Establishing the Fitness for Purpose of Mass Spectrometric Methods

Robert Bethem

Alta Analytical Laboratory, El Dorado Hills, California, USA

Joe Boison

Canadian Food Inspection Agency, Saskatoon, Saskatchewan, Canada

Jane Gale

Bristol-Myers Squibb, New Brunswick, New Jersey, USA

David Heller

FDA Center for Veterinary Medicine, Laurel, Maryland, USA

Steven Lehotay

USDA/ARS Eastern Regional Research Center, Wyndmoor, Pennsylvania, USA

Joseph Loo

Pfizer Global Research and Development, Ann Arbor, Michigan, USA

Steven Musser

FDA Center for Food Safety and Applied Nutrition, Washington, D.C., USA

Phil Price

Dow Chemical, South Charleston, West Virginia, USA

Stephen Stein

National Institute of Standards and Technology, Gaithersburg, Maryland, USA

This report is submitted by a working group sponsored by the ASMS Measurements and Standards Committee. The group responded to a 1998 opinion piece dealing with mass spectrometry in trace analysis (Bethem, R. A.; Boyd, R. K. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 643–648) which proposed that the concept of *fitness for purpose* addresses the needs of a wide range of analytical problems. There is a need to define *fitness for purpose* within the current context of mass spectrometry and to recommend processes for developing and evaluating methods according to suitability for a particular purpose. The key element in our proposal is for the interested parties to define in advance the acceptable degree of measurement uncertainty and the desired degree of identification confidence. These choices can serve as guideposts during method development and targets for retrospective evaluation of methods. A series of more detailed recommendations are derived from basic principles and also from reviews of current practice. This report highlights some areas where consensus is evident, but also revealed the need for further work in other areas. The recommendations are aimed primarily for the laboratory analyst but we hope they will be accessible to the non-scientist as well. Our goal was to provide a framework that can support informed decisions and foster discussion of the issues, because ultimately it is the responsibility of the analyst to make choices, provide supporting data, and interpret results according to scientific principles and qualified judgment. (*J Am Soc Mass Spectrom* 2003, *14*, 528–541) © 2003 American Society for Mass Spectrometry

---

Published online April 10, 2003

The opinions expressed herein do not represent an official position of the American Society for Mass Spectrometry (ASMS) or the home institutes of any of the contributors.

Address reprint requests to Mr. D. N. Heller, Center for Veterinary Medicine, US Food and Drug Administration, 8401 Muirkirk Rd., Laurel, MD 20708, USA. E-mail: dheller@cvm.fda.gov

Mass spectrometry is often called upon to address legal, regulatory, or other societal concerns that are outside the realm of academic inquiry. Such questions may include (but are not limited to):

- Is a suspect compound present or not?
- Is a suspect compound present above some decision point?
- What is the limit at which a method can answer either of these questions?

A successful outcome—for the analyst as well as for his or her client—depends on appropriate preparation, methodical laboratory work, and a defense of the data's validity in light of generally-accepted principles.

In 1996 the ASMS sponsored its annual Fall Workshop [1] on the theme of “Limits to Confirmation, Quantitation and Detection” in order to foster discussion of these issues. Many different viewpoints were presented on such topics as data requirements for confirming the presence of a suspect compound and for establishing a method's limit of quantitation. The conference organizers subsequently published two reports in the Journal of the ASMS. The first report summarized the presentations of the invited speakers [2] while the second report—an opinion piece—assessed the divergence of opinions and attempted to define a logical next step [3].

After the appearance of the second report in 1998, a working group was formed under the auspices of the ASMS Measurements and Standards Committee to continue this process. The 1996 Workshop revealed that it was not realistic to pursue a single, universal performance standard that, if met, would guarantee success in any situation. Different applications have different needs and therefore require different responses from analysts. However, we believe there is a need to bring unity to the divergence of opinions within the mass spectrometry community. By drawing ideas from many sources, we attempted to craft a generally-applicable process for conducting analyses in *adversarial* situations. [In the *adversarial context* there is a conflict over the interpretation of data. The conflict is resolved by choosing one side over another. The quest for scientific truth may be subordinate to the desire to prevail.]

This report is presented in sections of varying depths. The recommendations are described in an Executive Summary, which hopefully is appropriate for both scientists and non-scientists. Numbered endnotes cite references to publications listed in the Reference section. Detailed discussions are provided in the text, if brief, or as appendices (with Roman numerals) if extensive. Some concepts are illustrated in figures.

To make this discussion more accessible, we provided examples (Appendix I, *Examples*) of situations where analysts could make use of our recommendations. These examples are drawn from a variety of disciplines and activities. Readers may want to review these examples and to keep them in mind before reading further.

Ultimately it is the responsibility of the analyst to make choices, provide supporting data, and interpret results according to scientific principles and qualified judgment. We hope to provide a framework that can

support informed decisions and foster discussion of the issues. It is appropriate for the ASMS to contribute to this discussion, because many ASMS members provide, defend or review data in adversarial situations. Some recommendations were based on a survey we conducted at the 1999 ASMS Conference (Appendix II, *Survey*). Where some issues could not be fully resolved, we tried to describe the problems and point toward possible solutions.

Executive Summary of Recommendations

The unifying principles underlying this report are:

- **Analysts should use methods which are Fit for Purpose.**
- **Analysts should be able to show that their methods are Fit for Purpose.**

Analysts need to work with basic principles of Fitness for Purpose because in most cases we bear the burden of defending our methods and choices. No recommendations or guidance from any agency, advisory group, or professional society can fully remove this burden. Indeed, our own fitness as experts in our own field is dependent on familiarity with the issues described in this report (Appendix III, *Purpose*).

We advance the following definition:

- **Fitness for Purpose means that the uncertainty inherent in a given method is tolerable given the needs of the application area.**

Figure 1 is a flow chart showing a process for achieving and demonstrating method fitness. Basically the process consists of addressing the most important things first. Each stage consists of specific investigations of key factors. Detailed discussions of each stage are listed in Appendix IV, *Detailed Process*.

The first stage is a detailed examination of what is at stake. One should investigate and define the external factors that influence choices in methodology and reporting. The objective of this stage is to define the analytical purpose in a way that sets targets for acceptable uncertainty. The concept of *acceptable uncertainty* is key to working with these recommendations. The concept can be defined more fully:

- **Targets for measurement uncertainty describe how accurate and precise the measurements need to be.**
- **Targets for identification confidence describe how certain one needs to be that the correct analyte has been identified.**

The second stage is to define the method itself. This process is the conventional method development process familiar to all analytical chemists. The work should be conducted with the acceptable uncertainty range in mind.

