CHIMIA **2003**, *57*, No. 1/2

Chimia 57 (2003) 13–18

ISSN 0009–4293

Focal Point: Analytical Technology

October 15, 2002

Quality Assurance of Analytical Data: Measurement Uncertainty and Traceability in View of ISO 17025

Organized by: Dr. Bernhard Schreiber[a]* and Dr. Peter Radvila[b]*,
Division of Analytical Chemistry of the Swiss Chemical Society
Chairperson: Dr. Bernhard Schreiber,
Division of Analytical Chemistry of the Swiss Chemical Society

Definition, determination and consequences of analytical measurement uncertainty were discussed. After a thorough explanation of actual procedures to handle uncertainty, the focus was switched to practical applications in process control, the meaning of qualitative analysis near the limit of detection, a case study involving doping in sports events and the problems involved in setting legal limits (*e.g.* ILAC rules). Five experts from academia, industry, and governmental organizations as well as from R&D-driven companies gave their interpretation from different points of view.

Keywords: Analytical data · ILAC · Setting limits · Traceability · Uncertainty

The Relationship between Traceability and Measurement Uncertainty

Wolfhard Wegscheider
Institute for General and Analytical Chemistry, University of Leoben, A-8700-Leoben, Austria
E-Mail: wegschei@unileoben.ac.at

Although the requirement of traceability is undisputed in principle, the actual realization of traceability still constitutes problems at times. While the traceability for values of standards is mostly in the hands of the producers, the traceability of results is an everyday problem for the analytical laboratory. Although at first sight these two instances seem to be different sides of a coin it will be shown that actually they constitute an identical case, as no standard can be produced without due consideration of the measurement leading to the value of the standard. It is therefore warranted to concentrate without loss in generality on the traceability of results as the more immediate situation faced by any laboratory.

Measurement uncertainty formally comes into the picture when the official definition of traceability is considered: all links to 'higher' standards need to be associated with a statement of uncertainty and so do those established in our own laboratories when we perform measurements and 'produce data'. To quantify the strength of this last link to real world samples it is necessary to employ standards. Thus, any measurement involves a quantitative comparison with a standard, so that the result is expressed in multiples of this standard.

Accordingly, in order to claim traceability of our results it is not only required that these standard values are accompanied by an uncertainty statement, but also that our daily practice resulting in extra components of uncertainty is well enough characterized for obtaining reliable estimates. The latter contributions are generally large compared to the uncertainties of standard values. It is therefore fair to claim an enhanced role of validation of analytical procedures as this is usually the only route to reliable estimates of the uncertainty.

Apart from the chemical standards the role of physical standards in the laboratory was examined. In many cases the particular analytical protocol followed is devised in a way to keep the components of uncertainty from the physical standards negligibly small. This is either done by tight control of the permissible ranges (*e.g.* in ambient temperature) or by introducing adequate corrections for these physical influence quantities.

These principles were outlined on the basis of practical examples, part of which is taken from the upcoming CITAC/EURACHEM Guide on Traceability in Chemistry.

Correspondence: [a]Dr. B. Schreiber
Weidenstrasse 12
CH–4106 Therwil
Tel.: +41 61 721 45 87
Fax: +41 61 723 90 97
E-Mail: bschreiber@mus.ch
[b]Dr. P. Radvila
Rösslistrasse 27
CH–9056 Gais
Tel./Fax: +41 71 793 28 82
E-Mail: p.radvila@bluewin.ch

CHIMIA **2003**, *57*, No. 1/2

Problems of Defining Uncertainty at the Detection Limit – A Case Study in Doping Control

Michael Weber
Metrology in Chemistry,
EMPA, St. Gallen
E-Mail: michael.weber@empa.ch

Newspaper reports of athletes failing drug tests are on the increase [1]. However, the results of the chemical analyses on which decisions are based are not always as clear-cut as they seem. There are also some questionable aspects in connection with the limits set. This situation can be illustrated by an actual case.

