Pure Appl. Chem., Vol. 84, No. 9, pp. 1939–1971, 2012.
http://dx.doi.org/10.1351/PAC-REP-11-10-04

IUPAC/CITAC Guide: Investigating out-of-specification test results of chemical composition based on metrological concepts (IUPAC Technical Report)*

Ilya Kuselman[1,‡], Francesca Pennecchi[2], Cathy Burns[3], Aleš Fajgelj[4], and Paolo de Zorzi[5]

[1]National Physical Laboratory of Israel, Givat Ram, 91904 Jerusalem, Israel; [2]Istituto Nazionale di Ricerca Metrologica, 91 Strada delle Cacce, 10135 Turin, Italy; [3]Food and Drug Administration, 6th Avenue and Kipling Street, DFC-Bldg 20, Denver, CO 80225, USA; [4]International Atomic Energy Agency, Vienna International Centre, P.O. Box 100, A-1400 Vienna, Austria; [5]Istituto Superiore per la Protezione e la Ricerca Ambientale, Via Castel Romano, 100 – 00128, Roma, Italy

Abstract: A metrological background for investigating out-of-specification (OOS) test results of chemical composition is discussed. When an OOS test result is identified, it is important to determine its root causes and to avoid reoccurrence of such results. An investigation of the causes based on metrological concepts is proposed. It includes assessment of validation data of the measurement process and its metrological traceability chains, evaluation of measurement uncertainty, and related producer's and consumer's risks. This approach allows distinguishing between OOS test results that indicate an actual change in chemical composition of an analyzed object, and OOS test results that are metrologically related with a certain confidence probability, i.e., caused by measurement problems, while the analyzed object still meets the specification requirements at the time of testing.

Practical examples illustrating applications of the described approach in environmental and food analysis, as well in drug analysis and stability study of drug products, are described. Acceptance limits, warning and action lines for the test results, and corresponding producer's and consumer's risks are discussed.

Keywords: acceptance limits; IUPAC Analytical Chemistry Division; measurement uncertainty; out-of-specification test results; producer's and consumer's risks; warning and action lines.

CONTENTS

*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 on p. 1969.

‡Corresponding author: E-mail: ilya.kuselman@moital.gov.il

1. INTRODUCTION

Out-of-specification (OOS) test results of chemical composition are results that fall outside the specifications of acceptance criteria established in pharmaceutical industry [1], or do not comply with regulatory, legislation, or specification limits in other industries and such fields as environmental and food analysis.

The problem of OOS test results has been known for analysts working in quality control laboratories since the 1920s [2]. However, this problem attracted special attention in 1993, when Barr Laboratories (a generic-drug manufacturer) was sued by the U.S. government regarding a set of issues influencing the product quality, including the way the company dealt with OOS test results. Judge Wolin's ruling (the Barr Decision) was that following an OOS test result, an investigation must be initiated before any retesting can be done [3].

Identifying OOS test results is described in U.S. Food and Drug Administration (FDA) Guidance [1] as the laboratory (Phase 1) investigation. It includes responsibility of the analyst and his/her supervisor, conditions of the testing in the laboratory, etc. After identification of an OOS test result, it is important to determine its root causes with the purpose to avoid any reoccurrence of the conditions when appearance of a next OOS test result is possible or even inevitable. The FDA Guidance formulates recommendations for such incidences including production process review, additional laboratory testing using a predefined procedure, reporting testing results, and concluding the investigation with identification of the root causes. Thus, this document establishes an empirical organizational approach to the full-scale (Phase 2) investigation and decisions that can be accepted at the different stages of this investigation.

Currently, the majority of analysts realize that the measurement uncertainty concept is very important because of the necessity to balance the cost of measurements vs. the product quality risk [4,5]. For example, to assess compliance of a test result within legislation limits in food and feed in Europe, the analyst should report not only an analyte concentration, but also the associated measurement uncertainty [6]. When the compliance assessment is made on the basis of a measurement result accompanied by information on the uncertainty associated with the result, the rules developed in the EURACHEM/CITAC guide [7] can be used. Similar rules are included in the ILAC guidelines [8]. The JCGM guide [9] on the role of measurement uncertainty in conformity assessment has been recently developed.

An approach [10] based on measurement uncertainty and other metrological concepts, amplifying recommendations of the FDA guidance for the full-scale investigation of OOS test results, is detailed in the present guide.

1.1 Scope and field of application

This Guide is developed for implementation of metrological concepts for investigation of OOS test results of chemical composition. This includes assessment of validation data of the measurement process (of the test method) and its metrological traceability chains, evaluation of measurement uncertainty, and related producer's and consumer's risks.

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

1.2 Terminology

Terminology used in this Guide corresponds to ISO Guide 99 [11], ISO 3534 [12], and ISO 17000 [13].

1.3 Symbols and abbreviations

a	year (annus in Latin)
ANOVA	analysis of variance
AOAC	Association of Official Analytical Chemists International
c	amount-of-substance concentration of an analyte in a product or environmental object
$c_{\mathrm{a.l}}$	action line for c_{test}
$c_{\mathrm{l.a.l}}$	lower acceptance limit for c_{test}
$c_{\mathrm{l.s.l}}$	concentration lower specification limit
c_{r}	parameter expressing c_{test} in parts of corresponding MRL
c_{test}	concentration test result
c_{true}	concentration true value
$c_{\mathrm{u.a.l}}$	upper acceptance limit for c_{test}
$c_{\mathrm{u.s.l}}$	concentration upper specification limit
$c_{\mathrm{w.l}}$	warning line for c_{test}
CITAC	Cooperation on International Traceability in Analytical Chemistry
Codex Alimentarius Commission	international organization that develops food standards, guidelines, and related documents, named Food Book (Codex Alimentarius in Latin)
D_{OOS}	deviation of OOS test result from a specification limit
EP	European Pharmacopoeia
EPA	U.S. Environmental Protection Agency
EURACHEM	network of organizations providing a focus for analytical chemistry and quality related issues in Europe
F	observed frequency
$f(c_{\mathrm{true}})$	pdf of c_{true} distribution
$f(c_{\mathrm{test}})$	pdf of c_{test} distribution
$f(c_{\mathrm{test}}\|c_{\mathrm{true}})$	pdf of c_{test} distribution at a certain c_{true} (likelihood function)
f_0	fraction of tested samples in which no pesticide residues were found
FDA	U.S. Food and Drug Administration
GC	gas chromatography
H_0	null hypothesis
H_1	alternative hypothesis
HPLC	high-performance liquid chromatography
i	index number
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use

ILAC	International Laboratory Accreditation Cooperation
INPL	National Physical Laboratory of Israel
ISO	International Organization for Standardization
ISRAC	Israel Laboratory Accreditation Authority
JCGM	Joint Committee for Guides in Metrology
k	coverage factor
LC	liquid chromatography
m	mass
MRL	maximum residue limit
MS	mass spectrometry
$\hat{m}$	maximum likelihood estimate of a Weibull distribution shape parameter
n	number of OOS test results
N	total number of test results
NIST	National Institute of Standards and Technology
OOS	out of specification
P	probability; level of confidence
pdf	probability density function
Q_{av}	average rate of air drawn into a sampler during sampling
r	correlation coefficient
R_c	consumer's risk
R_p	producer's risk
s, s_c	standard deviation; index c is for confidence limit to regression line
SI	International System of Units
t	normalized variable having t distribution (Student's distribution)
$t(0.95, \nu)$	quantile of Student's distribution for level of confidence 0.95, one-sided interval, and ν degrees of freedom
TSP	total suspended particulates
u	standard measurement uncertainty
U	expanded measurement uncertainty
USP	U.S. Pharmacopoeia
UV	ultraviolet
V	volume
α	probability of Type I error
β	probability of Type II error
$\hat{\beta}$	maximum likelihood estimate of a Weibull scale parameter
δ_{an}	contribution to c_{test} caused by analytical error(s)
δ_{samp}	contribution to c_{test} caused by sampling error(s)
μ	population mean
ν	number of degrees of freedom
σ	population standard deviation
τ	time
τ_0	shelf life or retest period of a drug product
$\tau_{l.s.l}$	time of a product storage when c_{test} achieves $c_{l.s.l}$
τ_{OOS}	time of a product storage when OOS test results appear
$\tau_{u.s.l}$	time of a product storage when c_{test} achieves $c_{u.s.l}$
Φ	cumulative distribution function

2. METROLOGICAL APPROACH

Any OOS test result can indicate a product failure, or be caused by measurement (analytical), i.e., metrological problems. When a result of testing is quantitative and equal to the measurement result, the metrological approach requires, first of all, defining the measurand, i.e., the quantity intended to be measured. In an analytical quality control laboratory, it is amount-of-substance concentration c of an analyte in a product batch or an environmental object. The concentration true value is c_{true}. A model of the concentration test result c_{test} includes c_{true} and contributions caused by error(s) in sampling δ_{samp} and analysis δ_{an} as two stages of testing:

$$c_{test} = c_{true} + \delta_{samp} + \delta_{an} \quad (1)$$

Distribution functions associated with these contributions can be very different. In particular, distribution of c_{true} values depends on changes of conditions of the production process from batch to batch or changes of the tested environmental object depending on place, day, etc. (global distribution). However, for a well-studied and widely used measurement method including sampling and analysis, distributions of δ_{samp}, δ_{an}, and corresponding c_{test} distribution at one and the same c_{true} value (measurement distribution), are normal more often than not or can be transformed into normal. This c_{test} distribution caused by the measurement uncertainty can be characterized by a probability density function (pdf), general for any c_{true} value in the range under investigation (the likelihood function). Taking into account this model, the full-scale investigation of OOS test results based on the metrological concepts includes the following:

1) assessment of validation data for sampling and chemical analysis;
2) evaluation of contributions to the measurement uncertainty from different stages of the test;
3) assessment of the metrological traceability chains important for the measurement parameters and environmental conditions influencing the test results; and
4) evaluation of consumer's and producer's risks in interpretation of OOS test results.

Such an investigation should answer the question, whether the OOS test result is caused by unsatisfactory product (environment) quality, or this result is metrologically related. In other words, are the root causes of an OOS test result deviation from the specification limit found in the measurement/analytical process?

2.1 Assessment of validation data

Validation is verification, where the specified requirements are adequate for an intended use [11]. This is a widely used procedure in the pharmaceutical industry. There are the FDA guidance for industry process validation including validation of sampling procedures [14], the ICH guideline for validation of analytical procedures [15], recommendations for analytical method and measuring equipment validation [16], etc. In other industries and analytical fields, validation is regulated by the EURACHEM guide [17], AOAC validation programs, and other national and international documents [18].

The most common validation parameters are repeatability, reproducibility, trueness and bias, limit of detection, selectivity and sensitivity, as well as linearity and limit of quantification [15], robustnesss, and ruggedness [19].

Investigating OOS test results, one should verify where the specified requirements and the validation data are adequate for the intended use. Absence of the adequacy can be a root cause of the OOS test results. Another question to be checked is whether the validation data are complete enough to evaluate contributions of the associated measurement uncertainty.

2.2 Evaluation of measurement uncertainty contributions

Measurement uncertainty is a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used [11]. Evaluation of measurement uncertainty can be done using repeatability, reproducibility, and trueness estimates from the validation data [20]. A number of examples of uncertainty calculation in the field of environmental analysis are available in the handbook [21]. Other methods for quantifying uncertainty in analytical measurement are described in the EURACHEM/CITAC guide [22]. Methods and approaches for evaluating measurement uncertainty arising from sampling are discussed in the EURACHEM/CITAC guide [23].

There are two important measurement uncertainty aspects and questions in the full-scale investigation of OOS test results: 1) is the measurement uncertainty adequate for the intended use? and 2) are the contributions to the measurement uncertainty the values of the same order of magnitude? Any negative answer for one or both of these questions can indicate a cause of the OOS problem. If a dominant contribution is detected while answering the second question, this contribution should be studied thoroughly.

2.3 Assessment of metrological traceability chains

Metrological traceability is a property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. Traceability chain is a sequence of measurement standards and calibrations that is used to relate a measurement result to a reference [11]. There are the EURACHEM/CITAC guide [24] on this topic, the IUPAC technical report on metrological traceability of measurement results in chemistry [25], and many other documents and publications, e.g., [26].

Assessment of metrological traceability chains is important for measurement parameters and environmental conditions influencing the test results. For example, traceability chains of measurement results to SI units of mass (kilogram), of amount of substance (mole), and of thermodynamic temperature (kelvin) should be realized for practically every chemical test. The reason is that a test portion is quantified by mass, measuring instruments are calibrated by certified reference materials, and temperature is to be under control. In particular, the pharmaceutical industry's practice of using a one-point calibration raises questions regarding the traceability chain of the measurement result to mole. This calibration consists of comparison of responses of the measuring system obtained for the test portion and a working standard. The working standard is certified by comparison with a USP's or other reference standard. Commutability [11] between the reference and the working standards, and adequacy of the working standard to the substance or drug product under analysis should be a point for investigation as well, first of all for impurities and degradation products. Any broken metrological traceability chain can lead to OOS test results.

2.4 Metrologically related OOS test results and acceptance limits

When root causes of metrologically related OOS test results are found, corresponding corrective actions of the measurement/analytical process will be helpful.

