Journal of Analytical Sciences, Methods and Instrumentation, 2023, 13, 13-25
https://www.scirp.org/journal/jasmi
ISSN Online: 2164-2753
ISSN Print: 2164-2745

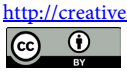
Consequences of Insufficient Selectivity in Quantitative and Qualitative Chemical Analysis

Mats Larsson, Göran Nilsson

Department of Physics, Alba Nova University Center, Stockholm University, Stockholm, Sweden
Email: mats.larsson@fysik.su.se

How to cite this paper: Larsson, M. and Nilsson, G. (2023) Consequences of Insufficient Selectivity in Quantitative and Qualitative Chemical Analysis. *Journal of Analytical Sciences, Methods and Instrumentation*, **13**, 13-25.
https://doi.org/10.4236/jasmi.2023.132002

Received: May 10, 2023
Accepted: June 27, 2023
Published: June 30, 2023



Abstract

A problem in chemical analysis in connection with measurements of a substance normally occurring in a sample, or identification of a substance which should not exist in a sample, is insufficient selectivity. In this article, we analyze this problem and propose remedies. We use a real doping case to illustrate how chemical noise causes a serious selectivity problem, probably causing a false positive outcome.

Keywords

Chemical Analysis, Quantitative and Qualitative Selectivity, Chemical Measurement Procedures, Measurement Errors, Chemical Noise

1. Introduction

Measurements are often the basis for decisions and for a measurement result to be useful, it must contain information about possible errors which are relevant and reliable. The relevant information depends on the application. In this article, we will consider measurements of the concentration of a substance (the analyte) in blood or urine and discuss the following situations:

1) Measurement of a substance normally occurring in samples (for instance insulin). The purpose is to investigate whether the level is normal (diagnosis) and the follow-up of treatments (quantitative analysis).

2) Identification of a substance which should not exist in a sample (qualitative analysis).

In the first situation, the ideal information is the uncertainty in relation to the true value. As we will see in clause 3, this information is often not possible to obtain. We have to be satisfied with information about the agreement between measurement procedures.

For the second situation, the relevant information is the specificity for the measurement procedure. Specificity is the probability to obtain a negative result when the substance is not present.

Reliable information about the specificity of a measurement procedure and the agreement between measurement procedures should be obtained from properly designed experiments. For the planning and statistical analysis of such experiments, we must first identify possible types of error. For earlier discussions on similar topics in the scientific literature, the reader is referred to references [1] [2].

2. Terms and Definitions

It is important that the terms used are clearly understood. Unfortunately, some terms are used with different definitions causing misunderstandings and confusions. In this clause we have defined and commented on some essential terms. When the definition is taken from a standard or guiding the reference is given.

Measurement procedure [3]

Detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result.

Note 1: The measurement result in this article is a concentration.

Note 2: The term measurement method is sometimes used as a synonym to measurement procedure but this term has according to VIM [3] a more generic definition.

Measurement uncertainty

An interval having a stated probability (usually 95%) to include the true value.

Note 1: If the measurement uncertainty is determined from statistical estimates of all possible error components, it is a confidence interval.

Note 2: In VIM [3] measurement uncertainty is defined as a "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used". According to VIM components contributing to the measurement uncertainty can be evaluated by means other than statistical analysis of observations, so called Type B evaluations. When Type B evaluations are based on assumptions they should be verified empirically.

Matrix

Other components of a sample than the analyte to be measured.

Note: The matrix often varies between the unknown samples to be measured.

Matrix effects

Effects from the matrix on the observed signal (response) in a measurement.

Selectivity

Refers to the extent to which the response in a measurement procedure is unaffected by the matrix.

Note 1: Perfect selectivity means that the response is not affected at all of other

substances.

Note 2: The term specificity is sometimes used for perfect selectivity. As specificity is defined for qualitative chemical analysis the use in this context is discouraged.

