IUPAC Technical Report

Ilya Kuselman* and Francesca Pennecchi

IUPAC/CITAC Guide: Classification, modeling and quantification of human errors in a chemical analytical laboratory (IUPAC Technical Report)

DOI 10.1515/pac-2015-1101
Received November 5, 2015; accepted April 20, 2016

Abstract: The classification, modeling, and quantification of human errors in routine chemical analysis are described. Classifications include commission errors (mistakes and violations) and omission errors (lapses and slips) in different scenarios at different steps of the chemical analysis. A Swiss cheese model is used to characterize error interaction with a laboratory quality system. The quantification of human errors in chemical analysis, based on expert judgments, i.e. on the expert(s) knowledge and experience, is applied. A Monte Carlo simulation of the expert judgments was used to determine the distributions of the error quantification scores (scores of likelihood and severity, and scores of effectiveness of a laboratory quality system against the errors). Residual risk of human error after the error reduction by the laboratory quality system and consequences of this risk for quality and measurement uncertainty of chemical analytical results are discussed. Examples are provided using expert judgments on human errors in pH measurement of groundwater, multiresidue analysis of pesticides in fruits and vegetables, and elemental analysis of geological samples by inductively coupled plasma mass spectrometry.

Keywords: chemical analysis; human error; metrology; quality; risk management.

CONTENTS

Article note: Sponsoring bodies: IUPAC Analytical Chemistry Division; IUPAC Interdivisional Working Party on Harmonization of Quality Assurance; Cooperation on International Traceability in Analytical Chemistry (CITAC): see more details in p. 513.

***Corresponding author: Ilya Kuselman,** Currently Independent Consultant on Metrology, 4/6 Yarehim St., 7176419 Modiin, Israel, e-mail: ilya.kuselman@bezeqint.net; National Physical Laboratory of Israel, Danciger A Bldg, Givat Ram, 91904 Jerusalem, Israel; and National University of Science and Technology "MISIS", 4 Leninsky Prospect, 119049 Moscow, Russia
Francesca Pennecchi: Istituto Nazionale di Ricerca Metrologica, 91 Strada delle Cacce, 10135 Turin, Italy

1 Introduction

The foundations of the study of human error as a kind of human performance were developed in the 1950s and 1960s. Both correct performance and human error follow from the same cognitive processes allowing us to be fast, to respond flexibly to new situations, and to juggle several tasks at once. They are "two sides of the same theoretical coin" [1–3].

There is an extensive literature on investigating factors leading to human error (human factors) in aviation, engineering, medicine, accident analysis, forensic science and criminal investigations, and other fields [4]. An understanding of human error and the necessity of including this topic in programs for teaching students is now also recognized in analytical chemistry [5–8]. Two international workshops were held recently to discuss the human error problems in chemical analysis [9, 10].

Human errors in a routine analytical laboratory may lead to atypical test results of questionable reliability. An important group of atypical results is out-of-specification test results that fall outside the established specifications in the pharmaceutical industry, or do not comply with regulatory, legislative, or specification limits in other industries and fields, e.g. environmental and food analysis [11, 12]. Such results may also not meet the established/agreed requirements for a non-regulated product under chemical analysis. Where no limits have yet been established (e.g. for a new material), human errors may lead to incorrect evaluation of the tested property. When an atypical test result is identified, it is important to determine the root causes of the event and to avoid its recurrence. About 80 % of the root causes may be human errors [13].

Preventing, avoiding, or blocking errors by a laboratory quality system is not easy, since *errare humanum est* (to err is human). Thus, a study of human factors is necessary in any field of analytical chemistry. It is required by the U.S. Food and Drug Administration, the UK Medicines and Healthcare Products Regulatory Agency, and by other regulators, as a part of quality risk assessment [14–16]. Laboratories demonstrating competence in analytical chemistry and conformity assessment according to ISO 17025 [17] should also be able to develop relevant human-error related corrective and preventive actions. Such a study includes human error classification, modeling, and quantification.

Currently there is no available data bank (database) containing empirical values of probabilities/frequencies of occurrence of human errors in analytical chemistry derived from relevant operating experience, experimental research, and simulation studies. Therefore, known mathematical techniques for human error prediction, i.e. calculation of their probabilities, cannot yet be applied for chemical analysis of a specific material or another object. On the other hand, experts in chemical analysis have necessary information accumulated during their work. Expert judgments are used in landscape ecology and biosecurity, counterterrorism, and many other fields [18]. In analytical chemistry and metrology, expert judgments are based on knowledge of the nature of the analyte and measurand, the chemical analytical procedure (measurement method) used, earlier observations, and common sense. Therefore, the judgments are not arbitrary [19] and can be helpful for the error quantification.

The classification, modeling, and quantification of human errors in a routine chemical analytical laboratory using expert judgments is detailed in the present Guide. Residual risk of human errors (not prevented or blocked by the laboratory quality system) and consequences of this risk for quality and measurement uncertainty of the analytical results are discussed.

1.1 Scope and field of application

This Guide is developed for the study of human errors in chemical analysis and the reduction of the risk of such errors.

The document is intended for quality control, measurement, and testing (chemical analytical) laboratories, for accreditation bodies, laboratory customers, regulators, quality managers, metrologists, and analytical chemists – analysts.

1.2 Terms and definitions

Terms and definitions used in this Guide correspond to JCGM 200 (VIM) [20], ISO 3534 [21], ISO 9000 [22] and ISO Guide 73 [23]. The most relevant definitions relating to human errors in chemical analysis are given here.

1.2.1

human error in chemical analysis
human error
error
any action or lack thereof that leads to exceeding the tolerances of the conditions required for the normative work of the chemical analytical (measuring/testing) system with which the human interacts

NOTE 1 Adapted from Ref. [24, 25].
NOTE 2 When the measuring system is dealing with sampling, “the human” may be a sampling inspector. On other steps of chemical analysis “the human” is an analyst/operator of the measuring system.
NOTE 3 “The tolerances of the conditions” are, for example, intervals of temperature and pressure values for sample decomposition, purity of reagents, pH values for an analyte extraction and separation, etc.

They are formulated in a standard operation procedure (SOP) of the chemical analysis describing the normative work, based on the results of the chemical analytical method validation study.

NOTE 4 This human error definition relates to "active errors" according to ISO/TS 22367 [15], clause 3.2.

NOTE 5 Human error should not be confused with measurement error, defined in VIM [20], clause 2.16, as a difference between measured and reference quantity values. Human error may cause measurement error. However, when a human error is identified in time and the measurement process is corrected or repeated, there is no influence on the measurement error. Therefore, the use of term "error" is possible when there is no ambiguity.

1.2.2

quality risk of human error
risk of human error
risk
the combination of likelihood (probability) of occurrence of human error (1.2.1) and the severity of that error for quality of chemical analytical results

NOTE 1 Adapted from ISO Guide 73 [23], clause 1.1, note 4, and ICH Q9 [14], clause 7.

NOTE 2 There is not only a quality risk of human error: financial, safety, and other risks may be discussed also. Therefore, the use of terms "risk of human error" and "risk" is possible when there is no ambiguity.

1.2.3

quality risk management
risk management
coordinated activities to direct and control a laboratory with regard to the risk (1.2.2)

NOTE 1 Adapted from ISO Guide 73 [23], clause 2.1.

NOTE 2 Such activities are components of the laboratory quality (management) system according to ISO 9000 [22], clause 3.2.3.

1.2.4

quality risk reduction
reduction of the risk
risk reduction
change of the risk (1.2.2) due to the laboratory quality system reducing the likelihood of the occurrence of human error (1.2.1) and/or the severity of that error for quality of chemical analytical results

NOTE Adapted from ICH Q9 [14], clause 7.

1.2.5

quality residual risk
residual risk
risk (1.2.2) remaining after its reduction (1.2.4) by the laboratory quality system

NOTE Adapted from ISO Guide 73 [23], clause 3.8.1.6.

1.3 Symbols

$\Delta_{jj'}^{(i)}$	synergy of two quality system components j and j' ($j' \neq j$) in blocking human error by scenario i
E^*	score of the effectiveness of the quality system in decreasing likelihood and severity of human errors
f_{HE}	fraction of the quality of the measurement/test results which may be lost due to residual risk of human errors
i	index of a human error scenario
I	number of human error scenarios possible in a chemical analytical process
j	index of a component (layer) of a laboratory quality system
J	number of components of a laboratory quality system
k	index of a kind of human error
K	number of kinds of human errors possible in a chemical analytical process
l_i	expert judgment on severity (loss of quality of test results) of human error scenario i
L^*	severity score
p_i	expert judgment on likelihood of human error scenario i
P^*	likelihood score
q_j	effectiveness of quality system component j in decreasing likelihood and severity of human errors
q_j^*	score of the effectiveness of component j
$\tilde{q}_m^*$	score of the quality system effectiveness at step m of a chemical analytical process
Q	quality
Q_{res}	resulting quality
m	index of a step of a chemical analytical process
M	number of steps of a chemical analytical process
n_{MC}	number of Monte Carlo trials
r_{ij}	reduction of likelihood and severity of error scenario i as a result of error blocking by quality system layer j, degree of their interaction
$\tilde{r}_{ij}$	reduction value taking into account the synergy factor
r^*	score of the risk reduction
R^*	score of the residual risk
s_{ij}	synergy factor of quality system component j with the other system components in blocking error scenario i
t	time
u	standard measurement uncertainty [20]
u_{HE}	standard uncertainty caused by residual risk of human error
$u_{\mathrm{HE\text{-}r}}$	relative standard uncertainty caused by residual risk of human error
u_{r}	relative standard measurement uncertainty [20]
u_{res}	resulting standard measurement uncertainty, including the human error contribution
$u_{\mathrm{res\text{-}r}}$	resulting relative standard uncertainty, including the human error contribution

1.4 Abbreviations

CAS number	unique numerical identifier assigned by Chemical Abstract Service (CAS) to every chemical substance
CITAC	Cooperation on International Traceability in Analytical Chemistry
CODEX	Alimentarius Commission – international organization developing food standards, guidelines and related documents, named Food Book (Codex *Alimentarius* in Latin)
CRM	certified reference material
FPD	flame photometric detector
GC	gas chromatography

ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICP-MS	inductively coupled plasma mass spectrometry
ISO/TS	International Organization for Standardization Technical Specification
JCGM	Joint Committee for Guides in Metrology
Lab	laboratory
LC	liquid chromatography
MRL	maximum residue limit
MS	mass spectrometer
MW	molar mass
NIST	National Institute of Standards and Technology, US
pmf	probability mass function
QuEChERS	Quick, Easy, Cheap, Effective, Rugged and Safe method for sample preparation in pesticide residue analysis
RSD	relative standard deviation
SD	standard deviation
SOP	standard operation procedure
VIM	International vocabulary of metrology – Basic and general concepts and associated terms [20]
XSD	halogen selective detector

2 Classification of human errors

2.1 Commission errors

Errors of commission (mistakes and violations) are inappropriate actions resulting in something other than what was intended [5]. An example is the choice of a chemical analytical method for a reference material homogeneity study having a reproducibility standard deviation larger than the standard deviation of the method for which the reference material is intended.

2.1.1 Mistakes

Mistakes occur when actions follow a plan, but the plan is wrong, as an analyst does not have appropriate or sufficient information for correct planning. They are also possible when an analyst does not completely understand the chemical analytical method and quality rules he/she works within, or applies the information incorrectly because of a lack of experience or knowledge. The classification of mistakes is based on the human behavior. The following three classes are widely accepted.

A *skill-based* mistake is an inadequate analyst performance of SOP, occurring from the overconfidence of the mentality, "I have done this a thousand times" [26].

A *rule-based* mistake happens when an analyst encounters some relatively familiar problem, but applies an unsuitable solution or rule. For example, it is an everyday mistake for an analyst operating an analytical instrument (spectrophotometer, chromatograph or another) to use wrong method conditions or part of the method conditions downloaded improperly from the instrument software.

A *knowledge-based* mistake occurs when an analyst faces a situation where his/her knowledge is not sufficient for making the right decision [27].

2.1.2 Violations

Deliberate mistakes, e.g. deviations from SOP with the purpose 1) to shorten the chemical analytical process, and 2) to improve it, where the possible harm is ignored, are the SOP *routine* and *reasoned violations*, respectively.

A *reckless* violation may be the result of a state of mind in which an analyst acts without caring about the consequences.

A *malicious* violation, including sabotage, is also possible as a result of a conflict between an analyst and the laboratory manager [28, 29].

2.2 Omission errors

Errors of omission (lapses and slips) are inactions contributing to a deviation from the intended path or outcome [5]. For example, such an error occurs when a column from a previous chemical analysis is forgotten in the chromatograph and not replaced by the column required in SOP.

2.2.1 Lapses

A lapse is an occurrence in which an analyst fails to act as required for a brief time, because of the absence of attention. Lapses are associated with the analyst's memory (lapses of memory, "senior moments" [30], etc.) and are generally not observable. For example: chromatographic vials for samples labeled as required, but filled inexplicably in an order contradictory to the labels.

2.2.2 Slips

Slips are associated with the execution phase of cognition. They are observable actions that are not in accordance with a plan. For example, an analyst is interrupted by a colleague while preparing a calibration solution in a volumetric flask, forgets that 1 mL of the certified reference material required by SOP has already been added into the flask, and adds another 1 mL.

2.3 Mapping human errors

Mapping human errors, as potential hazards which may occur at every step of chemical analysis, is necessary for quality risk management. Such a map is shown in Fig. 1. Kinds of commission errors – mistakes and violations – are listed and marked by brackets in the left part of Fig. 1. Omission errors are in the right part of the figure. Pointers show links between human errors and steps of the chemical analysis. There are, for example, the following main steps: 1) choice of the chemical analytical method and corresponding SOP, 2) sampling, 3) analysis of a test portion, and 4) calculation of test results and reporting. However, after sampling in many chemical analytical methods, sample preparation is required, including freezing, milling, and/or decomposition. The chemical analysis may start from an analyte extraction from a test portion, and separation of the analyte from other components of the extract. The analyte identification and confirmation are important in some cases. Then only calibration of the measuring system and quantification of the analyte concentration are relevant. On the other hand, choosing an analytical method and SOP may not occur in a laboratory where only one method and corresponding SOP are applied for a specific task. Many chemical analytical laboratories are not responsible for sampling, etc.