However, even though a problem is not found in the investigation, and an OOS test result c_{test} differs from specification limit in the range of expanded measurement uncertainty $U(c_{\mathrm{test}})$, it can be considered also as a metrologically related OOS test result because of the uncertainty. For example, when concentration upper specification limit $c_{\mathrm{u.s.l}}$ should be taken into account, the difference $D_{\mathrm{OOS}} = c_{\mathrm{test}} - c_{\mathrm{u.s.l}} \leq U(c_{\mathrm{test}})$ may be metrologically related with certain probability. When $D_{\mathrm{OOS}} > U(c_{\mathrm{test}})$, this difference is probably not caused by metrological problems and indicates violation of the product or environmental quality.

For interpretation of such results, acceptance limits for c_{test} set by a testing laboratory, manufacturer, or regulator, different from specification limits by measurement uncertainty, are applied according to the EURACHEM/CITAC guide [7].

Upper acceptance limit $c_{u.a.l}$ for test results smaller than upper specification limit $c_{u.s.l}$ by expanded measurement uncertainty can be used as a "warning line" $c_{w.l} = c_{u.a.l} = c_{u.s.l} - U(c_{test})$, which are seen used in quality control charts [27]. When a test result exceeds the warning line, i.e., $c_{test} > c_{w.l}$, the sampling and measuring systems should be checked and a decision to repeat the test may be made. At the same time, any decision about the product quality is still based on comparison of the test results with the upper specification limit $c_{u.s.l}$. Upper acceptance limit $c_{u.a.l}$ larger than $c_{u.s.l}$ by expanded measurement uncertainty can be used as an "action line" $c_{a.l} = c_{u.a.l} = c_{u.s.l} + U(c_{test})$, separating the metrologically related and violating OOS test results. The $c_{a.l}$ values resemble action lines that are used in quality control charts.

Similar acceptance limits $c_{l.a.l}$ for test results different from concentration lower specification limit $c_{l.s.l}$ by expanded measurement uncertainty, as well as warning and action lines to lower specification limit $c_{w.l} = c_{l.a.l} = c_{l.s.l} + U(c_{test})$ and $c_{a.l} = c_{l.a.l} = c_{l.s.l} - U(c_{test})$, respectively, are applicable also.

Examples from the environmental field, food quality control, and pharmaceutical analysis, including a stability study of drug products, are provided in Annex B.

3. HYPOTHESES ON A PRODUCT QUALITY AND OOS TEST RESULTS

Any decision on a product quality and its conformity assessment is based on comparison of null hypothesis H_0 that the quality is satisfactory and an alternative hypothesis H_1 about unsatisfactory product quality [27]. For example, when an upper specification limit $c_{u.s.l}$ is discussed, there are H_0: $c_{true} \leq c_{u.s.l}$ against H_1: $c_{true} > c_{u.s.l}$. True value c_{true} is unknown, and decisions are made using test results c_{test}. The distribution of c_{true} values in different batches of a product (the global distribution) and the measurement distributions of c_{test} values for two of these batches under testing according to the model in eq. 1 are illustrated in Fig. 1 as pdfs, i.e., $f(c_{true})$ and $f(c_{test})$, respectively, truncated normal for simplicity. Centers (means) of the c_{test} distributions are shown by vertical dotted pointers. These dotted pointers reach the true values c_{true} of the analyte concentration in the particular batches of the product under testing, i.e., considered coinciding with them. The upper specification limit $c_{u.s.l}$ is represented by a dotted vertical line.

The range of $c_{true} > c_{u.s.l}$ values corresponding to the product failure is shown by the horizontal dotted pointer. The shaded area under the $f(c_{true})$ curve to the right side of $c_{u.s.l}$ in Fig. 1a equals the probability P of the product failure. The range of OOS test results $c_{test} > c_{u.s.l}$ is shown by horizontal dotted pointers in Figs. 1b and 1c. The shaded area under the $f(c_{test})$ curve in Fig. 1b is the probability α of Type I error in the decision on the product quality. This error, named also "false positive", appears when $c_{true} \leq c_{u.s.l}$, while an OOS test result $c_{test} > c_{u.s.l}$ is obtained, hypothesis H_0 is rejected and hypothesis H_1 is not rejected (accepted). The probability α of Type I error is the producer's risk.

Type II error in the decision on the product quality, named also "false negative", is possible when product failure is analyzed, $c_{true} > c_{u.s.l}$, while $c_{test} \leq c_{u.s.l}$ and hypothesis H_0 is not rejected. This situation is illustrated in Fig. 1c. The shaded area under the $f(c_{test})$ curve to the left side of $c_{u.s.l}$ is probability β of Type II error. It is the consumer's risk.

3.1 Modeling a distribution

When a measurement distribution of c_{test} results (like in Fig. 1b and/or in Fig. 1c) is known and a number of tested batches are statistically significant, the global distribution of c_{true} values shown in Fig. 1a can be approximated by the empirical distribution of test results accumulated from batch to batch or from day to day of environmental monitoring, etc. The empirical distribution is fitted by a theoretical distribution (a model) with unknown, as a rule, parameters. Goodness-of-fit techniques for a control of

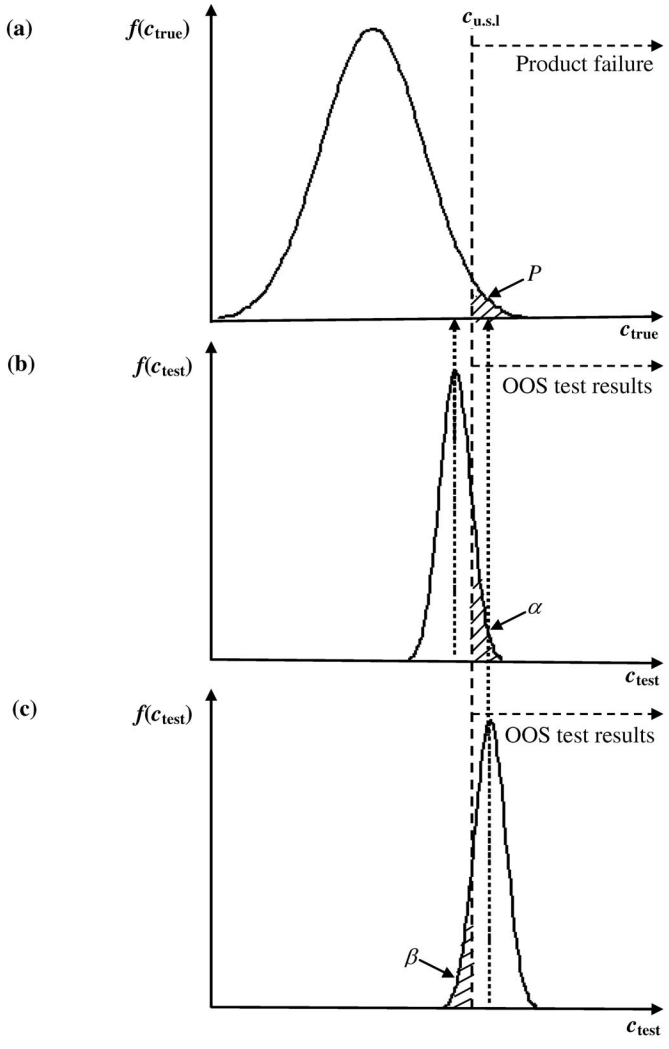

Fig. 1 OOS test results, producer's risk α and consumer's risk β. Functions $f(c_{true})$ in Fig. 1a and $f(c_{test})$ in Figs. 1b and 1c are pdf of c_{true} and c_{test}, respectively; vertical dotted pointers are means of c_{test} distributions equal to certain c_{true}; P is the probability of the product failure. Reproduced from ref. [10] by permission of Springer.

the model adequacy are described, e.g., in the textbook [28]. A model can be also chosen based on knowledge about the production process and properties of the product or about the environmental object under testing.

Examples of lognormal, Weibull, Student's, and normal models of global c_{true} distributions are discussed in Annex B, Examples 1–4, respectively.

3.2 Probability of OOS test results

When a global c_{true} distribution is approximated by a corresponding model, adequate to the c_{test} data, probability P of the product failure in Fig. 1a is transformed in the probability P_{OOS} of OOS test results. Therefore, P_{OOS} can be evaluated by the following equations at upper, lower, and both specification limits, respectively:

$$P_{OOS} = 1 - \Phi(c_{u.s.l}),\ P_{OOS} = \Phi(c_{l.s.l}) \text{ and } P_{OOS} = 1 - [\Phi(c_{u.s.l}) - \Phi(c_{l.s.l})] \quad (2)$$

where Φ is the cumulative distribution function modeling the global c_{true} distribution, i.e., the integral of the modeling pdf. The integrals $\Phi(c_{u.s.l})$ and $\Phi(c_{l.s.l})$ in eqs. 2 have the left integration limit equal to zero, since concentration of an analyte in a product or environmental object is a non-negative property.

Since calculation of the probability by eqs. 2 is per se integration of the distribution tails, results of such calculation can be larger than observed frequency values of OOS test results, when OOS test results appear mostly close to the specification limit(s), i.e., far from zero and infinity.

Examples of such calculations are provided in Annex B.

3.3 Global producer's and consumer's risks

While α and β are producer's and consumer's risks, respectively, for one and the same batch of a product, such risks evaluated in general for a statistically significant number of batches, forming a global c_{true} distribution, are named global producer's risk R_p and global consumer's risk R_c.

Equations for their calculation are presented in Annex A, examples of the calculations are in Annex B.

4. LIMITATIONS

The metrological approach does not spread in the present guide to cases of semi-quantitative and qualitative (e.g., organoleptic) testing and human errors.

If investigation of an OOS test result indicates a product failure caused by technological problems, this approach cannot be directly useful. However, when a significant contribution to measurement uncertainty arising from sampling is identified as a cause of OOS test results, an optimization of the technological parameters may be required to increase the product homogeneity.

The global producer's risk R_p and consumer's risk R_c do not take into account possible economical, health, and social consequences of false decisions on quality of a material or environment under testing.

ANNEX A. CALCULATION OF GLOBAL RISKS

There are three scenarios for calculation of the global risks according to the JCGM guide [9]: 1) around an upper specification limit, 2) around a lower specification limit, and 3) when both specification limits should be taken into account.

The global producer's risk R_p and consumer's risk R_c around an upper specification limit $c_{u.s.l}$ can be evaluated by the following equations:

$$R_p = \int_0^{c_{u.s.l}} \int_{c_{u.a.l}}^{\infty} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \quad (3)$$

$$R_c = \int_{c_{u.s.l}}^{\infty} \int_0^{c_{u.a.l}} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \quad (4)$$

where $f(c_{test}|c_{true})$ is the measurement pdf of c_{test} distribution at a certain c_{true} value (the likelihood function).

The risks around a lower specification limit $c_{l.s.l}$ are:

$$R_p = \int_{c_{l.s.l}}^{\infty} \int_0^{c_{l.a.l}} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \quad (5)$$

$$R_c = \int_0^{c_{l.s.l}} \int_{c_{l.a.l}}^{\infty} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \quad (6)$$

In the case when both specification limits should be taken into account simultaneously, the risks are:

$$R_{\mathrm{p}} = \left(\int_{0}^{c_{\mathrm{l.a.l}}} + \int_{c_{\mathrm{u.a.l}}}^{\infty} \right) \int_{c_{\mathrm{l.s.l}}}^{c_{\mathrm{u.s.l}}} f\left(c_{\mathrm{true}}\right) f(c_{\mathrm{test}} | c_{\mathrm{true}}) \mathrm{d}c_{\mathrm{true}} \mathrm{d}c_{\mathrm{test}} \tag{7}$$

$$R_{\mathrm{c}} = \left(\int_{0}^{c_{\mathrm{l.s.l}}} + \int_{c_{\mathrm{u.s.l}}}^{\infty} \right) \int_{c_{\mathrm{l.a.l}}}^{c_{\mathrm{u.a.l}}} f\left(c_{\mathrm{true}}\right) f(c_{\mathrm{test}} | c_{\mathrm{true}}) \mathrm{d}c_{\mathrm{test}} \mathrm{d}c_{\mathrm{true}} \tag{8}$$

The integrals are calculated numerically. Examples of such calculations are presented in Annex B:

1) around an upper specification limit, Examples 1–3;
2) around a lower specification limit, Example 3; and
3) for the case of both specification limits, Example 4.

ANNEX B. EXAMPLES

CONTENTS

EXAMPLE 1. INVESTIGATING OOS TEST RESULTS OF TOTAL SUSPENDED PARTICULATE MATTER CONCENTRATION IN AIR

B-1-1 Introduction

The objective of this example was an application of the metrological approach in the environmental field for investigating OOS test results of total suspended particulate (TSP) matter concentration in ambient air of some industrial zones [29].

