PDA Journal of Pharmaceutical Science and Technology



Identifying and Mitigating Errors in Screening for Organic Extractables and Leachables: Part 2—Errors of Inexact Identification and Inaccurate Quantitation

Piet Christiaens, Jean-Marie Beusen, Philippe Verlinde, et al.

PDA J Pharm Sci and Tech **2020**, 74 108-133 Access the most recent version at doi:10.5731/pdajpst.2018.009779

TECHNOLOGY/APPLICATION

Identifying and Mitigating Errors in Screening for Organic Extractables and Leachables: Part 2—Errors of Inexact Identification and Inaccurate Quantitation

PIET CHRISTIAENS¹, JEAN-MARIE BEUSEN¹, PHILIPPE VERLINDE¹, JAN BAETEN¹, and DENNIS JENKE^{1,*}

¹Nelson Labs, Romeinsestraat 12, 3001 Heverlee, Belgium © PDA, Inc. 2020

ABSTRACT: Patients can be exposed to leachables derived from pharmaceutical manufacturing systems, packages, and/ or medical devices during a clinical therapy. These leachables can adversely decrease the therapy's effectiveness and/or adversely impact patient safety. Thus, extracts or drug products are chromatographically screened to discover, identify, and quantify organic extractables or leachables. Although screening methods have achieved a high degree of technical and practical sophistication, they are not without issues in terms of accomplishing these three functions. In this Part 2 of our three-part series, errors of inexact identification and inaccurate quantitation are addressed. An error of inexact identification occurs when a screening method fails to produce an analyte response that can be used to secure the analyte's identity. The error may be that the response contains insufficient information to interpret, in which case the analyte cannot be identified or that the interpretation of the response produces an incorrect identity. In either case, proper use of an internal extractables and leachables database can decrease the frequency of encountering unidentifiable analytes and increase the confidence that identities that are secured are correct. Cases of identification errors are provided, illustrating the use of multidimensional analysis to increase confidence in procured identities. An error of inaccurate quantitation occurs when an analyte's concentration is estimated by correlating the responses of the analyte and an internal standard and arises because of response differences between analytes and internal standards. The use of a database containing relative response factors or relative response functions to secure more accurate analyte quantities is discussed and demonstrated.

KEYWORDS: Extractables, Leachables, Chromatographic analysis, Screening analysis, Identification, Database, Quantitation, Internal standard, Response factor.

Introduction

When drug products are manufactured, packaged, and administered, they unavoidably and inevitably contact items such as manufacturing components, packaging systems, and administration devices. During contact, substances present in or on these items can be transferred to the drug product where they become foreign impurities known as leachables. When a drug product is administered to a patient during clinical therapy, the patient is exposed to the leachables. As foreign impurities, leachables could adversely affect the

* Corresponding Author: Triad Scientific Solutions, LLC, 181 Peregrine Lane, Hawthorn Woods, IL 60047; Telephone: (1)-847-393-3151; e-mail: dennisjenke@triadscientificsolutions.com doi: 10.5731/pdajpst.2018.009779 drug product's suitability for its intended use, including patient and user health and drug product attributes such as quality, stability, efficacy, and compliance.

Thus, drug products are tested for foreign impurities (leachables), and extracts of contacted items are tested for extractables (as potential or probable foreign impurities) so that the foreign impurities can be identified, quantified, and ultimately assessed for potential adverse effects (1, 2).

When an extract is tested for organic extractables (or a drug product is tested for organic leachables), the desired outcome is to account for all extractables uniquely present in an extract (versus an extraction blank) above an established threshold or to establish all leachables uniquely present in a drug product above an established threshold. This desired outcome is achieved by analyzing the extract or drug product (and any associated blank or control) with chromatographic methods that are able to produce useful and interpretable responses for potential extractables or leachables (3– 5). If the extractables in an extract or leachables in a drug product are not or cannot be specified upfront, they must be discovered, identified, and quantified by an analytical process termed screening.

In Part 1 of this series (6), the process of chromatographically screening extracts or drug products for organic extractables or leachables was established as having three primary objectives:

- 1. accounting for (discovering) organic substances (extractables or leachables) present in a test sample at a concentration above a defined threshold;
- 2. identifying the discovered substances; and
- 3. quantifying the identified substances.

As was also noted previously, practical and scientific limitations of the chromatographic screening process impede the process's ability to fully accomplish these objectives. Thus, Part 1 of the series also considered errors of omission, where an error of omission involves the situation that the screening method fails to produce a recognizable response for one or more of the analytes present in a sample (extract or drug product). Thus, the omission error involves the discovery aspect of screening.