In general, the kind of human error $k = 1, 2, \ldots, K$, and the step of the chemical analytical measurement/testing process $m = 1, 2, \ldots, M$, in which the error may happen (location of the error), form the event scenario $i = 1, 2, \ldots, I$. Different scenarios at the same location are shown in Fig. 1 as a net. There are a maximum $I = K \times M$ scenarios of human errors on the map. As $K = 9$ in this Guide, $I = 9M$.

Examples of mapping human errors in pH measurement of groundwater, multi-residue pesticide analysis of fruits and vegetables, and ICP-MS analysis of geological samples are provided in Annex A. A part of the error scenarios adduced in this annex may seem trivial for professionals in pH-metry, food analysis,

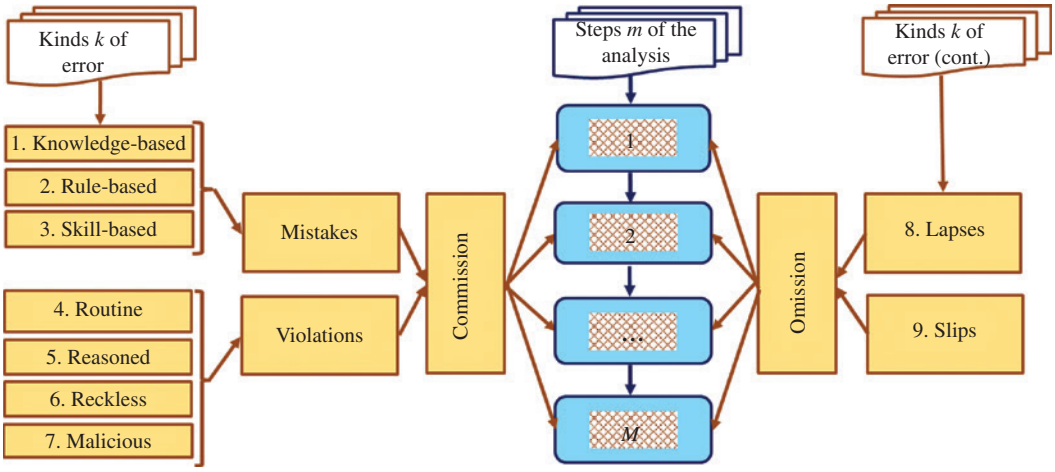

Fig. 1: A map of human errors in chemical analytical process. Kinds of mistakes and violations are marked by brackets. Pointers show links of errors to steps of the analytical process. Nets indicate error scenarios. Adapted from Ref. [4].

and/or chemical analysis in geology. However, the fact is that people make trivial errors in their day-to-day life.

3 Modeling human errors

3.1 Approach

Historically, the first approach to human error modeling was the *person* approach. It was focused on unsafe acts of an analyst (a "bad apple" in the laboratory staff) arising from his/her forgetfulness, inattention, poor motivation, carelessness, negligence, and recklessness. Reducing unwanted variability in the analyst's behavior includes disciplinary measures; naming, blaming and shaming; more rules; and more automation. The bad apple model is simple and easy to implement. However, it leads to conflicts in a laboratory and, finally, is not effective [3, 31]. The person approach is not used in this Guide.

The assumption of the later *system* approach is that analysts do not come to their laboratory to make errors. As no one can change human nature, error countermeasures in this approach are based on the laboratory quality system defense. The model, named "Swiss cheese", is widely applied in the system approach [1, 32] and is implemented in this Guide.

3.2 Swiss cheese model

A quality system in a chemical analytical laboratory should prevent, block, or impede errors of analysts. The system includes the following main defensive components/layers: 1) validation of the chemical analytical method and SOP formulation; 2) training analysts and proficiency testing; 3) quality control with monitoring Shewhart charts of results of reference material analysis, internal and external inspections, etc.; and 4) supervision. Other components are also possible. In general, there are $j = 1, 2, \ldots, J$ such quality system components/layers as shown in Fig. 2.

None of these layers alone can prevent all human errors. Each system layer is imagined in this model as a slice of Swiss cheese in Fig. 2. The holes in the layers, like in the cheese slices, are the layer's weak points, not blocking human errors. Unlike in the cheese, these holes can be opened, shut and shifted depending on analyte, matrix, analyst, and other conditions. The presence of holes in a layer will not lead to system failure,

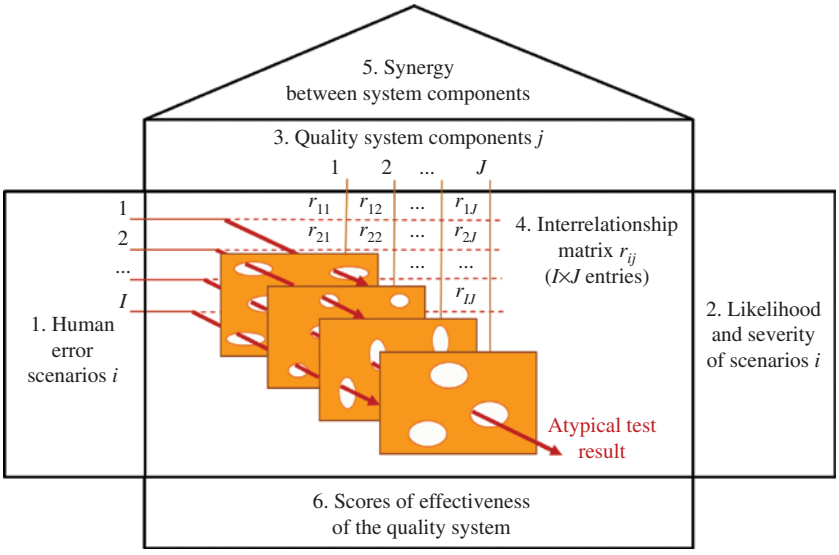

Fig. 2: A laboratory quality system against human errors in the house of security. Human error scenarios are indicated by pointers $i = 1, 2, \ldots, I$. Quality system components/layers are shown as the Swiss cheese slices $j = 1, 2, \ldots, J$. Estimates of reduction r_{ij} of likelihood and severity of error scenario i, as the result of interaction between the error and layer j, form here the interrelationship matrix. Adapted from Ref. [33].

as a rule, since other layers are able to prevent a bad outcome. That is shown in Fig. 2 as the pointers blocked by the layers. In order for an incident to occur and an atypical test result to appear, the holes in the layers must line up at the same time to permit a trajectory of incident opportunity to pass the system (through its defect), as depicted in Fig. 2 by the longest pointer.

Examples of modeling human errors are available in Annex A.

4 Quantification of human errors

A technique for the quantification of human errors in chemical analysis using expert judgments was formulated based on the Swiss cheese model and the house-of-security approach [33]. This approach, developed originally in the field of security and safety, combined the house-of-quality matrix transformation of customer requirements into quality design by generic quality function deployment, and some ideas of failure mode and effect analysis [34].

4.1 Swiss cheese in house-of-security

The house-of-security with the Swiss cheese for quantification of human errors in chemical analysis consists of the following elements illustrated in Fig. 2:

1) list of human error scenarios $i = 1, 2, \ldots, I$ – at the left wall of the house;
2) expert judgments on likelihood p_i of scenarios $i = 1, 2, \ldots, I$ and their severity (loss of quality of chemical analytical results) l_i – at the right wall of the house;
3) list of quality system components $j = 1, 2, \ldots, J$ considered as protective layers of the system – at the ceiling of the house;
4) interrelationship matrix of expert judgments about reduction r_{ij} of likelihood and severity of error scenario i as the result of interaction between the error and protective layer j – the main content of the house;
5) synergy $\Delta_{jj'}^{(i)}$ of two quality system components j and j' ($j' \neq j$) in blocking human error by scenario i – at the house roof;

6) scores of effectiveness q_j^* of quality system component j, scores of effectiveness $\tilde{q}_m^*$ of quality system at step m of the chemical analysis, and scores of effectiveness E^* of the quality system in whole in decreasing likelihood and severity of human errors – at the house floor.

4.2 Elicitation scale

The elicitation process of expert judgments [35, 36] on the likelihood of human error scenarios and other topics requires a scale able to transform the semi-intuitive personal knowledge and experience of the expert into discrete quantities. For example, the geometrical scale with three values (1, 3, 9) was used originally in the house-of-security approach [34] in order to emphasize the dramatic character of expert judgments for the security system of an organization. However, this scale may not be informative enough. On the other hand, when the scale from 1 to 9 is more detailed, the choice on the scale may be hampered. For example, at the arithmetic scale of nine values (1, 2, …, 9), the probability that an expert will choose the first value of the scale is about 30 %, while the probability of the choice of the last scale value is <5 %, according to Bernford's law, applied in forensic fraud detection [37].

As such a scale is intended in the house-of-security approach for the separation of important events from the less important, the scale division should be clearly visible, i.e. the number of scale values should be limited. Therefore, the mentioned geometrical scale (1, 3, 9) was extended by 0 only for an unfeasible scenario, and the known scale of four values (0, 1, 3, 9) [38] was considered to be the optimal solution for a chemical analytical task.

4.3 Likelihood and severity

An expert in the chemical analytical method may estimate the likelihood p_i of error scenario i by the scale (0, 1, 3, 9): likelihood of an unfeasible scenario as $p_i = 0$, weak likelihood as $p_i = 1$, medium as $p_i = 3$, and strong (maximal) likelihood as $p_i = 9$.

For the characterization of human errors in an analytical method as a whole, the likelihood score P^* (expressed in %), equal to the averaged and normalized likelihood value, is used:

$$P^* = (100\,\%/9)\sum_{i=1}^{I} p_i / I. \tag{1}$$

The P^* value can be interpreted as a kind of "intuitive" or "subjective" (mean) error probability [19] of human error by any scenario in measurement or testing by the chemical analytical method. For example, $P^* = 10\,\%$ means that human error may happen (on average) in one of 10 measurements by this method.

Severity of scenario i is estimated as expected loss l_i of quality of the test results, when the error by this scenario is not blocked. The same expert may estimate severity using the same scale: no severity as $l_i = 0$, light severity as $l_i = 1$, medium as $l_i = 3$, and high (maximal) severity as $l_i = 9$.

For the characterization of human error severity in the chemical analytical method overall, the following score L^* (%), equal to the averaged and normalized severity value is used:

$$L^* = (100\,\%/9)\sum_{i=1}^{I} l_i / I. \tag{2}$$

For example, $L^* = 50\,\%$ can be interpreted as the severity occurring when half of measurement results burdened with human errors cannot be corrected and the measurements should be repeated.

4.4 Interrelationship matrix

To characterize the quality system in detail, one should estimate the possible reduction r_{ij} of likelihood and severity of scenario i as a result of the error blocking by quality system layer j (degree of their interaction).

Such estimation is again the task of the expert in the chemical analytical method. The judgments about the interaction between an error by scenario i and layer j can be formulated using four ensuing degrees: no interaction as $r_{ij} = 0$, low as $r_{ij} = 1$, medium as $r_{ij} = 3$, and high (maximal) interaction as $r_{ij} = 9$. The r_{ij} values form the interrelationship matrix with $i = 1, 2, \ldots, I$ rows, and $j = 1, 2, \ldots, J$ columns. An empty row i of the matrix ($r_{ij} = 0$ for any j) means that the quality system is unable to prevent scenario i; an empty column j ($r_{ij} = 0$ for any i) indicates the uselessness of layer j.

Three dimensions are used in Fig. 2 to show how an error scenario may fall into a weak point of a quality system component (a hole of a cheese slice): two dimensions for the slices and another dimension for the error scenario pointers, perpendicular to the slices. However, the error scenario numbers have only one dimension i, and the slice numbers have also one dimension j. Therefore, the interrelationship matrix of r_{ij} values is two-dimensional. The total number of the matrix entries is $I \times J$.

4.5 Synergy of the quality system components

Blocking human error according to scenario i by a quality system component j can be more effective in the presence of another component j' ($j' \neq j$) because of their synergy $\Delta^{(i)}_{jj'}$. For example, such component j is the training of analysts for the correct performance of a measurement and for avoiding the routine method violation in scenario i (with the purpose of shortening the measurement process and decreasing the necessary time). The training is more effective when the analytical method is validated and the SOP is already formulated, i.e. in the presence of the component j'. In this case the synergy is $\Delta^{(i)}_{jj'} = +1$. When the synergy is absent, $\Delta^{(i)}_{jj'} = 0$. Theoretically, the synergy can also be negative and $\Delta^{(i)}_{jj'} = -1$ is possible. However, quality systems in chemical analytical laboratories do not contain, as a rule, antagonistic components.

The synergy of component j with the rest of the components of the quality system in blocking human error scenario i in whole can be characterized by the following factor:

$$s_{ij} = 1 + \sum_{j' \neq j}^{J} \Delta^{(i)}_{jj'} / (J - 1). \tag{3}$$

The synergy factor value is $1 \leq s_{ij} \leq 2$ when antagonistic components are absent and $J \geq 2$.

Taking into account the synergy factor, the interrelationship matrix is to be transformed, replacing r_{ij} by $\tilde{r}_{ij} = r_{ij}\, s_{ij}$ in every cell ij of the matrix.

4.6 Effectiveness

Effectiveness of quality system component j in decreasing likelihood and severity of human errors is evaluated as

$$q_j = \sum_{i=1}^{I} p_i l_i r_{ij} s_{ij}. \tag{4}$$

A score q^*_j of effectiveness of component j relative to other components of the quality system, i.e. priority or importance of component j, expressed in %, is:

$$q^*_j = (100\,\%) q_j / \sum_{j=1}^{J} q_j. \tag{5}$$

Calculation of the score values q^*_j allows the evaluation of the quality system components for all steps of the chemical analysis together. However, an analyst may be interested to know which step m is less protected from errors, with the intent to improve it. A score $\tilde{q}^*_m$ (%) of the quality system effectiveness at step m of the chemical analysis was developed for this purpose [39]:

$$\tilde{q}^*_m = (100\,\%) \tilde{q}_m / \sum_{m=1}^{M} \tilde{q}_m, \quad \text{where} \quad \tilde{q}_m = \sum_{i'=m}^{m+M(K-1)} \sum_{j=1}^{J} p_{i'} l_{i'} r_{i'j} s_{i'j}, \tag{6}$$

and $i' = m + M(k-1)$ are the scenario numbers related to the same error location (step m) for all kinds of error $k = 1, 2, \ldots, K$. For example, for $m = 1$, $M = 6$ and $K = 9$, there are $i' = 1, 7, \ldots, 49$.