According to ISO 17025 all testing laboratories must be able to evaluate measurement uncertainties for all measurement results [2]. The degree of rigor needed in an estimation of uncertainty of measurement results depends on the consequences this result may implicate. In ISO 17025 the reporting of measurement uncertainties is explicitly required when there are narrow limits on which decisions on conformance to a specification are based. Today documents are available to help the laboratory staff to evaluate measurement uncertainties [3][4]. However, very often no statement of uncertainty is given in analysis reports for the values determined.

When measurement results are compared to a limit value the reported uncertainty of the result can play a decisive role for the interpretation of the measurement result. Fig. 1 shows three different cases which are all easy to understand. Case *a* is clearly below the limit whereas case *b* exceeds the limit. In case *c* the result is equal to the limit and therefore no statement can be made because the result encompasses the limit value. In Fig. 2 both cases *d* and *e* have the same values and both are above the limit. Here the uncertainty plays a decisive role for the interpretation whether the limit value is exceeded or not. In case *e* we cannot be 100% sure that the reported value is above the limit. But guilt must be proven without doubt. In case *e* therefore the accused should be cleared of suspicion.

Of course this is usually done in practice. For instance, when speed checks are carried out in road traffic a tolerance of 3 to 5 km/h is always subtracted from the measured speed value. A tolerance is also deducted when the amount of alcohol in the blood is checked. The driver can therefore be certain that he will not be punished because of an uncertain measurement. But this type of reduction does not exist in doping analyses, even though in these cases the uncertainty is considerably larger than, for instance, in checks made on alcohol in the blood.

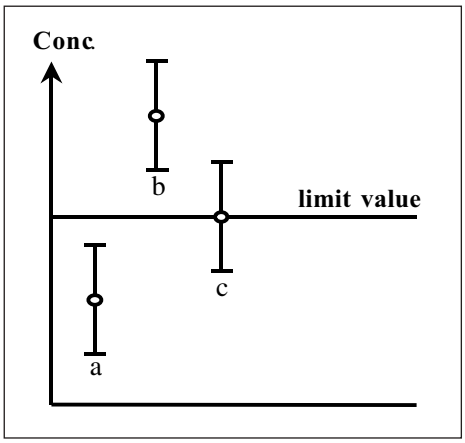

Fig. 1. Values and their uncertainties

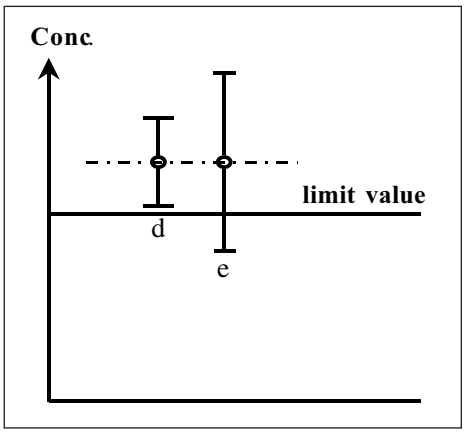

Fig. 2. Values and their uncertainties in relation to the limit value

When measurement uncertainties are reported, in most cases a standard deviation *s* is used to represent the variability of a test result. For a confidence of 99.7% statistically an interval of three times the standard deviation (+/– 3*s*) is necessary. This procedure might be adequate in many cases. But we should not forget that considering 3*s* as a tolerance still leads to 3 out of 1000 false positive observations. In addition it is often forgotten that the standard deviation is based on several assumptions which are not fulfilled in daily practice. On the one hand the standard deviation is only valid for a large number of observations. In normal laboratory practice only few replicates are made for economical reasons. Furthermore the standard deviation is based on the assumption that all observed values follow a normal distribution. But especially in trace analysis asymmetric distributions of observed values are often found (no negative concentrations). On the basis of these arguments the simple use of a standard deviation as an estimation of a measurement uncertainty is regarded as too optimistic.

Interlaboratory comparisons can also give some information about measurement uncertainty. Because such comparisons are only a snapshot in time they are not representative for individual measurement results. Therefore this approach only gives a rough estimate of measurement uncertainties. On the basis of about 2000 different interlaboratory tests Horwitz showed that the uncertainty in chemical analyses is dependent in the first approximation only on the concentration of the substance being analyzed: the lower the concentration the higher the uncertainty [5]. This relationship is shown in Fig. 3. Following the Horwitz curve uncertainties of much more than 50% in the range of a few ng/ml are more the norm than the exception. This relationship has not yet been shown to be not valid [6].