Commutability of a reference material [3]

Property of a reference material, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two given measurement procedures, and the relation obtained among the measurement results for other specified materials.

Note: The reference material in question is usually a calibrator and the other specified materials are usually routine samples.

Traceability [3]

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.

Note 1: For this definition a "reference" can be a definition of a measurement unit through its practical realization or a measurement procedure.

Note 2: When the matrix effects are not the same for references materials and the routine samples, traceability of measurements of routine samples can only be achieved to a combination of a reference measurement procedure and reference materials used for calibration, not only to reference materials.

Critical value of response or concentration [4]

Value of the observed response or concentration, the exceeding of which leads, for a given error probability, to the decision that the concentration is not zero.

Note: If the critical value is determined from measurements of calibrators or reference materials, the value may be underestimated due to matrix effects for the routine samples.

Minimum detectable value of the concentration [4]

True value of the concentration that with a specified probability will lead to rejection of the hypothesis that the concentration is zero (using the critical value above).

Note 1: The term *limit of detection* is often used instead of *minimum detectable value*, but usually without specification of a critical value.

Note 2: The term *sensitivity* is sometimes used to express how low one can detect the substance of interest. As sensitivity is defined for qualitative chemical analysis the use in this context is discouraged.

Note 3: If the minimum detectable value is determined from measurements of calibrators or reference materials, the value may be underestimated due to matrix effects for the routine samples.

Qualitative chemical analysis

The result of a qualitative chemical analysis is reported as negative or positive, where negative usually means absence of a certain substance and positive means presence of the substance.

Specificity

The probability that a qualitative chemical analysis reports a negative result when the substance is absent.

Note: If the decision is based on a quantitative measurement the probability is the given error probability in the definition of *critical value of response or concentration* (see above).

Sensitivity

The probability that a qualitative chemical analysis reports a positive result when the substance is present.

Chemical noise

Signals derived from components in the sample that are indistinguishable from the signal generated by the analyte of interest.

Note: This term is used in mass spectrometry and is comparable with matrix effects, but it can be assumed that chemical noise only increases the signal.

3. Possible Errors in Quantitative Measurements

3.1. Assumptions and Prerequisites

The measurand is the concentration of a substance (the analyte) in a liquid, for example blood or urine. A sample of the liquid can be partitioned into aliquots, which are supposed to be identical. An aliquot is used for one measurement.

By processing an aliquot according to a measurement procedure, a response is obtained. The response is supposed to be monotonously related to the size of the measurand (the true concentration). The relationship between the concentration and the response, the calibration curve, is estimated from measurements of calibrators with assigned values of the measurand. This calibration curve is then used to transform a response for an unknown sample to an estimate of the measurand. The calibration procedure may introduce two types of systematic errors: one caused by an error in the assigned values of the calibrators and one caused by an unsuitable model for the calibration curve.

3.2. Influencing Factors and Measurement Errors

If the response only depends on the measurand the calibration curve can be determined once and for all and there would be no measurement errors. This is, however, likely never the case and in practice the response is influenced by a lot of other factors. These factors are properties belonging to:

1) The measuring system (reagents, equipment, procedures, operational conditions);

2) The measured sample (unknown samples and calibrators).

The properties in the first group often vary between measurements and in order to reduce the variation of the observed concentrations one tries to identify subgroups of measurements within which the variation of the influencing factors is limited. A calibration curve is then determined for each subgroup. A series of measurements performed on one instrument and evaluated with the same cali-

bration curve is called a *run*.

In the following discussion we will assume that variation between repeated measurements within runs is completely random but this assumption should in practice be verified.

If a measurement procedure has a perfect selectivity, that is the response is not affected by other properties of the sample than the substance we want to measure, we should only have three error components: a random variation within runs, a random variation between runs and a systematic error (bias) caused by errors in the assigned values of the calibrators and an unsuitable model for the calibration curve. The random variation between runs is a calibration error caused by the random variation within runs.