EPA method IO-2.1 [30] for characterizing TSP uses a high-volume sampler for collection of particles with aerodynamic diameters of 100 µm or less. The sampler design causes air to be drawn into the sampler by means of a blower and through a glass or quartz fiber filter located downstream of the sampler inlet in order that the airborne particulate matter can be collected uniformly on the filter surface. A large volume, V, of (1600 to 2400) m^3 of air is typically sampled at an average rate Q_{av} of (1.13–1.70) $m^3 \cdot min^{-1}$ during sampling. Thus, $V = Q_{av} \cdot \tau$, where τ is the total elapsed sampling time in min. In order to determine a metrologically traceable value of the air volume, the flow rate measurement device should be calibrated and the total volume of sampled air corrected to V_{st} at the EPA's standard temperature and pressure by EPA method IO-2.4 [31]. The mass m of the matter in the sampled air volume is measured as the difference between the results of weighing the filter before and after sampling, in mg. The filter media characteristics and performance, as well as its conditions before and after sampling, are prescribed in EPA method IO-3.1 [32].

The TSP concentration in ambient air is the measurand in this method, while practically measured is the average value of the TSP concentration over the sampling/test period. Therefore, the test result is $c_{test} = m/V_{st}$.

There are national regulations of air quality including upper specification limits $c_{u.s.l}$ for TSP concentration depending on the period of averaging. For example, in Israel $c_{u.s.l} = 0.200$ $mg \cdot m^{-3}$ for 24 h. OOS test results of TSP concentration in ambient air of stone quarries located in Israel were investigated during a year as a case study.

B-1-2 Experimental

High-volume samplers and glass fiber filters were used. Three quarries were monitored by the National Physical Laboratory of Israel (INPL) according to EPA method IO-2.1 at all four points in the compass approximately (1–3) km from each quarry, four to five times per month, i.e., once per week or more frequently during the quarries' work. Each test lasted for 24 h. A total of 496 test results were obtained. The results were sorted with analysis of variance (ANOVA). The main found factor of variation was a quarry. Some details of the data distribution for every quarry, including number N of the test results, mean μ and standard deviation σ of the result natural logarithms, number n of OOS test results and their observed frequency $F = n/N$, are presented in Table 1. A total of 20 out of the 496 were OOS test results. The distributions are shown in Fig. 2 as histograms.

Table 1 Observed frequency F and probability P_{OOS} of OOS test results.

Quarry	N	μ	σ	n	F	P_{OOS}
1	220	–2.326	0.434	7	0.032	0.049
2	176	–2.031	0.280	11	0.063	0.066
3	100	–2.338	0.403	2	0.020	0.035

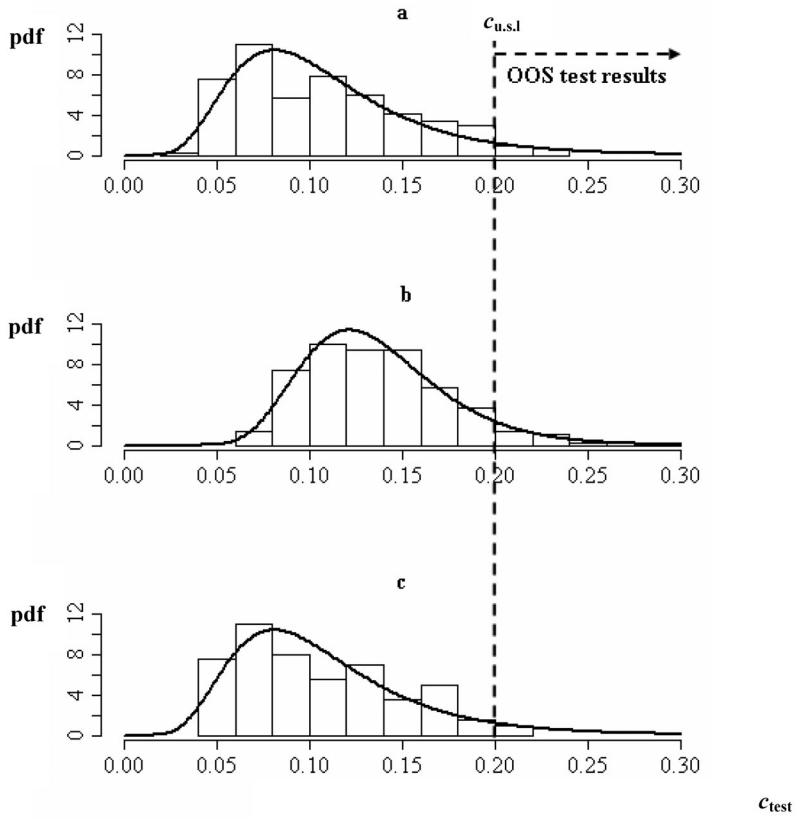

Fig. 2 Histograms and pdf of lognormal distributions of test results c_{test}/mg·m^{-3}. Data for quarry 1 are shown in Fig. 2a, quarry 2 in Fig. 2b, and quarry 3 in Fig. 2c. The upper specification limit $c_{u.s.l}$ is indicated by the dotted vertical line, the range of OOS test results by the dotted pointer. Reproduced from ref. [29] by permission of Springer.

The expanded relative measurement uncertainty was evaluated as $U(c_{test})/c_{test}$ = 0.14 to 0.21 or (14 to 21) % for normal distribution and the range of the levels of confidence P = 0.95 to 0.99, with the coverage factor k = 2 to 3.

B-1-3 Global distribution

Lognormal distributions were used for modeling pdf of the global c_{true} distributions for every quarry:

$$f(c_{true}) = \frac{1}{c_{true}\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln c_{true} - \mu)^2}{2\sigma^2}\right] \tag{9}$$

where μ and σ values for a quarry are from Table 1. These pdfs are shown in Fig. 2 for quarry 1 (Fig. 2a), quarry 2 (Fig. 2b), and quarry 3 (Fig. 2c) by solid lines smoothing the empirical c_{test} distributions, while the empirical distributions are presented here by the histograms. The upper specification limit $c_{u.s.l}$ = 0.200 mg·m^{-3} is shown by a dotted vertical line, common for all parts of Fig. 2. The range of OOS test results is indicated by dotted pointer.

Using the lognormal models of the c_{true} distributions by eq. 9, one can estimate probabilities P_{OOS} of OOS test results for every quarry by eq. 2 for an upper specification limit. The cumulative lognormal distribution functions Φ in eq. 2 were calculated for $c_{test} > 0$ and upper specification limit $c_{u.s.l} = 0.200$ mg·m^{-3}. Results of calculations are presented in Table 1. Comparison of the frequency values with probabilities of the OOS results shows that their annual numbers n may be larger than the observed ones.

B-1-4 Causes and probability of OOS test results

All observed OOS test results, their deviations D_{OOS} from the upper specification limit, and $U(c_{test})$ values for $P = 0.95$ to 0.99 are listed in Table 2. In this table $i = 1, 2, \ldots, n$ is the number of an OOS test result for a quarry. Answers to the question "is the OOS test result metrologically related?" are presented in the last column of the table. Only 2 out of 20 OOS test results indicate decidedly that the TSP concentration in ambient air violates the national regulations. The other 18 OOS test results may be caused by metrological problems.

Table 2 Deviations D_{OOS} of OOS test results c_{test} from the upper specifications limit $c_{u.s.l}$ in comparison with the expanded measurement uncertainty $U(c_{test})$.

Quarry	OOS test results		D_{OOS}/mg·m^{-3}	$U(c_{test})$/mg·m^{-3}		Metrologically
	i	c_{test}/mg·m^{-3}		$P = 0.95$	$P = 0.99$	related?
1	1	0.210	0.010	0.029	0.044	Maybe
	2	0.210	0.010	0.029	0.044	Maybe
	3	0.204	0.004	0.029	0.043	Maybe
	4	0.231	0.031	0.032	0.049	Maybe
	5	0.210	0.010	0.029	0.044	Maybe
	6	0.224	0.024	0.031	0.047	Maybe
	7	0.223	0.023	0.031	0.047	Maybe
2	1	0.223	0.023	0.031	0.047	Maybe
	2	0.288	0.088	0.040	0.060	No
	3	0.211	0.011	0.030	0.044	Maybe
	4	0.204	0.004	0.029	0.043	Maybe
	5	0.255	0.055	0.036	0.054	No
	6	0.215	0.015	0.030	0.045	Maybe
	7	0.211	0.011	0.030	0.044	Maybe
	8	0.216	0.016	0.030	0.045	Maybe
	9	0.226	0.026	0.032	0.047	Maybe
	10	0.225	0.025	0.032	0.047	Maybe
	11	0.232	0.032	0.032	0.049	Maybe
3	1	0.206	0.006	0.029	0.043	Maybe
	2	0.218	0.018	0.031	0.046	Maybe

B-1-5 Risks of stone producer and inhabitant

Global producer's risk R_p is here the probability that the satisfactory quality of air ($c_{true} \leq c_{u.s.l}$) will be determined falsely as violating the national regulations since $c_{test} > c_{u.s.l}$. Corresponding global consumer's/inhabitant's risk R_c is the probability that air quality violating the national regulations, when $c_{true} > c_{u.s.l}$, will be accepted falsely as conforming, since $c_{test} \leq c_{u.s.l}$. These risks were estimated by eqs. 3 and 4, where the $f(c_{true})$ was modeled by lognormal pdf by eq. 9 with parameters μ and σ shown in Fig. 1 and Table 1, whereas the likelihood function $f(c_{test}|c_{true})$ was approximated by a normal pdf with $\mu = c_{true}$ and corresponding $\sigma = u(c_{test})$ for every c_{true}:

$$f(c_{test}|c_{true}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(c_{test}-\mu)^2}{2\sigma^2}\right] \tag{10}$$

The warning lines $c_{w.l}$ to the upper specification limit were $c_{w.l} = c_{u.s.l} - U(c_{test}) = c_{u.s.l} - 0.14c_{test} = 0.200/(1 + 0.14) = 0.175$ mg·m^{-3} or $c_{w.l} = c_{u.s.l} - 0.21c_{test} = 0.200/(1 + 0.21) = 0.165$ mg·m^{-3} for $P = 0.95$ and 0.99, respectively. When a test result exceeds the warning lines, i.e., $c_{test} > c_{w.l}$, the sampling and measuring systems should be checked and a decision to repeat the test may be made.

The action lines $c_{a.l}$ to the upper specification limit were $c_{a.l} = c_{u.s.l} + U(c_{test}) = c_{u.s.l} + 0.14c_{test} = 0.200/(1 - 0.14) = 0.233$ mg·m^{-3} or $c_{a.l} = c_{u.s.l} + 0.21c_{test} = 0.200/(1 - 0.21) = 0.253$ mg·m^{-3} for $P = 0.95$ and 0.99, respectively. When a test result exceeds the action lines, i.e., $c_{test} > c_{a.l}$, the air quality is violated.

Results of R_p estimation for different $c_{u.a.l}$ are displayed in Fig. 3 by solid line 1, while R_c estimation results are shown by solid line 2. The upper specification limit is presented by a dotted line. The risks for quarry 1 are shown in Fig. 3a, quarry 2 in Fig. 3b, and quarry 3 in Fig. 3c. Acceptance limits for the range of the levels of confidence $P = 0.95$ to 0.99 are indicated by grey bars. The left one is the warning lines $c_{w.l}$, while the right one is the action lines $c_{a.l}$.

When acceptance limits are not in use and $c_{u.a.l} = c_{u.s.l}$, R_p and R_c are equal to 0.008 and 0.006, respectively, for quarry 1; 0.016 and 0.011 for quarry 2; and 0.006 and 0.005 for quarry 3. This means for quarry 1, for example, that the producer may be punished mistakenly in 8 cases of the ambient air testing from 1000, while violation of the national regulations may be not determined in 6 cases of the testing from 1000.

From Fig. 3, one can see as acceptance limits influence the global inhabitant's and producer's risks. For example, for the same quarry 1, when the level of confidence $P = 0.95$ was chosen and upper acceptance limit $c_{u.a.l}$ was equal to the warning line $c_{w.l} = 0.175$ mg·m^{-3}, risks $R_p = 0.043$ and $R_c = 0.0003$, while when acceptance limit $c_{u.a.l}$ was equal to the action line $c_{a.l} = 0.233$ mg·m^{-3}, R_p and R_c were already equal to 0.0001 and 0.026, respectively.