The limit for nandrolone in urine permitted by the IOC is 2 ng/ml. In the case of the player in the Dutch national soccer squad Edgar Davids (Fig. 4) values of 2.3 and 2.6 ng/ml were measured, which are both only slightly above the set limit of 2 ng/ml. Davids was subsequently banned. The uncertainty calculated from the scatter of both measured values amounts to at least 26%. If this amount is subtracted from both measured values, the results produced are 1.7 and 1.9 ng/ml. And if the figures are corrected employing the Horwitz function both values even fall below 1 ng/ml. Irrespective of which method is used to esti-

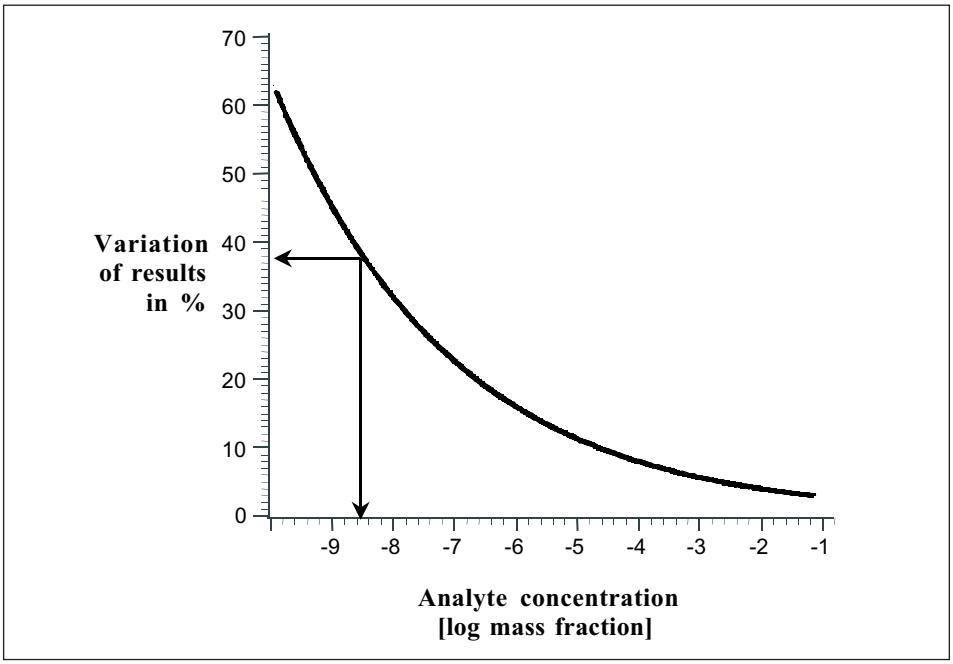

Fig. 3. Following the Horwitz function chemical measurements at the level of ng/ml normally lead to variations of *ca.* 40%. Therefore uncertainties are at least twice as high and can become up to 80% (at 95% confidence level).

mate the uncertainty, the corrected figures clearly fall below the limit value set. This means not guilty according to the principle «*In dubio pro reo*» – in the case of doubt, for the accused.

The situation is made even more difficult when one considers that on average the body itself produces approx. 0.6 ng/ml of nandrolone [7][8]. This concentration can increase several times during extreme sport performance. Other studies report of endogenous nandrolone production up to the IOC limit of 2 ng/ml. When these finding are taken into account, the IOC limit for nandrolone in urine of 2 ng/ml would in fact seem to have been set too low. Cyclists have a better situation. According to the cycle sport association the limit is 5 ng/ml. Therefore a cyclist with values similar to those of Edgar Davids would have not even been charged.

Finally we should accept that the declaration of uncertainties is a sign of security and not an indication of doubt. Whenever a measurement result is compared to a limit the uncertainty of the result must be taken into account. Furthermore, to avoid wrong decisions, uncertainties must be calculated according to established rules. Finally it is within the liability of the analyst to educate the executive bodies about these matters.

