J Forensic Sci, May 2005, Vol. 50, No. 3
Paper ID JFS2005013
Published 6 April 2005
Available online at: www.astm.org

Letter to the Editor—Fitness for Purpose of Mass Spectrometric Methods in Substance Identification

Sir:

Recently, the *Journal of the American Society of Mass Spectrometry* published an important report by a working group of the ASMS Measurements and Standards Committee on the fitness for purpose of mass spectrometry (MS) in trace analysis, with a special focus on substance identification (1). Indeed, MS is frequently called upon to address many important forensic, regulatory and other societal concerns, making it imperative that the applied methodology and the interpretation of the results are scientifically sound, legally defensible and practically feasible. Therefore, it is laudable that the working group has taken a closer look at this problem (also by examining a number of guidance documents issued by various regulatory agencies) and that it has endeavoured to craft a generally applicable process for conducting qualitative analyses in what it calls "adversarial" situations. This has resulted in proposing the following concept for mass spectrometric methods:

- A reference standard should be analyzed contemporaneously with the sample presumed to contain the suspect compound.
- Confirmation should be based on three or more diagnostic ions (except for exact mass measurements).
- Relative abundance matching tolerances should be used for selected ion monitoring (SIM).
- Adequate quality assurance/quality control should be carried out.
- Analytes should be separated by on-line chromatography.

Because these criteria were found in all guidance documents, it was assumed that they represent the core judgement of the mass spectrometry community.

However, the Report also seems to have some shortcomings. In the following, a number of important issues are being addressed for further consideration and to enhance the ongoing discussion on how to properly utilize mass spectrometric methods in substance identification.

Confirmation versus Identification—The Report and the studied guidance documents do not provide clear definitions or descriptions of these terms, using them indiscriminately. The following should be realized:

- Confirmation presumes the presence of substance Y in a sample, based on initial tests or prior information. The presence of Y is then "confirmed" by further tests (in this case MS).
- Identification does not make *a priori* presumptions based on initial tests or other information. Results on the sample after a number of tests are compared with reference data on all other substances that may come into consideration.

Moreover, it must be realized that a "positive" confirmation test thus obtained is *not* an unambiguous identification of Y. It only shows that the test result is *not against* the presumption. Other substances may be able to give results that are the same or indistinguishable from those of Y. Therefore, unambiguous identification of Y is achieved if all other (relevant) substances can be excluded, so that Y remains as the *only* possible candidate (I prefer to call the exclusion criterion the "reverse angle approach"). For example, if methamphetamine is suspected to be present and the mass spectrum in the sample matches that of a reference sample within acceptable limits, it still remains to be established that the test result in the sample cannot be due to any other relevant substance (e.g., other amphetamine-like substances, isomers, metabolites, endogenous compounds, omnipresent interferences, etc.). Obviously, the term "relevant" is important in this context. With tens of thousands of substances known to society, it is clearly unfeasible to consider them all. Yet, even if one focuses only on those that have some relevance to the field of analysis (e.g., forensic toxicology, doping, environmental pollution, drugs and driving), data on thousands of substances per field are necessary.

Contemporaneous Analysis of Reference Standards—All recommendations in the report are based on the assumption that a reference standard is available and that it be run contemporaneously with the suspected sample to achieve a high degree of repeatability. This approach may be feasible in directed analyses to assess whether a particular compound is present or not and if the number of target compounds is limited. However, the consequence of the exclusion approach in a correct identification process is that—in addition to references for the target compounds—thousands of other reference standards must be available as well. Not only should their analytical properties in the applied analytical methodologies be known in advance, they are also to be run contemporaneously with every suspect sample under the recommendations of the report. Yet, this requires huge banks of (precious) reference standards in every laboratory carrying out qualitative analyses, and it results in a high turnover rate. Obviously, this is neither desirable nor feasible in practice.

On the other hand, the concept of using a reference standard hinges on a presumption that the sample contains a particular compound. The Report does not indicate how to handle if no presumptions exist. This occurs frequently in clinical and forensic toxicology cases in which no presumption exists (e.g., in clinical and forensic toxicology, drugs and driving) in which an undirected

analytical approach is required. Moreover, undirected approaches may also be required as a follow-up to a directed analysis to see if additional suspect substances are present in the sample (e.g., multi-drug intoxications), or when the initial presumption turns out to be false.

Finally, what should be done if no reference standard is available for contemporaneous comparison. In this situation, which frequently occurs in practice, many laboratories then revert to computer-assisted data base searching. Yet, it is well known that the interlaboratory reproducibility of mass spectra is rather questionable and that there is ongoing debate as to what can be considered an acceptable 'match.' It is a pity, therefore, that the Report does not provide guidance on when and how to perform data base searching.

The 3-ion Criterion for Substance Identification—To assess the degree of matching between the MS results on the reference and those on the sample, the Report advocates the well known principle of examining a minimum of three diagnostic ions (at nominal mass accuracy), which have to be present in both spectra at distinct relative abundances and within acceptable tolerance windows. This criterion appeared to be the only broadly recognized standard for analyte identification (2,3) and it was found in all surveyed guidance documents. Hence, it was considered the shared professional judgement of the mass spectrometric community.

The scientific rationale of the 3-ion criterion comes from a study by Sphon in 1978 (4) and a reevaluation and expansion of this work at the 1996 ASMS Workshop (2). In brief, Sphon used diethylstilbestrol (DES) as test substance and a EI-GC-MS data base with some 30,000 spectra. Monitoring the relative intensities of three ions (viz., 268, 239 and 145) at discrete tolerance windows was sufficient to select DES as the only candidate. At the re-evaluation in 1996, the database contained some 270,000 spectra. Monitoring the same three ions, albeit with tighter windows of 10% (absolute) was still sufficient to select DES as the only candidate. However, a number of critical observations must be made:

- The approach by Sphon is certainly valuable , but so far it has been tested for only one substance.
- The mass spectrum of DES has all three ions at relatively high masses, where the diagnostic value is larger than at lower masses.
- The suitabiltiy of the databases for adversarial analyses remains uncertain (i.e. did they contain the various substances likely to be encountered in this field).
- To judge general applicability, tests with other relevant substances and classes are needed, especially when many structurally related compounds can be encountered (e.g., amphetamines, benzodiazepines, steroids, pesticides, opiates).
- The approach has not been checked for techniques other than EI-GC-MS.
- Thus, monitoring a minimum of three ions in EI-GC-MS is unwarranted as a general rule. Follow-up studies are needed with a much larger selection of target drugs, and also towards other techniques, such as CI-GC-MS, LC-MS and tandem MS.

Matching Tolerances for Selected ion Monitoring (SIM)—The working group found considerable variation in the tolerance windows of relative abundances (RA) in the guidance documents. It was concluded, therefore, that it may not be fruitful to put exhaustive attention on matching tolerance selection, other than to use matching criteria not dramatically different than the windows proposed until now. Indeed, the array of tolerance criteria all seem to be arbitrary, not based on scientific evidence and not checked towards their attainability in practice. Furthermore, it should be noted that, apart from the size of the windows, some documents work with one window for all relative abundances, whereas others allow larger windows for lower relative abundances. Also, there is no uniformity whether to express windows in terms of relative or absolute differences. Obviously, harmonization is needed here. It cannot be that, based on a given set of GC-MS results, the identification of e.g. clenbuterol will come out positive under the criteria the EU guidelines for residues in animals and negative under the criteria of the World Anti Doping Agency. In the end, uniform criteria are required, not based on arbitrary numbers but on scientific evidence that makes them properly defensible. Suitable roads to arrive at such criteria are available; yet, they require considerable time and effort.

Apart from the above observations, there is a more fundamental question whether the use of tolerance windows for RA is acceptable conceptually. Obviously, the idea is to let the windows compensate for small run-to-run variations in RA. However, what is being overlooked here is that, basically, the absolute ion abundances are measured in the MS instrument, which are then 'normalized' against the base peak. As with any ion, the absolute abundance of the base peak may also vary from run to run. Yet, the latter is phased out in the normalization process. This can have crucial consequences, as shown in Table 1, which is derived from reference data in (5) and using tolerance windows for tandem MS as defined by the European Union for the analysis of residues in live animals and animal products (6).

In Table 1, upper part, absolute and relative intensities are shown for a reference sample, together with the respective tolerance windows. For reasons of simplicity, the values for the corresponding absolute and relative intensities in this mock example have been made the same. In the middle part, suspect sample 1 shows an acceptable variation in the absolute intensity of the base peak of 20%; all other ions in the sample have the same absolute intensities as in the reference. Thus, when using absolute intensities, a positive match is obtained for all ions. However, when using relative intensities, three ions (viz., 364, 320, and 314) now give a negative

TABLE 1—*Matching of selected ion monitoring data from reference samples with suspect samples.*

Ion	Abs. Int. Arb. Units	Rel. Int. %	Tolerance Window	Match with Abs. Int.	Match with Rel. Int.
Reference sample					
440	100	100	080–120[a]		
364	027	27	022–032		
320	068	68	055–081		
314	083	83	067–099		
264	012	12	008–016		
Suspect sample 1					
440	080	100	80–120[b]	OK	OK
364	027	34	22–32	OK	NOK
320	068	85	55–81	OK	NOK
314	083	105	67–99	OK	NOK
264	012	15	08–16	OK	OK
Suspect sample 2					
440	115	100	80–120[b]	OK	OK
364	035	30	22–32	NOK	OK
320	068	59	55–81	OK	OK
314	083	72	67–99	OK	OK
264	018	15	08–16	NOK	OK

[a] based on absolute intensities.
[b] based on relative intensities.
NO = not OK.

match, despite the fact that all ions had variations in their absolute intensities that were within the allowable range. The conclusion based on relative intensities thus would be a false-negative result.

In the lower part of Table 1, the base peak in suspect Sample 2 shows an acceptable variation in absolute intensity of 15% in the opposite direction, but the absolute intensities of ions 364 and 264 are now outside their respective tolerance window, resulting in negative matches. Yet, if the relative intensities are considered, positive matches are obtained for all ions, yielding a false-positive result. Thus, the normalization process against the base peak may lead to scientifically untenable, arbitrary results and which do not allow equal justice.

Assessing Qualitative Uncertainty—The report contains an interesting discussion on this often-neglected topic but concludes that a generally accepted manner, either numerically or in prose, for describing the identification confidence of a given method is not yet available. In my opinion, expressing qualitative uncertainty as potential rates of false positives and false negatives is not enough. It must be combined with a parameter describing the probability of correctness of the identification (note that the latter may not only be based on mass spectral information, but that it may also include information from other techniques such as chromatography, UV absorption, etc.). Such a probability parameter can be obtained by applying the reverse angle approach as discussed above See for details (7). However, as also mentioned in the report, the large databases on mass and UV spectra, chromatographic retention parameters, etc. do not exist at the moment. Nevertheless, all laboratories involved in adversarial analysis are obliged to assess the uncertainty of their qualitative and quantitative results under the requirements of ISO/IEC 17025 (8).

Some other Remaining Issues—Understandably, the primary focus of the Report is on qualitative analysis by EI-GC-MS, because most of the presently available knowledge is in this area. Yet, other techniques, such CI-GC-MS, LC-MS with its ionization techniques, tandem MS and high-resolution MS are being rapidly introduced in many laboratories. It is a pity, therefore, that their potentials and limitations are not being addressed in more detail.

Other issues that receive little attention are computer-assisted spectral matching via databases and the need to assess peak purity.

Conclusions—After the above discussions, what can be said about the fitness for purpose of mass spectrometric methods in substance identification? The Report states: “Fitness for purpose means that the uncertainty inherent in a given method is tolerable given the needs of the application area.” Yet, this is not the only issue. Prior to assessing its uncertainty, it should be established that a given method is scientifically sound (i.e., based on appropriate scientific evidence), legally defensible (providing equal justice), and feasible in practice. The above paragraphs demonstrate why current MS does not meet latter three requirements. Furthermore, the means to assess the qualitative uncertainty and/or probability of correctness of an identification are still inadequate. Thus, one must conclude that the mass spectrometric methods for substance identification recommended in the Report and in the surveyed guidance documents are not fit for purpose. Hopefully, the above comments and caveats may stimulate further discussions and actions that will overcome the various shortcomings in the near future, because correct substance identification is of pivotal importance in many adversarial situations in modern society.

References

1. Bethem R, Boison J, Gale J, Heller D, Lehotay S, Loo J, et al. Estimating the fitness for purpose of mass spectrometric methods. J Am Soc Mass Spectrom 2003;14:528–41. [PubMed]
2. Baldwin R, Bethem RA, Boyd RK, Budde WL, Cairns T, Gibbons RD, et al. Limits to confirmation, quantitation and detection. J Am Soc Mass Spectrom 1997;8:1180–90.
3. Bethem RA, Boyd RK. Mass spectrometry in trace analysis. J Am Soc Mass Spectrom 1998;9:643–8.
4. Sphon JA. Use of mass spectrometry for confirmation of animal drug residues. J Assoc Off Anal Chem 1978;61:1247–52. [PubMed]
5. André F, De Wasch KKG, De Brabander HF, Impens SR, Stolker LAM, Van Ginkel L, et al. Trends in the identification of organic residues and contaminants: EC regulations under revision. Trends Anal Chem 2001;8:435–44.
6. European Union Decision 2002/657/EC 17.08.2002. Commission decision laying down performance criteria for the analytical methods to be used for certain substances and residues thereof in live animals and animal products according to Council Directive 96/23/EC. Off J Eur Commun 2002;221:8–36.
7. Hartstra J, Franke JP, De Zeeuw RA. How to approach substance identification in qualitative bioanalysis. J Chromatogr B 2000;739:125–37.
8. International Organization for Standardization. SO/IEC 17025. General requirements for the competence of testing and calibration laboratories. Geneva: International Organization for Standardization, 1999.

Rokus A. de Zeeuw, Ph.D.
Emeritus Professor of Toxicology and Bioanalysis
De Dobbe 4
9901 BV Appingedam
The Netherlands
E-mail: r.a.dezeeuw@wolmail.nl