A score E^* of effectiveness of the quality system in whole can be calculated as relative to an ideal (virtual/imagined) quality system tending to zero defects. This system has the maximal degree of interaction, $r_{ij} = 9$, of every human error by scenario i and every system component/layer or slice of the Swiss cheese, j. Thus, the effectiveness score, expressed in %, is:

$$E^* = (100\,\%/9)\sum_{j=1}^{J} q_j / \sum_{j=1}^{J}\sum_{i=1}^{I} p_i l_i s_{ij}. \tag{7}$$

Examples of the quantification are available in Annex A.

5 Risk evaluation of human errors

5.1 Risk reduction

As the risk of human error is a combination of the likelihood and severity of that error, their reduction $\tilde{r}_{ij}$ is the risk reduction. The $\tilde{r}_{ij}$ value can be normalized by dividing its multipliers r_{ij} and s_{ij} by their maximal values, 9 and 2, respectively. Averaging the normalized risk reduction values for the interrelationship matrix over all the error scenarios and quality system components leads to score r^* characterizing the (mean) risk reduction by the laboratory quality system, expressed in %:

$$r^* = (100\,\%/18IJ)\sum_{j=1}^{J}\sum_{i=1}^{I}\tilde{r}_{ij}. \tag{8}$$

5.2 Residual risk and its consequences

A score of residual risk of human errors R^* (%), which are not prevented/blocked or reduced/mitigated by the quality system, is:

$$R^* = 100\,\% - r^*. \tag{9}$$

The fraction of the quality f_{HE} of the analytical results which may be lost due to residual risk of human errors, expressed in %, is:

$$f_{\mathrm{HE}} = (P^*/100\,\%)(L^*/100\,\%)R^*. \tag{10}$$

When an ideal laboratory quality system is able to prevent or block human errors completely, one has $R^* = 0\,\%$ and $f_{\mathrm{HE}} = 0\,\%$: there is no loss of quality, the quality system effectiveness score is $E^* = 100\,\%$. If a quality system is not effective at all or absent, $E^* = 0\,\%$, as $\tilde{r}_{ij}=0$ for all i and j. Then $R^* = 100\,\%$ and $f_{\mathrm{HE}} = (P^*/100\,\%)L^*$. The extreme case of a complete loss of quality is theoretically possible when, in absence of a quality system ($R^* = 100\,\%$), scores P^* and L^* reach also 100 %, i.e. human errors are inevitable and destructive. Thus, $f_{\mathrm{HE}} = 100\,\%$ as well.

In practice, $0\,\% < f_{\mathrm{HE}} < 100\,\%$, and the residual risk of human errors can be interpreted as a source of measurement uncertainty when a human being is involved in the measurement process and this human interaction with the measuring system is taken into account. Such interpretation is discussed in Annex B. Examples of the calculation of the risks, their consequences for quality of analytical results, and corresponding contributions to uncertainty budget are available in Annex A.

6 Limitations

6.1 Variability

Any expert is also a human being and the elicitation process (by which the expert is prompted to provide error likelihood, severity, and other estimates) is influenced by aleatory and epistemic uncertainty [40], intrapersonal conflicts [41], etc. Therefore, the evaluation of variability of the error quantification scores and relative risks due to the inherent expert's hesitancy is important also.

The expert may feel a natural doubt choosing one of close values from the proposed scale: 0 or 1? 1 or 3? 3 or 9? One change of an expert judgment, for example, on the likelihood of scenario i from $p_i = 0$ to $p_i = 1$ and vice versa ($p_i = 0 \leftrightarrow 1$) leads to a 0.2 % change (shift) of the likelihood score P^* according to formula (1) at $I = 54$ scenarios, as in Annex A, Example 2. The change of the expert judgment $p_i = 1 \leftrightarrow 3$ results in a 0.4 % shift of the P^* value. The maximum correction of the likelihood score P^* for $I = 54$ scenarios caused by one change $p_i = 3 \leftrightarrow 9$ is 1.2 %.

The influence of a judgment change on the score value increases with decreasing I [39]. Thus, for $I = 36$ scenarios, as in Annex A, Examples 1 and 3, changing $p_i = 3 \leftrightarrow 9$ leads to a 1.9 % shift in the P^* value. The same is true for severity score L^*. However, the evaluation of the variability of other scores, depending simultaneously on more than one expert judgment for the same scenario i, is more complicated.

A detailed analysis of the score variability, as well as the variability of the corresponding loss of quality f_{HE}, based on Monte Carlo simulations, is presented in Annex C.

6.2 Specificity

The score values and the residual risk of human errors are specific for the chemical analytical measurement/test method and for the laboratory conditions, i.e. they may be different in different laboratories active in the same field and using the same method. On the other hand, it is impossible to expect an equal risk of human errors in measurement or testing using different methods, even in the same laboratory.

Changes in the laboratory environment, as well as in any quality system component or staff, require a re-evaluation of the quality of the analytical results, which may be lost due to residual risk of human errors f_{HE}.

The re-evaluation may indicate either an f_{HE} increase (e.g. due to the retirement of an experienced supervisor) or its decrease (e.g. due to the acquisition of a new, more accurate and more automated measuring system) [42].

6.3 Latent errors

Errors due to poor laboratory design, a defect in the equipment, or an unsuccessful management decision, not depending on the sampling inspector and/or the analyst/operator, are defined in ISO/TS 22367 [15], clause 3.5, as "latent errors". Latent errors are not considered in this Guide.

6.4 Positive human factors

In non-routine chemical analysis (scientific research, development of new chemical analytical methods and instruments, etc.) human errors are also possible. At the same time, the most successful way of solving problems arising in such analysis is human as well. The knowledge and experience of analytical chemists, their activity, creativity, and other abilities, are the positive human factors that can help to overcome the problems [43].

Neither human errors in non-routine chemical analysis, nor positive human factors are discussed in this Guide.

7 Implementation remarks

Classification, modeling, and quantification of human errors in a routine laboratory show the ways to increase the quality system effectiveness and subsequently reduce the risk of these errors in the laboratory.

In particular, the results of the human error study would be useful for validating the analytical method and formulation of the SOP, as well as for training (how to avoid the errors), and for supervision. The map of possible human error scenarios, included in the validation report, may be useful also as a checklist for the prior assessment of an analyst before assigning the task, etc.

Annex A. Examples

CONTENTS

Example 1. Human errors in pH measurements of groundwater

A-1-1 Introduction and main steps of the pH measurements

There is a wide spectrum of methods available using different measuring equipment, from pH indicator strips and routine pH meters with thermo-compensation to multifunctional instruments (allowing measurements of pH together with some other water properties without probe changes) and pH opto-sensing flow injection analysis. In any pH measurement, human error may influence the quality of the measurement results. However, first of all, this is important for such an object as groundwater [33]. The problem is that pH of groundwater is influenced by the partial pressure of dissolved carbon dioxide, which is much larger than the atmospheric one. CO_2 degassing continues during the measurement process due to the water stirring. Therefore, time is necessary (about 10 min in a case study [44]) for obtaining the stable response of a measuring instrument when the drift of the response does not exceed 0.02 pH unit in 1 min.

Thus, the step $m = 1$ in pH-metry is the choice of the method with corresponding equipment and SOP. The step $m = 2$ is the sampling of groundwater. Then there is the proper pH measurement, step $m = 3$. The last step, $m = M = 4$, is the calculation of the test results and reporting.

A-1-2 The map of human error scenarios

A total of $I = 9M = 36$ scenarios of human error in pH measurement of groundwater are discussed below, based on the optimistic assumption that an analyst knows what the quantity pH is.

A-1-2-1 Knowledge-based mistakes, $k = 1$

Scenario $i = 1$ in choice of the method, $m = 1$. Knowledge-based mistakes may happen when an analyst does not have enough knowledge about the methods and instruments and cannot choose them correctly.

Scenario $i = 2$ in sampling, $m = 2$. This kind of error may occur when an analyst (in the role of sampling inspector) does not have the necessary knowledge regarding the criticality of using a suitably cleaned container for sampling and does not take into account that this container may be contaminated by acidic or basic substances.

Scenario $i = 3$ in measurement, $m = 3$. A knowledge-based mistake in the pH measurement happens when an analyst does not have relevant knowledge about the electrode calibration. Such an analyst may not know that the calibration results are not universal concerning the measurement range and other conditions. Thus, the calibration can be performed in a different range than required, at a different temperature, etc.

Scenario $i = 4$ in calculation and reporting, $m = 4$. Such a mistake in the last step of the measurement process may happen as a result of a lack of knowledge of statistics (e.g. averaging of regular and outlier results, incorrect rounding of significant figures, etc.).

A-1-2-2 Rule-based mistakes, $k = 2$

Scenario $i = 5$ in choice of the method, $m = 1$. It may be the simplest, very likely choice of the pH measurement method and SOP which has "always" been used in the laboratory.

Scenario $i = 6$ in sampling, $m = 2$. A rule-based mistake of an analyst, who interferes with sampling water for elemental analysis, is the addition of nitric acid to a sample for pH measurement.

Scenario $i = 7$ in pH measurement, $m = 3$. This kind of mistake may occur when an electrode, used earlier for pH measurement in another media, is used for current measurements of pH in groundwater without necessary preparation (cleaning, etc.).

Scenario $i = 8$ in calculation and reporting, $m = 4$. A rule-based mistake may happen when the data from yesterday's measurements, remaining in the worksheet, are reported instead of the current measurement results which have not yet been recorded in the file.

A-1-2-3 Skill-based mistakes, $k = 3$

Scenario $i = 9$ in choice of the method, $m = 1$. A possible consequence of overconfidence is the erroneous choice of a method and SOP developed for another subject, e.g. surface water. The chosen method and SOP may set inadequate time for stirring the test portion and pH measurement. In this way a criterion for decision, the time when the measurement process can be stopped (e.g. when a difference between replicate results does not exceed 0.02 pH units), may not be taken into account or be set incorrectly.

Scenario $i = 10$ in sampling, $m = 2$. A skill-based mistake is possible when a previous experience with sampling surface water, for example, is not applicable. In particular, sample temperature is, practically, not important for surface water sampling for pH, whereas it influences the CO_2 concentration in a groundwater sample.

Scenario $i = 11$ in pH measurement, $m = 3$. This kind of mistake may occur when an analyst measures pH after 1–2 min according to his previous experience, whereas about 10 min are necessary for groundwater, in order that the difference between replicate results does not exceed 0.02 pH units.

Scenario $i = 12$ in calculation and reporting, $m = 4$. A skill-based mistake may lead to reporting a measurement result when the drift of the measuring instrument used has not been checked, as it is not usually present.

A-1-2-4 Routine violations, $k = 4$

Scenario $i = 13$ in choice of the method, $m = 1$. A routine violation may happen when an analyst wishes to shorten the pH measurement process, in spite of the understandable risk of erroneous results by the chosen method.

Scenario $i = 14$ in sampling, $m = 2$. The violation is possible when an analyst does not rinse the container by the sampled water in order to shorten the sampling time.

Scenario $i = 15$ in pH measurement, $m = 3$. A similar error may occur when an analyst shortens the measurement process and stops it, in spite of the response drift.

Another possibility is when an analyst decreases the number of replicates, in spite of the SOP requirement.

Scenario $i = 16$ in calculation and reporting, $m = 4$. A routine violation may occur when not all data are taken into account, especially outliers. In such a case the measurement time is shortened, since an outlier investigation is neglected.

A-1-2-5 Reasoned violations, $k = 5$

Scenario $i = 17$ in choice of the method, $m = 1$. A reasoned violation with the purpose of improving the pH measurement results may lead to an incorrect choice of the measurement method and SOP.

Scenario $i = 18$ in sampling, $m = 2$. An error may occur when an analyst changes the conditions of the container filling required by SOP (influencing the sample degassing) with the purpose of improving the sampling process.

Scenario $i = 19$ in pH measurement, $m = 3$. A reasoned violation may happen when an analyst increases stirring of a test portion required by SOP with the purpose to improve the instrument response. This increasing may provoke turbulence of the test portion.

Scenario $i = 20$ in calculation and reporting, $m = 4$. The same kind of violation is neglecting outliers with the purpose to improve reporting.

A-1-2-6 Reckless violations, $k = 6$

Scenario $i = 21$ in choice of the method, $m = 1$. An error related to reckless violation may lead to disregarding the specificity of the analyzed object (CO_2 presence in groundwater) and to an incorrect choice of the method and SOP.

Scenario $i = 22$ in sampling, $m = 2$. A reckless violation may lead to sampling a lower volume than necessary for cleaning the electrode and other equipment and for replicate measurements according to SOP.

Scenario $i = 23$ in pH measurement, $m = 3$. A reckless violation is use of an electrode unprepared for the pH measurement, as required by SOP.

Scenario $i = 24$ in calculation and reporting, $m = 4$. That may be the reckless use of not relevant data or a report form.

A-1-2-7 Malicious violations, $k = 7$

Scenario $i = 25$ in choice of the method, $m = 1$. A malicious violation may occur as a deliberate choice of a method and SOP, contrary to the requirements of the laboratory manager.

Scenario $i = 26$ in sampling, $m = 2$. Sampling from another water source than the required one is a malicious violation.

Scenario $i = 27$ in pH measurement, $m = 3$. The measurement of the pH of another sample than the required one is also a violation.

Scenario $i = 28$ in calculation and reporting, $m = 4$. A malicious violation may be a report on another measurement/testing than the required one.

A-1-2-8 Lapses, $k = 8$

Scenario $i = 29$ in choice of the method, $m = 1$. A lapse on the step of choosing a method and SOP is infeasible. Therefore, this scenario is treated further with likelihood $p_{29} = 0$.

Scenario $i = 30$ in sampling, $m = 2$. The error may happen, for example, when an analyst does not use the designated container.

Scenario $i = 31$ in pH measurement, $m = 3$. A lapse may occur when an analyst forgets to check the measuring instrument with a reference solution before the pH measurement.

Scenario $i = 32$ in calculation and reporting, $m = 4$. Incorrect introduction of the data into the file may be caused by a lapse.

A-1-2-9 Slips, $k = 9$

Scenario $i = 33$ in choice of the method, $m = 1$. This scenario is infeasible as scenario $i = 29$. Thus, likelihood $p_{33} = 0$ also.

Scenario $i = 34$ in sampling, $m = 2$. A sample may be spilled by a slip.

Scenario $i = 35$ in pH measurement, $m = 3$. A slip occurs when an analyst forgets to clean the electrode after previous measurements or allows it to dry out rather than leaving it in the solution recommended by the electrode producer.