If the selectivity of the measurement procedure is not perfect some factors in group 2 may influence the response. If these influencing factors have the same values for samples and calibrators, the relationship between response and concentration should be the same for samples and calibrators and there should be no further error components.

If the influencing factors in the second group have different values for the unknown samples and the calibrators, the relationship between response and concentration will not be the same for the calibrators and the samples and a consequence will be a systematic error, which often depends on the concentration. This error will probably also depend on the population of unknown samples we are interested in.

Often the influencing factors in the second group also differ between samples and we will then have errors that are specific for the individual samples. If the factors in group 2 interact with the factors in group 1 the sample specific error consists of two components: one contributing to the variation between runs and one contributing to average systematic error for a sample. The contribution to the variation between runs may depend on the sample.

When there are sample specific errors, the common systematic error concerns the average error in the population of unknown samples.

It should be noticed that choosing calibrators, which are more similar to the samples, can reduce a common systematic error, but a change of calibrators has no effect on the sample specific errors.

Thus, when a measurement procedure does not have a perfect selectivity, the following types of error must be considered: a random variation within runs, a random variation between runs, a bias (probably depending on the concentration) and sample specific errors.

When a difference in relationship between response and concentration for reference materials (calibrators) and unknown samples cannot be assumed to be negligible, traceability of measurements of unknown samples can only be achieved to a reference measurement procedure, not to a reference material.

3.3. Agreement between Measurement Procedures

Ideally the agreement with an accepted reference measurement procedure should

be assessed, but usually we have to be satisfied with agreement between different routine measurement procedures.

Information about the agreement between two measurement procedures can be obtained from a comparison according to the procedure described below.

Experimental design for comparison of measurement procedures

At least 30 samples covering the concentration interval of interest are measured in for instance triplicate in one run with each of two measurement procedures denoted MP*x* and MP*y*. Note that it is essential that individual samples are used. If sample pools are used, we have physically created means of a number of samples and we will underestimate the individual sample specific differences.

Statistical analysis

A procedure for statistical analysis of this type of experiments is described by Nilsson [5] and Nilsson *et al.* [6] and is here illustrated by a fictitious example. For each sample and measurement procedure the means, *x* and *y*, and standard deviations are calculated. In this example the standard deviations are approximately proportional to the concentration level and then it is suitable to perform the analysis for ln(concentration). A difference in ln(concentration) corresponds approximately to the relative difference and the standard deviation of ln(concentration) corresponds approximately to the coefficient of variation.

The differences in ln(concentration) between MP*y* and MP*x* are plotted against the mean concentrations on a logarithmic scale, see **Figure 1**. For this difference plot two characteristics can be identified: a continuous function fitted to the center of the scatter, and the standard deviation around this continuous function.

The continuous function is an estimate of the systematic difference between the MPs. The standard deviation around a continuous function is estimated from the successive differences according to [5] and [6] and is here denoted by

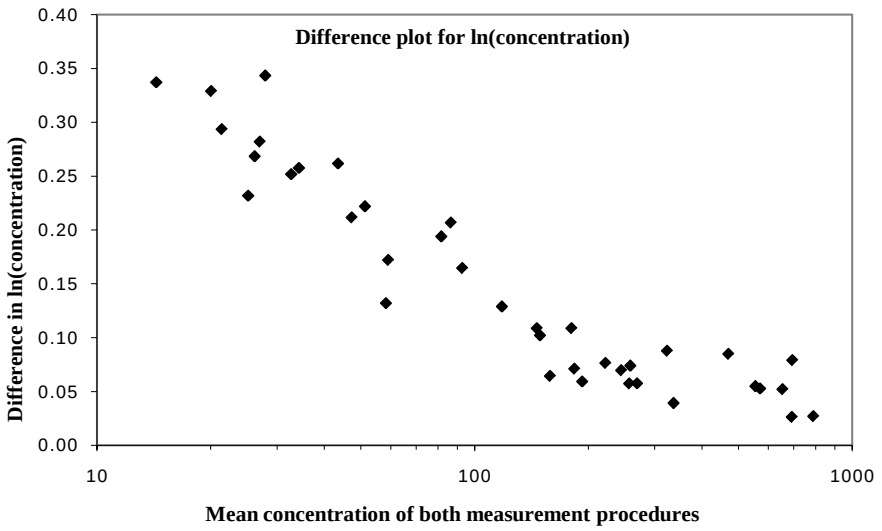


Figure 1. The differences in ln(concentration) between MP*y* and MPx are plotted against the mean concentrations of both measurement procedures on a logarithmic scale.