Figure 1. Recommended process for achieving fitness for purpose.

The third stage is to assess the uncertainty of the method. There are a variety of possible ways to assess measurement uncertainty and identification confidence. However, it is important to recognize that there are two divergent approaches to assessing uncertainty. These two approaches are applicable for both quantitative methods and qualitative methods:

- **Empirical or top-down approaches work with data acquired within the method's working range and use a holistic view of method performance.**
- **Statistical or bottom-up approaches look to differentiate signals from background and consider method performance as a combination of individual steps.**

These two basic approaches apply to either qualitative or quantitative methods, and may be carried out in various ways in each case. Analysts must select from a variety of options to assess analytical uncertainty.

The conclusion of the process is to evaluate whether the method is acceptable:

- **Establishing method fitness consists of showing that the targets for measurement uncertainty and identification confidence have been met.**

The process may be iterative if the targets cannot be met initially, in which case the method should be revised and re-evaluated.

Our group deliberately chose not to write prescriptive recommendations. On the other hand, our inclusive process enabled us to assess trends within the mass spectrometry community. Since the 1996 Workshop a growing number of agencies and bodies worldwide have codified the standards that are considered appropriate within their jurisdiction or discipline for confirming the presence of suspect compounds (qualitative analyses). We surveyed draft guidance documents [4] for common elements that could be considered a set of core concepts and definitions within the mass spectrometry community (Appendix V, *Core Concepts*).

- **Certain confirmation criteria were common to all surveyed guidance documents:**

- Reference standard [all recommendations in this Report are based on the assumption that a reference standard is available. The structural identification of unknowns is not addressed here] should be analyzed contemporaneously with unknowns;
- Identification should be based on three or more diagnostic ions (at nominal mass accuracy).
- Relative abundance matching tolerances should be used for selected ion monitoring.
- There should be a quality assurance/quality control program.
- Analytes should be separated by on-line chromatography prior to analysis.

The sets of glossaries examined revealed a number of terms that could be defined for a particular method, as well as a few ambiguous terms that should be avoided or carefully defined. The term *Diagnostic Ion* was the only term that appeared in all documents surveyed, which highlights the importance of this issue (Appendix VI, *Diagnostic Value of Data*):

- **The diagnostic value of selected signals should be assessed and described.**

The lower limit of method performance is usually the most hotly debated aspect of methods used in adversarial situations. Our recommendations lead to a particular understanding of method limits:

- **A method's limit is the point where the targets for acceptable measurement uncertainty or identification confidence can no longer be met.**
- **Given this definition, the empirical or top-down approach is preferable for describing method limits in adversarial analyses.**

Since the working range of the method is defined as the range where analytical uncertainty is acceptable, the

limit of the working range is the point where measurement uncertainty or identification just fails the targets set in stage 1 (Figure 1). The virtue of this understanding of method limits is that data supporting the validity of a method are acquired within the working range of the method. Descriptions of method limits are not based on analyses outside the method's true capability, for example, by measuring the apparent concentration of background noise (Appendix VII, *Method Limits*).

There are many areas where analysts must make important choices in selecting methods. Sometimes analysts must make choices in the absence of clear guidance from clients; such choices should be documented. There may be an assumption that a "conventional" analysis is appropriate but such assumptions need to be examined for method fitness in the given application. Our basic recommendation for handling these choices is

- **All choices that are inherent in the chosen methodology should be documented.**

Certain choices are especially critical:

- **Identification criteria are neither absolute nor arbitrary: they arise from the level of identification confidence that is considered acceptable for a given application.**
- **Method limits may be determined according to either qualitative or quantitative considerations; the point where identification confidence is unacceptable could be different from the point where precision and accuracy are unacceptable.**
- **For qualitative methods, the desired balance between the acceptable rates of false positives and false negatives should be described.** (Appendix VIII, *Objectives of Qualitative Analyses*).

Issues specific to qualitative methodology. The greatest need for improved clarity concerns *qualitative methodology*, i.e., methods of identification. There is not yet a generally-accepted manner for describing, either numerically or in prose, the identification confidence associated with a given method. At present we can identify several strategies for describing this qualitative uncertainty, but even these are far from ideal (Appendix IX, *Assessing Qualitative Uncertainty*).

Figure 2 shows a stepwise proposal for demonstrating the fitness for purpose of qualitative methods. Specific steps are outlined according to the stages of our general recommendations (Figure 1). For certain steps, the underlined items could benefit from further clarification:

- **Certain additional resources are still needed in this area:**
  - A glossary of terms for expressing the degree of identification confidence.
  - Updated guidelines for the use of exact mass measurement in qualitative methods.

What's
at stake

- Set targets for identification confidence
- Set targets for false positives/false negatives
- Set target for decision level (minimum level of full performance)

Define the method

- Define terms to be used in reporting
- Define the identification criteria to be used in the method
- Define the diagnostic power of the identification criteria
  - Ions, ion ratios, mass losses, transitions, isotopes, patterns
  - Mass accuracy—exact mass sufficient to exclude alternatives
- Validate in working range
  - Analyze control tissues—exclude endogenous interferences
  - Interference testing—exclude similar compounds
- Characterize lower limit of performance
  - Extrapolate from working range, fortify through lower limit range
  - Lower limit is where false negative / false positive target is exceeded

Assess
uncertainty

- Show that empirically-determined lower limit meets target for decision level
- Show that validation data support claim for diagnostic power of method identification criteria, rates of false outcomes
  - Harmonize with core confirmation criteria
  - Compare diagnostic spectrum with relevant library of spectra
  - Estimate selectivity index of total method
- Report identification confidence in numeric or prose terms

Report
results

- YES/NO outcome based on whether criteria met, identification confidence achieved.
- YES/NO outcome based on interpretation by qualified analyst
- ACCEPTABLE based on review by qualified expert
- Suitable methodology based on precedent, references

Figure 2. Demonstrating fitness for purpose of qualitative methods. Underlined items could benefit from further clarification or additional resources in the mass spectrometry community.

- Updates on "core confirmation criteria" gleaned from documents worldwide.
- Links to libraries of mass spectra and information on library-search algorithms.
- A consensus-based approach to estimating the total selectivity of a method.
- A bibliography of publications on method validation.
- A training or qualification process for analysts, reviewers and experts.
- A bibliography of legal precedents, guidance documents, and compendia of reference methods.

Finally, the manner of reporting results is critical for translating technical results to a form desired by the client. If one assumes the client prefers a simple YES/NO answer, then the question must be defined precisely. The first two questions are features of the technique, and can be addressed objectively. However, the third question requires an interpretation of data, and may always contain some degree of subjectivity.