Scenario $i = 36$ in calculation and reporting, $m = 4$. A slip is possible when an analyst calculates and reports data obtained in previous (e.g. yesterday) measurements.

A-1-3 Elicited data and error quantification scores

The elicited data for the mapped 36 human error scenarios and the four components of quality system listed in clause 3.2 of this Guide, including the interrelationship matrix with $36 \times 4 = 144$ entries, are presented in Table 1.

Table 1: The elicited expert judgments on human errors in pH measurements of groundwater.

Scenario i	Likelihood p_i	Severity l_i	Degree of interaction r_{ij}				Synergy factor s_{ij}			
			Quality system layer j				Quality system layer j			
			1	2	3	4	1	2	3	4
1	3	9	0	9	0	9	1.67	1.33	1.33	1
2	3	3	1	9	3	9	1.67	1.33	1.33	1
3	3	3	9	9	9	9	1.67	1.33	1.33	1
4	1	3	9	9	9	9	1.67	1.33	1.33	1
5	9	9	0	9	0	9	1.67	1.33	1.33	1
6	3	3	1	9	3	9	1.67	1.33	1.33	1
7	3	3	9	9	9	9	1.67	1.33	1.33	1
8	1	3	9	9	9	9	1.67	1.33	1.33	1
9	9	9	0	9	0	9	1.67	1.33	1.33	1
10	3	3	1	9	3	9	1.67	1.33	1.33	1
11	3	3	9	9	9	9	1.67	1.33	1.33	1
12	1	3	9	9	9	9	1.67	1.33	1.33	1
13	1	9	0	9	0	9	1.67	1.33	1.33	1
14	3	3	1	9	3	9	1.67	1.33	1.33	1
15	9	9	9	9	9	9	1.67	1.33	1.33	1
16	1	3	3	9	9	9	1.67	1.33	1.33	1
17	1	9	0	9	0	9	1.67	1.33	1.33	1
18	1	3	1	9	3	9	1.67	1.33	1.33	1
19	3	9	3	9	3	9	1.67	1.33	1.33	1
20	1	3	3	9	9	9	1.67	1.33	1.33	1
21	1	9	0	9	0	9	1.67	1.33	1.33	1
22	1	3	1	9	3	9	1.67	1.33	1.33	1
23	3	9	3	9	3	9	1.67	1.33	1.33	1
24	1	3	3	9	9	9	1.67	1.33	1.33	1
25	1	9	0	3	0	9	1	1	1	1
26	1	9	0	3	3	9	1	1	1	1
27	1	9	0	3	3	9	1	1	1	1
28	1	9	0	3	9	9	1	1	1	1
29	0	9	0	3	3	9	1	1	1	1
30	3	9	0	3	3	9	1	1	1	1
31	3	3	0	3	3	9	1	1	1	1
32	1	3	0	3	9	9	1	1	1	1
33	0	9	0	1	3	9	1	1	1	1
34	1	9	0	1	3	9	1	1	1	1
35	3	9	0	1	3	9	1	1	1	1
36	1	3	0	1	9	9	1	1	1	1

The synergy of the validation of the measurement method and SOP formulation as component $j = 1$ of the quality system, and training of analysts as component $j' = 2$ of this system, is $\Delta_{12}^{(i)} = +1$ for any $i = 1$–24. The same is for validation and quality control, $j' = 3$: $\Delta_{13}^{(i)} = +1$ for any $i = 1$–24. A synergy in the cases of malicious violations and omission errors (lapses and slips) is unreliable, i.e. $\Delta_{jj'}^{(i)} = 0$ for any $i = 25$–36 and any $j' \neq j$, both are from 1 to 4. The supervision ($j = 4$) synergy with other quality system components is also absent here, $\Delta_{4j'}^{(i)} = 0$ for any $i = 1$–36 and any $j' = 1$–3. The synergy factor s_{ij} values calculated by formula (3) are presented in Table 1 as well.

The likelihood score calculated by formula (1) is $P^* = 26\,\%$. The human error severity in this method, evaluated by formula (2), is $L^* = 67\,\%$. Note, here the obtained values of P^* and L^* differ slightly from the 27 % and 65 %, respectively, published for the same pH measurements in Ref. [33]. A similar difference is also seen in other score values. The reason is that the unfeasible scenarios ($i = 29$ and $i = 33$) were not taken into account in Ref. [33] and the total number of scenarios was limited by $I = 34$.

The values of score q_j^* of effectiveness of the quality system components $j = 1$–4 calculated by formulas (4) and (5) are $q_1^* = 14\,\%$, $q_2^* = 37\,\%$, $q_3^* = 15\,\%$ and $q_4^* = 33\,\%$. Thus, the most effective here is training of analysts, while the least effective is validation of the measurement method and SOP formulation.

Effectiveness scores $\tilde{q}_m^*$ of the quality system at different steps $m = 1$–4 of the pH measurements, calculated by formulas (6) are $\tilde{q}_1^* = 31\,\%$, $\tilde{q}_2^* = 12\,\%$, $\tilde{q}_3^* = 50\,\%$ and $\tilde{q}_4^* = 7\,\%$. From these score values one can understand that the quality system is the most effective at the step of proper measurements and the least effective at the step of calculation and reporting of the measurement results. The score of effectiveness of the quality system in whole, calculated by formula (7), is $E^* = 59\,\%$.

A-1-4 Residual risk and measurement uncertainty

The score of residual risk of human errors by formulas (8) and (9) is $R^* = 62\,\%$. The percentage of the quality of the measurement/test results which may be lost due to residual risk of human errors is $f_{\mathrm{HE}} = 10.8\,\%$ by formula (10).

The standard measurement uncertainty reported for test item preparation in proficiency testing of pH measurement of groundwater [44] was $u = 0.10$ (pH units). The contribution to the uncertainty budget caused by residual risk of human errors by formula (14), Annex B, is $u_{\mathrm{HE}} = 0.05$. Thus, u_{HE} is not negligible and not the dominant contribution in the uncertainty budget. The resulting uncertainty value by formula (13), Annex B, is $u_{\mathrm{res}} = 0.11$.

Example 2. Human errors in multi-residue pesticide analysis of fruits and vegetables

A-2-1 Introduction and main steps of the analysis

Investigation of atypical results of pesticide residue analysis in fruits and vegetables, in particular out-of-specification results exceeding maximum residue limits (MRL), based on metrological concepts showed that only a minor part of these results can be defined as metrologically-related, i.e. caused by measurement problems. The newest applications of chromatography and mass spectrometry using detailed libraries of mass spectra of pesticides, their metabolites and derivatives, cannot exclude human errors in the analysis. Moreover, human errors are the greatest source of failures in pesticide identification and confirmation, even if performed by the most diligent and intelligent analysts. Theoretical calculations of probabilities of false positive and false negative results of pesticide identification in real samples, based on the chemical structure of analytes and measurement uncertainty, are practically unacceptable, as they do not take into account mislabeling, contamination of a sample, inappropriate spikes, etc.

There are $M = 6$ main steps $m = 1, 2, \ldots, 6$ of the analysis: 1) sampling, 2) sample processing, 3) sample preparation, further named "extraction", 4) identification and confirmation of pesticides, further "identification", 5) measurement of their amount in the extract, further "quantification", and 6) calculation of the pesticide concentrations in the analyzed sample and reporting, further "reporting".

Sampling is conducted by certified inspectors for official control according to the CODEX guidelines directly from the field, packing houses, and logistics centers before sending the product to the market. Laboratory (Lab) samples of different fruits and vegetables (1 kg usually) undergo the same general procedure at the Lab, starting from processing, i.e. homogenization/blending.

Sample preparation for gas chromatography (GC) is performed by the Mini-Luke method based on extraction of analytes with acetone from a test portion of 15 g taken from the homogenized laboratory sample. For liquid chromatography (LC) the QuEChERS method of sample preparation is used, employing extraction with acetonitrile from 10 g test portions.

The extracts are analyzed by GC system with mass spectrometer (MS). Electron ionization is applied in the MS in full scan mode. Two other GC systems, equipped with flame photometric (FPD) and halogen selective (XSD) detectors, respectively, are used to verify the results obtained by GC/MS screening, and also to detect those analytes that exhibit a better response to either XSD or FPD than can be achieved with MS in full scan mode. LC-amenable pesticides are determined by combining an ultra-performance LC system with an advanced tandem quadrupole MS system, operated in multiple reaction monitoring mode with electrospray ionization [39].

A-2-2 The map of human error scenarios

A total of $I = 9M = 54$ scenarios of human errors in multi-residue pesticide analysis of fruits and vegetables are discussed below.

A-2-2-1 Knowledge-based mistakes, $k = 1$

Scenario $i = 1$ in sampling, $m = 1$. For example, in sampling grapes the mistake is when an inspector picks grapes from an outer part of a bush, which is usually sprayed by pesticides much more than the internal part of the bush.

Scenario $i = 2$ in sample processing, $m = 2$. Grinding fresh grapes is a knowledge-based mistake, since this leads to an inhomogeneous mixture of the grape rinds and pulp, which have different concentrations of pesticide residues. Therefore, the correct processing requires freezing the sample before grinding.

Scenario $i = 3$ in extraction, $m = 3$. An attempt to extract analytes with any solvent without wetting a dry sample (e.g. of tea or rice) is a knowledge-based mistake: water wets the surface of sample particles and allows transportation of pesticide residues, their metabolites, and derivatives from the particles to the solvent phase.

Scenario $i = 4$ in identification, $m = 4$. The mistake may occur when the analyte mass spectrum is compared with an inadequate standard spectrum from a database library. For example, this happened when a mass spectrum of a product of diuron degradation was compared with the spectrum of an iprodione degradation product.

Scenario $i = 5$ in quantification, $m = 5$. Lack of relevant knowledge about a matrix effect, able to change the response of the measuring system, may lead to a mistake in quantification. For example, this effect was described in GC/FPD determination of methamidophos in tomato and thyme matrices.

Scenario $i = 6$ in reporting, $m = 6$. Reporting may be mistaken because of a deficiency of knowledge of the residue definitions. For example, according to CODEX sum of fenthion, its oxygen analog and their sulfoxides and sulfones, should be expressed as fenthion (fat-soluble). Therefore, it is a mistake when any of the

following six residues is reported individually: fenthion, fenthion sulfoxide, fenthion sulfone, fenthion oxon, fenthion oxon sulfone, and fenthion oxon sulfoxide.

A-2-2-2 Rule-based mistakes, $k = 2$

Scenario $i = 7$ in sampling, $m = 1$. The mistake is possible when an inspector, having experience in control of plant illness, picks only damaged fruits for a routine sampling.

Scenario $i = 8$ in sample processing, $m = 2$. For example, to peel a banana (as we usually do in order to eat it) before grinding is a mistake, unless banana pulp is qualified.

Scenario $i = 9$ in extraction, $m = 3$. The mistake may happen when an analyst, routinely using 10 g of tomato test portion in a 50 mL test tube for extraction, tries to follow the same procedure for herbs. However, such a test portion is too large for extracting pesticide residues from herbs in the applied volume: the extraction cannot be completed.

Scenario $i = 10$ in identification, $m = 4$. An analyte may not be included in the local list of monitored pesticides (target analytes), as it happened with isofenphos-methyl, unauthorized in Europe, but detected in Germany in sweet peppers from Spain [39]. Attempts at identification of such an analyte using the local pesticide database may lead to a rule-based mistake.

Scenario $i = 11$ in quantification, $m = 5$. When a measuring system does not have a linear calibration curve with the necessary correlation coefficient for the analyte range of concentrations, the system switches automatically to one-point calibration. If an analyst does not detect a problem and continues quantification, a mistake may occur.

Scenario $i = 12$ in reporting, $m = 6$. A mistake is possible when a mass of the test portion used for extraction is unusual (differs from 10 g for extraction with acetonitrile or 15 g for extraction with acetone) and that is not taken into account.

A-2-2-3 Skill-based mistakes, $k = 3$

Scenario $i = 13$ in sampling, $m = 1$. A mistake may appear when an inspector, usually sampling 1 kg of fruits and/or vegetables, must take a sample of cabbages or watermelons, wherein not <5 units are required.

Scenario $i = 14$ in sample processing, $m = 2$. According to SOP a Lab blender (its internal thimble) used in sample processing is cleaned with water and acetone after completion of any task. Therefore, it is not necessary to clean the blender before the next task. Straying from this rule may lead to a skill-based mistake. For example, once a blender had been rented by a neighboring laboratory for preparation of fly bait containing malathion, and was not cleaned thereafter as required. Subsequently, a significant malathion content was found in a sample of strawberry, grown using biopesticides and natural enemies of pests, analyzed in the Lab.

Scenario $i = 15$ in extraction, $m = 3$. Interaction of an analyte with natural components of the sample matrix may lead to incomplete analyte extraction. For example, when chlorothalonil is extracted from leek or garlic, the analyte binds to the sulfur derivative groups (present in the matrixes) and is trapped thereby, which may even cause a false negative result. This interference was observed especially with acetonitrile extraction by the QuEChERS method, while extraction with acetone by the Mini-Luke method overcame this skill-based mistake.

Scenario $i = 16$ in identification, $m = 4$. To decrease matrix effects in analysis of fresh fruits and vegetables one uses diluted solutions. For example, the matrix of tea depresses the methomyl analytical signal, and dilution of the extract with acetonitrile is useful. However, identification of methomyl in an excessively diluted solution may be difficult: a false negative result is likely here. Therefore, the dilution for compensation of the matrix effect should be balanced.

Scenario $i = 17$ in quantification, $m = 5$. Separation of a pesticide from another analyte may be unsuccessful, and instead of the pesticide's individual concentration, a sum of several analyte concentrations is

obtained. For example, peaks of azinphos-methyl and phosmet were not separated in a LC/MS/MS run, which was only discovered in GC/FPD analysis of the same sample.

Scenario $i = 18$ in reporting, $m = 6$. A skill-based human error is possible when the analyte is written incorrectly in the report and read as another one. Such an error is understandable, as any report provides information on a number of pesticides out of hundreds of substances permitted or forbidden for use in agriculture.

A-2-2-4 Routine violations, $k = 4$

Scenario $i = 19$ in sampling, $m = 1$. An example of a routine violation is sampling from a corner of the field (or from only one packing-box) with the intention to shorten the sampling.

Scenario $i = 20$ in sample processing, $m = 2$. When a cutter is not cleaned of the previous sample, the processing time is shortened, but contamination is possible.