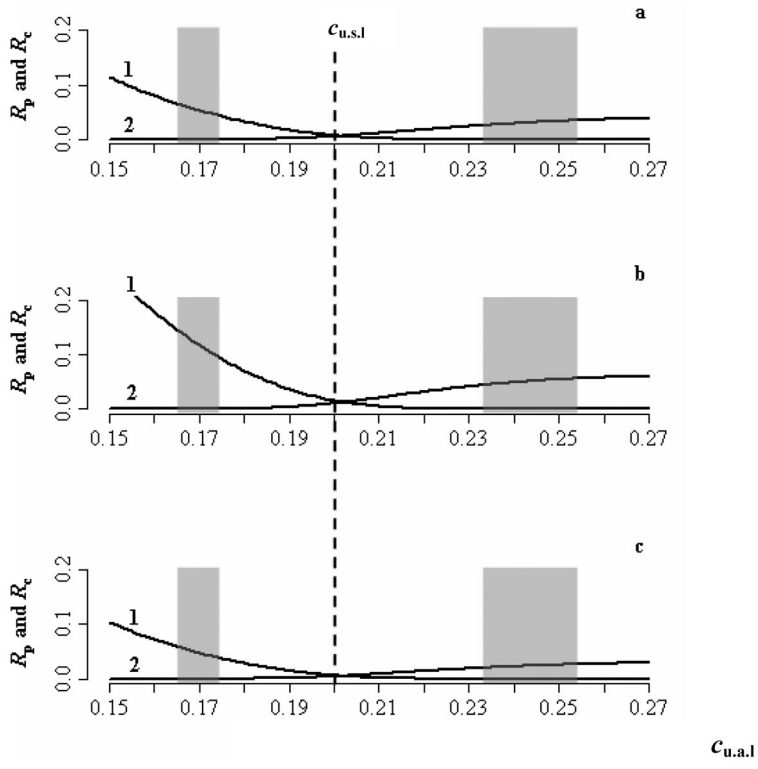

Fig. 3 Global producer's risk R_p and inhabitant's risk R_c in dependence on upper acceptance limit $c_{u.a.l}$/mg·m^{-3}. The risks for quarry 1 are shown in Fig. 3a, quarry 2 in Fig. 3b, and quarry 3 in Fig. 3c. R_p is displayed by solid line 1, and R_c by line 2. The upper specification limit is shown by the dotted line. Acceptance limits in the range of the levels of confidence $P = 0.95$ to 0.99 are indicated by grey bars. The left one is "warning lines", while the right one is "action lines". Reproduced from ref. [29] by permission of Springer.

EXAMPLE 2. MULTICOMPONENT OOS TEST RESULTS: PESTICIDE RESIDUES IN TOMATOES

B-2-1 Introduction

The objective of this example was an application of the metrological approach in the field of food analysis for investigating multicomponent OOS test results of pesticide residuals in tomatoes [33]. In this field, OOS test results are $c_{test} > MRL$, where MRL is a national legal maximum residue limit expressed in mg of the residue in kg of tomatoes (mg·kg^{-1}) [34], known also in the United States as a tolerance [35].

Investigated data were obtained during a year by the Israel Laboratory for Pesticide Residue Analysis. The laboratory has participated successfully in 22 proficiency testing rounds of the Food Analysis Performance Assessment Scheme for pesticide residues and has been accredited by the national laboratory accreditation authority (ISRAC). Periodically, ISRAC assesses the test method validation data, metrological traceability chains of the measurement/test process, and measurement uncertainty. Nevertheless, when an OOS test result is obtained, the question about causes of the result arises: is it because of a laboratory's metrological problem or a problem of a farmer/producer of the commodity violating the national legal limits?

The European guidance [34] requires, in the case of official food control by regulatory authorities, to check compliance with MRL assuming the lower limit of the uncertainty interval $c_{test} - U(c_{test})$

with $U(c_{test})/c_{test}$ = 0.50 or 50 %, if a laboratory proves its own calculated uncertainty to be less than 50 %. In other words, this requirement sets an acceptance limit $c_{u.a.l}$ for test results, meaning that the concentration of pesticide residues in a sample does not violate the national legal limits when $c_{test} \leq c_{u.a.l} = 2MRL$ (mg/kg) at the 0.95 level of confidence. What are the global risks to the farmer/producer and buyer/consumer with such an acceptance limit? How can the risks be changed at the same measurement uncertainty? Answers to these questions are discussed further using the case study of tomatoes.

B-2-2 Experimental

Sampling was conducted by certified inspectors according to the Codex Sampling Guidelines [36] directly from the field, packing houses, and logistic centers before sending the product to the market. Laboratory samples of tomatoes were in the amount of 1 kg.

Sample preparation for gas chromatography (GC) was performed by the known method based on extraction of analytes with acetone from a test portion of 15 g. For liquid chromatography (LC), sample preparation was performed by the method employing acetonitrile extraction from 10 g test portions. The test portions were taken from the homogenized (blended) laboratory samples.

The extracts were analyzed by GC methods with flame photometric and halogen selective detectors, as well as with mass spectrometry (MS). Electron ionization was applied in the MS in full scan.

LC of the extracts was performed by LC/tandem MS method with a triple quadrupole instrument and electrospray ionization.

A total of 217 reference standards (reference materials) for calibration of the chromatographs and for quality control purposes (125 for GC, 45 for LC, and 47 for both GC and LC methods) were used for simultaneous determination of the pesticides, as well as some of their metabolites and degradation products in the samples.

The analytical methods used were validated by the validation technique of the European guidance [34]. Relative expanded uncertainty $U(c_{test})/c_{test}$, including sample preparation and measurement/analytical components, was evaluated from the ongoing validation data. When averaged for all analytes, it was about 39 % with the coverage factor 2 at the 0.95 level of confidence. Therefore, the intralaboratory value of $U(c_{test})/c_{test}$ = 39 % was replaced for 50 % according to the European guidance [34] and used in the following discussion when the product is not yet marketed.

B-2-3 Global distribution

In 46 out of 169 tested samples, i.e., in f_0 = 46/169 = 0.272 or 27.2 % of them, no pesticide residues were found. A total of 39 analytes from 130 pesticides, authorized for use in Israel for tomatoes cultivation, were determined in $1 - f_0$ = 0.728 or 72.8 % of samples (123 out of 169). The analyte names, their occurrence (numbers of samples in % of 169), test results, and *MRL* values by the national regulations are listed in Table 3. Five (n = 5) out of N = 169 were OOS test results, and three of them violated the national legal limits by the 2*MRL* criterion of the European guidance [34].

Table 3 Analytes, their occurrence, test results c_{test} and the national *MRL* values in tomatoes.

Analyte	Occurrence/%	c_{test}/mg·kg^{-1}	*MRL*/mg·kg^{1}
Azoxystrobin	2.4	0.04–0.11	0.5
Bifenazate	3.0	0.02–0.04	0.05
Boscalid	3.6	0.01–0.10	0.2
Carbendazim	0.6	0.41	0.1
Carbosulfan	0.6	0.01	0.1
Chlorothalonil	18.3	0.01–1.33	5
Chlorpyrifos	1.8	0.01–0.39	0.5
Clofentezine	0.6	0.13	1
Cymoxanil	0.6	0.02	0.05
Cypermethrin	0.6	0.08	0.5
Cyprodinil	2.4	0.02–0.31	0.5
Diafenthiuron	0.6	0.05	0.05
Diazinon	0.6	0.03	0.5
Diethofencarb	1.2	0.01–0.04	0.1
Difenoconazole	0.6	0.09	0.1
Dimethoate	0.6	0.01	1
Endosulfan	0.6	0.08	0.5
Fenazaquin	1.2	0.02–0.33	0.1
Fenhexamid	0.6	0.03	0.5
Fludioxonil	1.8	0.01–0.11	0.3
Folpet	0.6	0.20	0.5
Iprodione	3.0	0.16–0.76	5
Iprovalicarb	0.6	0.04	0.05
Lufenuron	0.6	0.02	0.05
Mepanipyrim	0.6	0.11	0.1
Metalaxyl	3.0	0.01–0.09	0.5
Metominostrobin	1.2	0.01–0.14	0.2
Myclobutanil	1.2	0.06–0.08	0.3
Novaluron	0.6	0.02	0.2
Penconazole	3.0	0.03–0.11	0.2
Propargite	0.6	0.10	2
Pyrimethanil	0.6	0.03	0.05
Spiromesifen	3.6	0.01–0.28	1
Tebuconazole	1.8	0.03–0.11	0.2
Tebufenpyrad	0.6	0.03	0.05
Tetradifon	1.2	0.01–0.07	1
Thiamethoxam	0.6	0.03	0.02
Triadimenol	5.9	0.01–0.19	0.5
Trifloxystrobin	2.4	0.03–0.75	0.2

In order to enable analysis of multiresidue data as a statistical sample from the same population for different pesticide residues, the test results c_{test} were expressed in parts of corresponding *MRL* using a new dimensionless parameter c_r:

$$c_r = c_{test}/MRL \tag{11}$$

Such a transformation led to universal characterization of a concentration of any analyte in a sample from the point of view of the concentration adjacent to *MRL*. When $c_{test} = 0$, $c_r = 0$ also, and when $c_{test} = MRL$, $c_r = 1$. Any $c_r > 1$ indicates an OOS test result, and $c_r > 2$ denotes an OOS test result violating regulations by the European guidance [34], etc. The histogram of the experimental data in Fig. 4

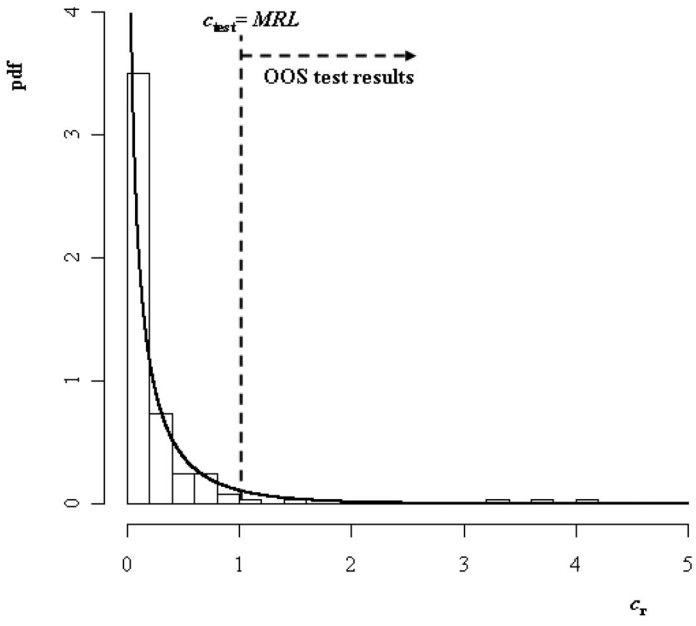

Fig. 4 Histogram of c_r values and pdf of the Weibull distribution. The pdf of c_r calculated by eq. 12 is shown by the solid line; *MRL* by the dotted vertical line; the range of OOS test results by the dotted pointer. Reproduced from ref. [33] by permission of Springer.

was plotted for $c_r > 0$, i.e., for 123 samples in which one or more pesticide residues were detected, identified, and quantified.

The pdf of the Weibull distribution used for modeling the global distribution of true c_r values c_{true}, shown in Fig. 4 by solid line, was

$$f(c_{true}) = \frac{\hat{m}}{\hat{\beta}}\left(\frac{c_{true}}{\hat{\beta}}\right)^{\hat{m}-1} \exp\left[-\left(\frac{c_{true}}{\hat{\beta}}\right)^{\hat{m}}\right] \tag{12}$$

where $\hat{m} = 0.652$ and $\hat{\beta} = 0.204$ are the maximum likelihood estimates of the shape and scale parameters, respectively. The MRLs are displayed in Fig. 4 by a dotted vertical line at $c_r = 1$, common for all analytes. The range of OOS test results is indicated by a dotted pointer.

B-2-4 Causes and probability of OOS test results

Analytes and other details of the observed OOS test results are presented in Table 4.

Note, occurrences of these analytes (Table 3) were minimal. Any OOS test result can indicate a pesticide concentration in the tomato sample violating the national legal limits, or be caused by measurement (metrological) problems, i.e., be metrologically related, especially when a specific analyte (a kind of measurement) is rare relatively to others.

Dividing both terms of the condition of a metrologically related OOS test result $c_{test} - MRL \leq U(c_{test})$ by the *MRL* leds to the following requirement: $c_r \leq 1/[1 - U(c_{test})/c_{test}]$. Therefore, if $U(c_{test})/c_{test} = 0.50$ in compliance with the European guidance [34], OOS test results can be classified as metrologically related at the level of confidence $P = 0.95$, when $c_r \leq 2$.

For the same relative standard uncertainty $u(c_{test})/c_{test} = 0.50/2 = 0.25$ or 25 %, the expanded uncertainty $U(c_{test})/c_{test}$ achieves $0.25 \cdot 3 = 0.75$ (75 %) at the coverage factor 3 corresponding to the

level of confidence $P = 0.99$. In such a case, a metrologically related OOS test result may appear up to $c_r \leq 4$. When $c_r > 1/[1 - U(c_{test})/c_{test}]$, the OOS test result was not caused by metrological problems. Answers to the question "is the OOS test result metrologically related?" are presented in the last column of Table 4. Only carbendazim residue in the sample can be classified as definitely (with more than 0.99 confidence) caused by a farmer's/producer's problem violating the national legal limit. The other OOS test results may be metrologically related with different probabilities.

Table 4 Comparison of the OOS test results c_{test} with *MRL* (c_r) and the expanded measurement uncertainty at different levels of confidence *P*.

OOS test results			Metrologically related?*	
Analyte	c_{test}/mg·kg^{-1}	c_r	$P = 0.95$	$P = 0.99$
Carbendazim	0.41	4.1	No	No
Fenazaquin	0.33	3.3	No	Maybe
Mepanipyrim	0.11	1.1	Maybe	Maybe
Thiamethoxam	0.03	1.5	Maybe	Maybe
Trifloxystrobin	0.75	3.8	No	Maybe

*"No" is for $P = 0.95$ when $c_r > 2$, and for $P = 0.99$ when $c_r > 4$.