- **Were the decision criteria met?**
- **Was the method capable of meeting the targets for identification confidence and measurement uncertainty?**
- **Does the analyst conclude that a suspect analyte is present above some limit or some concentration?**

We hope this overview of the issues will foster clarity wherever mass spectrometry is used to support decision-making in the greater society.

Acknowledgments

The authors thank John Chakel, Ed Houghton, William Horwitz, and Michael Thompson.

Appendices

I. Examples

- You work in the central research laboratory of a chemical manufacturer. One lot of your company's product fails to meet a basic QC test. Contamination of a raw material is suspected, but the certificates of analysis from your supplier seem in order. You screen reserve samples of the raw material and identify a contaminant. On the basis of your results, your company files suit against the supplier to recover losses due to the contamination. The supplier disputes your results.
- You are an employee of a contract research organization. An agrochemical company contracts with your employer to develop a surveillance method that will meet the needs of a regulatory agency. You strictly follow the language of the contract. Halfway through the project you begin to suspect that the language of the contract asks your employer to do something different from what the agency regulations seem to call for.
- While running a routine method for a customer, an automated software program fails to flag an analyte that you believe is clearly present in the sample. You can substantiate your assessment that the analyte is present by examining the data. However, the customer's method only allows for automated data processing, not manual processing.
- Your new Standard Operating Procedure for confirmatory analyses requires an independent review before any results are acted upon. However, no one in your Quality Assurance Unit is familiar with mass spectrometry. How does the QAU receive appropriate training and certification to perform independent data reviews?
- You work for a regulatory agency. Routine monitoring turns up a compound that might be at a violative level. You are asked to describe how confident you are that your measurement exceeds the threshold level for enforcement action, and how confident you are that the suspect compound has been identified correctly. Depending on your answer, the agency may or may not begin a lengthy legal process.
- You have run a particular method routinely for a long time, and you know that a certain compound will either be present in the sample at an easily detectable level or it will not be present at all. You

find a positive result for the compound at a barely detectable level.

- A client requires three ions be used for MS confirmation of a set of compounds, but some of the analytes only give two major ions. Analysis of blanks and spiked samples indicate that the method consistently identifies those analytes correctly on the basis of only two ions. Using only two ions for monitoring, your analyses of the client's samples result in identifying those particular analytes only in matrices from a certain location.
- You work for a contract laboratory running samples in a competitive area. Your competitors can do the analysis for \$50/sample, but you can trim your costs to \$40/sample by using a method with less definitive identification criteria.
- You use a simple mass spectrometric method to screen a set of samples for a set of target analytes. The method is fast and cheap because the method is not very rigorous in its diagnostic power. A few samples are flagged for follow-up analysis by a more rigorous and costly confirmatory method. The rigorous method fails to identify any target analytes in the samples flagged by the screening method. Your results are challenged because the two methods gave conflicting results for the same samples.

II. Survey of Attendees at the 1999 ASMS Conference

The survey aimed to investigate the form and content of a resource document that might help analysts deal with adversarial situations. Questions concerned (1) analytical purpose, (2) document format, (3) factors that influence the respondent's work environment, and (4) the disciplines, employers and personal opinions of the respondents. Respondents ranked the importance of each issue and the potential value of a resource document in their setting. Salient results of the survey are listed below, roughly in order of the clarity or strength of the responses. *This summary is not a formal position of the Society or its membership.*

- Every purpose we listed was considered highly important to at least some respondents.
- These issues are primarily of concern to analysts in government or private enterprise, not academia. Some non-ASMS members are also concerned with these issues.
- No consensus emerged on most issues; i.e., there were no issues clearly relevant or critical to every group. Some issues were of great importance to some but of little importance to others. The relative importance of each specific issue varied with the analyst's employer or discipline.
- A resource document was perceived to be more valuable if flexible or innovated methods were

needed, and less valuable if conventional criteria or an established precedent were applicable.

- All respondents were subject to some form of oversight. Ranked from most to least frequently mentioned, these included: Regulatory agencies, peers, inspectors in QA or GLP programs, internal auditors, public interest groups, clients, and juries. Every source of oversight was important to at least some respondents.
- There were some points of general agreement:
  - The Society has the support of analysts in fostering discussion of this topic.
  - The language of a resource document should be scientifically precise but understandable to non-scientists.
  - The preferred formats for a resource document included general recommendations, a decision tree, or a checklist. Prescriptions or a standard procedure were not favored.
  - The resource document was deemed of greatest value at the inception and conclusion of work (i.e., for help in defining the purpose, reporting results, and defending choices).
  - Analysts typically need to minimize time more than to minimize cost.
- The more important factors were identified as follows:
  - Availability of reference standard.
  - Validation of methods before their application.
  - Thorough consideration of the issues related to uncertainty in measurements.
  - Ability to modify existing methods and performance requirements.
  - Demonstration of consistent matrix background, lack of interference.
  - Oversight of regulatory agencies.
- Certain factors were moderately important to analysts. These included the use of corroborating data, whether the concentration range was at trace levels or higher, the perceived threat posed by suspect compound, and the use of library matching or accurate mass measurements.

The survey respondents ranked certain external issues (oversight, analyte risk) as less important than many technical issues. This was in contrast to the opinion of our working group members. Our view held that all external issues should be considered before technical issues, so that method development and application address the needs of a pre-defined purpose. Our recommendations are a blend of the survey's indications and our group's views.

III. Purpose

At the 1996 Fall Workshop, M. A. Kaiser advanced the following definition: "Fitness for purpose refers to the

magnitude of the uncertainty associated with a measurement in relation to the needs of the application area." Thompson and Ramsey [5] defined fitness for purpose as "the property of data produced by a measurement process that enables a user to make technically correct decisions for a stated purpose." We felt that by comparing the magnitude of uncertainty associated with a method to the degree of certainty needed for the application, it might be possible to put this entire process on an objective basis.

The two principles stated in our Executive Summary (Analysts should use methods which are Fit for Purpose. Analysts should be able to show that their methods are Fit for Purpose) correspond in a general way to the first two of six principles of best practice recommended by a Eurachem working group on the *Fitness for Purpose of Analytical Methods*: "Analytical measurements should be made to satisfy an agreed requirement, i.e., to a defined objective. Analytical measurements should be made using methods which have been tested to ensure they are fit for purpose" [6].