Scenario $i = 21$ in extraction, $m = 3$. Evaporation of solvents should be performed at (45 ± 5) °C. To shorten the evaporation time an analyst could increase the evaporation temperature. However, content of volatile analytes in the extract, e.g. dichlorvos, is decreased.

Scenario $i = 22$ in identification, $m = 4$. An analyst tries to shorten the identification time and uses GC/MS only, whereas the co-eluted analytes may not be separated. In such a case the identification result (without additional information by GC/FPD and/or GC/XSD) will be incorrect.

Scenario $i = 23$ in quantification, $m = 5$. Economy of the quantification time can be achieved when an analyst does not perform re-calibration of the measuring instrument and uses old calibration data. However, that may lead to biased quantification results.

Scenario $i = 24$ in reporting, $m = 6$. When a report is not checked in order to save time, a twist of the data during their transformation to the final file may not be noted and not corrected.

A-2-2-5 Reasoned violations, $k = 5$

Scenario $i = 25$ in sampling, $m = 1$. An inspector may take a sample from the field treated with a pesticide earlier than prescribed after the treatment, with the purpose to obtain more accurate test results. However, the violation of the prescribed waiting time will lead to increased results of the test, not comparable with the corresponding limits.

Scenario $i = 26$ in sample processing, $m = 2$. Almonds should be cleared of their peel before processing a sample. However, an analyst may decide erroneously to use all the material for processing (including the peel) with an effort to be more accurate.

Scenario $i = 27$ in extraction, $m = 3$. The required duration of extraction of analytes from a test portion is 30 sec. An analyst may decide that a 1 min extraction is more complete. In such a case the extract contains a higher concentration of matrix components and even particles, which complicate the analysis.

Scenario $i = 28$ in identification, $m = 4$. Identification of an analyte requires matching of 50 % or more of the spectrum of the analyte and the standard spectrum. When <50 % of the spectra are matching and the concentration of the analyte is lower than the MRL, the identification can be stopped: reporting is not required. If nevertheless an analyst will report the analyte detection (without its confirmation with an orthogonal method), the identification result may be a false positive.

Scenario $i = 29$ in quantification, $m = 5$. For preparation of a calibration solution, 1 mL of each standard solution (reference material) should be introduced into a 100 mL volumetric flask, filled then with solvent to the mark. An attempt to put the flask into ultrasonic bath for mixing instead of inverting the flask is a reasoned violation, since the solution cannot be mixed as necessary in the bath.

Scenario $i = 30$ in reporting, $m = 6$. An outlier may be removed from the data without investigation with the purpose to obtain a more “accurate” test result.

A-2-2-6 Reckless violations, $k = 6$

Scenario $i = 31$ in sampling, $m = 1$. A reckless inspector may confuse the farmer name on a sample, and the analytical/test results will be related to another farmer (and another field), i.e. will be not correct.

Scenario $i = 32$ in sample processing, $m = 2$. When a reckless analyst does not shake off the soil from onions before the sample processing (homogenization), the obtained test results are biased.

Scenario $i = 33$ in extraction, $m = 3$. Addition of 1 mL of solvent to dry residue of a test portion after extraction and evaporation is required. A no exact solvent volume may lead to erroneous test results.

Scenario $i = 34$ in identification, $m = 4$. An analyst may perform identification based on matching of the spectrum of the analyte and the spectrum of its standard by masses of ions and ratio of their peaks, not taking into account the peak retention times. In this way a false positive identification is possible, e.g. of etoxyquin in parsley, similar to a component of the sample matrix (7H-Furo[3,2-g][1]benzopyran-7-one, 4-hydroxy, C11H6O4, MW 202.027, CAS No. 486-60-2, 80 % match with NIST 2002).

Scenario $i = 35$ in quantification, $m = 5$. Recklessness may cause an incomplete injection of the extract into the chromatograph, when contamination of the syringe needle is not seen.

Another example is when an analyst does not check the calibration curve, which has a belly and corresponding low correlation coefficient.

Scenario $i = 36$ in reporting, $m = 6$. Recklessness may lead to confusing names of pesticides with different MRLs, e.g. bifenazate instead of bifenthrin. Their MRLs in pepper are 0.050 $mg \cdot kg^{-1}$ and 0.200 $mg \cdot kg^{-1}$, respectively.

Another example is reporting on dimethylamine instead of diphenylamine in apples.

A-2-2-7 Malicious violations, $k = 7$

Scenario $i = 37$ in sampling, $m = 1$. A malicious violation may be reflected in filling sample labels in a confusing manner. Thus, the labels cannot be read simply and unambiguously.

Scenario $i = 38$ in sample processing, $m = 2$. A written identification number of a laboratory sample may be mistaken in order to vex the laboratory manager.

Scenario $i = 39$ in extraction, $m = 3$. An extraction may be knowingly performed for GC instead of LC and vice-versa.

Scenario $i = 40$ in identification, $m = 4$. A chromatographic file of another sample may be introduced for identification under the current sample title.

Scenario $i = 41$ in quantification, $m = 5$. In this step of the analysis the violation may consist of replacement of the calibration date instead of re-calibration.

Scenario $i = 42$ in reporting, $m = 6$. Any falsification of the data is a malicious violation.

A-2-2-8 Lapses, $k = 8$

Scenario $i = 43$ in sampling, $m = 1$. Like in scenario $i = 31$ a farmer name on a sample may be confused and test results related to another farmer (and another field) will be incorrect. However, in the current case that may happen by an inspector senior moment, not by the recklessness.

Another example is when an inspector, tired after a working day, stored plastic containers for sampling at home in the box room, treated earlier with dichlorophos against cockroaches. As a result, a sample of organic oranges put in the container the next morning was contaminated.

Scenario $i = 44$ in sample processing, $m = 2$. An example of a lapse in sample processing is confusing numbers of samples, again by a senior moment.

Scenario $i = 45$ in extraction, $m = 3$. For extraction of pesticide residues for GC with acetone, addition of dichloroethane and petroleum ether is recommended for the phase separation. When an analyst forgot that the addition was already done and repeated it again, the conditions of the extraction were changed.

Scenario $i = 46$ in identification, $m = 4$. Numbers of samples for identification may not correspond to the physical order of the samples on the table of an analyst. In such a case the sample numbers may be confused and the identification results will be inadequate.

Scenario $i = 47$ in quantification, $m = 5$. Stones of mango and/or avocado should be taken into account in the mass of a fruit sample, though only the pulp is analyzed. When this requirement is forgotten, the test result is erroneous.

Scenario $i = 48$ in reporting, $m = 6$. A lapse may happen during the introduction of the data into the file: incorrect test results may be reported.

A-2-2-9 Slips, $k = 9$

Scenario $i = 49$ in sampling, $m = 1$. When a box is dropped on the ground and its content (a sample of fruits or vegetables) is collected again, the sample may be contaminated.

Scenario $i = 50$ in sample processing, $m = 2$. A slip in sample processing is possible, for example, when one sample is put in the cutter instead of another.

Scenario $i = 51$ in extraction, $m = 3$. On-going validation requires in every set of more than 5 samples inclusion of an early tested clean sample (where corresponding pesticide residues were not detected) with standard additions of the analytes/spikes, for evaluation of their recovery. When an analyst forgets to introduce the spikes into the sample, the recovery is zero.

Scenario $i = 52$ in identification, $m = 4$. A slip in identification leading to a false-negative result may happen when a test portion is not centrifuged after extraction and therefore a syringe needle is fouled.

Scenario $i = 53$ in quantification, $m = 5$. A test result concerning one analyte may be confused with another one by a slip.

Scenario $i = 54$ in reporting, $m = 6$. The daily number of analyzed/tested samples (sample throughput) is large, and reporting results related to another sample, as a slip, is possible.

A-2-3 Elicited data and error quantification scores

The elicited data for the mapped 54 human error scenarios, and the same four components of quality system as in Example 1, are presented in Table 2. The interrelationship matrix in Table 2 has $54 \times 4 = 216$ entries. The synergy between quality system components was taken into account similar to that in Example 1. The only exceptions are the cases of reckless violations, malicious violations, and omission errors, for which any synergy was considered unreliable in this analytical task. In other words, $\Delta_{jj'}^{(i)} = 0$ for any $i = 31$–54 and any $j' \neq j$.

The likelihood score calculated by formula (1) here is $P^* = 19\,\%$. The human error severity, evaluated by formula (2), is $L^* = 84\,\%$. The following effectiveness score values q_j^* of the quality system components are calculated by formulas (4) and (5): for validation – $q_1^* = 22\,\%$, for training – $q_2^* = 26\,\%$, for quality control – $q_3^* = 25\,\%$, and for supervision – $q_4^* = 27\,\%$. Thus, the most effective/important component of the quality system is supervision, followed by training, quality control, and validation.

The $\tilde{q}_m^*$ values calculated by formula (6) are: for sampling – $\tilde{q}_1^* = 7\,\%$, for sample processing – $\tilde{q}_2^* = 9\,\%$, for extraction – $\tilde{q}_3^* = 23\,\%$, for identification – $\tilde{q}_4^* = 22\,\%$, for quantification – $\tilde{q}_5^* = 29\,\%$, and for reporting – $\tilde{q}_6^* = 11\,\%$. These values show that the ability of the quality system to prevent human errors at sampling is minimal. This situation is caused by the fact that sampling is performed in the field, i.e. inspectors work mostly out of the Lab. The sample processing and reporting steps also require a more attention for the improvement of the quality system. Effectiveness of the whole quality system for all steps of the analysis by formula (7) is $E^* = 71\,\%$.

Table 2: Results of the expert judgments on human errors in multi-residue analysis of fruits and vegetables.

Scenario i	Likelihood p_i	Severity l_i	Degree of interaction r_{ij}				Synergy factor s_{ij}			
			Quality system layer j				Quality system layer j			
			1	2	3	4	1	2	3	4
1	3	9	3	9	1	9	1.67	1.33	1.33	1
2	1	3	9	3	3	9	1.67	1.33	1.33	1
3	3	9	9	9	9	9	1.67	1.33	1.33	1
4	3	9	9	9	3	9	1.67	1.33	1.33	1
5	3	9	9	9	9	9	1.67	1.33	1.33	1
6	1	9	9	9	9	9	1.67	1.33	1.33	1
7	1	3	3	1	1	9	1.67	1.33	1.33	1
8	1	9	3	9	3	9	1.67	1.33	1.33	1
9	3	3	3	9	9	9	1.67	1.33	1.33	1
10	1	9	3	9	9	9	1.67	1.33	1.33	1
11	3	9	1	9	9	9	1.67	1.33	1.33	1
12	1	9	3	3	9	9	1.67	1.33	1.33	1
13	3	3	3	9	1	3	1.67	1.33	1.33	1
14	1	9	3	9	3	9	1.67	1.33	1.33	1
15	3	9	9	9	9	9	1.67	1.33	1.33	1
16	3	9	9	9	9	9	1.67	1.33	1.33	1
17	3	9	9	9	9	9	1.67	1.33	1.33	1
18	1	3	3	3	9	9	1.67	1.33	1.33	1
19	3	3	3	9	1	3	1.67	1.33	1.33	1
20	1	9	3	9	3	9	1.67	1.33	1.33	1
21	3	9	9	9	9	9	1.67	1.33	1.33	1
22	3	9	9	9	9	9	1.67	1.33	1.33	1
23	3	9	9	9	9	9	1.67	1.33	1.33	1
24	3	9	3	3	9	9	1.67	1.33	1.33	1
25	1	9	3	3	1	3	1.67	1.33	1.33	1
26	1	9	3	3	3	9	1.67	1.33	1.33	1
27	1	3	1	9	9	9	1.67	1.33	1.33	1
28	1	3	3	9	9	9	1.67	1.33	1.33	1
29	1	9	3	9	9	9	1.67	1.33	1.33	1
30	1	3	3	3	9	9	1.67	1.33	1.33	1
31	1	9	1	1	1	3	1	1	1	1
32	1	3	1	3	3	9	1	1	1	1
33	1	9	1	3	9	9	1	1	1	1
34	1	9	1	3	9	9	1	1	1	1
35	3	9	1	3	9	9	1	1	1	1
36	1	9	1	3	9	9	1	1	1	1
37	1	3	0	1	1	1	1	1	1	1
38	1	9	0	3	3	9	1	1	1	1
39	1	9	0	3	3	9	1	1	1	1
40	1	9	0	3	3	9	1	1	1	1
41	1	9	0	3	3	9	1	1	1	1
42	1	9	0	3	3	9	1	1	1	1
43	3	9	0	1	1	1	1	1	1	1
44	3	9	0	3	3	9	1	1	1	1
45	1	9	0	3	3	9	1	1	1	1
46	1	9	0	3	3	9	1	1	1	1
47	1	9	0	3	3	9	1	1	1	1
48	1	9	0	3	3	9	1	1	1	1
49	1	3	0	1	1	1	1	1	1	1
50	1	9	0	3	3	9	1	1	1	1
51	1	3	0	3	3	9	1	1	1	1
52	1	9	0	3	9	9	1	1	1	1
53	1	9	0	3	9	9	1	1	1	1
54	1	9	0	3	9	9	1	1	1	1

A-2-4 Residual risk and measurement uncertainty

The score of residual risk of human errors by formulas (8) and (9) is $R^* = 65\,\%$. The percentage of the quality of the measurement/test results which may be lost due to residual risk of human errors is $f_{HE} = 9.9\,\%$ by formula (10).

The relative standard measurement uncertainty reported for tomatoes, for example, averaged for all analytes and expressed in % of an analytical result, was $u_r = 20\,\%$ [45]. The contribution of the uncertainty budget caused by residual risk of human errors by formula (14), Annex B, is $u_{HE\text{-}r} = 10\,\%$ relative. Thus, $u_{HE\text{-}r}$ is not negligible and not the dominant contribution in the uncertainty budget, similar to the case of pH measurements of groundwater in Example 1. The resulting uncertainty value by formula (13), Annex B, is $u_{res\text{-}r} = 22\,\%$ relative.

Example 3. Human errors in ICP-MS analysis of geological samples

A-3-1 Introduction and main steps of the analysis

ICP-MS is used widely in many laboratories for the chemical analysis of geological and other samples. There are typically four main steps m of the analysis with ICP-MS: 1) sample preparation, 2) calibration of the ICP-MS measuring system, 3) measurement of analyte concentrations in the prepared solutions, and 4) calculation of elemental mass fractions in analyzed samples and reporting ($M = 4$).

