 1063–1072, https://doi.org/10.1080/01621459.2012.682850.
[24] P.A. Ferrari, A. Barbiero, Simulating ordinal data, Multivariate Behav. Res. 47 (2012) 566–589, https://doi.org/10.1080/00273171.2012.692630.
[25] International Business Machines Corporation (IBM), IBM SPSS Software, 25 May 2025. https://www.ibm.com/analytics/spss-statistics-software.
[26] A. Barbiero, P.A. Ferrary, Package 'GenOrd'. Simulation of discrete random variables with given correlation matrix and marginal distributions, 2015. https://doi.org/10.32614/CRAN.package.GenOrd.