There are important benefits to defining analytical purpose after a complete examination of what is at stake:

- This process can result in a better definition of the technical problem.
- A definition of purpose can be used to develop performance criteria that make sense for the situation at hand.
- The process of defining method fitness prospectively (i.e., in advance) draws the interested parties together in a way that can bring out conflicting expectations or assumptions. We believe it is preferable to reveal and resolve such conflicts before the fact, because if hidden conflicts emerge when analyses are complete, the value of even rigorous work will be diminished.

In the present era, the business model for analytical laboratories emphasizes the relationship to the *customer* for whom analyses are conducted. However, the notion of *customer* can be resolved into several distinct aspects. *Clients* use technical results as the basis for a decision. *Patrons* pay for analyses. *Beneficiaries* are those whose welfare may depend on the decision. The analyses may be subject to *oversight* from a variety of internal or external groups. In a strictly contractual arrangement between private parties, the *client*, *patron*, *beneficiary*, and *overseer* may be the same person or corporation. However, when government agencies, international bodies, courts, non-profit organizations, accreditation bodies, third-party reference laboratories, etc. are involved, the concept of *customer* can be quite fragmented. It is important for analysts to identify and document the assumptions and expectations of each interested party when defining analytical purpose. If there are conflicting interests, it is preferable to address these early in the process, before laboratory work begins.

IV. Detailed Process

First, examine what's at stake. If the client cannot offer guidance in the first stage, then document the assumptions built into choices made on the client's behalf.

- *Define the analytical problem.* Every analytical problem includes both qualitative and quantitative aspects.
- *Evaluate the perceived risk associated with the analyte and the potential consequences associated with the analytical result.* The effort put into planning needs to be proportional to the potential impact of the result more than the technical difficulty. In adversarial situations, the perceived risk and potential consequences could be quite out of proportion to the technical complexity. Analysts should prepare for each possible outcome, not just the preferred outcome.
- *Consider the needs of all interested parties.* The interested parties may include some or all of the following: Regulatory agencies, peers, inspectors in QA or GLP programs, internal auditors, public interest groups, clients, or juries. These parties could be technically unsophisticated or openly hostile. However, the act of asking for clarification from people with something at stake may well increase their understanding and decrease their hostility. As a final consideration, it may be necessary to consult with legal advisers before proceeding.
- *Balance the need for thoroughness versus time and resources available.* Analytical laboratories may find that time is more valuable than money, or that funds are inadequate to address every issue completely. If time and money do not permit, it may not be possible to achieve a solution that is technically rigorous in every aspect.
- *Define the acceptable uncertainty.* This is the key issue in establishing method fitness. The analyst and client should jointly set limits on the precision and accuracy that are acceptable for measurement purposes, as well as limits on false identifications that are acceptable for qualitative purposes.
- *Identify the expected concentration range.* Certain tasks can be handled differently if the analyte is suspected to be present at trace levels (near the method's lower limits) or considerably above trace levels.
- *Consider the sophistication of the client and report.* Reporting needs may vary among clients and situations. Without making any universal recommendations, it is necessary to agree in advance of what form the report should take, and how results will be reported.

Next, define the method. This stage comprises conventional method development and validation. These issues have been thoroughly discussed in many prior publications, and will not be treated in depth here:

- *Obtain reference standard.*
- *Determine if a reference method is required or available.*

- *Define the needed mass measurement accuracy.*
- *Characterize the matrix, background, and possible interferences.*
- *Establish analyst expertise and proficiency.*
- *Check system suitability and compliance with SOPs.*
- *Balance the expected concentration range in the light of qualitative and quantitative issues.*
- *Validate the method and/or perform appropriate Quality Assurance.*
- *Determine if corroborating data will be used [non-MS methods or MS libraries].*

Then, assess uncertainty. Quantitative uncertainty is conventionally described in terms of precision and accuracy. Qualitative uncertainty may be described as potential rates of false positive and false negative results, or as some measure of spectral uniqueness or probability of correctness.

Qualitative analyses aimed at unambiguous identification have historically been called *confirmation* methods. In qualitative analyses, the rate of false negatives is the percent of samples known to contain the suspect analyte that fail to meet the method's confirmation criteria. Conversely, the rate of false positives is the percent of samples known to contain the suspect analyte that meet the method's confirmation criteria anyway.

Various algorithms and data libraries exist for computing spectral uniqueness or the probability that an identification is correct. Proper use of these measures of identification confidence depends on grasping the assumptions built into the algorithms, as well as considering the size and quality of the data library.

V. Core Concepts

Core confirmation criteria. If there is to be a core set of confirmation criteria and definitions that evolve within the mass spectrometry community, this core should appear as elements that are common to all criteria documents now in force or under development. We compiled such common elements from a limited survey of guideline documents, whether in draft or final form [4]. Common elements were interpreted as the shared professional judgment of the mass spectrometry community. Elements that varied among the documents were considered as areas in which analysts must make key choices when developing methods that fit their purpose.

It is critical to realize that no document described its recommendations as absolute. The regulatory bodies set firm requirements, but still might allow in-house variation if supported by justification, validation, and/or pre-approval. The advisory body offered clear guidance, but would allow customized methods if appropriate scientific expertise, judgment, and review were applied. Therefore, even when method developers rely on guidance provided by an agency or advisory group,

Figure 3. Comparison of abundance matching tolerances, EI-GC/MS. The legend indicates which document listed under reference [4] recommends each particular matching tolerance.

they still have a certain burden of proof to justify their choices.

Certain criteria appeared in all documents. It is in the spirit of our working group's view that these elements represent the core judgment of the mass spectrometry community:

- Use of reference standard analyzed contemporaneously with unknowns.
- Three or more diagnostic ions (except for exact mass measurements).
- Use of relative abundance matching tolerances for selected ion monitoring (SIM).
- Quality assurance/Quality control program.
- On-line chromatography prior to MS analysis.

Although the last two elements are not directly concerned with mass spectrometry, it is clear that quality management and chromatographic separations are widely considered critical to the success of analytical methods used in the adversarial context.

The following elements appeared in four of the five documents, and should be seriously considered when developing confirmatory methods:

- Signal-to-noise minimum of 3:1.
- Characterization of performance at lower concentration limit.
- Method validation.
- Demonstration of method specificity (absence of interferences) through analysis of negative controls.
- Equivalence of precursor-product ion pair to two diagnostic ions.

- Choice of scan function, ionization technique.

The use of a negative control (i.e., blank matrix) is very important for quality assurance and validation purposes. The lack of bona fide control matrix could act to reduce identification confidence and increase the limit of quantitation. A similar or surrogate matrix, or even a sample which shows no analyte signals, is preferable to using no control matrix at all.