$s_{Scatter}$. This standard deviation contains contributions from the random variation within runs and sample specific errors of both measurement procedures. The standard deviations of the random variation within runs can be estimated from the repeated measurements (triplicates) and the pooled estimates from all samples are denoted $s_{e(x)}$ and $s_{e(y)}$ respectively. If there are no sample specific errors, $s_{Scatter}$ can be estimated from $s_{e(x)}$ and $s_{e(y)}$ and, thus, the standard deviation of the sample specific differences between the measurement procedures is estimated by

$$s_d = \sqrt{s_{Scatter}^2 - \frac{s_{e(x)}^2 + s_{e(y)}^2}{3}}.$$

The formula is valid with the error components identified in this article. In the example in **Figure 1** $s_{Scatter}$ is 0.027 and if the estimates $s_{e(x)}$ and $s_{e(y)}$ are 0.013 and 0.015 respectively we have $s_d = 0.024$.

Thus, we now have estimates of the errors contributing to the disagreement between two measurement procedures:

- The standard deviation of the random variation within runs, $s_{e(x)}$ and $s_{e(y)}$ respectively.
- The random calibration errors caused by the variation within runs. The standard deviation for each measurement procedure should be of the size $s_e/\sqrt{k}$ (or less, depending on the calibration curve model). k is the number of replicates for the calibrators. This standard deviation can be reduced by increasing k, but usually only a few replicates are necessary, as other error components often are dominating.
- The systematic difference between the measurement procedures, estimated by a continuous curve fitted to the difference plot. This difference can be reduced by choosing calibrators which are more similar (commutable) to the samples we want to measure.
- The standard deviation of the sample specific differences between the measurement procedures is estimated by s_d. To reduce the sample specific errors the selectivity of the measurement procedures must be improved.

If reference materials are included in a comparison of two measurement procedures, see [6], the commutability of a reference material can be estimated by the difference in bias between the reference material and the samples. If reference materials are used as common calibrators of the two measurement procedures the differences in bias will be estimates of bias between the measurement procedures.

The estimates are based on only one run with each measurement procedure and are of course very uncertain, but they should at least support a decision whether the agreement between measurement procedures may be sufficient or not.

Comparison of measurement procedures evaluated according to the principle presented here has shown that sample specific errors often are not negligible [7].

4. Chemical Noise, a Deleterious Effect in Doping Cases

A common problem in substance identification is testing the hypothesis that the

concentration of an analyte is zero in a certain sample. A typical example is doping control. For a doping test the important information about the test performance is the specificity, *i.e.*, the probability to get a negative result when the analyte is absent. This probability shall be sufficiently high.

Let us first consider the situation when the response of the test is not a numeric value but only negative or positive. The specificity can only be estimated by testing a large number of samples, which are representative for the athletes but can be assumed to be free from the target analyte. If the specificity is not 100%, the results are obviously influenced by other factors than the target analyte. It is then also reasonable to assume that there may be differences between populations, for example between males and females.

If the test gives a response with a numerical value having a positive relationship with the analyte concentration, it should be possible to determine the critical value that gives a wanted specificity. If the response is not influenced by other substances than the analyte (perfect selectivity), the critical value can be determined from the response distribution estimated from repeated tests of a reference material not containing the analyte. If perfect selectivity cannot be assumed, the critical value must be determined from the response distribution for a population that is representative for the athletes but can be assumed to be free from the target analyte. The essential difference in comparison with the previous situation is that we now can determine a critical value giving the wanted specificity.