Fig. 4. Edgar Davids

[1] Actual information on the internet *http://www.sportgericht.de*
[2] ISO/IEC FDIS 17025 – General requirements for the competence of testing and calibration laboratories, April 20, **1999**.
[3] Guide to the Expression of Uncertainty in Measurement, issued by BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, revised **1995**.
[4] M. Rösslein, S. Ellison, A. Williams; EURACHEM/CITAC Guide 'Quantifying Uncertainty in Analytical Measurement', 2nd Edition **2000**, QUAM:2000.P1. Free download on the internet at *http://www.measurementuncertainty.org*
[5] RSD (%) = 2 exp[1 – 0.5 × log(c)] by R. Albert, W. Horwitz, *Anal. Chem.* **1997**, *69*, 789–790.
[6] M. Thompson, P.J. Lowthain, *J. AOAC Int.* **1997**, *80*, No. 3.
[7] B. Le Bizec, F. Monteau, I. Gaudin, F. André, *J. Chromatogr. B* **1999**, *688*, 157–172.
[8] G. Debruyckere *et al.*, Proc. of 4th Symposium on the Analysis of Steroids **1990,** 363–370.

Problems in Establishing Legal Limits Considering Measurement Uncertainty

Hans Peter Ischi
Schweiz. Akkreditierungsstelle Bern
E-Mail: hanspeter.ischi@metas.admin.ch

Introduction

The concept of measurement uncertainty is not new. The old standard EN 45001 already required the calculation of measurement uncertainties where relevant. The standard ISO/IEC 17025 goes beyond this and specifies that measurement uncertainty must be indicated each time two results are compared with each other or a result is compared with specifications.

Accredited laboratories are able to judge the quality of their methods with view to their use. They know the measurement uncertainty of results obtained. This uncertainty is an essential quality mark of validated test methods used. In principle, informed laboratory clients are interested in knowing the reliability of testing results received.

Definitions

According to the 'vocabulary of basic and general terms in metrology' (VIM) measurement uncertainty is an estimate characterizing the range of values within which the true value of a measurand lies. This estimate may be a standard deviation or any other indication suited to define a valid range of confidence.

Measurement Uncertainty

Basically, measurement uncertainty embraces all components of a test contributing to it. Measurement uncertainty therefore considers contributions of the whole testing process, also including the sample itself.

Testing results state the best approximation to the true value. Statistical and systematical factors nevertheless contribute to the measurement uncertainty of testing results. Systematical factors should, whenever possible, be eliminated, *e.g.* by appropriate correction factors.

The following factors may contribute to measurement uncertainty:
- The property of the sample
- The statistical sampling (AQL: acceptable quality level tables may be used)
- Transport, storage and handling of samples
- Preparation of samples
- Environmental conditions around the measuring desk
- Influence of the personnel executing tests
- Variations during the testing process
- Influence of the measuring instruments (the indication of measurement uncertainty given by calibration laboratories is very useful and often adequate!)
- Calibration standards and reference materials (the indication of measurement uncertainty given by calibration laboratories is very useful and often adequate!)
- Software
- The uncertainty of factors to correct systematical deviations

The Policy of EURACHEM, EUROLAB and EA

The three organizations consider the indication of measurement uncertainty where relevant (see par. 1) as good practice. They also state that, for the moment, measurement uncertainty should primarily be indicated for quantitative testing results. Reflections on the estimation of measurement uncertainty for qualitative statements should be made later on.

EUROLAB, EURACHEM und EA agree that:
- The indication of the measurement uncertainty should meet the purpose, *e.g.* it should give enough information to allow comparison with requirements set;
- The GUM (Guide for the estimation of Uncertainty in Measurements) is the basis;
- Available knowledge shall be included as much as possible. This means that only a minimum of supplementary investigation should be done, and experimentally evaluated data, as findings based on control cards, results of method validation, direct-comparisons, proficiency tests or resulting from the use of (certified) reference materials, should be used;
- How far the evaluation of measurement uncertainty should go depends on the technical field of application;
- In some cases an indication concerning reproducibility is enough.

EA is producing guidelines in order to help accredited laboratories implementing the criteria given by the standards in the frame of this policy.