Variable confirmation criteria. At the other extreme, there was considerable variation in the acceptance windows (matching ranges) for acceptable relative abundances in SIM, when using electron ionization. The matching ranges were different in virtually every case. Clearly there is no generally-accepted approach to relative-abundance matching in SIM, although there is a moderate degree of conformity among the various approaches (Figure 3). The matching tolerances defined by various groups diverge because they must oversimplify what is actually a complex situation. De Boer et al. showed that simple matching tolerances (as in Figure 3) do not reflect the reality of abundance variation for certain drugs analyzed repeatedly by GC/MS [7]. Stein showed that ion abundance variability is best modeled by the square root of abundance multiplied by a compound-specific factor [8]. Consequently, it may not be fruitful to put exhaustive attention on matching tolerance selection, other than to use matching criteria that are not dramatically different than the ranges evident in Figure 3.

Core definitions. We compiled a list of terms from the glossaries included in five draft guideline documents. There was not much overlap among terms defined in

various glossaries. The only term that was defined in every document was *diagnostic ion* (or, the alternative terms *structurally-specific* or *structurally-significant*).

The composite glossary was rich with key terms and concepts that could be addressed in a detailed description of a confirmatory method. These included: Analyte (or target compound), blank matrix (or negative control), carryover, failure to confirm, limit of confirmation (or limit of identification), noise characteristics, peak resolution, presence, qualifier ion, quantitation ion, rate of false negatives (or β error), rate of false positives (or α error), reference material (or standard), reference method, screening method, specificity, suitability, traceable (or certified), validation, violative result.

Variable definitions. In the draft documents examined, *confirmation* was used to mean either verification of a prior test or verification of the presence of a suspect compound. This ambiguity reflects a historical trend. Mass spectrometry was first used non-routinely to “confirm” prior results from less specific detectors. As capable and user-friendly instruments proliferated, *confirmation* was applied to firmly identifying the presence of target analytes, with mass spectrometric data primarily standing alone.

There was a range of definitions for *detection*. There are several critical problems with this term that make it undesirable in adversarial analyses. *Detection* is understood by non-scientists as an identification with high confidence. However, *detection* is also used to mean an event in which a signal can be differentiated from noise. This mode has the lowest qualitative confidence possible, because a single signal at this level has little diagnostic value.

It is in the spirit of our working group’s view that *detection* is a term that should be avoided, in favor of some definition of identification or confirmation criteria. Identification with the utmost certainty is better described as confirmation, where multi-ion identification criteria are fully met and each diagnostic signal is fully distinguished from noise.

Certain terms are understood differently between scientists and non-scientists. Furthermore, scientists from different disciplines may not use the same term in the same way. These terms include (in addition to *detection* and *confirmation*) *absence*, *confidence*, *negative*, *positive*, *probability*, *screening*, and *uncertainty*. Definitions should be spelled out clearly if terms with multiple meanings are used. It has been recommended that the terms *selectivity* be used to express a degree of confidence, while *specificity* applies only to a quality of exclusiveness or uniqueness. That is, if a method is selective enough, it can become specific [9].

VI. Diagnostic Value of Data

Diagnostic ions indicate the presence of the target compound. As more diagnostic data is acquired, the confidence in the identification increases. It is recommended that quantitative methods also include identi-

fication criteria, because quantitation typically makes use of only a single data channel. Furthermore, the diagnostic power of conventional quantitative methods are frequently based on assumptions that translate to 95% confidence (1 in 20 chance of false identification). Analysts must determine if this qualitative confidence is adequate. Some purposes (e.g., confirmation) may require considerably higher confidence in the identification, and this can be accomplished by using multiple diagnostic ions to increase selectivity. The diagnostic information in a mass spectrum is essentially a composite of various elements. Depending on the analyst or application, the following elements may be combined in various weightings:

- *m/z Values.* Each observed ion may be indicative of a specific structure.
- *Exact mass.* For lower-mass ions (roughly <500 Da) measured with sufficient accuracy (e.g., <5 ppm), unique elemental formulae may be indicated.
- *Mass differences.* For two observed ions, the difference may indicate loss of a specific moiety. This mass difference may be non-specific (e.g., loss of 18 Da) or unique, depending on value.
- *MS/MS transition.* One ion is shown to be a discrete fragment of another ion.
- *Abundance ratios.* Ion abundance ratios are a function of unique ion chemistry, and are useful because their values can be compared and/or treated statistically.
- *Isotopes.* Some abundance ratios may indicate presence of specific elements.
- *Molecular ion.* Presence of certain adduct ions may give strong support for nominal molecular weight.
- *Spectral pattern.* The entire mass spectrum can be treated mathematically as a unique pattern, without regard to structural interpretation as above.

VII. Method Limits

Method limits define the transition between acceptable and unacceptable performance. The distinctions between bottom-up and top-down approaches were described previously. The strengths and weaknesses of each approach can be evaluated in various situations. The top-down approach is more consistent with the Working Group’s recommendations, because acceptability criteria are defined in the first stage of method development, and then data are acquired which indicate a point where these acceptability criteria can no longer be met.

In the bottom-up approach, a *limit of detection* corresponds to the minimum signal that can be differentiated from noise. Unfortunately, this approach makes an assumption that the signal is derived from the target analyte, because in mass spectrometry there is little diagnostic value in such a signal. This approach is not recommended, especially in adversarial circumstances. In the top-down approach, the limit of “detection” corresponds to the lowest level at which the weakest

diagnostic signal is still evident. The pre-determined identification criteria are met, so the predetermined level of acceptable qualitative confidence is met. [Furthermore, due to ambiguity in the meaning of *detection*, it is preferable to define qualitative limits in other terms.] Alternatively, a qualitative limit can be defined as the point where the rate of false positives is unacceptable, or where the rate of false negatives is unacceptable. In each of these empirical approaches the limit is keyed to a target determined during the initial examination of what's at stake.

In the bottom-up approach, a *limit of quantitation* corresponds to the minimum signal that can be differentiated from noise at some quantifiable level of confidence. In the top-down approach, a limit of quantitation corresponds to the level in the working range below which the precision and accuracy are no longer acceptable. The latter viewpoint ties the method's LOQ to the initial examination of "what's at stake" and thus is quite in line with our recommendations. Furthermore, this level can be established with true signals that fall within the working range of the method. Extrapolations based on noise levels or signals with little diagnostic value are not required.