To determine the critical value from a reference population is of course complicated. To avoid this problem measurement procedures based on mass spectrometry has been developed for doping control. But how selective are these measurement procedures?

It is well known that urine, particularly from a sexually active woman, contain thousands of bacteria of many species [8]. Even if only a small number of contaminating microorganisms are present in the warm medium of freshly voided urine, they may multiply rapidly to high concentrations [9]. Critical values for detection of the target analyte in mass spectrometry of complex mixtures are generally influenced by chemical noise [10], which is defined in Section 2.

We will use a real doping case to illustrate the problems caused by chemical noise. We will also discuss the misinterpretation of the minimum criteria for chromatographic mass spectrometric confirmation of the identity of analytes for doping control purposes, as expressed in the technical document TD2021IDCR of the World Anti-Doping Agency [11].

Figure 2 shows the mass spectrum for a metabolite of 17α-methyltestosterone with two trimethylsilanol added so that the molecule can be vaporized. In selected reaction monitoring (SRM), commonly performed with a triple quadrupole instrument, a precursor ion is selected by the first quadrupole. The precursor ion continues to the collision cell, where it collides with a target gas, for example nitrogen. The laboratory collision energy is chosen to provide maximum

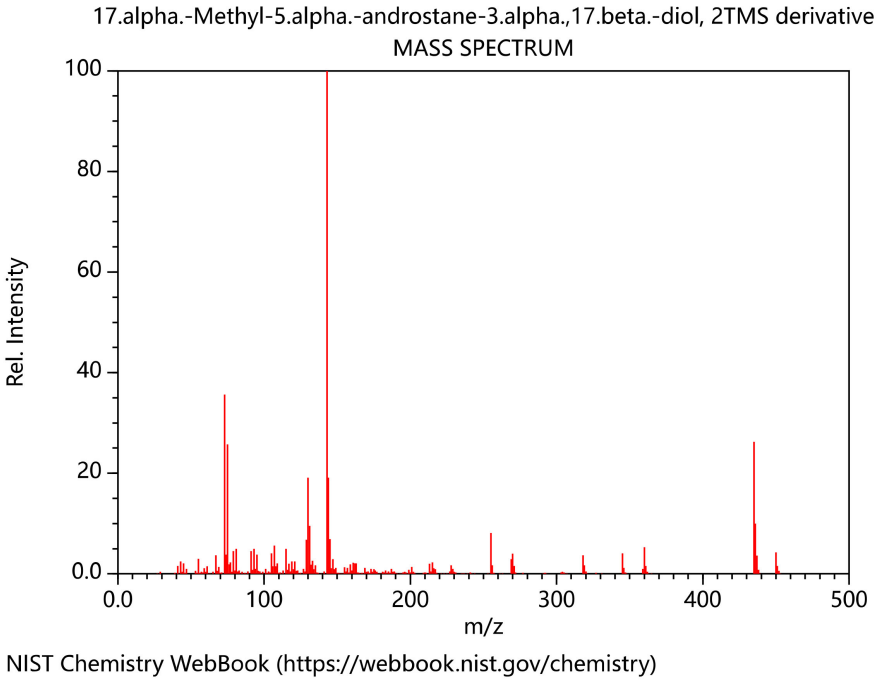


Figure 2. Mass spectrum for the 17a-methyl-5a-androstane-3a,17b-diol metabolite of 17a-me-thyltestosterone from the NIST Chemistry WebBook. The *x*-axis (mass over charge) is given in Dalton. The small peak at m/z = 270 Da is the molecular M^+ at 450 Da minus 2TMSiOH (mass = 180 Da). This ion was chosen as Reference Diagnostic Ion by the doping laboratory for the 17α-methyl-5β-androstane-3α,17β-diol metabolite. The only structural difference between the two metabolites is the orientation of the H-atom on the fifth carbon atom.

analytical information at medium transmission. If the pressure in the collision cell is too high, the result is losses due to scattering, charge exchange and charge stripping processes instead of delivering additional structural information. In the final step, product ions are selected by the second quadrupole. If the quadrupole is scanned, a product ion spectrum is obtained, whereas in SRM one ion is selected at the time.