Customers' Advantages from Uncertainty Estimation by Testing Laboratories

Laboratories will experience that several advantages are linked with a statement of measurement uncertainty:
- Customers such as governmental authorities, product certification bodies, industries and others need information to establish correct statements of compliance with specifications. ILAC provides guidance on reporting on compliance with specifications.
- The statement of a measurement uncertainty can be considered as an added value to customers, helping them to interpret results correctly.
- The knowledge how to calculate measurement uncertainty is crucial to optimize test procedures for a given purpose (*e.g.* in the frame of contract review) and to plan proficiency tests.
- Measurement uncertainty is helpful when selecting the most appropriate testing procedures, optimized to both economic and quality aspects.

Recommendations for the Definition of Specifications

It is important that customers develop a certain knowledge of the meaning of measurement uncertainty and of how to use it. Laboratory organizations are recommended to develop some guidance on how to read and to use uncertainty statements in testing reports.

Specifications should be established in a way to allow the application of the ILAC principles mentioned above. This means that quantitative specifications should be established at a level that allows products tested with specified methods to fulfill them, and estimated to be in conformity with a certain probability.

EA and its partners Eurolab and Eurachem, asked ISO and CEN to consider this issue when producing new standards. Both organizations gave positive answers.

Uncertainty in Process Control: The Limit of Setting Limits

Bernd Renger
Baxter AG, Industriestrasse 67,
1221 Wien, Austria
E-Mail: bernd_renger@baxter.com

The process of setting limits must be considered one of the most interesting aspects of analytical testing and regulations in the pharmaceutical industry. In pharmaceutical analysis – as in many other regulated industries – no individual uncertainties of analytical steps are estimated because the entire procedure is subject to an extensive, highly formalized validation process prior to being submitted and approved.

But this approach underestimates the procedure's true long-term uncertainties. In addition, there is a common misconception that using the latest analytical technology will automatically lead to even 'better', *i.e.* more reliable results. However, analytical variability has remained nearly unchanged over the last decade, as the major sources of uncertainty such as sampling, sample pretreatment or sample injection steps are either not equipment related or independent of the measurement technique. Interestingly, reported uncertainties from analytical laboratory experience clearly indicate that the most precise techniques are still the unspecific ones like titration *etc.* (Table 1).

Table 1. Intermediate precision of selected analytical techniques [1–3]

Titration, potentiometric	~0.3%
Titration, visual	~0.5%
Gravimetry	~0.5%
UV/Vis	1.0–2.5%
HPLC, photometric with derivat.	1.0–2.0%
GC (direct)	1.5–2.5%
CE	1.5–2.5%
Quant. TLC (Scanner)	2.0–5.0%
GC (Headspace)	>3.0%
Fluorimetry	>3.5%
Microbiological Assays	>5.0%
Biological Assays	>8.0%

When defining and setting specifications, the main questions to be answered are
- Which product attributes and limits are critical to establish product quality?
- How can these attributes be assessed?
- What variance or uncertainty levels are acceptable?

Very often, there is inconsistency between the specified product limits and both the systematic and random variability of the analytical procedure. It is apparent that many chromatographic procedures, run as routine applications to monitor processes or to assess final products, are unlikely to be capable of controlling the expected tight specification limits.

As a consequence the European Pharmacopoeia has introduced the concept of maximum permitted standard deviation (RSD_{max}). The RSD_{max} is based on the upper specification limit of the described substance and on the number of replicate

injections made during system suitability testing. For a 101.0% upper limit – corresponding to an estimated analytical uncertainty of +/– 1.0% – the RSD_{max} must not exceed 0.42% for a series of six replicate injections of the reference solution [4].

A pragmatic approach to define meaningful specification limits taking into account both the analytical procedures inherent variability and the manufacturing process variability has recently been proposed: the 3σ approach. Assay data from all available batches produced by the same process are basis for the calculation of the mean and the 3σ limits that define the range that the process and the analytical procedure are capable to control (Fig. 5)

However, the resulting wide limits are very often subject to non approval by regulatory authorities or non acceptance by internal regulatory departments. A proven – yet incorrect – way to define tighter limits is to rely on smoother processes in later routine manufacturing and testing phases as well as on OOS procedures [5].