VIII. Various Objectives for Qualitative Analyses

There are two clearly distinct and opposed qualitative analytical purposes—*enforcement* and *risk assessment*. Each is characterized by the need to avoid false outcomes, either by minimizing false positives (enforcement) or false negatives (risk assessment). For example, a false positive outcome could result in falsely accusing an innocent person. On the other hand, mass spectrometry is often used to search for the presence of a potentially dangerous substance. In such cases, it is critical to not overlook the presence of a "bad actor". These cases call for methods with a low rate of false negative outcomes. The need for such methods frequently arises from *risk assessment*, such as in health and environmental monitoring.

There are other analytical purposes that are intermediate, or somewhat more tolerant of false outcomes. Frequently these cases result from *dosing* experiments, where a known substance is added to a system, and the compound (or a metabolite or degradant) is identified elsewhere in the system. This prior knowledge about the system can contribute to acceptance of less rigorous confirmation criteria or intermediate degrees of confidence. In other cases maximum rigor may not be considered necessary. For example, Geerdink et al. have described a form of identification with intermediate confidence as an "indication" [10].

These objectives can be compared by constructing a distribution function. A given method and a set of confirmation criteria can be applied to blank matrices fortified with various amounts of analyte and a plot can be created showing the percent of samples meeting criteria as a function of concentration. At very

Figure 4. Comparison of various objectives for qualitative methods. This scheme demonstrates the relative rates of false positives for various purposes. The solid line is extrapolated from experimental data. The diagram is not necessarily to scale.

high concentrations the method will virtually always give a positive outcome. At very low concentrations there will virtually never be a positive outcome. There will always be some transition zone, or a range of uncertainty, within which the method may give either outcome for repeat analyses of the same sample. This is an undesirable situation, especially because the unsophisticated client or layperson may assume that methods always return a negative result below some limit, and a positive result at or above the same value.

Figure 4 compares the rate of false positives among methods with various objectives. The area above the curve and to the right represents the percent rate of false positives for a given set of identification criteria. Risk assessment methods might tolerate up to a modest rate of false positives (perhaps 1 in 20 is reasonable in this case), but enforcement methods require a far smaller rate of false positives. Dosing methods are intermediate, since other information may be used in addition to the mass spectra.

Mass spectrometry may also be used for screening, although this may be less cost-effective than other techniques. Screening methods are designed to have a very low rate of false negatives, to remove true negatives from the sample stream submitted for confirmation (to improve efficiency and reduce cost). The rate of false positives for a screening method may be quite high. Therefore, screening methods do not provide substantive information about what might be present in a sample. Screens should be part of a multistep process, combined with a follow-up method with a low rate of false positives.

IX. Assessing Qualitative Uncertainty

Numerical expressions of identification confidence. Qualitative uncertainty may be expressed as potential rates of false positives and false negatives. It is also possible to describe qualitative uncertainty through a numerical expression of spectral uniqueness or probability of

correct identification. However, these numerical expressions require a sufficiently large and appropriately chosen library of mass spectra, because uniqueness values can only be calculated in relation to such a library. For electron ionization, such libraries exist, but libraries of standard MS/MS spectra or spectra acquired by LC/MS with atmospheric pressure ionization have not been compiled to the same degree. As a result, there is less basis for generally-acceptable confirmation criteria in MS/MS or LC/MS. Therefore the burden of proof is greater on analysts who would use MS/MS or LC/MS instead of EI-MS methods in adversarial situations. On the other hand, library-searching can be a reasonable means to exclude possible incorrect identifications.

Our recommendations call for the selected confirmation criteria to lead to an acceptable degree of identification confidence. It is conventional in mass spectrometry to describe identification confidence in terms of predicted rates of false positives and/or false negatives. However, there are considerable challenges in showing that this has been achieved in practice. Nevertheless, it is important to provide data in support of the claims for identification confidence. There are auxiliary benefits to an approach to method fitness based on assessment of analytical uncertainty. In 1993 the U.S. Supreme Court clarified the Federal Rules of Evidence specifying what counts in court as science. In *Daubert versus Merrell Dow Pharmaceuticals*, the court said that a technique or methodology qualifies as science only if it can be tested, is subject to peer review, **possesses known rates of error**, and is generally accepted as science [11].

Data linking a set of confirmation criteria with a claimed rate of false *negative* outcomes are readily obtained, as follows. Replicate analyses of standards or fortified samples provide relative abundances from multiple diagnostic ions. Statistical analysis of the relative abundance data yields the mean and standard deviation for each ion. Analysts may define an abundance matching range that achieves a specified confidence limit. For example, a 95% confirmation rate is achieved with an abundance matching range ± 2 standard deviations about each abundance mean. The predicted false negative rate is then 5% or 1 in 20. De Boer et al. provide a very complete example of this approach applied to a GC/MS analysis of veterinary drugs [7].

Data linking a set of criteria with a claimed rate of false positives is critically important in supporting data used for enforcement purposes. There are several possible strategies:

- *Interference testing.* In theory, exhaustive analysis of blank matrices and similar compounds demonstrates the inability of the method to misidentify endogenous or exogenous interferences, respectively.
- *Validation.* Conventional method validation yields empirical evidence of performance.
- *Library-searching.* A numerical measure of spectral uniqueness can be calculated by comparison to a

large and appropriate library of mass spectra, if one is available.

- *Selectivity scoring.* An overall confidence level of a given method can be compiled from the estimated selectivity of all steps. Each step of the analysis has some degree of resolving power, including extraction, clean-up, chromatography, and separation by mass through one or more stages. This resolving power is cumulative, and in theory can become quite high. Some steps are widely recognized as contributors to high selectivity, such as monitoring multiple diagnostic ions, applying limits on abundance ratios, and using good chromatography to separate matrix components from target analytes.

As alternatives to these approaches, analysts have historically fallen back on the following strategies to support their conclusions:

- *Qualified professional judgment.* If analysts are sufficiently trained, experienced or knowledgeable in the field, their judgment may be considered reliable to decide that confirmation has been achieved.
- *Independent expert review.* An independent audit by a qualified reviewer may establish the acceptability of the data.
- *Generally-recognized precedent.* There may have been a similar case in the past in which the method or data quality was found acceptable.

These three strategies are dependent on interpretation of qualitative data, rather than a strict numeric procedure to draw a conclusion.

- *Best practice.* There is a tendency to assume that a confirmation is acceptable if it uses the best mass spectrometer and method we have available at the time. Within the framework of our discussion, this is not an advisable strategy. **Establishing fitness for purpose should not be reduced to accepting whatever data are already at hand.**

Interpreting qualitative data for utmost certainty. Expert judgment is as hard to define as qualitative uncertainty. We hope that one's expertise will be based on familiarity with adversarial analyses through reports like this one. Our recommendation is to be aware of the strengths and weaknesses of various strategies because many or all of them may be combined to defend a method's validity.