The terminology concerning the product ions differs. In TD2021IDCR, the terms are "Reference Diagnostic Ion" and "Diagnostic Ion" and are illustrated in Figure 3. In an earlier WADA technical document, the Reference Diagnostic Ion was referred to as "base peak", but without being given a stringent definition.

In order to establish by means of SRM whether a forbidden substance is present in an Athlete's urine sample, the relative abundances of the Diagnostic Ions are calculated with respect to the Reference Diagnostic Ion for the sample and compared with the relative abundances for a positive control sample. The concentration of the analyte in the positive control sample should be comparable to the concentration in the Athlete's sample. If the relative abundances of the Diagnostic Ions fall within certain intervals of the relative abundances of the same ions in the control sample, the doping laboratory has identified the analyte in the Athlete's sample and an Adverse Analytical Finding shall be reported. The

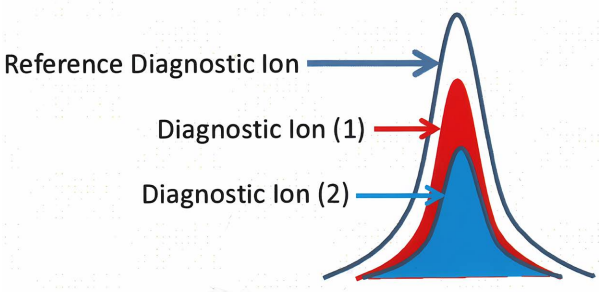


Figure 3. Chromatogram illustrating the definitions used by WADA in their technical document TD2021IDCR. The retention time through the gas (or liquid) chromatograph is on the *x*-axis and the ion intensity on the *y*-axis. An alternative terminology used in commercial mass spectrometry software is "target ion" (Reference Diagnostic Ion) and "qualifier ions" (Diagnostic Ions). The Reference Diagnostic Ion, usually most abundant diagnostic ion, is used to calculate the relative abundances of the Diagnostic Ions.

intervals are specified TD2021IDCR [11], but the authors are unaware of any scientific motivation for the numbers in Table 1 in [11].

A critical question is how many transitions a laboratory should monitor. In a tutorial article by Maurer [12]: "A minimum of two transitions [in SRM] (e.g. one target ion and two qualifier ions) per analyte is requested". The importance of more than one transition is pointed out; with only one transition (*i.e.* one target ion and one qualifier ion) the desired selectivity is not obtained. WADA is also clear on the number of transitions that shall be acquired: "When using multiple-stage MS (e.g. MS/MS) at least two (2) Diagnostic Ions [*i.e.* two precursor-product ion transitions (SRM transitions)] shall be acquired" [11].

It would seem like there is total harmony between references [11] and [12], but this is not the case. In a real doping case reporting the two metabolites of 17α-methyltestosterone, the Swedish doping laboratory, in addition to the Reference Diagnostic Ion (435 → 255, *i.e.* the intensity of m/z = 255 Da), used only one precursor-product ion transition for the 17α-methyl-5α-androstane-3α,17β-diol metabolite (435 → 199). This would seem to contradict the TD2021IDCR, but that was not the opinion of the doping laboratory, nor WADA's Senior Deputy Director of Science, who in a communication to the doping laboratory claimed that the doping laboratory had indeed monitored two transitions, of which 435 → 255 was counted as one transition. In reality, the doping laboratory only acquired two ions, one Reference Diagnostic Ion and one Diagnostic Ion. The National Anti-Doping Organization did not object.