These problems increase in the case of biological products that require complex and error-prone biological testing. Assays and potency determinations may easily show relative standard deviations (RSDs) in precision (expressed as repeatability or intermediate precision) ranging from 8% up to 10% or more. As the RSD must not be mistakenly substituted for 'range' or 'confidence interval' it becomes apparent that in these cases it is hard to control even 'liberal' limits of +/– 20%, thus giving rise to problems especially when assay data are the basis of further production steps with subsequent testing.

Out of Specification procedures with their usually accepted common approach 'No out of Specification result in a series of six retests' do not really compensate for too tightly set specification limits. The graph of the probability as shown in Fig. 6 clearly indicates that gained improvement is small compared to the still existing probability of rejecting a good product of *ca.* 20%.

This situation can become even more complex in situations where different alternative analytical techniques may be used.

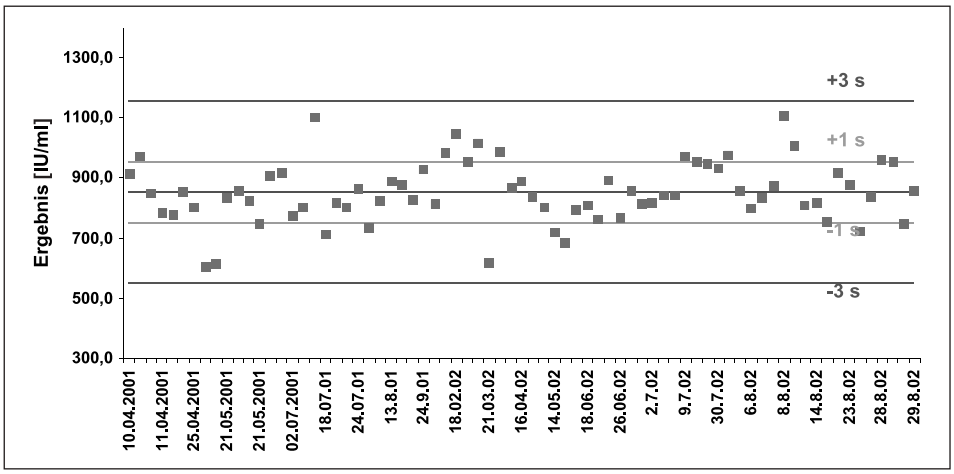

Fig. 5. 3s Approach – Potency of different lots of a biological material

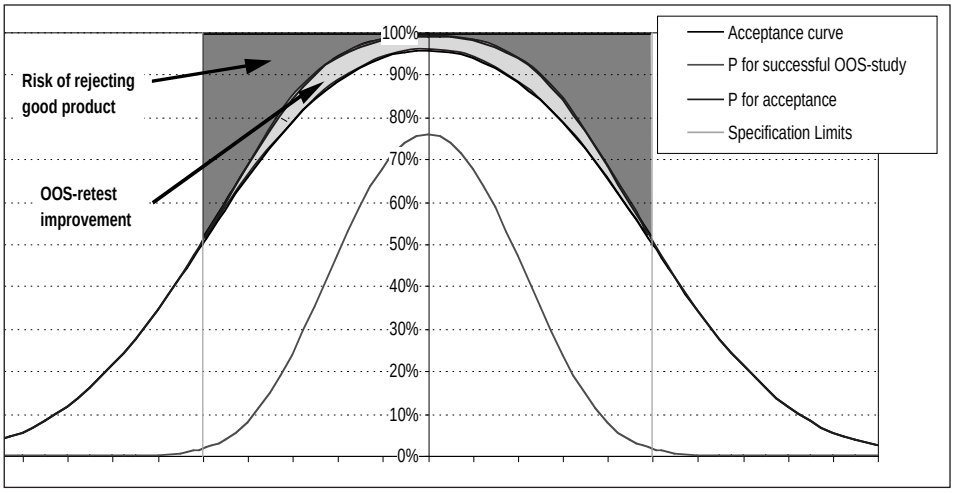

Fig. 6. Acceptance function. In the case of batches with assays near the specification limits, the probability of rejecting good product is not significantly reduced by a succesful OOS procedure.