Using the Weibull distribution modeling the global empirical c_r distribution in Fig. 4, one can calculate probability P_{OOS} of OOS test results by eq. 2 for an upper specification limit, where Φ is the cumulative Weibull function and $c_{u.s.l}$ is equivalent to $c_r = 1$. When the distribution parameters are $\hat{m} = 0.652$ and $\hat{\beta} = 0.204$, the probability is $P_{OOS} = 0.06$. Therefore, the annual number of OOS test results may be larger than the observed frequency $F = n/N = 5/169 = 0.03$.

B-2-5 Risks of tomato producer and consumer

Since pesticide residues were not detected in every sample, the global risk R_p of tomato producer/farmer and the global buyer's/consumer's risk R_c were calculated here after the following modification of eqs. 3 and 4:

$$R_p = (1 - f_0)\int_0^{MRL} \int_{c_{u.a.l}}^{\infty} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \tag{13}$$

$$R_c = (1 - f_0)\int_{MRL}^{\infty} \int_0^{c_{u.a.l}} f(c_{test}|c_{true}) f(c_{true})\, dc_{test}\, dc_{true} \tag{14}$$

where $1 - f_0 = 0.728$ is the frequency/probability of a pesticide residue detection in a tomato sample ($c_{test} > 0$). The global pdf $f(c_{true})$ was modeled by the Weibull pdf with eq. 12, whereas the likelihood function $f(c_{test}|c_{true})$ was approximated by a normal pdf as in eq. 10 having the mean $\mu = c_{true}/MRL = c_r$ and the standard deviation $\sigma = u(c_r) = u(c_{test})/c_{test}$ for every c_{true}. Simultaneously, *MRL* in the role of integration limits in eqs. 13 and 14 was replaced for $c_r = 1$, and acceptance limits $c_{u.a.l}$ were expressed also in c_r values.

The values of the upper acceptance limit $c_{u.a.l}$ for test results lower than *MRL* under the measurement expanded uncertainty were used as warning lines: $c_{w.l} = c_{test} = MRL - U(c_{test}) = MRL - 0.5c_{test} = MRL/1.5 = 0.67MRL$ mg·kg^{-1} or $c_r = 0.67$ for $P = 0.95$, and $c_{w.l} = c_{test} = MRL - 0.75c_{test} = MRL/1.75 = 0.57MRL$ mg·kg^{-1} or $c_r = 0.57$ for $P = 0.99$. When a test result exceeds the warning lines, i.e., $c_{test} > c_{w.l}$, the sample preparation and measurement/analytical systems should be checked and a decision to repeat the test may be made.

Acceptance limit values larger than *MRL* by the measurement expanded uncertainty were used as action lines: $c_{a.l} = c_{test} = MRL + U(c_{test}) = MRL + 0.5c_{test} = 2MRL$ mg·kg^{-1} or $c_r = 2$ for $P = 0.95$, as required in the European guidance [34], and $c_{a.l} = c_{test} = MRL + 0.75c_{test} = 4MRL$ mg·kg^{-1} or $c_r = 4$ for $P = 0.99$. When a test result exceeds the action lines, i.e., $c_{test} > c_{a.l}$, the quality of tomatoes is violated with corresponding probabilities 0.95 or 0.99.

Results of R_p estimation for different $c_{u.a.l}$ are displayed in Fig. 5 by solid line 1, while R_c estimation results are shown by solid line 2. The $c_{u.a.l}$ values were expressed here in parts of *MRL*, i.e., as c_r. Acceptance limits in the range of the levels of confidence $P = 0.95$ to 0.99 were indicated by grey bars. The left one is warning lines $c_{w.l}$, while the right one is action lines $c_{a.l}$.

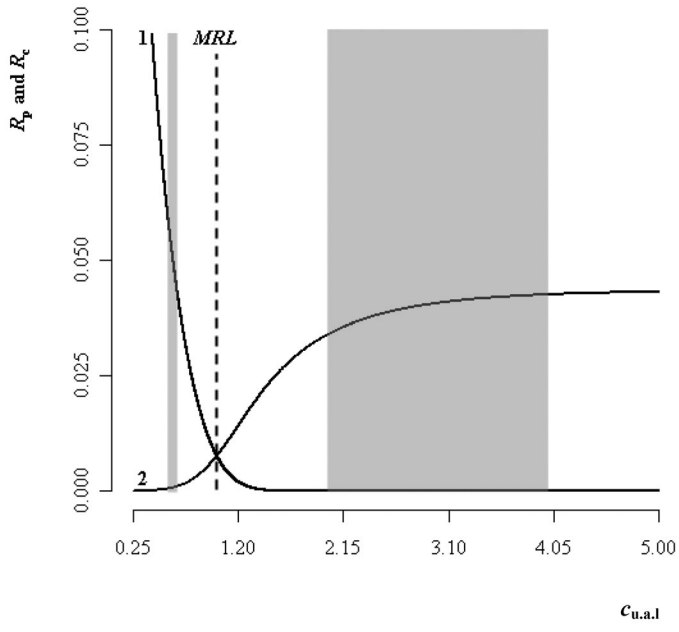

Fig. 5 Global producer's risk R_p and consumer's risk R_c in dependence on upper acceptance limit $c_{u.a.l}$. R_p is displayed by solid line 1, and R_c by solid line 2. The $c_{u.a.l}$ values are expressed in parts of *MRL* (as c_r). Acceptance limits in the range of the levels of confidence $P = 0.95$ to 0.99 are demonstrated by grey bars. The left one is "warning lines", while the right one is "action lines". Reproduced from ref. [33] by permission of Springer.

From Fig. 5, one can see as acceptance limits influence the global producer's and consumer's risks. For example, when the level of confidence $P = 0.95$ is chosen and acceptance limit $c_{u.a.l}$ is equal to the warning line $c_{w.l} = 0.67MRL$ mg·kg^{-1} or $c_r = 0.67$, the risks are $R_p = 0.040$ and $R_c = 0.001$, while when acceptance limit $c_{u.a.l}$ is equal to the action line $c_{a.l} = 2MRL$ mg·kg^{-1} or $c_r = 2$, R_p and R_c are already $1 \cdot 10^{-7}$ and 0.034, respectively.

When acceptance limits are not in use, $c_{u.a.l} = MRL$ and $c_r = 1$ (dotted line in Fig. 5), e.g., according to the system of tolerances of EPA [35], R_p and R_c are equal to 0.008 and 0.007, respectively. That means the farmer/producer may be punished mistakenly in 8 cases of the tomatoes testing from 1000, while violation of the national regulations may be not determined in 7 cases of the testing from 1000.

EXAMPLE 3. OOS TEST RESULTS IN LONG-TERM STABILITY STUDY OF DRUG PRODUCTS

B-3-1 Introduction

The objective of this example was an application of the metrological approach in pharmaceutical field for investigating OOS test results in a stability study of drug products [37].

When stability of a stored drug product is studied, it is important to establish a retest period or shelf life of the product, during which its properties are not influenced as yet and the drug can be used according to a physician prescription. International harmonized guideline ICH Q1E [38] recommends the establishment of a retest period or shelf life for a drug product using regression analysis of stability data (e.g., assay results vs. time) accumulated during long-term storage of the product. For a measured attribute (property) of the product known to increase with time, the regression one-sided upper 0.95 confidence limit should be compared to the acceptance criterion. The retest period or shelf life is estimated as the earliest time at which the confidence limit intersects the criterion. A similar rule is recommended for a measured property of the product known to decrease with time. The regression one-sided lower 0.95 confidence limit should be compared in such a case to the acceptance criterion.

The acceptance criterion may be formulated as a requirement to an amount-of-substance analyte concentration in a product not to exceed the upper specification limit $c_{\mathrm{u.s.l}}$, or not to be less than the lower specification limit $c_{\mathrm{l.s.l}}$. However, true values of the concentration c_{true} are unknown, and test results c_{test} are affected by the measurement uncertainty. Therefore, OOS test results $c_{\mathrm{test}} > c_{\mathrm{u.s.l}}$ or $c_{\mathrm{test}} < c_{\mathrm{l.s.l}}$ in a stability study can indicate an actual change (e.g., degradation) of the product or be metrologically related with a certain confidence probability P, i.e., be caused by the measurement problems, though the product still meets the quality requirements at the time of testing.

As examples, the test results of sodium chloride injection in plastic containers and epinephrine injection in ampoules, accumulated in the Research & Quality Control Laboratory of the Medical Corps, the Israel Defense Forces, are discussed.

The sodium chloride assay specification limits are $c_{\mathrm{l.s.l}}$ = 95.0 % and $c_{\mathrm{u.s.l}}$ = 105.0 % of the labeled amount [39], whereas the labeled amount is, for example, 0.9 % weight per volume, i.e., 0.9 g of sodium chloride in 100 mL of the solution. During long-term storage, an amount of water permeates from inside the container into the over-wrap space due to evaporation through the plastic. The water loss increases the sodium chloride concentration with time. Therefore, the test results c_{test} of sodium chloride concentration in the stored product (relative also to the labeled amount) were compared with the upper specification limit $c_{\mathrm{u.s.l}}$ = 105.0 %.

The assay specification limits for epinephrine injection in ampoules are $c_{\mathrm{l.s.l}}$ = 90.0 % and $c_{\mathrm{u.s.l}}$ = 115.0 % of the labeled amount of L-adrenaline, i.e., (L)-4-[1-hydroxy-2-(methylamino)ethyl]benzene-1,2-diol [40], whereas the labeled amount is, for example, 1 mg•mL^{-1}. L-adrenaline in the solution is subject to degradation during long-term storage, caused by oxidation, sulfonation, and racemization. Products of these reactions, including D-adrenaline and adrenaline sulfonate, do not have pharmacological activity comparable to L-adrenaline [41]. Therefore, the test results c_{test} of L-adrenaline concentration in the stored product (relative to the labeled amount) were compared with the lower specification limit $c_{\mathrm{l.s.l}}$ = 90.0 %.

Besides the assay results, concentration of some impurities and other properties of these products should be under control. Thus, the examples with sodium chloride and L-adrenaline assay were used here as a model only for discussion of OOS test results in both situations of the measured product property changes (increasing and decreasing) specified in guideline ICH Q1E [38].

B-3-2 Experimental

Samples of 18 batches of sodium chloride injection in 500 mL plastic containers (labeled as 0.9 %) were manufactured by B. Braun Melsungen AG, Germany, and Teva Medical Ltd., Israel. Samples of 93

batches of epinephrine injection in 1 mL ampoules (labeled as 1 mg·mL^{-1}) were manufactured by Teva Pharmaceutical Industries Ltd., Israel. The samples were stored under controlled conditions recommended by their manufacturers. Choice of these samples does not mean any preference or criticism.

Sodium chloride assay was performed by titration with silver nitrate of test portions sampled from a bag. The titration end-point was determined potentiometrically [42] with an automated titrator.

L-adrenaline chiral HPLC assay with UV–vis detection was performed as described in the paper [41].

The expanded relative measurement uncertainty associated with a routine sodium chloride assay result was evaluated as $U(c_{test})/c_{test}$ = 0.008 to 0.012 or 0.8 to 1.2 % for normal distribution and the range of the levels of confidence P = 0.95 to 0.99, with the coverage factor k = 2 to 3, respectively. The measurement uncertainty for L-adrenaline assay results evaluated for the same conditions was $U(c_{test})/c_{test}$ = 0.060 to 0.090 or 6.0 to 9.0 %.

B-3-3 Regression analysis and shelf life of the products

Results of regression analysis of the accumulated data c_{test} vs. time τ of the product storage are demonstrated in Fig. 6 for the sodium chloride injection and in Fig. 7 for the epinephrine injection.

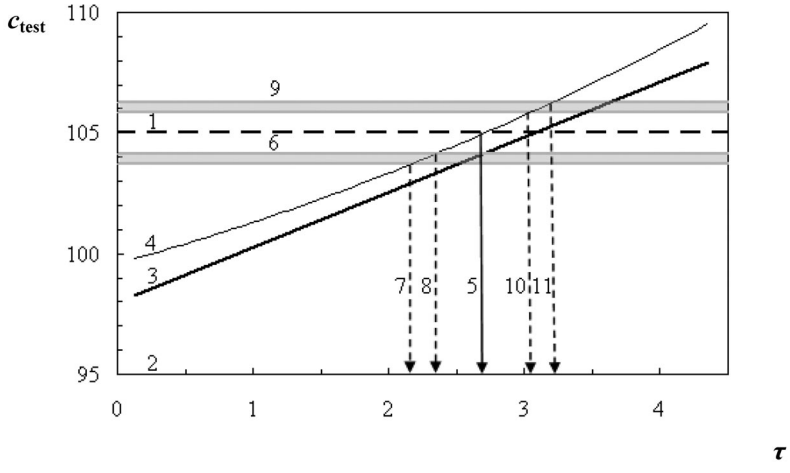

Fig. 6 Specification limits and shelf life of the sodium chloride injection. The ordinate is c_{test}/% axis; $c_{u.s.l}$/% is displayed by the dotted line 1; $c_{l.s.l}$/% by line 2 coincided with the abscissa, i.e., storage time τ/a (years); the regression is shown by solid line 3; the one-sided upper 0.95 confidence limit to this line is indicated by thin line 4. The product shelf life is shown by solid pointer 5. Grey bar 6 illustrates the corridor of test results c_{test} for the levels of confidence P = 0.95 to 0.99. Dotted pointers 7 and 8 indicate corresponding τ values. Grey bar 9 demonstrates the corridor of OOS test results at P = 0.95 to 0.99. Dotted pointers 10 and 11 show the storage time values corresponding to the corridor borders. Reproduced from ref. [37] by permission of Springer.