Suspect compounds are identified by comparing data from a reference standard with those of a sample acquired using the same method. If the data can be shown to be the same, within some degree of confidence, signals from the unknown can be said to arise from the presence of the suspect compound. This is the Null Hypothesis of statistics, meaning that the difference between the two data sets can be considered null. The Alternative Hypothesis states that the Null Hy-

pothesis is incorrect, i.e., the two data sets are statistically different from one another. From the analytical chemistry standpoint, the Null Hypothesis means that the suspect compound can be confirmed as present, while the Alternative Hypothesis means the suspect compound cannot be confirmed as present. This is not the same as proving the compound is absent; one can only prove it is not present above the limit of confirmation.

Neither viewpoint enables one to *absolutely* prove that a suspect compound is present. In statistics, the Null Hypothesis can never be proved. The closest approach is to fail to disprove the Alternative Hypothesis. In analytical chemistry, one can never absolutely prove that the signals detected from the unknown are due to the suspect compound, because there will always be some chance (perhaps very remote) that the signals have arisen from some hitherto unknown compound or phenomenon. Science does not offer a technical means of proving absolutely, beyond all possible doubt, that a suspect compound is present.

It is necessary to address *qualitative uncertainty* in order to show that qualitative methods are fit for purpose. The Null Hypothesis is reasonable if the two data sets are the same within some degree of confidence. A suspect compound can be considered confirmed if the criteria are drawn so the possibility of error is small enough to meet the needs of the analysis.

If the analyst or reviewer concludes on the basis of such data that a suspect compound is *absolutely* present, this conclusion is an interpretation and not a feature of the method or data. Interpretations are a matter of debate. Our recommendation in this regard is to not to make too great a leap from the data. Shallow asserts that any subjectivity on the part of the analyst's interpretation of the results is a potential basis for rejecting scientific testimony [12]. The defense may raise a reasonable doubt if it can be shown that the analyst overstated the objective basis for confidence in the identification. This might arise from making an "absolute" identification on the basis of a weak library-match or using data from highly variable or non-specific ions. *Approaches to assessing identification confidence.* Analytical uncertainty may be assessed either in a holistic manner (empirical or top-down) or as an accumulation of individual steps (statistical, or bottom-up). The approaches differ, but each can be used to support a method's fitness for purpose.

In the bottom-up approach, each step of a method contributes to overall specificity. This approach yields the "selectivity index" advocated by Stephany and van Ginkel in several papers [13–16]. It can be a basis for comparing the resolving power of various detection schemes, and for allowing different confirmation criteria depending on detection technology. This approach is the underlying principle behind a recently-proposed system of *Identification Points* for confirmatory analysis [17]. In principle this is reasonable, but assigning a value for Selectivity Index of a given method requires

Figure 5. Towards consensus selectivity indices. The scheme attempts to illustrate two inter-related points: (a) That a consensus of experienced analysts might consider individual steps of various methods to possess comparable selectivity (reference [14]); (b) that the cumulative selectivity of multiple steps in a single method might be described with Selectivity Indices (references [14–16]) or Identification Points (reference [17]).

assumptions, or else a considerable effort to build a consensus estimate. Figures 2, 5, and 6 are inspired by the Selectivity Index concept. In the absence of any other objective measure of identification confidence, this approach is always available, although it is necessarily imprecise. There has been at least one attempt to compare various methodologies on a consensus-basis, as is suggested in Figure 5 [14].

Prose	Numeric
	(highest selectivity)
Identified with utmost certainty	↑
Confirmed	⋮
Identified with confidence	⋮
Identified	⋮
Indicated	⋮
Tentative identification	⋮
Suspected	⋮
Presumptive	⋮
Non-Negative	↓
	(lowest selectivity)

Note: *Detected, Positive, Absent* are terms to avoid because of potential miscommunication of the associated degree of confidence.

Figure 6. Toward a lexicon for expressing identification confidence. The assignment of a numeric scale to these terms is not yet possible. However, the selectivity index of a screening method could be in single digits, while that of a method possessing "utmost certainty" could be some orders of magnitude.

In the top-down approach, empirical data is used to demonstrate acceptable identification confidence. Blank matrices are fortified from well below the predicted limit to well above it. The percent of such samples (true positives) found positive is plotted as a function of concentration to create a distribution function (in Figure 3, the heavy line). The selection of diagnostic signals determines the exact location of the transition zone. The chosen set of confirmation criteria need to have an acceptable diagnostic power for the target compound. It is also necessary to refer to the target level for full performance described in stage 1 of method development. If the acceptable rate of false positives is achieved at a concentration lower than the target level, Fitness for Purpose has been demonstrated empirically.

Another empirical approach to assessing identification confidence was described by Sphon in 1978, in a seminal publication that predates most other discussions of method fitness. When three ions from the EI spectrum of diethylstilbestrol (DES) were used to search a library of 30,000 spectra, and abundance criteria were applied to those ions, the only matching spectrum was that of DES [18].

X. Measurement Uncertainty

In the bottom-up approach to measurement uncertainty, total method error is calculated by propagating uncertainty through each step of the method. This calls for a thorough theoretical review of the method, but ultimately may not account for all of the experimental error. In the top-down approach, the method's precision and accuracy are determined experimentally by replicate analyses of true positives. These data result from a conventional method validation. A working range is determined according to conventional procedures for precision and accuracy.

Assessing quantitative uncertainty at the lower limit. Quantitative analytical uncertainty has been closely examined over the years. These examinations depend on constructing a theoretical *distribution function*. Without some model for the distribution functions, quantitative uncertainty cannot be calculated. The distribution function could be defined by repetitive measurements of samples where the analyte is either absent (case 1), present at the detection limit (case 2), or present at some quantitation limit (case 3).

The classical "detection limit" is based on repetitive measurements of the background signals when analyte is absent. Statistical theory based on an assumed normal (Gaussian) distribution holds that the mean plus some multiple k of the standard deviation gives the detection limit, and k defines the confidence level that a signal has been differentiated from noise. For example, $k = 2$ gives 95% confidence, $k = 3$ gives >99% confidence, etc. These confidence levels translate to the probability that a given signal has been differentiated from noise

One may assume that analyte is present at the level equivalent to this limit of detection. However, to calcu-

late the distribution functions for this or the first case, it would be necessary to obtain values for measurements below the detection limit or below zero. This paradox creates practical problems in determining what represents a true signal. As an alternative, one may assume that the analyte is present at the quantitative limit determined from one of the first two cases. In this scheme the quantitation limit corresponds to $k = 6$, so that the confidence level (based on $k = 3$) addresses whether a true signal has been detected. There is a defined confidence that the measurement is above the detection limit, and the distribution function is defined with true signals. Our recommendations are that the uncertainty at the lower limit is better found by extrapolation from the working range, not by measurements at or below the lower limit.