Measurement Uncertainty in the Validation of Quantitative PCR Methods in Food Analysis

Jürg Rentsch
Swiss Quality Testing Services (SQTS),
Federal Cooperative of Migros,
CH–1784 Courtepin
E-Mail: juerg.rentsch@sqts.ch

Switzerland and the EC have adopted labeling regulations for foods and food ingredients derived from GM plants to guarantee consumers a choice between GM and non-GM products. The principle of substantial equivalence is decisive for GM food labeling in the United States, whereas the criterion for food labeling in the EC is the presence of proteins or DNA resulting from genetic modification. However, when the presence of GM material is adventitious and represents only a small amount, *e.g.* as a result of commingling during cultivation, harvesting, transport, or processing, labeling becomes non-informative for the consumer. Therefore, *de minimis* threshold values have been introduced to distinguish adventitious contamination of GM materials from food produced from GM material. In Switzerland and in the EC, the threshold value was set by the legislative body to 1% of GM material on the basis of ingredients.

Consequently, quantitative methods have been evolved mostly based on the polymerase chain reaction (QPCR) to target specifically and sensitively transgenic nucleic acid derived from various genetically modified crops.

Real-time quantitative PCR is a powerful and versatile gene quantification technique (Fig. 7). It uses threshold measurements during the exponential phase of amplicon production instead of end points. As the PCR proceeds (in the TaqMan 5' nuclease assay), fluorescence is released in direct proportion to the accumulation of PCR product. Through the use of a CCD camera, fluorescence production is continually monitored in each reaction well. The cycle number (Ct) at which the fluorescence reaches a threshold value is determined and used to calculate the amount of starting material by comparison to known standards.

[1] T. Layloff, *American Genomic/Proteomic Techn.* **2002**, Jan/Feb, 14–17.
[2] B. Renger, *J. Chromatography B*, **2000**, *745*, 167–176.
[3] W. Horwitz, *J. AOAC Int.* **1977**, *60*, 1355–1363.
[4] European Pharmacopoeia 2002 Monograph 2.2.46 'Chromatographic Separation Techniques', EDQM, Strasbourg, **2002**.
[5] Draft Guidance for Industry: 'Investigating Out Of Specification Test Results for Pharmaceutical Production', US FDA, CDER, Rockville MD, **1998**.

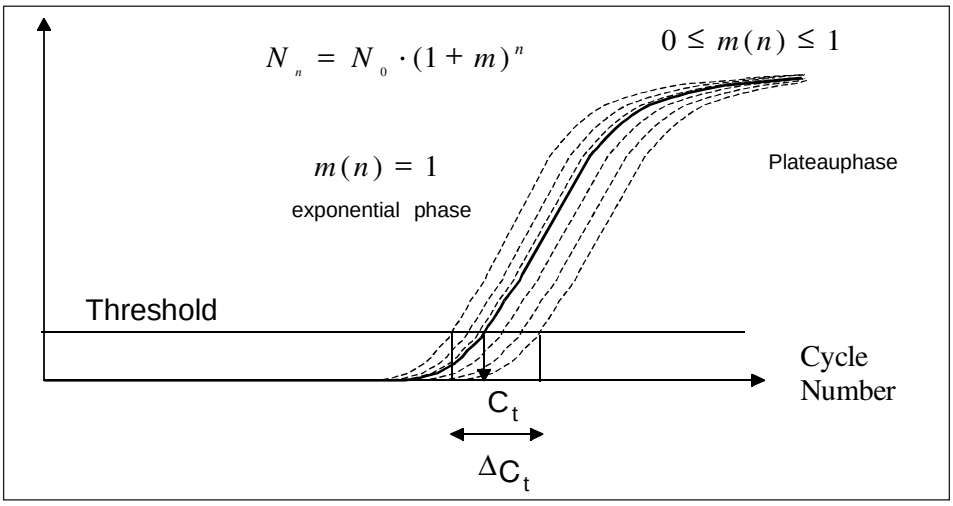

Fig. 7. During PCR amplification in the TaqMan 5' nuclease assay, fluorescence is released in direct proportion to the accumulation of PCR product. Fluorescence production (Rn) is continually monitored in each reaction well. The cycle number (Ct) at which the fluorescence reaches a threshold value is determined and used to calculate the amount of starting material by comparison to known standards. During the exponential growth stage the relationship of amplified PCR product to initial template amount can be recorded in the Eqn.: $N_n = N_0 \times (1+m)^n$, where N_n is the concentration of amplified product at cycle n, N_0 is the initial concentration of target template and m is the efficiency of PCR amplification.