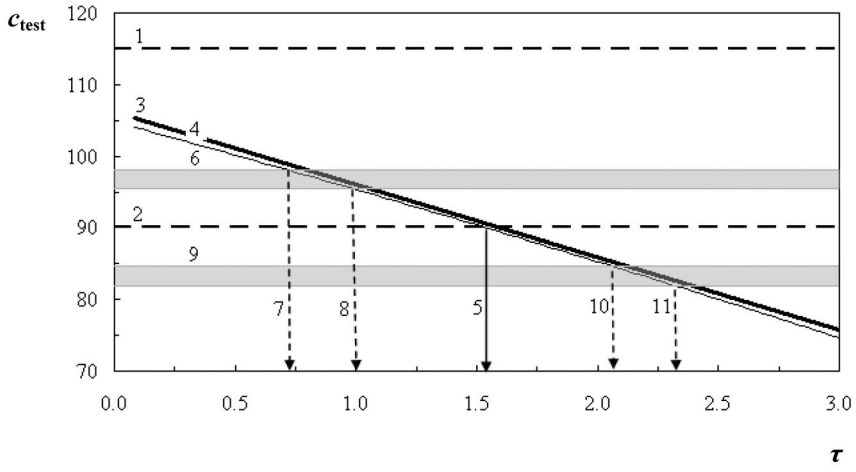

Fig. 7 Specification limits and shelf life of the epinephrine injection. $c_{u.s.l}$ and $c_{l.s.l}$ are displayed by dotted lines 1 and 2, respectively. Other symbols and signs are the same as in Fig. 6. Reproduced from ref. [37] by permission of Springer.

One-sided 0.95 confidence limit to the linear regression lines $\hat{c}_{test}(\tau)$ was calculated by the known formulas:

$$c_{c.l} = \hat{c}_{test} \pm t(0.95, \nu)s_c \tag{15}$$

$$\text{where } s_c = \sqrt{s_0^2\left[\frac{1}{N} + \frac{(\tau - \bar{\tau})^2}{\sum_{i=1}^{N}(\tau_i - \bar{\tau})^2}\right]} \text{ and } s_0^2 = \sum_{i=1}^{N}\left(c_{test} - \hat{c}_{test}\right)_i^2 \Big/ \nu$$

$t(0.95, \nu)$ is the quantile of t distribution (Student's distribution) for the level of confidence 0.95, one-sided interval, and the number of degrees of freedom $\nu = N - 2$, N is the number of observed test results c_{test} used in the regression analysis, s_c is the standard deviation of the predicted $\hat{c}_{test}$, and $\bar{\tau}$ is the mean of the τ range (the mean storage time).

The optimal range of storage time τ values for the study was estimated from eq. 15 as the range providing the minimal s_c when $\tau = \tau_0 = \bar{\tau}$, where τ_0 is the shelf life of the product. Such a range at any symmetrical distribution of τ is from 0 to $2\tau_0$. According to the manufacturer recommendations (known "a priori"), the shelf life for sodium chloride injection was 3 a and the optimal range of the storage time was 2 × 3 = 6 a. In practice, there were not any test results during 4.5 to 6 a of storage, and the studied range was limited by 4.5 a as shown in Fig. 6.

The manufacturer recommendation concerning the shelf life of the epinephrine injection was 1.5 a, and the optimal range of the studied storage time was 2 × 1.5 = 3 a as in Fig. 7.

The problem also is, eq. 15 does not have any direct solution $\tau(c)$ for calculation of the actual shelf life τ_0 in a general form, as the earliest time at which the confidence limit intersects the critical c_{test} value. Therefore, the confidence limit was approximated by parabola $\hat{c}_{c.l} = b_2\tau^2 + b_1\tau + b_0$. This approximation allowed calculation of the shelf life using the parabola roots:

$$\tau_0 = \left[-b_1 \pm \sqrt{b_1^2 - 4b_2\left(b_0 - \hat{c}_{c.l}\right)}\right] / 2b_2 \tag{16}$$

The sign "+" was used for the square root in the case of upper specification limit, while the sign "–" was necessary in the case of lower specification limit.

B-3-3-1 Shelf life of sodium chloride injection

The upper specification limit $c_{u.s.l}$ is displayed in Fig. 6 by dotted line 1. The lower specification limit $c_{l.s.l}$ (line 2) coincides with the abscissa. The linear regression for N = 18 observations was $\hat{c}_{test} = 2.2837\tau + 97.982$ with the squared correlation coefficient $r^2 = 0.7192$ (shown by solid line 3). The one-sided upper 0.95 confidence limit $c_{c.l}$ to the regression line 3, shown as thin line 4 in Fig. 6, was approximated by parabola $\hat{c}_{c.l} = 0.1717\tau^2 + 1.5319\tau + 99.587$, $r^2 = 0.9999$.

The τ_0 value was calculated by eq. 16 where the parabola coefficients were $b_2 = 0.1717$ %·a^{-2}, $b_1 = 1.5319$ %·a^{-1}, and $b_0 = 99.587$ %, while $\hat{c}_{c.l} = c_{test} = c_{u.s.l} = 105.0$ %. Thus, the actual product retest period or shelf life for the sodium chloride injection in the storage conditions of the Defense Forces was $\tau_0 = 2.7$ a. It is shown in Fig. 6 by solid pointer 5. The calculated τ_0 value is close to the 3 a recommended by both the injection manufacturers.

However, c_{test} may be less than $c_{u.s.l}$ because of the measurement uncertainty. For example, the grey bar 6 in Fig. 6 illustrates the corridor of values $c_{test} = 105.0 - U(105.0)$ from 104.2 to 103.7 % at the levels of confidence $P = 0.95$ to 0.99, respectively. Dotted pointers 7 and 8 indicate the storage time values of 2.4 and 2.2 a corresponding to these levels of confidence (the bar borders). In other words, there is a chance of a significant change of the injection quality after 2.2 a of the storage.

Metrologically related OOS test results $c_{test} > c_{u.s.l}$ appear up to $c_{test} = 105.0 + U(105.0)$. The grey bar 9 in Fig. 6 demonstrates the corridor of such OOS test results from 105.8 to 106.3 % at the levels of confidence $P = 0.95$ to 0.99, respectively. Dotted pointers 10 and 11 show the storage time values of 3.0 and 3.2 a corresponding to the corridor/bar borders. Thereby, even after 3.2 a of the storage, the injection quality may be satisfactory. OOS test results greater than 106.3 % would unlikely be classified as metrologically related but are, as a rule, evidence of the product change.

B-3-3-2 Shelf life of epinephrine injection

The assay upper and lower specification limits are shown in Fig. 7 by dotted lines 1 and 2, respectively. The linear regression for $N = 93$ observations displayed as solid line 3 in Fig. 7 was $\hat{c}_{test} = -10.157\tau + 106.18$ with $r^2 = 0.8632$. The parabolic approximation of the lower one-sided 0.95 confidence limit (line 4) was $\hat{c}_{c.l} = 0.2950\tau^2 - 9.1646\tau + 104.76$, $r^2 = 0.99999$. Substituting the coefficients of this parabola in eq. 16 for $\hat{c}_{c.l} = c_{test} = c_{l.s.l} = 90.0$ %, one can calculate the actual time value τ_0 indicated in Fig. 7 by solid pointer 5. The product retest period or shelf life calculated for the epinephrine injection ampoules was $\tau_0 = 1.5$ a for the Defense Forces storage conditions. This is exactly what was recommended by the manufacturer.

As in the situation with the sodium chloride injection, c_{test} may exceed here $c_{l.s.l}$ because of the measurement uncertainty. For example, the grey bar 6 in Fig. 7 demonstrates the corridor of such test results $c_{test} = 90.0 + U(90.0)$ from 95.4 to 98.1 % at the levels of confidence $P = 0.95$ to 0.99, respectively. Dotted pointers 7 and 8 show the storage time values of 1.0 and 0.7 a corresponding to the corridor borders. Therefore, there is the necessity in increasing quality control measures after 0.7 a of the product storage.

Metrologically related OOS test results appear for epinephrine injection when $c_{test} < c_{l.s.l}$. The grey bar 9 in Fig. 7 illustrates the corridor of values $c_{test} = 90.0 - U(90.0)$ from 84.6 to 81.9 % at the levels of confidence $P = 0.95$ to 0.99, respectively. Dotted pointers 10 and 11 show the storage time values of 2.1 and 2.3 a corresponding to these levels of confidence (the bar borders). Thus, even after 2.3 a of the storage, the injection quality may be satisfactory. OOS test results smaller than 81.9 % would unlikely be classified as metrologically related but provide an indication of the product degradation.

B-3-4 Risks of setting a shelf life

B-3-4-1 When a measured attribute increases with time

When a decision is to be made about a shelf life of a product in which the measured attribute increases with time, such as the assay of the sodium chloride injection, there are two risks. One of them is a prob-

ability that the product still meeting the quality requirements ($c_{true} \leq c_{u.s.l}$) will be falsely determined as violating the upper specification limit, since $c_{test} > c_{u.s.l}$. This is the global manufacturer's/producer's risk R_p, and a product owner (e.g., the Defense Forces) is in the role of "producer". There is also a probability that the product quality violating the upper specification limit ($c_{true} > c_{u.s.l}$) will be falsely accepted as conforming, since $c_{test} \leq c_{u.s.l}$. This is the global consumer's risk R_c, and a patient is in the role of "consumer".

As in the previous examples, the global producer's and the consumer's risks were estimated by eqs. 3 and 4. However, the global c_{true} pdf was modeled in this case by the Student's pdf with location parameter $\hat{c}_{test}$ and scale parameter s_c calculated from the regression data $\hat{c}_{test}(\tau)$ at corresponding time τ:

$$f\left(c_{true}\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{s_c\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{17}$$

where Γ is the Gamma function, $t = (c_{test} - c_{true})/s_c$ is the normalized variable c_{test}, and $c_{true} = \hat{c}_{test}$. The likelihood function is here a normal pdf of measurement/test results c_{test} for a product with the true value of the measured property c_{true}, mean value of the test results equal to each c_{true} which can be assumed by the c_{true} pdf (in practice to each $c_{true} = \hat{c}_{test}$) and corresponding standard deviation $s = u(c_{test})$:

$$f(c_{test}|c_{true}) = \frac{1}{s\sqrt{2\pi}}\exp\left[-\frac{(c_{test} - c_{true})^2}{2s^2}\right] \tag{18}$$

R_c and R_p values vs. $c_{u.a.l}$ for sodium chloride injection assay and the c_{true} Student's pdf with $\nu = 16$ degrees of freedom, referring to time $\tau_0 = 2.7$ a, are displayed in Fig. 8 by solid lines 1 and 2, respectively. The upper specification limit is presented by a vertical dotted line. The range of OOS test results is shown by a horizontal dotted pointer. Acceptance limits for the levels of confidence $P = 0.95$ to 0.99 are indicated by grey bars 3 and 4, similar to bars 6 and 9 in Fig. 6, respectively. These limits can be interpreted and used respectively as warning and action lines in quality control charts. When $c_{u.a.l} = c_{u.s.l}$, R_c and R_p are equal to 0.015 and 0.067, respectively. This means that the violation of the upper specification limit may be not determined in 15 cases of the testing from 1000, while the sodium chloride injection may be mistakenly found as not acceptable for use in 67 cases of the product testing from 1000.

When the level of confidence $P = 0.95$ is chosen, for example, and the acceptance limit is equal to the warning line $c_{w.l} = 104.2$ %, the risks are $R_c = 0.0004$ and $R_p = 0.460$. Acceptance limit equal to the action line $c_{a.l} = 105.8$ % leads to R_c and R_p of 0.043 and 0.001, respectively. Different risk values R_c and R_p correspond also to every product storage time τ, since c_{true} and c_{test} change with τ, as discussed above.

The dependences of R_c and R_p on τ for the sodium chloride injection are displayed in Fig. 9 by solid lines 1 and 2, respectively, where the R_c and R_p values were calculated with respect to an acceptance limit equal, for each τ, to the relevant one-sided upper 0.95 confidence limit to the regression line.

Time corresponding to the upper specification limit, $\tau_{u.s.l} = \tau_0 = 2.7$ a, is shown by vertical dotted line, while the range of time values τ_{OOS} leading to OOS test results is indicated by horizontal dotted pointer. Grey bars 3 and 4 present the time acceptance limits for the levels of confidence $P = 0.95$ to 0.99 (time warning and action lines, respectively) corresponding to bars 6 and 9 in Fig. 6. It is clear that setting the shelf life $\tau_0 > 2.7$ a increases significantly the consumer's risk R_c.