The sample preparation of rocks and sediments is based on the fusion of the sample with lithium metaborate or sodium peroxide flux. Then the obtained bead is dissolved in nitric acid in an ultrasonic bath. The solution should be filtered and diluted with water to a sample/solution weight ratio in the range from 1:1000 to 1:4000. Samples of peridotites and a number of types of magma can be prepared by digestion of a sample with an $HF–HNO_3$ mixture in an ultrasonic bath. When samples contain resistant phases, e.g. zircon, the applied temperature and pressure are increased using microwaves or digestion bombs. Then samples are evaporated to incipient dryness, refluxed in nitric acid, evaporated and dissolved again, filtered, and diluted with water. For analysis of trace and rare earth elements the sample digestion with an $HF–HClO_4$ mixture under pressure can be applied. In any case, an analytical blank is prepared identically to the samples.

Synthetic and natural certified reference materials (CRMs) are used for the preparation of matrix matched calibrators of ICP-MS. The concentration of the acids and flux quantity in such calibrators should be the same as in the samples prepared for analysis. This is in addition to the known requirement of CRMs to have a composition close to the composition of the analyzed samples in order to minimize matrix effects. The CRMs (not the same as for calibration) are used also as internal standards and quality control samples.

A-3-2 The map of human error scenarios

In spite of achievements in instrument development there are still a number of human error scenarios which should be taken into account in a routine laboratory for quality risk management. A total of $I = 9M = 36$ scenarios of human errors in ICP-MS analysis of geological samples are discussed below [18].

A-3-2-1 Knowledge-based mistakes, $k = 1$

Scenario $i = 1$ in sample preparation, $m = 1$. A sample containing an excessively high quantity of an analyte (not diluted as necessary) may produce too low of a response since not all the quantity will be ionized, resulting in an incorrect recovery factor.

Scenario $i = 2$ in ICP-MS calibration, $m = 2$. Application of an inadequate calibrator (with a difference in the matrix in comparison to the samples) may lead to a bias in the test results.

Scenario $i = 3$ in measurement with ICP-MS, $m = 3$. Use of an improper blank solution (did not pass all the steps of the sample preparation) may also cause biased results.

Scenario $i = 4$ in calculation and reporting, $m = 4$. Mistaken interpretation of interferences (e.g. due to diatomic molecules) may influence the test result.

A-3-2-2 Rule-based mistakes, $k = 2$

Scenario $i = 5$ in sample preparation, $m = 1$. An analyst, using as a rule sample preparation by digestion of a sample with an HF-HNO_3 mixture in an ultrasonic bath, may not take into account that a sample contains a resistant phase, which requires application of a microwave or digestion bombs.

Scenario $i = 6$ in ICP-MS calibration, $m = 2$. Usual dilution of reference materials for preparation of calibrators, when another dilution is necessary, may cause atypical test results.

Scenario $i = 7$ in measurement with ICP-MS, $m = 3$. When drift of the instrument response is usually controlled for specific ion masses, whereas another analyte is under determination, the control may be not sufficient and the results shifted.

Scenario $i = 8$ in calculation and reporting, $m = 4$. Unusual sample mass applied in an analysis (e.g. to increase quantity of an analyte) may be forgotten by an operator and the regular mass introduced erroneously in the file for calculations.

A-3-2-3 Skill-based mistakes, $k = 3$

Scenario $i = 9$ in sample preparation, $m = 1$. Dissolution of a sample in an acid mixture containing HF in a Teflon beaker (not in a digestion bomb) as usually done for determination of minor elements and/or traces, wherein silicon is an analyte, may lead to a loss of silicon.

Scenario $i = 10$ in ICP-MS calibration, $m = 2$. Use of the same calibrator as previously, when its container is not closed hermetically and the element concentrations change due to water evaporation, is a mistake.

Scenario $i = 11$ in measurement with ICP-MS, $m = 3$. Flux-fusion sample solutions may form a gel, not always immediately visible, but clogging the nebulizer and leading to inhomogeneity of the analyte distribution in the test portion. Measurements of the analyte concentrations in such solutions (in regular conditions) may lead to mistaken results.

Scenario $i = 12$ in calculation and reporting, $m = 4$. A skill-based human error is possible when an analyst uses a certain order of samples, whereas an assisting operator arranged the samples in another way.

A-3-2-4 Routine violations, $k = 4$

Scenario $i = 13$ in sample preparation, $m = 1$. A decision of an analyst after dissolution (based on visual inspection) that the filtration is not necessary and may be ignored to shorten the procedure, is a routine violation.

Scenario $i = 14$ in ICP-MS calibration, $m = 2$. To prepare a calibrator, a small value of concentrated CRM solution may be diluted to a large volume by one step, to avoid spending time for a longer procedure with more steps of dilution. The calibrator prepared in this way will be not accurate.

Another example is reducing the number of calibrators in order to shorten the work.

Scenario $i = 15$ in measurement with ICP-MS, $m = 3$. A routine violation is when result reading is started immediately after introduction of a sample into the instrument, without waiting at least 1 min for a stable response.

Scenario $i = 16$ in calculation and reporting, $m = 4$. When a report is not checked with purpose to save time, a twist of the data during their transformation to the final file may not be noted, and therefore not corrected, as in any other determination of a number of analytes in a number of samples.

A-3-2-5 Reasoned violations, $k = 5$

Scenario $i = 17$ in sample preparation, $m = 1$. An analyst may use more flux for fusion than required by the procedure in order to improve a sample preparation. However, it will lead to an increased concentration of salts, not appropriate for the blank in the run (for a set of samples).

Scenario $i = 18$ in ICP-MS calibration, $m = 2$. To improve a method, an analyst may wish to increase a calibration range (which is anyway wide in ICP-MS) in spite of limitation at both minimal and maximal analyte concentrations.

Scenario $i = 19$ in measurement with ICP-MS, $m = 3$. When a flow-injection system is used for the sample introduction, a limited number of analyte concentrations can be measured simultaneously (in the same run). An attempt to increase the number of analytes is a routine violation, as some of the analytes may not be detected accurately.

Scenario $i = 20$ in calculation and reporting, $m = 4$. An example of reasoned violation is the "reference materials syndrome", when an analyst reports analyte concentration values close to those in CRM certificates (applied as control samples) which are subsequently found to be incorrect [46].

A-3-2-6 Reckless violations, $k = 6$

Scenario $i = 21$ in sample preparation, $m = 1$. When cleaning of crucibles for fusing or glassware for dilution is performed improperly, a sample may become contaminated.

Scenario $i = 22$ in ICP-MS calibration, $m = 2$. Use of a CRM after the expiration date may lead to a biased calibration curve.

Scenario $i = 23$ in measurement with ICP-MS, $m = 3$. If an inadequate blank (from a previous analysis run) is taken recklessly, the measurement results may be biased.

Another example is when an analyst does not notice that a blank may also produce a response caused or influenced by contamination.

Scenario $i = 24$ in calculation and reporting, $m = 4$. Recklessness may lead to confusing names of samples.

A-3-2-7 Malicious violations, $k = 7$

Scenario $i = 25$ in sample preparation, $m = 1$. Filling sample labels in a confusing manner may lead further to their mistaken reading.

Scenario $i = 26$ in ICP-MS calibration, $m = 2$. The violation may consist of the use of previous calibration data instead of re-calibration.

Scenario $i = 27$ in measurement with ICP-MS, $m = 3$. Confusing names of samples may be caused not only because of recklessness, as in scenario $i = 24$ above, but also intentionally.

Scenario $i = 28$ in calculation and reporting, $m = 4$. Falsification of data is a malicious violation in any analysis.

A-3-2-8 Lapses, $k = 8$

Scenario $i = 29$ in sample preparation, $m = 1$. An analyst may forget to dry a sample before weighing.

Scenario $i = 30$ in ICP-MS calibration, $m = 2$. A lapse is also when an analyst forgets to stir a prepared calibrator.

Scenario $i = 31$ in measurement with ICP-MS, $m = 3$. Cleaning of a nebulizer and/or glassware used between runs may be forgotten because of a lapse.

Scenario $i = 32$ in calculation and reporting, $m = 4$. A lapse may happen during the introduction of the data into the file.

A-3-2-9 Slips, $k = 9$

Scenario $i = 33$ in sample preparation, $m = 1$. A sample may be incompletely transferred into a crucible, when poured out by a slip after weighing.

Scenario $i = 34$ in ICP-MS calibration, $m = 2$. Preparing a calibrator, an analyst may push, because of a slip, the arm of an automatic pipette stronger than necessary to achieve the stop. Then the taken volume is larger than required and the concentration of the analyte in the calibrator is not correct.

Table 3: The elicited expert judgments on human errors in ICP-MS analysis of geological samples.

Scenario i	Likelihood p_i	Severity l_i	Degree of interaction r_{ij}				Synergy factor s_{ij}			
			Quality system layer j				Quality system layer j			
			1	2	3	4	1	2	3	4
1	3	9	3	9	1	3	1.67	1.33	1.33	1
2	1	3	3	3	3	9	1.67	1.33	1.33	1
3	3	3	3	9	9	9	1.67	1.33	1.33	1
4	1	1	9	9	3	9	1.67	1.33	1.33	1
5	3	3	9	3	3	3	1.67	1.33	1.33	1
6	1	1	3	3	9	3	1.67	1.33	1.33	1
7	3	3	1	9	9	3	1.67	1.33	1.33	1
8	3	9	1	3	3	3	1.67	1.33	1.33	1
9	1	3	3	3	3	3	1.67	1.33	1.33	1
10	3	1	1	3	3	3	1.67	1.33	1.33	1
11	3	9	3	3	9	3	1.67	1.33	1.33	1
12	3	9	3	3	1	1	1.67	1.33	1.33	1
13	3	3	3	3	9	3	1.67	1.33	1.33	1
14	3	3	9	9	3	9	1.67	1.33	1.33	1
15	3	3	1	3	9	9	1.67	1.33	1.33	1
16	3	9	3	3	3	3	1.67	1.33	1.33	1
17	1	1	9	9	3	3	1.67	1.33	1.33	1
18	1	3	9	9	3	3	1.67	1.33	1.33	1
19	3	3	3	9	3	3	1.67	1.33	1.33	1
20	3	3	3	3	3	3	1.67	1.33	1.33	1
21	1	3	3	3	9	3	1.67	1.33	1.33	1
22	3	3	3	3	3	3	1.67	1.33	1.33	1
23	1	3	1	9	3	3	1.67	1.33	1.33	1
24	3	9	0	3	3	3	1.67	1.33	1.33	1
25	1	9	0	1	1	1	1	1	1	1
26	1	9	0	1	3	3	1	1	1	1
27	1	9	0	1	3	3	1	1	1	1
28	1	9	0	1	3	3	1	1	1	1
29	3	3	0	1	1	1	1	1	1	1
30	3	3	0	3	3	3	1	1	1	1
31	1	9	1	3	3	3	1	1	1	1
32	1	9	0	1	3	3	1	1	1	1
33	1	3	0	1	1	1	1	1	1	1
34	1	3	0	3	3	3	1	1	1	1
35	3	3	0	3	3	3	1	1	1	1
36	1	9	0	3	3	3	1	1	1	1

Scenario $i = 35$ in measurement with ICP-MS, $m = 3$. If a capillary used for sample introduction is set inaccurately by a slip, and air is passed with the liquid to the nebulizer, the measurement results may be erroneous.

Scenario $i = 36$ in calculation and reporting, $m = 4$. Reporting results related to one sample as results of another sample, by a slip, is possible.

A-3-3 Elicited data and error quantification scores

Elicited expert judgments on human errors by the described 36 scenarios and the same quality system components and their synergy, as in Example 1, are presented in Table 3. The interrelationship matrix here is also of $36 \times 4 = 144$ entries.

The likelihood score, summarizing the elicited judgments, is $P^* = 22\,\%$. The human error severity score is $L^* = 56\,\%$. The most effective/important component of the quality system by formulas (4) and (5) in this analysis is quality control ($q_3^* = 27\,\%$), followed by training ($q_2^* = 26\,\%$), validation ($q_1^* = 24\,\%$), and supervision ($q_4^* = 23\,\%$).

The $\tilde{q}_m^*$ calculation by formulas (6) for different steps of the analysis shows that the ability of the quality system to prevent human errors at the ICP-MS calibration ($\tilde{q}_2^* = 14\,\%$) is minimal. At the measurement step $q_3^* = 25\,\%$, at sample preparation $q_1^* = 30\,\%$, and at calculation and reporting $q_4^* = 32\,\%$. The effectiveness of the entire quality system for all steps of the analysis is characterized by $E^* = 55\,\%$, calculated using formula (7).

A-3-4 Residual risk and measurement uncertainty

The score of residual risk of human errors by formulas (8) and (9) here is $R^* = 65\,\%$. The percentage of the quality of the measurement/test results which may be lost due to residual risk of human errors is $f_{HE} = 8.1\,\%$ by formula (10).

The standard measurement uncertainty reported, for example, for determination of 10 ng·g^{-1} of ^{60}Ni in aqueous samples by ICP-MS [47] was $u = 0.75$ ng·g^{-1}. The contribution to the uncertainty budget caused by residual risk of human errors by formula (14), Annex B, is $u_{HE} = 0.32$ ng·g^{-1}. The resulting uncertainty value by formula (13), Annex B, is $u_{res} = 0.82$ ng·g^{-1}.

As in Examples 1 and 2, u_{HE} here is also not negligible and not the dominant contribution in the uncertainty budget.

Annex B. Contribution to measurement uncertainty

Considering standard measurement uncertainty u, evaluated according to guidelines [19, 48], as a quality parameter of a measurement or test result, one can say that quality Q is better when u is smaller, i.e. $Q = 1/u$. This is the simplest model $Q(u)$ and its simplicity is the main model advantage [42]. More complicated models could also be investigated and applied in specific cases.