Table 2. Influence of plant genome size on theoretical limit of quantitation.

Common name	Genome size[a] [in billion bp]	Genome copies [per 200 ng]	1%	0.1%	0.01%	Limit of quantitation[b]
corn	5.0	36000	360	36	4	0.1 %
potato	3.5	53000	530	53	5	0.07 %
rape	2.4	77000	770	77	8	0.05 %
rice	0.9	210000	2100	210	21	0.02 %
soybean	2.2	82000	820	82	8	0.04 %
tomato	1.9	96000	960	96	10	0.04 %
wheat	31.9	6000	60	6	–	0.6 %

[a] Published genome sizes (per 2C) were taken from Arumuganathan and Earle [2].
[b] The theoretical limit of quantitation is expressed as the fraction (in %) of 36 copies divided by the number of copies of the corresponding plant species within 200 ng DNA.

Table 3. Experimental determination of reproducibility for PCR-based GMO detection systems. The food samples (#A-D) were analyzed by four participant laboratories.

Sample ID		% Soybean-DNA lectin-DNA[a]	%GMO content 35S (norm)[b]	%GMO content RRS (norm)[c]
#A	mean	71	8.6	8.0
	std. dev.	13	1.4	1.8
	RSD [%]	19	17	23
#B	mean	69	11.4	12.1
	std. dev	11	3.0	3.3
	RSD [%]	15	26	28
#C	mean	75	0.33	0.30
	std. dev.	29	0.06	0.08
	RSD [%]	39	18	27
#D	mean	70	34	36
	std. dev.	7	9.4	13.1
	RSD [%]	10	28	36

[a] A taxon-specific real-time PCR targeting the soya-lectin-1 gene.
[b] GMO-screening by real-time PCR targeting the Cauliflower mosaic virus 35S-promoter element. The given values are normalized with the amount of soybean DNA present in the sample.
[c] A Roundup Ready® soybean specific real-time PCR. The given values are normalized with the amount of soybean DNA present in the sample.

The quantity of starting material, either RNA or DNA, is determined by comparison with known external standards. Absolute quantitation requires the use of carefully quantified external standards such as certified calibrators.

In the current political situation where most European food producers and distributors avoid the propagation of GM food, important decisions to meet legislative and trading requirements are based on results of quantitative PCR detection methods.

Recently, experiences from QPCR method validations have been accumulated on the part of enforcement, standardization and industrial bodies to establish and harmonize realistic performance characteristics. The limit of quantitation (LOQ) of GMO-specific, real-time PCR has been reported to reach 30–50 target molecules, which is close to theoretical prediction. Starting with template concentrations of 200 ng genomic plant DNA, the limit of quantitation depends primarily on the genome size of the target plant and ranges from 0.02% for rice to 0.7% for wheat [1] (Table 2).

Evaluations from proficiency studies provided measures of repeatability of quantitative PCR detection methods as expressed as a relative standard deviation (RSD) in the range of 10 to 30%.

Results of a small collaborative trial (method performance test) are compiled in Table 3 using commercial food-products containing soya ingredients that have been tested positive for RRS before without knowledge of the true GMO content.

Major sources of uncertainty arise from sampling and from matrix effects. Because quantitative GMO detection methods measure GMO contents of samples in relation to reference material, high priority must be given to international agreements and standardization on certified reference materials.

[1] P. Hübner, H.-U. Waiblinger, K. Pietsch, P. Brodmann, 'Validation of PCR Methods for Quantitation of Genetically Modified Plants in Food', *J. AOAC Int.* **2001**, *84*(6), 1855–1864.
[2] K. Arumuganathan, E.D. Earle, 'Nuclear DNA Content of Some Important Plant Species', *Plant Molecular Biology Reporter* **1991**, *9*(3), 211–215.