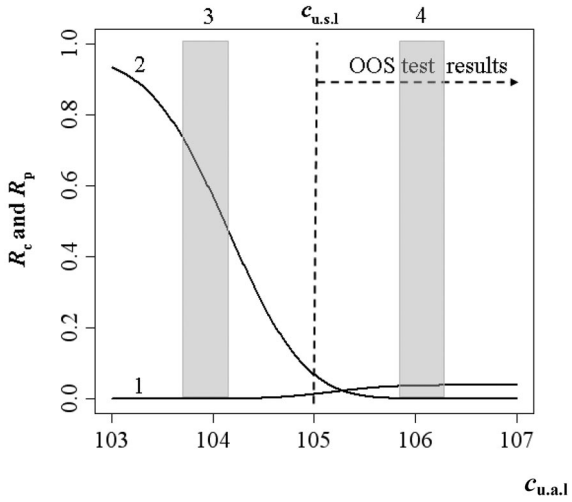

Fig. 8 Global risks of consumer R_c and of producer R_p of the sodium chloride injection vs. upper acceptance limit $c_{u.a.l}$/% for test results. R_c and R_p are displayed by solid lines 1 and 2, respectively; $c_{u.s.l}$ is presented by the vertical dotted line. The range of OOS test results is shown by the horizontal dotted pointer. Acceptance limits for the levels of confidence P = 0.95 to 0.99 are indicated by grey bars 3 (warning lines) and 4 (action lines). Reproduced from ref. [37] by permission of Springer.

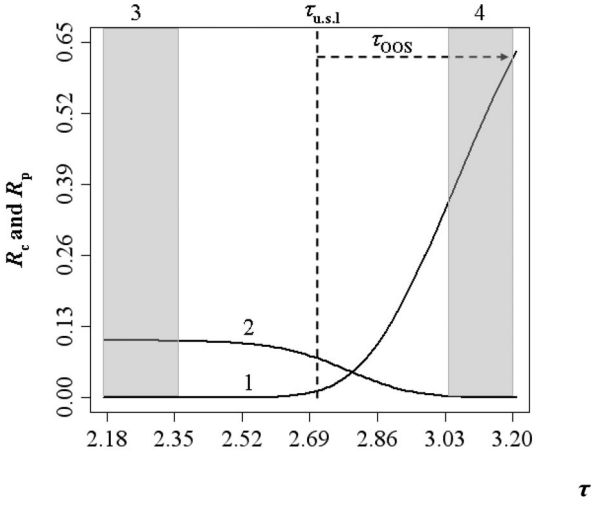

Fig. 9 Global risks of consumer R_c and producer R_p of the sodium chloride injection vs. storage time τ/a. R_c and R_p are displayed by solid lines 1 and 2, respectively. Time $\tau_{u.s.l}$ corresponding to $c_{u.s.l}$ is shown by the vertical dotted line, while the range of time values τ_{OOS} led to OOS test results indicated by the horizontal dotted pointer. Grey bars 3 (warning lines) and 4 (action lines) present the time acceptance limits for the levels of confidence P = 0.95 to 0.99. Reproduced from ref. [37] by permission of Springer.

B-3-4-2 When a measured attribute decreases with time

When a decision is to be made about the shelf life of a product whose measured attribute decreases with time, like the epinephrine injection assay, the risks can be estimated by eqs. 5 and 6. Results of calculation of R_c and R_p vs. lower acceptance limit $c_{l.a.l}$ for epinephrine injection assay and the global c_{true} Student's pdf with ν = 91 degrees of freedom, relevant to time τ_0 = 1.5 a, are shown in Fig. 10. The

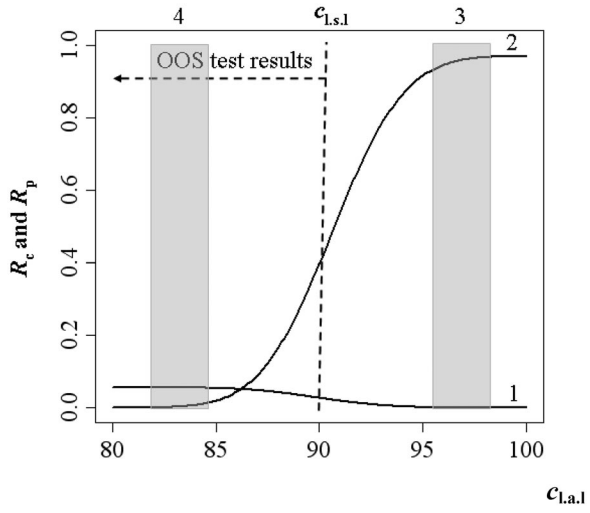

Fig. 10 Global risks of consumer R_c and producer R_p of the epinephrine injection vs. lower acceptance limit $c_{l.a.l}$/% for test results. Symbols and signs are the same as in Fig. 8. Reproduced from ref. [37] by permission of Springer.

pointer of OOS test results in Fig. 10 is directed in the opposite verse to the corresponding in Fig. 8: for the epinephrine injection, the direction is $c_{test} < c_{l.s.l}$, whereas for the sodium chloride injection it is $c_{test} > c_{u.s.l}$. Corresponding consumer's risk (solid line 1) decreases for the epinephrine injection in Fig. 10 when the acceptance limit increases, in contrast to the situation with the sodium chloride injection in Fig. 8. A similar difference in the behavior of the producer's risks (solid lines 2) is also observed.

The risk values vs. τ are demonstrated in Fig. 11 with respect to an acceptance limit equal for each τ to the corresponding value of the one-sided lower 0.95 confidence limit to the regression line.

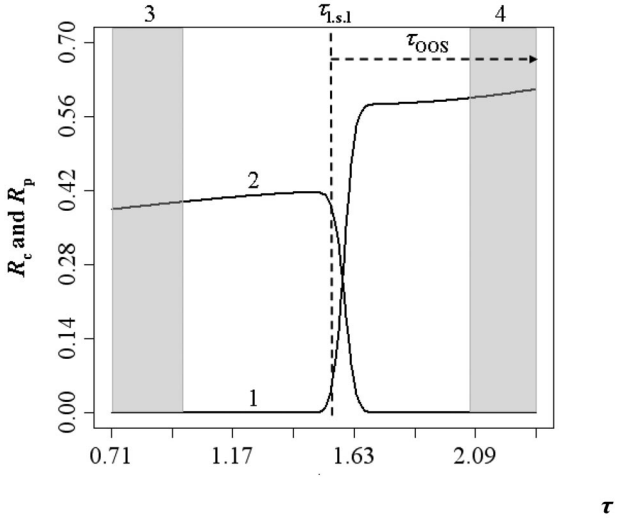

Fig. 11 Global risks of consumer R_c and producer R_p of the epinephrine injection vs. storage time τ/a. Symbols and signs are the same as in Fig. 9. Reproduced from ref. [37] by permission of Springer.

The same conventional signs for lines and pointers are used as in Figs. 8 and 9, respectively. Grey bars 3 and 4 present the time acceptance limits for the levels of confidence $P = 0.95$ to 0.99 (time warning and action lines, respectively) corresponding to bars 6 and 9 in Fig. 7. The direction of the time τ_{OOS} of OOS test results in both Figs. 9 and 11 is the same, since any product varies with time, not depending on the kind of change. In spite of different details of the dependences of risks on time for the two products, they are similar. In particular, the consumer's risk R_c sharply increases and the producer's risk R_p decreases after the shelf life τ_0 corresponding to the specification limits $\tau_{u.s.l}$ in Fig. 9 and $\tau_{l.s.l}$ in Fig. 11.

EXAMPLE 4. OOS RESULTS OF CETIRIZINE DIHYDROCHLORIDE ASSAY

B-4-1 Introduction

The objective of this example was an application of the metrological approach in the pharmaceutical field for investigating OOS test results of cetirizine, i.e., (±)-(2-{4-[(4-chlorophenyl)phenylmethyl]-1-piperazinyl}ethoxy)ethanoic acid, dihydrochloride assay, when both the lower and upper specification limits should be taken into account simultaneously.

EP sets the lower specification limit $c_{l.s.l}$ = 99.0 % and the upper specification limit $c_{u.s.l}$ = 100.5 % of cetirizine dihydrochloride content in bulk material (dried substance). The test/assay method is the acid–base potentiometric titration of acetone-water solution of the analyte with sodium hydroxide to the second point of inflexion [43]. A test result c_{test} is acceptable when it is in the specification limits, i.e., when $c_{l.s.l} \le c_{test} \le c_{u.s.l}$. OOS test results may appear when the true content c_{true} of cetirizine dihydrochloride is really less than the lower specification limit ($c_{true} < c_{l.s.l}$) because of impurities. However, OOS test results may be caused also by measurement problems and, for example, exceed the upper specification limit ($c_{test} > c_{u.s.l}$), whereas $c_{true} < c_{u.s.l}$.

As a case study, data described in ref. [44] are discussed here for investigation of OOS test results and evaluation of global producer's and consumer's risks.

B-4-2 Experimental

A total of 114 assay results c_{test} in the range from 98.7 to 101.2 % were obtained during a year by Chemagis Ltd., Israel, according to the EP titration method [43] with automated titrators. The results mean was μ = 99.7 %, and the standard deviation was σ = 0.4 %.

The standard measurement/assay uncertainty $u(c_{test})$ was evaluated as 0.2 %. The expanded uncertainty was $U(c_{test})$ = 0.4 to 0.6 % for normal distribution and the range of the levels of confidence P = 0.95 to 0.99, with the coverage factor k = 2 to 3.

B-4-3 Global distribution

The global normal c_{true} distribution modeling the empirical batch-to-batch c_{test} distribution is shown in Fig. 12 by a solid line.

The c_{true} pdf was approximated by

$$f(c_{true}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(c_{true} - \mu)^2}{2\sigma^2}\right] = \frac{1}{0.4\sqrt{2\pi}} \exp\left[-\frac{(c_{true} - 99.7)^2}{2 \cdot 0.4^2}\right] \tag{19}$$

The measurement distribution of c_{test} at one and the same c_{true} value was modeled by normal also distribution with standard deviation $u(c_{test})$ = 0.2 % and a mean equal to any relevant c_{true}. Thus, the likelihood function was approximated by

$$f(c_{\text{test}}|c_{\text{true}}) = \frac{1}{u(c_{\text{test}})\sqrt{2\pi}} \exp\left\{-\frac{(c_{\text{test}} - c_{\text{true}})}{2[u(c_{\text{test}})]^2}\right\} = \frac{1}{0.2\sqrt{2\pi}} \exp\left\{-\frac{(c_{\text{test}} - c_{\text{true}})}{2 \cdot 0.2^2}\right\} \quad (20)$$

For example, such distributions are shown in Fig. 12 by dashed lines at both the lower and the upper specification limits, when $c_{\text{true}} = c_{\text{l.s.l}} = 99.0$ % and $c_{\text{true}} = c_{\text{u.s.l}} = 100.5$ % (vertical dotted lines). Ranges of OOS test results are shown by pointers.

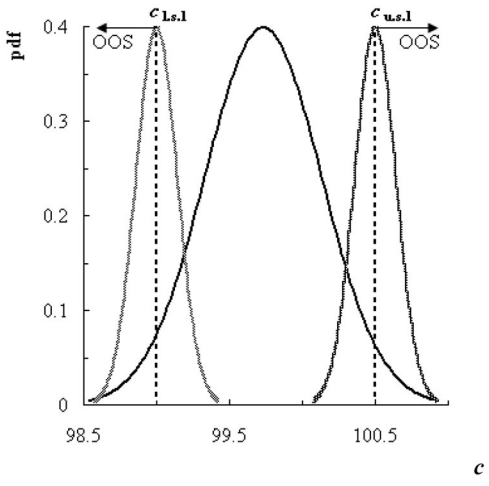

Fig. 12 Normal distributions modeling global c_{true} distribution (solid line) and measurement distributions (dashed lines) at the lower and upper specification limits $c_{\text{l.s.l}}$ and $c_{\text{u.s.l}}$, respectively (vertical dotted lines); c/% is the analyte concentration. Ranges of OOS test results are shown by pointers. Reproduced from ref. [44] by permission of Elsevier.

B-4-4 Causes and probability of OOS test results

Four OOS test results ($n = 4$), presented in Table 5, out of total $N = 114$ results were obtained during the year. Their deviations from the lower specification limit $D_{\text{OOS}} = c_{\text{l.s.l}} - c_{\text{test}}$ and from the upper specification limit $D_{\text{OOS}} = c_{\text{test}} - c_{\text{u.s.l}}$, and answers to the question "is the OOS test result metrologically related?" are also presented in the table. The answer is negative for the level of confidence $P = 0.95$ when the deviation D_{OOS} exceeded the expanded measurement uncertainty $U(c_{\text{test}}) = 0.4$ %.

Table 5 OOS test results and their deviation D_{OOS} from the upper and lower specification limits.

Batch	OOS test result/%	Specification limit/%	D_{OOS}/%	Metrologically related?*	
				$P = 0.95$	$P = 0.99$
1	101.2	100.5	0.7	No	No
2	101.1	1005	0.6	No	Maybe
3	101.0	100.5	0.5	No	Maybe
4	98.7	99.0	0.3	Maybe	Maybe

*"No" is for $P = 0.95$ when $D_{\text{OOS}} > 0.4$ %, and for $P = 0.99$ when $D_{\text{OOS}} > 0.6$ %.