The doping laboratory reported the following concentrations for the A sample: 17α-methyl-5α-androstane-3α,17β-diol concentration ≅ 2 ng/mL; 17α-methyl-5β-androstane-3α,17β-diol concentration ≅ 6 ng/mL. The doping laboratory used the precursor ion m/z = 270 Da for the 5β metabolite. It was clear from the 270 → 213 and 270 → 157 transitions that the chromatograms for 213 and 157 Da were not symmetric, indicating the presence of chemical noise. The presence of chemical noise became even more obvious when the laboratory documentation package for the B sample became available. **Figure 4** shows the

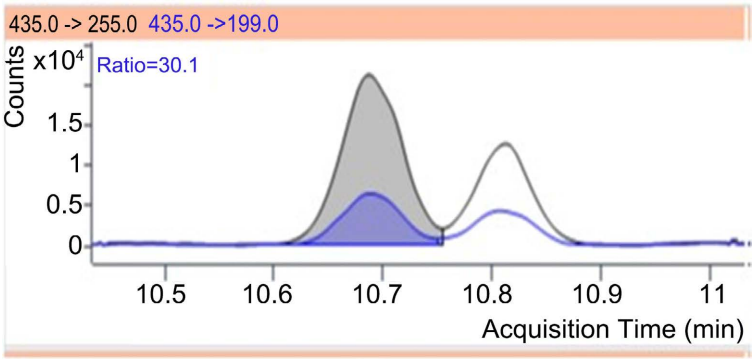


Figure 4. Chromatograms for the B sample. The peaks at retention (acquisition) time t = 10.687 minutes derives from the 17α-methyl-5α metabolite, whereas the two peaks at t = 10.811 minutes derives from the 17α-methyl-5β metabolite.

chromatograms for the 435 → 255 and 435 → 199 transitions for *both* metabolites (those transitions were not shown for the 5β metabolite for the A sample). It is obvious from a visual inspection that 1) the concentrations reported for the A sample are erroneous and 2) the m/z = 255 Da peak (used by the doping laboratory as Reference Diagnostic Ion for the 17α-methyl-5α metabolite) is asymmetric and hence affected by chemical noise. It should be pointed out that the chromatogram for the deuterated control substance was perfectly symmetric with Gaussian shape.

Unfortunately, TD2021IDCR is not very clear concerning the required shape of a chromatogram. It is stated that the doping laboratory is allowed to ignore a Diagnostic Ion that is being interfered by a partially co-eluding substance in the sample, but this is not a "shall" requirement. In the present case, it is noteworthy that it is the Reference Diagnostic Ion (m = 255 Da) that is interfered by a co-eluding substance in the sample with almost but not identical retention time in the gas chromatograph. Thus, already the normalizing chromatogram becomes more intense owing to chemical noise

The presence of chemical noise in the exemplified doping case sample is not surprising. It was an out-of-competition doping test performed in the private home of the athlete. The sample was then handled by the doping control officer, first transported 110 km and then stored in the officer's private home for about 5.5 hours. It was then transported by a commercial transport company for another 540 km to the doping laboratory, where it arrived at least 24 hours after the athlete had delivered the sample. There is no documentation of the conditions under which the sample was transported, although WADA requires that samples should be transported in a manner which minimizes the potential for sample degradation due to factors such as time delays and extreme temperature variations. Whereas one can assume that the doping control officer had some training in transporting samples, the commercial transport company most certainly did not.

5. Conclusions

Even if matrix effects are identified as possible causes to a systemic error, sample

specific errors are usually not included in the error models. In for example ISO 5725 [13] only variation under repeatability conditions, variation between laboratories and a bias are considered. The bias shall be evaluated at one level at a time by testing materials with known values of the measurand and with a matrix as close as possible to the matrix of the unknown samples we want to measure. However, it is not mentioned that the unknown samples often do not have the same matrix and that sample specific errors cannot be neglected. Thus, neglecting the possibility of sample specific errors will lead to an underestimation of the measurement uncertainty. The best we can do in practice is often to estimate the agreement between measurement procedures according to Section 3.3.