Possible loss of quality because of residual risk of human errors, as an absolute value, is $Q\,f_{HE}/(100\,\%)$. Therefore, the resulting quality according to the proposed model is

$$Q_{res} = Q - Qf_{HE}/(100\,\%) = (1/u)[1 - f_{HE}/(100\,\%)]. \tag{11}$$

As $Q_{res} = 1/u_{res}$, where u_{res} is the resulting (combined) standard uncertainty including the human error contribution, from formula (11) one has

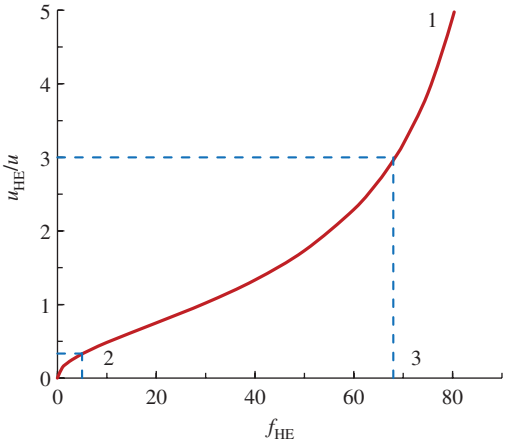

Fig. 3: Ratio of the uncertainty due to residual risk of human errors u_{HE} to the measurement uncertainty u in dependence on % of the quality loss f_{HE} (line 1). The cases $u_{HE} = 1/3u$ and $u_{HE} = 3u$ are indicated by dotted lines 2 and 3, respectively.

$$u_{res} = u/[1 - f_{HE}/(100\,\%)]. \tag{12}$$

In the view of Ref. [19, pp. 24–25] concerning uncertainty evaluation based on judgment "as for standard deviations derived by other methods", the contribution of the uncertainty u_{HE} (caused by residual risk of human errors) into the budget of the resulting uncertainty can be approximated by the conventional expression:

$$u_{res} = (u_{HE}{}^2 + u^2)^{1/2}. \tag{13}$$

Thus, it follows from formulas (12) and (13) that

$$u_{HE} = u\{[1 - f_{HE}/(100\,\%)]^{-2} - 1\}^{1/2}. \tag{14}$$

Because uncertainty caused by human errors u_{HE} is determined as a fraction of u by formula (14), both u_{HE} and u are expressed with the same units and the same number of digits.

Dividing u by the absolute value of the measured quantity value one can obtain relative standard measurement uncertainty u_r [20] and express it in %. In that case u_{HE} and u_{res} divided by the same value ($u_{HE\text{-}r}$ and $u_{res\text{-}r}$) are expressed also in %, as in Example 2, Annex A.

When $f_{HE} = 0\,\%$, the uncertainty contribution due to human errors $u_{HE} = 0$ and $u_{res} = u$. When f_{HE} increases in the range $0\,\% < f_{HE} < 100\,\%$, values of u_{HE} increase also, as shown in Fig. 3 by solid line 1. In particular, u_{HE} achieves $1/3u$ at $f_{HE} = 5\,\%$ (dotted lines 2) and begins to be a significant component of the uncertainty budget by formula (13). At $f_{HE} = 68\,\%$, value $u_{HE} = 3u$ dominates in the budget: this point is indicated by dotted lines 3 in Fig. 3. When f_{HE} exceeds 68 %, u_{HE} increases with f_{HE} dramatically. In the theoretical case of $f_{HE} = 100\,\%$, formulas (12) and (14) tend to infinity. However, such a contribution of human error to uncertainty is not realistic, inasmuch as the error becomes apparent: it will be identified and treated.

It is known that the largest contribution/component of a combined uncertainty needs to be investigated more thoroughly [19, p. 49]. Such a contribution may be overestimated and, hence, simply improved after investigation, or it may be the subject of a corrective action requiring an investment. Identified human errors can usually be reduced [49–51]. Thus, a good risk management result is when human errors are treated enough by the quality system to avoid their dominance in the uncertainty budget, as in Examples 1–3, Annex A.

Annex C. Monte Carlo simulations

CONTENTS

C-1 An expert judgment as a discrete quantity

An expert judgment for human error quantification is a discrete quantity that can take any scale value among (0, 1, 3, 9) according to the judgment probability mass function (pmf). When a value is chosen on the scale, the expert may still feel a doubt concerning neighboring scale values as pointed out in clause 6.1 of this Guide. Choosing 0, this expert thinks about 1 as a value which is also possible with equal or lower pmf. Choosing 1, the expert necessarily takes into account 0 and 3, but with equally lower pmf, etc. However, more distant scale values are not relevant. Otherwise, the expert is not experienced in the field and should not participate in the elicitation process.

On the other hand, the score values calculated directly from the elicited data can be interpreted as obtained when the expert judgments are completely confident, i.e. when a Dirac delta function, centered at a specific expert estimate on the scale, is applied as the pmf.

The following distributions modeling an expert behavior are studied below: 1) of completely confident expert judgments: the pmf at a chosen value is 1.00, whereas the remaining values on the scale have a total pmf equal to zero; 2) of confident expert judgments: the pmf at a chosen scale value is 0.90, whereas close values on the right and/or on the left on the scale have a total pmf equal to 0.10; 3) of reasonably doubting

Table 4: Probability mass functions (pmfs) of expert judgments.

Expert judgments	Chosen scale value	Scale			
		0	1	3	9
Completely confident	0	**1.00**	0.00	0.00	0.00
	1	0.00	**1.00**	0.00	0.00
	3	0.00	0.00	**1.00**	0.00
	9	0.00	0.00	0.00	**1.00**
Confident	0	**0.90**	0.10	0.00	0.00
	1	0.05	**0.90**	0.05	0.00
	3	0.00	0.05	**0.90**	0.05
	9	0.00	0.00	0.10	**0.90**
Reasonably doubting	0	**0.70**	0.30	0.00	0.00
	1	0.15	**0.70**	0.15	0.00
	3	0.00	0.15	**0.70**	0.15
	9	0.00	0.00	0.30	**0.70**
Irresolute	0	**0.50**	0.50	0.00	0.00
	1	0.25	**0.50**	0.25	0.00
	3	0.00	0.25	**0.50**	0.25
	9	0.00	0.00	0.50	**0.50**

The pmf at a chosen value is shown by bold.

expert judgments: the pmf at a chosen value is 0.70, whereas close values on the scale have a total pmf equal to 0.30; and 4) of irresolute expert judgments: the pmf at a chosen value is 0.50, and the close values on the scale have a total pmf equal to 0.50 also. More pmf details are shown in Table 4. These four pmfs represent properly the whole range of cases from the most to the least confident expert judgments in the framework of the proposed modeling.

Sampling from the distributions for random generation of expert judgments as discrete values was performed using a code developed in R [18].

C-2 Algorithm of simulations

Since human error quantification scores P^*, L^*, q_j^*, $\tilde{q}_m^*$ and E^*, are calculated as algebraic combinations of the elicited expert judgments p_i, l_i, r_{ij}, and synergy factors s_{ij} (which, in the present context, are considered as entirely known), the probability distributions for these scores depend on the distributions of the expert judgments. Monte Carlo simulations of the score distributions were performed based on the following algorithm inspired by JCGM 101 [52]:

1. input of the elicited estimates p_i, l_i and r_{ij}, synergy factors s_{ij}, numbers K of kinds of human error, M of steps of the chemical analysis, I of human error scenarios, J of the laboratory quality system components, and the number of the Monte Carlo trials $n_{MC} = 100\,000$;
2. assignment of pmfs to the expert judgments p_i, l_i, r_{ij};
3. simulation of possible values of expert judgments on human error by scenario i according to the chosen pmf on the scale values (0, 1, 3, 9) for $i = 1$ to I: the matrix of simulated values is of dimension $I \times n_{MC}$;
4. determination of the numerical distributions for scores P^*, L^*, q_j^*, $\tilde{q}_m^*$ and E^* by propagating the simulated distributions of the expert judgments into the relevant equations discussed in this Guide and evaluation of the score mean, median, and standard deviation (mean and median can be different because of a possible asymmetry in the simulated distributions);
5. plotting histograms for the distributions of the scores.

Note that for completely confident judgments with pmf by the Dirac delta function (for the scores calculated directly from the elicited data) the mean and median values obviously coincide, with the standard deviation of the simulated values being zero.

C-3 Examples

The elicited expert judgments on human errors in ICP-MS analysis of geological samples, Annex A, Example 3, Table 3, are used for the illustration of distributions of score values for error quantification with the Monte Carlo method [18].

Distributions of quality loss values due to residual risk of human errors are also studied with the Monte Carlo simulations [42]. All three sets of expert judgments on human errors from Annex A, Examples 1–3, are used here for the illustration of this application.

C-3-1 Distribution of score values for quantification of human errors

Results of the direct score calculations in comparison to the mean, median, and standard deviation of relevant score distributions simulated with the Monte Carlo method for ICP-MS analysis of geological samples are presented in Table 5.

One can see from Table 5 that the mean score values of a confident expert are very close to the score values calculated directly (of a completely confident expert). The mean values, as well as the median values,

Table 5: The score values (%) calculated directly from the elicited data for ICP-MS analysis in comparison to those obtained by Monte Carlo simulations.

Score	Calculated directly	Monte Carlo simulations								
		Confident expert			Reasonably doubting			Irresolute expert		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
P^*	22	24	23	2	26	26	3	29	29	4
L^*	56	55	55	3	53	53	5	51	51	5
q_1^*	24	25	24	2	26	25	4	27	26	5
q_2^*	26	26	26	2	26	26	3	26	26	4
q_3^*	27	27	27	2	26	26	3	26	26	4
q_4^*	23	23	23	2	22	22	3	22	21	4
$\tilde{q}_1^*$	30	29	28	6	26	25	9	25	23	10
$\tilde{q}_2^*$	14	15	14	4	16	15	6	18	17	8
$\tilde{q}_3^*$	25	25	25	5	26	25	9	27	26	11
$\tilde{q}_4^*$	32	32	31	6	31	30	9	30	29	11
E^*	55	54	54	3	53	53	4	51	51	5

SD is standard deviation of a simulated score value from its mean.

change systematically depending on the confidence of the expert judgments, i.e. depending on the relevant pmfs. Accordingly, it makes sense that the corresponding standard deviations increase as the expert judgments become less confident. However, the fact is that all the mean (and the median) values of the simulated scores remain consistent with the score values calculated directly from the data within two such standard deviations.

It appears from Table 5 that a less confident expert may lead to larger estimates for likelihood score P^*, from 22 % to 29 %, on average. Hence, the less confident the expert is, the more underestimated the P^* value directly calculated from the data. When P^* is increasing, the standard deviation is also increasing, from zero to 4 %.

In spite of the equivalence of P^* mean and median (rounded) values for a reasonably doubting expert, a minor asymmetry of the histogram in Fig. 4 is visible, probably due to the non-equidistant scale of the expert estimates/judgments through which the input pmfs were propagated. However, there are also other possible reasons for the asymmetry, e.g. when an expert unconsciously avoids one of the extreme choices on the scale (0 and 9).

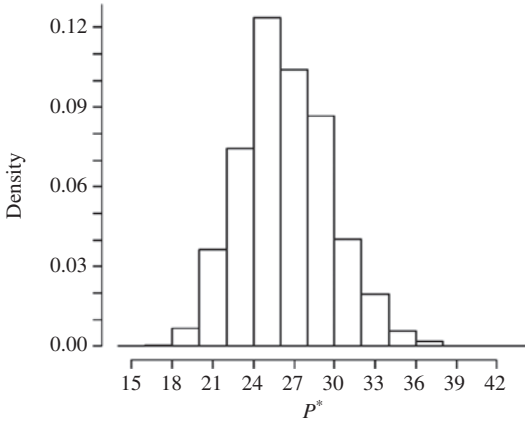

Fig. 4: A histogram of the likelihood score P^* (%) in ICP-MS analysis, corresponding to judgments of a reasonably doubting expert, simulated by Monte Carlo method. Reproduced from Ref. [18] with permission from Elsevier.

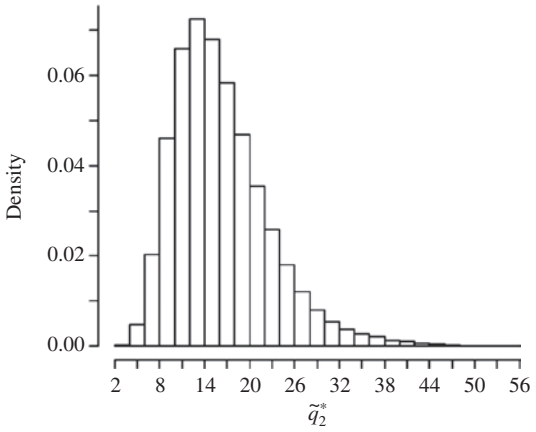

Fig. 5: A histogram of the score $\tilde{q}_2^*$ (%) of effectiveness of the quality system at second step of the ICP-MS analysis (the instrument calibration). This histogram corresponds to judgments of a reasonably doubting expert, simulated by Monte Carlo method. Reproduced from Ref. [18] with permission from Elsevier.

A less confident expert leads to a reduction in the estimates of the severity from mean $L^* = 56\,\%$ to 51 %, but again with an increasing standard deviation, from zero to 5 %. There is no difference between mean and median for a reasonably doubting expert, as in the likelihood score.

From q_j^* values in Table 5 one can find that the most effective/important component of the quality system (quality control) is characterized by the q_3^* score values from 27 % to 26 % with a standard deviation from zero to 4 %. There is no difference between the mean and the median of the simulated values for this score.

The $\tilde{q}_m^*$ simulated values for different steps of the ICP-MS analysis support the conclusion in Annex A, Example 3, that the ability of the quality system to prevent human errors at instrument calibration ($m = 2$) is minimal. The $\tilde{q}_2^*$ score values are from 14 % to 18 % with a standard deviation up to 8 %, depending on the expert's confidence. The variability of $\tilde{q}_m^*$ scores is the largest in comparison to other scores in Table 5. A difference of (1–2) % between the mean and the median of the $\tilde{q}_m^*$ score values and an evident histogram asymmetry, as in Fig. 5 for a reasonably doubting expert, are observed.

Effectiveness E^* of the whole quality system for all steps of the ICP-MS analysis is from 55 % to 51 % with a standard deviation varying from zero to 5 %. In general, E^* tends to be overestimated in direct calculation from the elicited data, similar to L^* (as opposed to P^* tending to be underestimated) when the confidence of expert judgments is decreasing.

C-3-2 Distribution of possible loss of quality

The results of f_{HE} direct calculations and simulations for the model of reasonably doubting expert judgments are shown in Table 6.

The f_{HE} values, calculated directly by formula (10) and interpreted as obtained from completely confident expert judgments, are close for the three examples. For all examples, the mean of the simulated f_{HE} values for the model of reasonably doubting expert judgments is a little larger than the f_{HE} calculated directly (not more than for one standard deviation of the simulated values). In other words, the estimated possible loss of quality due to residual risk of human errors is larger when an expert's doubt is taken into account. A similar effect was noted in clause C-3-1 concerning scores of likelihood, severity, and quality system effectiveness: less confident expert judgments lead to less optimistic score values.