Thereby, the OOS test result obtained for batch 1 was not metrologically related, for batch 4 it may be related, whereas OOS test results obtained for batches 2 and 3 were recognized as metrologi-

cally related when $P = 0.99$ was taken into account. For $P = 0.99$, the same answer is correct when D_{OOS} exceeded the expanded measurement uncertainty $U(c_{test}) = 0.6$ %.

The probability of OOS test results P_{OOS} by eqs. 2 and 19 was

$$P_{OOOS} = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_{c_{l.s.l}}^{c_{u.s.l}} \exp\left[-\frac{(c_{true} - \mu)^2}{2\sigma^2}\right] dc_{true} = 1 - \frac{1}{0.4\sqrt{2\pi}} \int_{99.0}^{100.5} \exp\left[-\frac{(c_{true} - 99.7)^2}{2 \cdot 0.4^2}\right] dc_{true} = 0.06 \quad (21)$$

This probability is a little larger than the frequency $F = n/N = 0.04$ of the observed OOS test results shown in Table 5, as in Examples 1 and 2 above. Such situation is caused by the fact that OOS test results were far from zero and infinity (the integration limits): in practice, the range of the obtained cetirizine dihydrochloride assay/test results (including OOS test results) was from 98.7 to 101.2 %.

B-4-5 Risks of producer and consumer

The global risk R_p of the cetirizine dichloride producer and the global risk R_c of its consumer were evaluated by eqs. 7 and 8 with the global pdf $f(c_{true})$ by eq. 19 and the likelihood function $f(c_{test}|c_{true})$ by eq. 20. The results of R_c calculation for different acceptance limit values are displayed in Fig. 13 by solid line 1. The results of R_p calculation are shown by solid line 2. The lower and upper specification limits, $c_{l.s.l}$ and $c_{u.s.l}$, respectively, are indicated by dotted lines. Acceptance limits in the range of the levels of confidence $P = 0.95$ to 0.99 (warning and action lines, $c_{w.l}$ and $c_{a.l}$, respectively) are demonstrated by grey bars. The warning lines were calculated as $c_{w.l} = c_{l.s.l} + U(c_{test})$ and $c_{w.l} = c_{u.s.l} - U(c_{test})$. The action line were calculated in the similar way as $c_{a.l} = c_{l.s.l} - U(c_{test})$ and $c_{a.l} = c_{u.s.l} + U(c_{test})$.

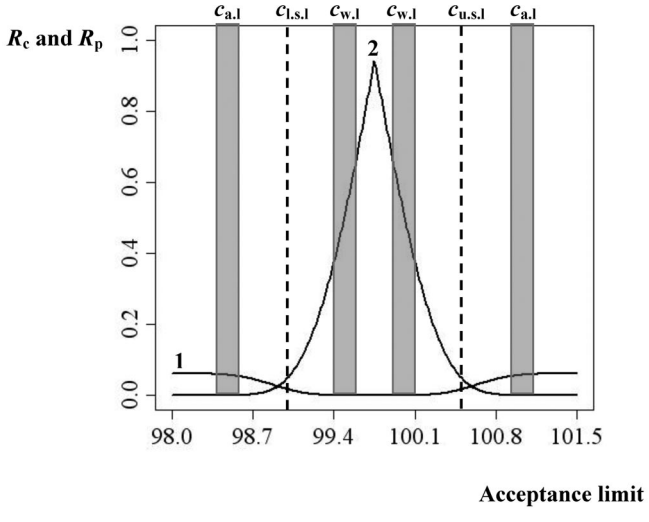

Fig. 13 Global consumer's risk R_c and producer's risk R_p vs. lower and upper acceptance limit values (%). R_c is displayed by solid line 1, and R_p by solid line 2. The specification limits are indicated by dotted lines. Acceptance limits in the range of the levels of confidence $P = 0.95$ to 0.99 (warning and action lines) are demonstrated by grey bars.

The plot is practically symmetric with respect to the mean of the specification limits: $(c_{l.s.l} + c_{u.s.l})/2 = 99.75$ %. Definitely, the producer's risk R_p of rejecting the null hypothesis about satisfactory quality of a batch with such an assay result is maximal: the probability that this decision is false achieves 0.937. The maximum consumer's risk R_c is not more than 0.060 (at the acceptance limit of 98.4 %).

When the acceptance limit coincides with the lower specification limit, $R_c = 0.017$ and $R_p = 0.050$. The same is at the upper specification limit. It means that the false decision that the assay result corresponds to the quality requirements was made for 17 batches out of 1000, whereas the false decision about the assay result not satisfing the quality requirements was made for 50 batches out of 1000.

For correct understanding and interpretation of the risks R_p and R_c discussed above, it is also important that expenses of a cetirizine dihydrochloride producer from the false decisions on a batch quality, as well as possible consequences of these decisions in the drug production using cetirizine dihydrochloride as a raw material, were not taken into account.

MEMBERSHIP OF SPONSORING BODIES

Membership of the IUPAC Analytical Chemistry Division Committee for the period 2012–2013 is as follows:

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

Membership of the IUPAC Interdivisional Working Party on Harmonization of Quality Assurance for the period 2012–2013 is 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 2011–2013 is as follows:

***Chair*:** W. Louw (South Africa); ***Vice Chair*:** L. Samuel (New Zealand); ***Secretary*:** S. Wunderli (Switzerland); ***Past Chair*:** I. Kuselman; ***Members*:** C. Puglisi (Argentina); A. Squirrell (Australia); L. Besley (Australia); A. Fagjelj (Austria); W. Wegscheider (Austria); P. De Bièvre (Belgium); O. P. de Oliveira Jr. (Brazil); V. Poncano (Brazil); G. Massiff (Chile); Y. Yadong (China); M. Suchanek (Czech Republic); I. Leito (Estonia); T. Hirvi (Finland); I. Papadakis (Greece); C. M. Lau (China); P. K. Gupta (India); M. Walsh (Ireland); K. Chiba (Japan); H. Y. So (Korea); Y. M. Nakanishi (Mexico); V. Baranovskaya (Russia); Y. Karpov (Russia); C. Cherdchu (Thailand); R. Kaarls (Netherlands); S. L. R. Ellison (UK); M. Milton (UK); C. Burns (USA); V. Iyengar (USA); W. F. Koch (USA); W. May (USA); J. D. Messman (USA); D. W. Tholen (USA); P. S. Unger (USA); W. Wolf (USA).

ACKNOWLEDGMENTS

The Task Group would like to thank S. Shpitzer, P. Goldshlag, I. Schumacher, and A. Weisman (Israel) for their data used and help in preparation of Examples 1–4, respectively, in Annex B of the Guide; Springer and Elsevier for permission to use material from the published papers cited in the Guide.

REFERENCES

1. U.S. FDA. *Guidance for Industry. Investigating Out-of-Specification (OOS) Test Results for Pharmaceutical Production*, U.S. Food and Drug Administration, Rockville, MD (2006).
2. S. Kuwahara. *BioPharm Int.* Nov. 1, 1 (2007).
3. A. M. Hoinowski, S. Motola, R. J. Davis, J. V. McArdle. *Pharm. Technol.* Jan., 40 (2002).
4. European Medicines Agency. *Quality Risk Management* (ICH Q9) (2005).
5. European Medicines Agency. *Pharmaceutical Quality System* (ICH Q10) (2008).
6. European Commission. *Report on the Relationship between Analytical Results, Measurement Uncertainty, Recovery Factors and the Provisions of EU Food and Feed Legislation* (2004). <http://ec.europa.eu/food/food/chemicalsafety/contaminants/report-sampling_analysis_2004_en.pdf>
7. EURACHEM/CITAC Guide. *Use of Uncertainty Information in Compliance Assessment* (2007).
8. ILAC G8. *Guidance on the Reporting of Compliance with Specification*, International Laboratory Accreditation Cooperation, Sydney (2009).
9. JCGM 106 Guide. *Evaluation of Measurement Data – The Role of Measurement Uncertainty in Conformity Assessment* (2012). <http://www.bipm.org/en/publications/guides/gum.html>
10. I. Kuselman, F. Pennecchi, C. Burns, A. Fajgelj, P. de Zorzi. *Accred. Qual. Assur.* **15**, 283 (2010).
11. BIPM. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (*VIM*), 3rd ed., Bureau International des Poids et Mesures, Geneva (2012). <http://www.bipm.org/en/publications/guides/vim.html>
12. ISO/IEC 3534. *Statistics – Vocabulary and Symbols – Part 1: General Statistical Terms and Terms Used in Probability*, International Organization for Standardization, Geneva (2006).
13. ISO 17000. *Conformity Assessment – Vocabulary and General Principles*, International Organization for Standardization, Geneva (2004).
14. U.S. FDA. *Guidance for Industry. Process Validation: General Principles and Practices*, U.S. Food and Drug Administration, Rockville, MD (2011).
15. ICH Q2(R1). *Validation of Analytical Procedures: Text and Methodology*, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2005).
16. L. Huber. *Validation and Quantification in Analytical Laboratories*, Interpharm Press, Buffalo Grove, IL (1999).
17. EURACHEM Guide. *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics* (1998).
18. P. De Bièvre, H. Günzler (Eds.). *Validation in Chemical Measurement*, Springer, Berlin (2005).
19. M. M. W. B. Hendriks, J. H. de Boer, A. K. Smilde (Eds.). *Robustness of Analytical Chemical Methods and Pharmaceutical Technological Products*, Elsevier, Amsterdam (1996).
20. ISO 21748. *Guidance for the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation*, International Organization for Standardization, Geneva (2010).
21. B. Magnusson, T. Näykki, H. Hovind, M. Kryssel. *Handbook for Calculation of Measurement Uncertainty in Environmental Analysis*, NORDTEST Report TR 537, NORDTEST Tekniikantie 12, FIN-02150 Espoo, Finland (2003).
22. EURACHEM/CITAC Guide. *Quantifying Uncertainty in Analytical Measurement*, 3rd ed. (2012).
23. EURACHEM/CITAC Guide. *Measurement Uncertainty Arising from Sampling. A Guide to Methods and Approaches* (2007).
24. EURACHEM/CITAC Guide. *Traceability in Chemical Measurement. A Guide to Achieving Comparable Results in Chemical Measurement* (2003).
25. P. De Bièvre, R. Dybkaer, A. Fajgelj, D. B. Hibbert. *Pure Appl. Chem.* **83**, 1873 (2011).
26. P. De Bièvre, H. Günzler (Eds.). *Traceability in Chemical Measurement*, Springer, Berlin (2005).

27. H. J. Mittag, H. Rinne. *Statistical Methods of Quality Assurance*, pp. 119–150, Charman & Hall, London (1993).
28. R. B. D'Agostino, M. A. Stephens (Eds.). *Goodness-of-Fit Techniques*, Marcel Dekker, New York (1986).
29. I. Kuselman, S. Shpitzer, F. Pennecchi, C. Burns. *Air Qual. Athmos. Health* (2010). <http://dx.doi.org/10.1007/s11869-010-0103-6>
30. U.S. EPA. EPA Method IO-2.1. *Sampling of Ambient Air for Total Suspended Particulate Matter (SPM) and PM_{10} Using High Volume (HV) Sampler*, Cincinnati (1999). <http://www.epa.gov/ttnamti1/inorg.html>
31. U.S. EPA. EPA Method IO-2.4. *Calculations for Standard Volume*, Cincinnati (1999). <http://www.epa.gov/ttnamti1/inorg.html>
32. U.S. EPA. EPA Method IO-3.1. *Selection, Preparation and Extraction of Filter Material*, Cincinnati (1999). <http://www.epa.gov/ttnamti1/inorg.html>
33. I. Kuselman, P. Goldshlag, F. Pennecchi, C. Burns. *Accred. Qual. Assur.* **16**, 361 (2011).
34. SANCO Document No. 10684/2009. *Method Validation and Quality Control Procedures for Pesticide Residues Analysis in Food and Feed* (2009). <http://ec.europa.eu/food/plant/protection/resources/qualcontrol_en.pdf>
35. U.S. EPA. *Setting Tolerances for Pesticide Residues in Foods* (2009). <http://www.epa.gov/pesticides/factsheets/stprf.htm#tolerances/>
36. Codex Alimentarius Commission. *Recommended Methods of Sampling for the Determination of Pesticide Residues* (1993).
37. I. Kuselman, I. Schumacher, F. Pennecchi, C. Burns, A. Fajgelj, P. de Zorzi. *Accred. Qual. Assur.* **16**, 615 (2011).
38. ICH Q1E. *Evaluation for Stability Data* (2003).
39. USP 34. *Sodium Chloride Injection*, Vol. 3, p. 4242 (2011).
40. USP 34. *Epinephrine Injection*, Vol. 2, p. 2701 (2011).
41. D. Stepensky, M. Chorny, Z. Dabour, I. Schumacher. *J. Pharm. Sci.* **93/4**, 969 (2004).
42. EP 6. *Sodium Chloride*, Vol. 2, p. 2897 (2008).
43. EP 6. *Cetirizine Dihydrochloride*, Vol. 2, p. 3715 (2008)
44. A. Weisman, I. Kuselman. *Int. J. Pharm.* **221**, 159 (2001).