Various assumptions about detection limits define two regions of false outcomes. The α -error defines false positive measurements, where the found value is above the limit although analyte was not present. The β -error defines false negative measurements, where the measurement was below the limit although analyte was present.

Assessing quantitative uncertainty at the lower limit calls for defining the distribution functions with true signals, along with an explicit understanding of k that is fit for the purpose. Our recommendations on method fitness would hold that limits are determined by (1) agreeing on a given uncertainty target; (2) choosing the appropriate k value for that uncertainty target; (3) defining the distribution function with true signals; (4) setting the limit at a value corresponding to a predetermined level of uncertainty.

References

1. Limits to Confirmation, Quantitation, and Detection, ASMS Fall Workshop; Alexandria, VA, Nov. 1996.
2. Baldwin, R.; Bethem, R. A.; Boyd, R.; Budde, W. L.; Cairns, T.; Gibbons, R. D.; Henion, J. D.; Kaiser, M. A.; Lewis, D. L.; Matusik, J. E.; Sphon, J. A.; Stephany, R. W.; Trubey, R. K. Report on 1996 ASMS Fall Workshop: Limits to Confirmation, Quantitation, and Detection. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 1180–1190.
3. Bethem, R. A.; Boyd, R. K. Mass Spectrometry in Trace Analysis. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 643–648.
4. Documents used in the 2001 survey of draft guidance documents. (a) COMMISSION OF THE EUROPEAN COMMUNITIES, draft SANCO/1805/2000 Rev. 1, 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. (b) Document C43-P, Vol. 20, No. 9, Gas Chromatography/Mass Spectrometry (GC/MS) Confirmation of Drugs; Proposed Guideline, Draft prepared for comment period ending November 2000, by National Committee for Clinical Laboratory Standards, Wayne, PA. (c) U.S. Department of Agriculture, Agricultural Marketing Service, Pesticide Data Program (www.ams.usda.gov/science/pdp/SOPs.htm) SOP no. PDP-QC-06, Minimum Requirements for QC/Quadrupole Mass Spectrometer Confirmation in EI mode, Effective 04/01/00. SOP no. PDP-QC-09, Minimum Requirements for

- the Operation of a Mass Spectrometer with MS/MS Capabilities, Effective 04/01/00. SOP no. PDP-QC-10, Determination of LOD and LOQ for Chromatographic Methods, Effective 02/01/96. SOP no. QC-11, Ion Trap Confirmation Guidelines, Effective 3/2/94. (d) Reports no. KOA 98.222 and KOA-00.100, Identification criteria for the GC-MS analysis of environmental contaminants in various matrices, prepared by KIWA N.V. for Netherlands laboratories that monitor for environmental contaminants in water and soil. Also described by Geerdink [10]. (e) FDA Center for Veterinary Medicine, Draft Guidance Document no.118, Mass Spectrometry for Confirmation of the Identity of Animal Drug Residues, version of June 2001.
- Thompson, M.; Ramsey, M. H. Quality Concepts and Practices Applied to Sampling: An Exploratory Study. *Analyst* **1995**, *120*, 261–270.
  - Eurachem Guide. *The Fitness for Purpose of Analytical Methods, a Laboratory Guide to Method Validation and Related Topics*; LGC: Teddington, UK, 1998; p 1.
  - De Boer, W. J.; van den Voet, H.; de Ruid, W. G.; van Rhijn, J. A.; Cooper, K. M.; Kennedy, D. G.; Patel, R. K. P.; Porter, S.; Reuvers, T.; Marcos, V.; Munoz, P.; Bosch, J.; Rodriguez, P.; Grases, J. M. Optimizing the Balance Between False Positive and False Negative Error Probabilities of Confirmatory Methods for the Detection of Veterinary Drug Residues. *Analyst* **1999**, *124*, 109–114.
  - Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770–781.
  - Vessman, J.; Stefan, R. I.; Van Staden, J. F.; Danzer, K.; Lindner, W.; Burns, D. T.; Fajgelj, A.; Muller, H. Selectivity in Analytical Chemistry. *Pure Appl. Chem.* **2001**, *73*, 1381–1386.
  - Geerdink, R. B.; Niessen, W. M. A.; Brinkman, U. A. T. Mass Spectrometric Confirmation Criterion for Product-Ion Spectra Generated in Flow-Injection Analysis: Environmental Application. *J. Chromatogr. A* **2001**, *910*, 291–300.
  - Daubert versus Merrell Dow Pharmaceuticals (92–102), 509 U.S. 579 (1993).
  - Shellow, J. M. The End of a Confidence Game—a Possible Defense to the Impossible Drug Prosecution. *Champion* **2000**, *24*, 22–32.
  - Stephany, R. W.; van Ginkel, L. A. Quality Criteria for Residue Analysis and Reference Materials. Relationship Between Legal Proceedings and Materials. *Fresenius J. Anal. Chem.* **1990**, *338*, 370–377.
  - Van Ginkel, L. A.; Stephany, R. W. Experimental Chemometrics, an Alternative Way for Estimating Overall Reliability. *Proceedings of the EuroResidue II Conference on Residues of Veterinary Drugs in Food*; Haagsma, N., Ed.; Veldhoven, Netherlands, May, 1993; pp 303–307.
  - Stephany, R. W. Validated Methods or Valid Test Results for Residue Analyses. *Proceedings of the EuroResidue IV Conference on Residues of Veterinary Drugs in Food*; Veldhoven, Netherlands, May, 2000; pp 137–147.
  - Stephany, R. W. Hormones Residue Testing: An update in Research and Approaches. *Proceedings of the EuroConference: Safety Assurance During Food Processing*; Vienna, Austria, October, 2001.
  - André, F.; Wasch, K.K.G.; De Brabander, H.F.; Impens, S.R.; Stolker, L.A.M. Trends in the Identification of Organic Residues and Contaminants EC Regulations Under Review. *Trends Anal. Chem.* **2001**, *20*, 435–445.
  - Sphon, J. A. Use of Mass Spectrometry for Confirmation of Animal Drug Residues. *J. Assoc. Off. Anal. Chem.* **1978**, *61*, 1247–1252.