A histogram, for example, of f_{HE} simulated values for reasonably doubted expert judgments on human errors in elemental analysis of geological samples with ICP-MS is shown in Fig. 6. This histogram is practically symmetric, its mean and median values differing insignificantly.

Table 6: Quality loss f_{HE} values (%) calculated directly from the elicited data in comparison to those obtained by Monte Carlo simulations for reasonably doubting expert judgments.

Chemical analytical method	Calculated directly	Monte Carlo simulations		
		Mean	Median	SD
pH metry of groundwater	10.8	11.2	11.1	1.6
Pesticide residues analysis	9.9	10.4	10.4	1.3
ICP-MS of geo-samples	8.1	9.5	9.4	1.4

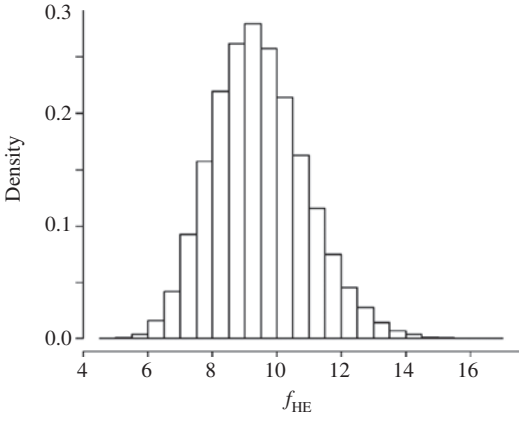

Fig. 6: A histogram of simulated f_{HE} values (%) in ICP-MS analysis, based on reasonably doubting expert judgments. ©Bureau International des Poids et Mesures. Reproduced from Ref. [42] by permission of IOP Publishing. All rights reserved.

C-4 Robustness

The score robustness for the quality risk management and improvement of a laboratory quality system can be considered satisfactory when a score's relative variability, expressed as relative standard deviation RSD (a ratio of standard deviation to mean), does not exceed 0.4, rounded up from 1/3. In other words, a score is robust when the standard deviation of the expert judgments can be defined as insignificant in comparison to the score mean. Such a rule of 1/3 is used in metrology, e.g. for verification of weights and preparation of test items for proficiency testing. Practically the same rule is applied in spectroscopy for determination of limit of detection, as an analyte concentration equal to three standard deviations of the measuring system response for a blank (noise).

It is important also that the score's relative range, i.e. the difference between the maximal and the minimal score values (calculated directly from elicited data and simulated values) related to their average, does not exceed the same 0.4 [18].

C-4-1 Score values for quantification of human errors

The requirement to robustness for E^* score, for example, implies RSD = SD/$E^* \leq 0.4$. In the case of the elemental analysis by ICP-MS the RSD of E^* is <0.1 for all models of the expert behavior in Table 5. Therefore, one can assume that the robustness of this score is satisfactory.

Other scores in Table 5 could also be assessed as robust enough. The $\tilde{q}_m^*$ scores, especially $\tilde{q}_2^*$, are less robust. However, even for irresolute expert judgments, the robustness of $\tilde{q}_2^*$ is still acceptable, as corresponding relative standard deviation is RSD = 8/18 = 0.4.

The largest relative range in Table 5 is also that of score $\tilde{q}_2^*$. However, this range satisfies the proposed criterion: (18–14)/16 = 0.25 < 0.4. Thus, the results of the human error quantification obtained in the case study are not dependent significantly on the kinds of calculation and simulation performed, i.e. they are robust from this perspective as well.

C-4-2 Possible loss of quality

Similar evaluation of quality loss f_{HE} due to residual risk of human errors in Table 6 shows that the f_{HE} relative standard deviations are in the range 0.12–0.15, i.e. smaller than 0.4. The same is true if one compares the maximal difference between the f_{HE} calculated directly by formula (10) and the mean of the simulated values with their common average. In Table 6 there is shown the case of ICP-MS, where $f_{HE} = 8.1\,\%$ by formula (10), while the simulated mean f_{HE} is 9.5 %, and their average is (8.1 % + 9.5 %)/2 = 8.8 %. Since (9.5–8.1)/8.8 = 0.15 < 0.4, the f_{HE} estimates are also robust enough to variability of expert judgments.

Membership of sponsoring bodies

Membership of the IUPAC Analytical Chemistry Division Committee for the period 2014–2015 was as follows:

President: D. B. Hibbert (Australia); **Vice-President:** J. Labuda (Slovakia); **Secretary:** Z. Mester (Canada); **Past President:** M. F. Camões (Portugal); **Titular Members:** C. Balarew (Bulgaria); Y. Chen (China); A. Felinger (Hungary); H. Kim (Korea); M. C. F. Magalhães (Portugal); H. M. M. Siren (Finland); **Associate Members:** R. Apak (Turkey); P. Bode (Netherlands); D. Craston (UK); Y.H. Lee (Malaysia); T. A. Marutina (Russia); N. Torto (South Africa); **National Representatives:** O. Chande Othman (Tanzania); L. Charles (France); P. De Bievre (Belgium); M. N. Eberlin (Brazil); A. Fajgelj (Austria); K. Grudpan (Thailand); J. Hanif (Pakistan); D. Mandler (Israel); P. Novak (Croatia); D. Shaw (USA).

Membership of the IUPAC Interdivisional Working Party on Harmonization of Quality Assurance for the period 2014–2015 was as follows:

Chair: A. Fajgelj (Austria); **Members**: P. Bode (Netherlands); P. de Zorzi (Italy); P. De Bièvre (Belgium); R. Dybkaer (Denmark); S. L. R. Ellison (UK); D. B. Hibbert (Australia); I. Kuselman (Israel); J. Y. Lee (Korea); L. Mabit (Austria); P. Minkkinen (Finland); U. Sansone (Austria); M. Thompson (UK); R. Wood (UK).

Membership of the Cooperation of International Traceability in Analytical Chemistry (CITAC) for the period 2014–2015 was as follows:

Chair: L. Samuel (New Zealand); **Vice Chairman:** A. Fagjelj (Austria); **Secretary:** S. Wunderli (Switzerland); **Past Chairman:** W. Louw (South Africa); **Members:** A. Squirrell (Australia); W. Wegscheider (Austria); P. De Bièvre (Belgium); O. P. de Oliveira Junior (Brazil); V. Poncano (Brazil); M. Suchanek (Czech Republic); T. Hirvi (Finland); I. Papadakis (Greece); C. M. Lau (Hong Kong, P. R. China); P. K. Gupta (India); M. Walsh (Ireland); I. Kuselman (Israel); M. Sega (Italy); T. Fujimoto (Japan); O. Zakaria (Malaysia); Y. M. Nakanishi (Mexico); V. Baranovskaya (Russia); Y. Karpov (Russia); T. K. Lee (Singapore); M. Weber (Switzerland); R. Kaarls (the Netherlands); R.J.C. Brown (UK); S. L. R. Ellison (UK); C. Burns (USA); V. Iyengar (USA), S.A. Wise (USA); J. D. Messman (USA); P. S. Unger (USA).

Membership of the Task Group was as follows:

Chair: I. Kuselman (Israel); **Members**: F. Pennecchi (Italy); A. Fajgelj (Austria); S. L. R Ellison (UK); Y. Karpov (Russia); M. Epstein (Israel).

Acknowledgments: The Task Group would like to thank E. Bashkansky (Israel) for the idea of the applicability of the house-of-security approach to human error quantification; E. Kardash and P. Goldshlag (Israel) for their expert judgments and help in preparation of Examples 1 and 2, respectively, in Annex A of this Guide; W. Bich (Italy) for useful discussions; Springer Science+Business Media (www.springer.com), Elsevier

(www.elsevier.com), Bureau International des Poids et Mesures, and IOP Publishing (www.ioppublishing.org) for permissions to use material from the published papers cited in this Guide.

References

[1] J. Reason. *Human Error*, Cambridge University Press, New York, USA (1990).
[2] B. Strauch. *Investigating Human Error: Incidents, Accidents and Complex Systems*, Ashgate, Farrnham, UK (2004).
[3] S. Dekker. *The Field Guide to Understanding Human Error*, Ashgate, Farrnham, UK (2006).
[4] I. Kuselman, F. Pennecchi, A. Fajgelj, Yu. Karpov. *Accred. Qual. Assur.* **18**, 3 (2013).
[5] E. Hellier, J. Edworthy, A. Lee. *Int. J. Cognit. Ergon.* **5**, 445 (2001).
[6] M. Plebani. *Clin. Chem. Lab. Med.* **44**, 750 (2006).
[7] I. Kuselman. *CITAC News* **2009**, 1 (2009).
[8] S. L. R. Ellison, W. A. Hardcastle. *Accred. Qual. Assur.* **17**, 453 (2012).
[9] I. Kuselman, A. Fajgelj. *Chem. Int.* **35**, 30 (2013).
[10] I. Kuselman. *Chem. Int.* **37**, 30 (2015).
[11] US FDA. *Guidance for Industry. Investigating Out-of-Specification (OOS) Test Results for Pharmaceutical Production* (2006).
[12] I. Kuselman, F. Pennecchi, C. Burns, A. Fajgelj, P. de Zorzi (Eds.). *Pure Appl. Chem.* **84**, 1939 (2012).
[13] P. B. Szecsi, L. Qdum. *Clin. Chem. Lab. Med.* **47**, 1253 (2009).
[14] ICH Harmonized Tripartite Guideline. *Quality Risk Management Q9* (2005).
[15] ISO/TS 22367. *Medical Laboratories – Reduction of Error Through Risk Management and Continual Improvement* (2008).
[16] ISO 10012. *Measurement Management Systems – Requirements for Measurement Processes and Measuring Equipment* (2003).
[17] ISO/IEC 17025. *General Requirements for the Competence of Testing and Calibration Laboratories* (2005).
[18] I. Kuselman, F. Pennecchi, M. Epstein, A. Fajgelj, S. L. R. Ellison. *Talanta* **130**, 462 (2014).
[19] S. L. R. Ellison, A. Williams (Eds.). *Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement*, 3rd ed., pp. 24–25 (2012).
[20] JCGM 200. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*, 3rd ed. (2012); http://www.bipm.org/en/publications/guides/vim.html.
[21] ISO/IEC 3534. *Statistics – Vocabulary and Symbols – Part 1: General Statistical Terms and Terms Used in Probability* (2006).
[22] ISO 9000. *Quality Management Systems – Fundamentals and Vocabulary* (2005).
[23] ISO Guide 73. *Risk Management – Vocabulary* (2009).
[24] J. J. Rooney, L. N. V. Heuvel, D. K. Lorenzo. *Quality Progress*, 27 (2002); http://www.capapr.com/docs/reducing%20human%20error%20QP.pdf.
[25] D. K. Lorenzo. *A Manager's Guide to Reducing Human Errors: Improving Human Performance in the Chemical Industry*, American Chemistry Council, Washington DC (1990).
[26] S. W. Lin, V. M. Bier. *Reliab. Eng. Syst. Safe.* **93**, 711 (2008).
[27] B. W. Marguglio. *Quality Digest Magazine*, 1 (2009); http://www.qualitydigest.com/print/8374.
[28] National Patient Safety Agency. *Root Cause Analysis Tool Kit 7 Steps. Guidance: An Introduction to Human Error Theory* (2006); http://www.csip.org.uk/silo/files/guidanceintroductiontohumanerrortheorydoc.doc.
[29] C. K. W. de Dreu, A. Evers, B. Beersma, E. S. Kluwer, A. Nauta. *J. Organiz. Behav.* **22**, 645 (2001).
[30] S. P. Carmien, F. I. Cavallaro, R. A. Koene. In: *PETRA '09 Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*. Article 44, ACM, New York, USA, DOI: 10.1145/1579114.1579158 (2009).
[31] J. Reason. *Managing the Risks of Organizational Accidents,* Ashgate, Aldershot, UK (1997).
[32] J. Reason. *Brit. Med. J.* **320**, 768 (2000).
[33] I. Kuselman, E. Kardash, E. Bashkansky, F. Pennecchi, S. L. R. Ellison, K. Ginsbury, M. Epstein, A. Fajgelj, Y. Karpov. *Accred. Qual. Assur.* **18**, 459 (2013).
[34] S. Dror, E. Bashkansky, R. Ravid. *Int. J. Safety Secur. Eng.* **2**, 317 (2012).
[35] R. M. Cooke, L. H. G. Goossens. *J. Risk Res.* **7**, 643 (2004).
[36] A. O'Hagan. *Metrologia* **51**, S237 (2014).
[37] M. J. Nigrini. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*, John Wiley & Sons, USA (2012).
[38] E. Bashkansky, S. Dror. *Qual. Reliab. Eng. Int.* **32**, 535 (2016).
[39] I. Kuselman, P. Goldshlag, F. Pennecchi. *Accred. Qual. Assur.* **19**, 361 (2014).
[40] J. C. Helton, M. Pilch, C. J. Sallaberry. *Reliab. Eng. Syst. Safe.* **124**, 171 (2014).
[41] T. Kelly. *Disagreement and the Burdens of Judgment*, Princeton University, NJ, USA (2011); https://www.princeton.edu/~tkelly/datbj.pdf.

[42] I. Kuselman, F. Pennecchi. *Metrologia* **52**, 238 (2015).
[43] H. M. Ortner. *Accred. Qual. Assur.* **5**, 130 (2000).
[44] E. Kardash, I. Kuselman, I. Pankratov, S. Elhanany. *Accred. Qual. Assur.* **18**, 373 (2013).
[45] I. Kuselman, P. Goldshlag, F. Pennecchi, C. Burns. *Accred. Qual. Assur.* **16**, 361 (2011).
[46] M. S. Epstein. *Talanta* **80**, 1467 (2010).
[47] V. J. Barwick, S. L. R. Ellison, B. Fairman. *Anal. Chim. Acta* **394**, 281 (1999).
[48] JCGM 100. *Evaluation of Measurement Data – Guide to the Expression Uncertainty in Measurement* (2008); http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.
[49] X. Fuentes-Arderiu, D. Dot-Bach. *Clin. Chem. Lab. Med.* **47**, 112 (2009).
[50] ISO 31000. *Risk Management – Principles and Guidelines* (2009).
[51] IEC/ISO 31010. *Risk Management – Risk Assessment Techniques* (2009).
[52] JCGM 101. *Evaluation of Measurement Data – Suppl. 1 to the "Guide to the Expression of Uncertainty in Measurement" – Propagation of Distributions Using a Monte Carlo Method* (2008); http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf.