Taking a closer look.

relana®

c/o Lach & Bruns Partnerschaft
Tempowerkring 1
21079 Hamburg
Germany

Phone +49(0)40-79012-232/-233
relana@relana-online.de

www.relana-online.com

POSITION PAPER No. 19 - 03

Differing results of competent laboratories: reasons and what is "common"?

Version 2019/06/25

Abstract

Differences in results from different laboratories can be considered as normal.
A 50% expanded measurement uncertainty (expMU) of accredited laboratories with good ring test performances is an overall accepted uncertainty. Correct application of this expMU will allow a statistical and practical comparison of results.

However, although the laboratory is producing the analytical results, the analyses are only one part of the "entire" process. Primary steps like f. ex. "sampling" and "sample preparation" contribute to the total measurement uncertainty in a much more significant way than the analytical procedure itself.

Consequently, the performance of the sampling and sample preparation steps are crucial for reliable and comparable results. The sampling step is often not under control of the laboratories. Different analytical results must therefore not be per definition linked directly to the laboratories' performances.

1. Introduction

An important aspect taken into consideration in the safety of food products is related to analytical results, which require a high level of reliability and accuracy.
Differences of results performed at different laboratories can have several reasons.
Not all of these reasons are under the control of the involved analytical laboratories.
Possible reasons for deviating results as well as approaches of handling these results in a meaningful way are discussed in this position paper, focussing on analyses for pesticide residues in food products and related matrices.

2. Expectable ("normal") differences between results of "identical" samples

Although modern analyses for pesticides make use of sophisticated instruments and well-trained personnel, differences between lab results are unavoidable and must therefore be expected.
Assuming that identical samples (which do not exist in reality) are analysed, a statistical variance is expectable and plausible, following the application of statistical models.

2.1 Gauss distribution

The “Gauss” distribution (also called “normal” or “bell” distribution) shows the continuous probable distribution of results: If a same sample is analysed repeatedly for several hundred times (at least theoretically), the individual analytical results will deviate. The results measured follow the so-called Gauss distribution.

The distribution of several / many measurements (f. ex. 20 times measurement of orange homogenate with a spiked level of 20 ppb chlorpyrifos-ethyl) follows the Gauss distribution. And the larger the number of measurements is, the better approximation is:

From the Gauss distribution it can be deduced that all values are symmetrical around the value μ. In addition, the frequency of the values is greatest at the point μ. The frequency decreases evenly in both directions. The distance between the value of the largest frequency (= mean of the population = μ) and the inflection point (the point at which the curve changes from the convex to the concave form - and vice versa) is the standard deviation of the population σ.

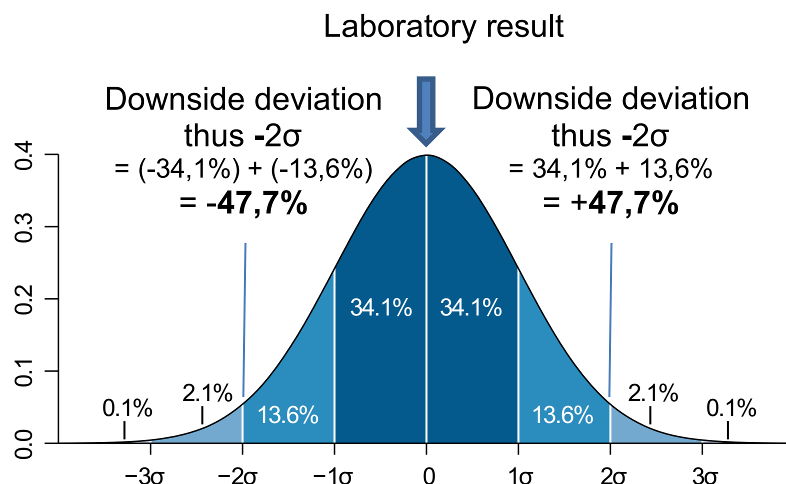

Figure 1. Gauss distribution of analytical results

95,4 % of all results (= confidence interval) are covered within the twofold standard deviation (± 2 σ).

Analytical labs strive to cover 95% confidence level. In multi-method pesticide residue analyses a deviation of +/- 50 % of the reported result is accepted [1, 2] when comparing results of different laboratories - see also following chapter.

2.2. SANTE 11813/2017 [3]

The SANTE document 11813/2017 states under point E10:
*"A **default expanded MU (measurement uncertainty) of 50 %** (corresponding to a 95 % confidence level and a coverage factor of 2) has been calculated from EU proficiency tests. In general, this 50 % value covers the inter-laboratory variability between the European laboratories and is recommended to be used by regulatory authorities in cases of enforcement decisions (MRL-exceedances)."* (MRL: maximum residue levels*)*

This expanded measurement uncertainty

- applies to each produced result,
- relates also to results between labs, having analysed the same (homogenous) sample,
- on condition that the involved labs are accredited, have acceptable performance in Proficiency Tests AND demonstrate that their internal expanded MU is less than 50%.
- is intended to be used ONLY in case of violations of maximum residue levels by regulatory authorities ("MRL-exceedances").

This default 50 % expanded measurement uncertainty is within the pesticide residue world generally accepted related to multi-residue-method-approaches. For pesticides covered by single residue methods and for several contaminants other maximum measurement uncertainties have to be applied. With regard to some contaminants, legal requirements exist. For these contaminants, each lab is obliged to publish its individual MU on the report.

2.3. Focus of quality management measures

It should be noted that most quality management (QM) measures such as validation, quality control cards or ring trials focus on the analytical performance, thereby neglecting the crucial steps of sampling, sample transport and sample preparation (homogenisation).

3. Reality check: Expectable differences between results of real samples

In reality, homogenates of the same sample ("identical samples" like f. ex. during ring tests with purées resp. powders or analysing retained homogenates of original sample material) are not often analysed.
The following laboratory routines are more common and realistic:

- The exporter takes a sample of (f. ex.) apples and orders an analysis in a lab of his choice.
- If the sample meets legal requirements and the specification, the sample is exported to the client.
- After receiving the apples, the client takes a new sample from the same lot and orders a residue analysis too, in a laboratory of his choice.

Such a procedure results in two samples which are taken from the same lot but can hardly be called identical. A few factors will affect the lack of comparability:

3.1. Influence of Sampling

The mentioned two samples cannot be called "identical", although they were taken from the same lot [4]:

- the respective samples are made up of different units (apples),
- the pesticides are likely to be distributed unequally across the entire lot, especially non-systemic pesticides. Even with the best sampling procedures, the samples will be different.
- Time gaps between dates of sampling (see also 3.2.).

Apples growing on the outer branches closer to the tractors spraying pesticides are likely to receive a higher pesticide load than those ones growing close to the stem or on the opposite side of the crown.

The **sampling procedure** may vary:

- number of sub-samples,
- scheme of sampling (z-scheme, start-middle-end, randomly picked samples, etc.),
- **lots are in general not homogeneous**, they may consist of products from different suppliers, different fields or plantations etc.,

⇒ **In "real-life lots" inhomogeneities are the rule, not the exception!**

3.2. Stability of analytes

Due to the time shift between both samplings, pesticide levels might have changed (especially labile pesticides are partly degraded or evaporated). On the other hand, some metabolites (as part of the residue definition) may show up more in the latest sample.

3.3. Sample transport and sample preparation

The **sample transport** to the lab is another crucial factor, especially the appropriate temperature control is important for the validity of the analysis result.

As discussed in the relana® sample preparation project (see relana® position paper No. 19-01), the normative pre-settings for **sample preparation** (choice of sample parts, degree of homogenisation etc.) are not very detailed, although any variation at this stage will lead to differing results, even if the analytical procedure is well comparable.

The effect of sampling and sample preparation is visualised in figure 2, showing the influence of sampling ("sampling") and sample preparation ("sample clean-up") on the variance of the final result – which is much higher than the influence of the analytical procedure ("instrumental analyses").

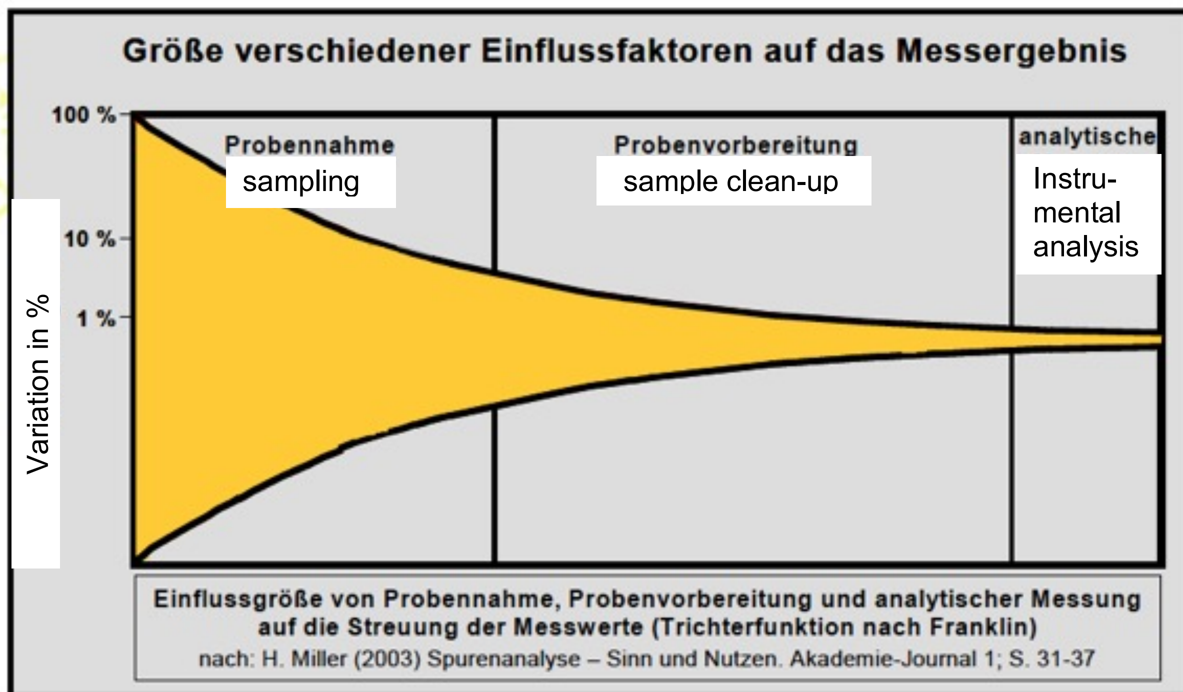

Figure 2. Influence of "sampling", "sample clean-up" and "instrumental analysis" on the variation of results
(For a closer discussion on the **influence of sampling**, please see relana® position paper no. 19-02)

3.3.1. Homogeneity of samples

Food and feed stuff and the samples derived from them might show a wide variance of homogeneity, with figure 3 showing some examples.

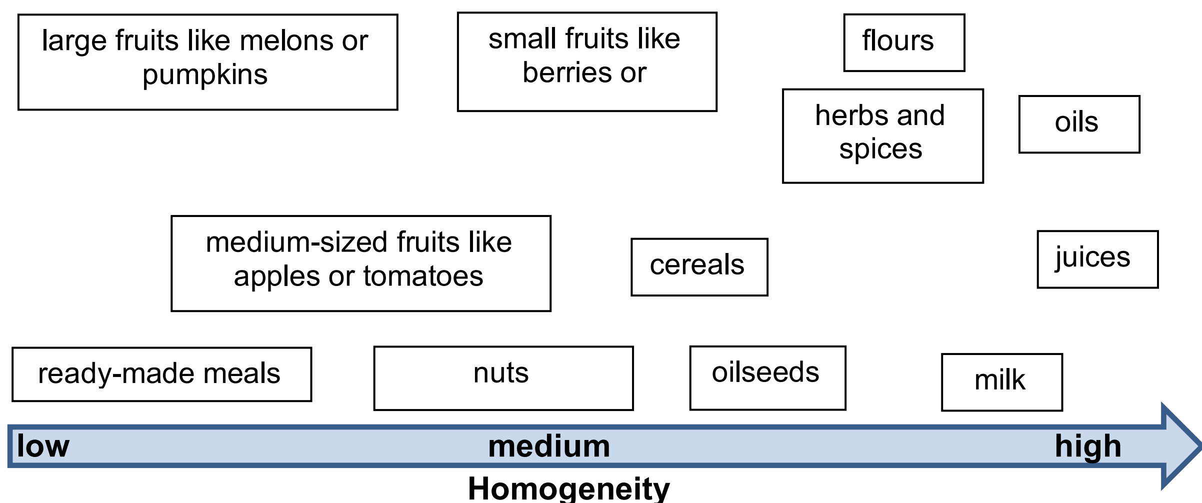

Figure 3. Classification of product according to their homogeneity

Generally speaking, samples are more homogeneous the lower the particle size is.
A lot also depends on the way of applying the pesticides. For contaminants other situations may occur (f.ex. mycotoxins resulting from post-harvest fungi infection may occur in so called hot-spots, where extremely high concentration may be present).

3.4. Analytical method compromises

The common approach to pesticide analyses is the application of a so-called "multi-method", which usually covers 400 to 600 compounds and metabolites and usually has to achieve a limit of quantification (LOQ) of 0.010 mg/kg for each analyte.
As the chemical properties of the analytes vary to a large extend, this analytical approach is a compromise [5] with limitations related to the sensitivity, the selectivity, the robustness and the entire scope.

A lot of factors contribute to the analytical variability. A few of them are listed here:

Loss during sample preparation including homogenisation
Some analytes are sensitive to higher temperatures, which means that they might get lost during homogenisation due to the induced friction energy leading to an increase in temperature. In case of sensitive analytes, it is recommended to add dry ice or liquid nitrogen to the sample in order to avoid an unfavourable temperature rise.

Other analytes are volatile and can get lost (partially or completely) during sample preparation and homogenisation.

Further aspects leading to a loss of substances may be pH-sensitivity, sticking to surfaces or to applied clean-up agents.

Insufficient homogenisation / impact on extraction efficiency
In case of systemic pesticides, it can be necessary to break up the cell walls in order to be able to extract the compounds locked in the cells. Therefore, a sufficient degree of homogenisation is crucial, before the extraction step is carried out.

Stability and quality of standard substances
Some pesticides are not stable, even in pure solutions, deep-frozen and under exclusion of light. A degradation of standard substances leads to an overestimation of pesticide concentrations.
In rare cases, even the standard materials supplied by acknowledged suppliers of standard substances prove to not contain the labelled purity (Omethoate, DDAC).
Variability of standard concentrations contributes to the variability of results (within and between labs).

Quantification methods
For the quantification of pesticides, several approaches are applied, particularly:

- external quantification with solvent standard solutions
- matrix-matched quantification (standard substances dissolved in matrix extracts)
- quantification via isotopically labelled internal standards (ILIS)
- standard addition.

All methods have their advantages and challenges. The choice of quantification method depends on the properties of the analyte and of the sample. As the listed quantification techniques influence the final result, differences between results may occur due to the applied quantification approach.

Measurement range
Pesticide methods usually achieve a limit of quantification (LOQ) of 0.010 mg/kg (10 ppb) for each one of about 500 substances.
This concentration is equal to 1 part in 100.000.000 parts, which must be quantifiable with a normal degree of confidence. On top of that, this low concentration needs to be analysed simultaneously for about 500 analytes (multi-methods).
When discussing the uncertainty of pesticide results, it should be kept in mind that the measured concentrations are quite low and require trained analysts and high-tech instruments.

Complex matrices
In case of very complex matrices (complex in terms of the physical-chemical properties of the nutrients of the matrix) such as spices, oleoresins, dried herbs, flavours, lecithin etc., the analytical challenge is significantly higher than for less complex products. Consequently, higher measurement uncertainties can be expected.

Properties of analytes
As the evaluation of ring trials shows [Bruns], another factor for the measurement uncertainty is directly linked to the properties of the analyte itself: stability during analyses, extractability, sticking to surfaces, matrix interaction, related instrumental sensitivity, etc.
Thus, some analytes can be “more complicated” to analyse than others, resulting in higher interlaboratory variabilities. Sometimes shifting an analyte from a multi residue method to a single residue method might be the only solution for a reliable analysis.

Analytical Scope
The scope of pesticides (in particular the scope of multi-methods) must not be equal between analytical laboratories. Although huge overlaps might exist, the scope may show some variations from lab to lab. The scope depends on the lab´s experience and its customers’ demand, covering legal requirements, known problems, etc.
As every analysis can only cover an excerpt from the number of compounds known worldwide, it is possible that a pesticide detected by one lab is not within the scope of another lab – or not part of the multi-method it applies, at least.

4. Comparability of results

In practise, clients are sometimes faced with two analytical results (of the same lot), which seem to be contradictory. As explained in part 3, differing results must be expected. These differences can be attributed to

- sampling,
- stability of analytes,
- sample transportation and sample preparation,
- analytical uncertainty.

4.1. Examples for the application of the measurement uncertainty

Under the assumption that two (almost) identical samples were analysed by two different competent labs, the following examples highlight the expectable variations:

Example 1:

Lab 1: 0,071 mg/kg Azoxystrobin

Lab 2: 0,037 mg/kg Azoxystrobin

Those two results seem contradictory, but are they contradictory from the statistical point of view as well?

Applying the expanded measurement uncertainty of 50 %, the ranges reflecting 95 % confidence (SANTE 11813/2017 [2]) are as follows:

Lab no.	Result	Range*
1	0,071 mg/kg	**0,035** – 0,11 mg/kg
2	0,037 mg/kg	0,018 – **0,055** mg/kg

*± 50 % expanded measurement uncertainty (SANTE 11813/2017) rounded to 2 significant figures

The confidence ranges of the two results have a significant overlapping range from 0,035 – 0,055 mg/kg. As the measured values lie inside ± 50 % range of the result of the other, the probability that the results are contradictory is < 5 %.

Or in other words: Applying a confidence interval of 95 %, the two results in question are analytically NOT CONTRADICTORY.

Example 2:

Lab 1: 0,071 mg/kg Azoxystrobin

Lab 2: 0,024 mg/kg Azoxystrobin

Applying the expanded measurement uncertainty of 50 % again leads to the following ranges:

Lab no.	Result	Range*
1	0,071 mg/kg	**0,035** – 0,11 mg/kg
2	0,024 mg/kg	0,012 – **0,036** mg/kg

*± 50 % expanded measurement uncertainty (SANTE 11813/2017) rounded to 2 significant numbers

The confidence levels of the two results have a very small overlapping range of 0,035 – 0,036 mg/kg.

As there still is an overlap between the range of lab no. 1 and lab no. 2 (considering the expanded measurement uncertainty of +/-50 %), the two results in question are analytically still COMPARABLE.

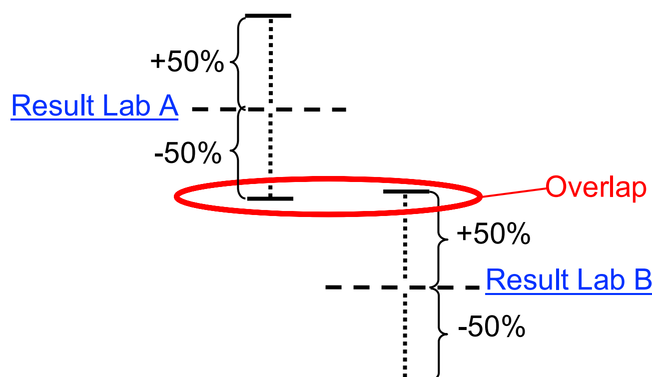

Figure 4: Comparability of analytical results considering the expanded measurement uncertainty of +/- 50 %

Example 3:

Lab 1: 0,090 mg/kg Azoxystrobin

Lab 2: 0,024 mg/kg Azoxystrobin

Applying the expanded measurement uncertainty of 50 % again leads to the following ranges:

Lab no.	Result	Range*
1	0,090 mg/kg	**0,045** – 0,135 mg/kg
2	0,024 mg/kg	0,012 – **0,036** mg/kg

*± 50 % expanded measurement uncertainty (SANTE 11813/2017) rounded to 2 significant numbers

The confidence levels of the two results show NO overlapping range. As a consequence, these two results are NOT comparable. Possible reasons for these strong differences are:

- analytical error / mistakes in one of the two (or in both) labs,
- reasons outside the labs (f.ex. differing samples although they are described to be identical; bad conservation of one of the two samples like un-cooled transport to the lab, ….).

4.2. General remark

As Preuss [6] points out, mistakes made in the lab (like mixing up of samples, technical troubles, dilution and pipetting mistakes) are responsible only for low part of differing results – most reasons can be identified earlier in the process chain, especially concerning sampling and homogeneity [7, 8].

5. Recommendations

In case of analytical results differing to a higher degree than expected, and an error within the responsibility of the laboratory could not be identified, it is recommended to

- check the identity of the samples in question: Do they belong to the same homogenate? When and how was the sampling carried out? What does the laboratory sample really represent?
- How was the sampling performed? How about the sample transport (cooled)?
- Check for other reasons for differing results (inhomogeneous lot, long time between samplings, differing LOQs or reporting limits, differing scopes of analytes etc.)
- In case no obvious reason for the differing results can be identified, a sample exchange between the involved labs (exchange of sample homogenates) may be helpful to search for analytical differences between the lab – as long as the analytes are stable over storage time and “survive” the freezing AND the thawing process.

6. Literature

[1] S. Bruns: Bewertung von rückstandsanalytischen Laborkompetenzen anhand experimenteller Daten, PhD thesis, Technische Universtität München, 2012.

[2] W. Horwitz, L. V. Kamps, K. W. Boyer (1980): Quality assurance in the analysis of foods for trace constituents, J. Assoc. Off. Anal. Chem., 63, 6, 1344-1354.

[3] European Commission, DG SANTE: Guidance document on analytical quality control and method validation procedures for pesticide residues and analysis in food and feed, SANTE/11813/2017.

[4] Directive 2011/91/EU of the European Parliament and of the Council of 13 December 2011 on indications or marks identifying the lot to which a foodstuff belongs, OJ L 334, 16.12.2011, p. 1–5.

[5] C. Hemmerling, B. Augustyniak, A. Maye, G. Seidl, G. Warschewske: Multirückstandsmethoden zur Pestizidanalytik in der Lebensmittelkontrolle – Möglichkeiten und Grenzen – Teil 1: Methodenvergleich. Deutsche Lebensmittel-Rundschau, Heft 4., Apr. 2009, 245-252.

[6] A. Preuß: Der wahre Wert – zur Bewertung analytischer Befunde, Deutsche Lebensmittel-Rundschau, Heft 7, Juli 2008, 313-319.

[7] D. E. Raynie: Know Your Sample: Size Matters, in LCGC North America, Aug 01, 2016, Volume 34, Issue 8, pg 530–532, available online: http://www.chromatographyonline.com/know-your-sample-size-matters-0, [as of Apr. 2018).

[8] M. W. Toews – own work, based (in concept) on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=1903871.

Acknowledgment

Important contributions to this relana® position paper are provided by:
Hans Braekman, PRIMORIS Belgium cvba

relana® Position Papers

This Position Paper is a literary property of relana®, based on the contribution and the knowledge of the members of the relana® laboratory quality circle. The aim of this publication is to increase knowledge and to provide expertise to all relevant and interested stakeholders in order to achieve best practices on analytical services related to food and feed testing. Everybody is invited to make use of this Position Paper and to circulate it wherever meaningful. While using this Position Paper please make the reference as:
relana® Position Paper No. 19-03 “Differing results of competent laboratories” version 2019/06/25.