Proficiency testing is often based on measurements of artificial samples but we are interested in the agreement between laboratories for the real samples we want to measure and, thus, real samples should be used also in proficiency testing. A goal should be to identify groups of measurement procedures within which the agreement is sufficient (compatible measurement procedures). How to design proficiency testing with this purpose is, however, not obvious.

A further consequence of insufficient selectivity is that an observed effect of for example a medical treatment on the concentration of a substance in blood may be caused by a change of the matrix in the blood. How to investigate this possibility is a challenging topic for further studies.

Substance identification was established as the weak link in analytical toxicology already almost 20 years ago [2]. In this article, we have shown that the problem still remains with urine drug testing, and it is surprising how simple remedies have not been implemented by WADA. It would be very simple to require that chemical noise shall not lead to false-positives by implementing strict rules for the shape of chromatograms. Chemical noise is a problem in particular for female athletes.

Acknowledgements

This work was supported by the Swedish Research Council. The authors are grateful to the anonymous athlete for access to data from the laboratory documentation packages for the A and B samples.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] van der Veen, A.M.H. (2003) Measurement Uncertainty and Doping Control in Sport. *Accreditation and Quality Assurance*, **8**, 334-339. https://doi.org/10.1007/s00769-003-0644-6

[2] de Zeeuw, R.A. (2004) Substance Identification: The Weak Link in Analytical Toxicology. *Journal of Chromatography B*, **811**, 3-12. https://doi.org/10.1016/j.jchromb.2004.07.043

[3] Joint Committee for Guides in Metrology, JCGM 200 (2012) VIM: International Vocabulary of Metrology. 3rd Edition.

[4] ISO 11843-1 (1997) Capability of Detection—Part 1: Terms and Definitions. ICS.

[5] Nilsson, G. (1991) Comparison of Measurement Methods Based on a Model for the Error Structure. *Journal of Chemometrics*, **5**, 523-536. https://doi.org/10.1002/cem.1180050605

[6] Nilsson, G., *et al.* (2018) IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. *Clinical Chemistry*, **64**, 455-464. https://doi.org/10.1373/clinchem.2017.277541

[7] Miller, W.G., Thienpont, L.M., Van Uytfanghe, K., Clark, P.M., Lindstedt, P., Nilsson, G. and Steffes, M.W. (2009) Toward Standardization of Insulin Immunoassays. *Clinical Chemistry*, **55**, 1011-1018. https://doi.org/10.1373/clinchem.2008.118380

[8] Honour, J.W. (1996) Testing for Drug Abuse. *The Lancet*, **348**, 41-43. https://doi.org/10.1016/S0140-6736(96)05336-6

[9] Tsivou, M., Livadara, D., Georgakopoulos, D.G., Koupparis, M.A., Atta-Politou, J. and Georgakopoulos, C.G. (2009) Stabilization of Human Urine Doping Control Samples: II Microbial Degradation of Steroids. *Analytical Biochemistry*, **388**, 146-154. https://doi.org/10.1016/j.ab.2009.02.013

[10] Hassell, K.M., LeBlanc, S.A. and McLuckey, S.A. (2011) Chemical Noise Reduction via Mass Spectrometry and Ion/Ion Charge Inversion: Amino Acids. *Analytical Chemistry*, **83**, 2352-2355. https://doi.org/10.1021/ac200439k

[11] https://www.wada-ama.org/sites/default/files/resources/files/td2021idcr_final_eng_0.pdf

[12] Maurer, H.H. (2020) Pitfalls in Drug Testing by Hyphenated Low- and High-Resolution Mass Spectrometry. *Drug Testing and Analysis*, **12**, 172-179. https://doi.org/10.1002/dta.2744

[13] ISO 5725 (1994-1998): Accuracy (Trueness and Precision) of Measurement Methods and Results (6 Parts). There Are New Versions of Part 2 (2019) and Part 4 (2020).