Once all of the extractables in an extract or leachables in a drug product at levels above a justified reporting threshold have been accounted for (discovered), the identities of the individual extractables or leachables must be established as identity links an extractable or leachable to that information which enables its assessment. Considering safety, for example, it is the extractable's identity that links the extractable to its relevant toxicological safety information. Clearly, if an identity cannot be secured or if the secured identity is incorrect (errors of inexact identification), then either the assessment cannot be performed or the assessment that is performed is faulty. Additionally, the discovered extractables must be quantified, as it is the quantity of an extractable in an extract (or a leachable) in a drug product that establishes a patient's exposure (or potential exposure) to the substances. Clearly, inaccurate quantitations lead to erroneous safety assessments that either underestimate or overestimate the safety hazard.

Errors of Inexact Identification

Identification Hierarchy

As was noted in Part 1 of this series, although a screening assay produces a response that contains information that can be used to infer an identity, the response itself is not an identity. It is only with further processing and/ or interpretation that the response's information can lead to an identification. Thus, screening methods do not identify substances; rather, the screening method produces data that are further interpreted to provide an identity.

Rigorously speaking, an error of inexact identification occurs when (1) the response contains no identifying information, (2) an identity cannot be inferred from the response data, or (3) the inferred identity is not the correct identity. Understanding and addressing errors of inexact identification is facilitated if one understands the data interpretation process that is most commonly used for compound identification. The process is based on the observation that an identification derived from information represents a guess at the identity, where the confidence one has in the guess depends on the amount and the nature of the information that suggests, supports, and ultimately confirms the identification. The more information that is available, and the more rigorous the available information, the greater confidence one can have that the inferred identity is the correct identity. Thus, one can "grade" an identification based on the level of confidence one has that the identification is correct. Such a "grading" is captured, for example, in USP <1663> Assessment of Extractables Associated with Pharmaceutical Packaging/Delivery Systems (5). In this monograph, four "grades" of identifications are proposed: unidentified, tentative, confident, and confirmed. These various "grades" are ranked in terms of conciseness; thus, for example, a tentative identification specifies the chemical class of a substance, a confident identification specifies a specific structure that precludes all but the most closely related structures, and a confirmed identification specifies an exact identification. Moreover, USP <1663> describes what type of supporting analytical information is required to move from one identification "grade" to a higher "grade".

An alternate, albeit similar, approach to identification "grading" is illustrated in Figure 1, which introduces a fifth identification "grade", partial. Furthermore, the tentative and confirmed grades are divided into two





Identification hierarchy.

subgrades depending on the means by which the tentative or confirmed identification is secured. Understanding this "grading" scale is facilitated if one considers the identification process most commonly employed with the chromatographic methods used in extractions and leachables (E&L) screening.

In understanding the identification process, one notes that the information most commonly collected and used for E&L identification purposes is a mass spectrum. The chromatographic methods used for screening employ mass spectral detectors and thus the resultant response to an eluted analyte is its mass spectrum. In certain circumstances, the mass spectrum may provide enough information to infer the structural characteristics of the compound of interest (i.e., the spectrum may contain certain diagnostic masses) but will not provide enough definitive information to link the mass spectrum to a specific compound (secure a tentative identity). This level of identification is thus termed a partial identification. For example, reporting that a compound is a phthalate, but not being able to specify the specific phthalate means that the identity is partial. Although clearly a rigorous toxicological safety assessment cannot be based on a partial identity, partial identities may be sufficient to facilitate some level of safety assessment. For example, quantitative structure-activity relationship (QSAR) analysis of a compound's structural characteristics (e.g., via DEREK or SARAH), can be used to establish whether the structural characteristics are associated with an increased risk of an adverse safety effect (e.g., mutagenicity). Compounds without QSAR-alerting structures represent less of a safety hazard than do compounds with QSAR-alerting structures.

If we have the situation that the mass spectrum is sufficiently robust that a name can be proposed for the compound, such a tentative identification can be secured in one of two ways: (1) the mass spectrum can be linked to a mass spectral library to find a matching spectrum (which implies that the compound that produced the library spectrum is the compound that produced the analytical spectrum) or (2) the mass spectrum can be interpreted from first principles to infer the analyte's structure (and thus its identity), a process that is termed structure elucidation. Thus, a compound can be identified, in a tentative way, based on one-dimensional data analysis such as matching or elucidation (interpretation).

If an external mass spectral database or library (i.e., a database constructed and populated by a third party) exists, then mass spectral matching can be performed. For example, the NIST/Wiley MS libraries are often utilized to secure identities in gas chromatography–

mass spectroscopy (GC-MS) analyses. In this case, the mass spectrum obtained for a compound via a screening method is compared to the mass spectra contained in the database, establishing those compounds in the database whose mass spectra closely match the spectrum of interest. Closely matching spectra are assigned a "match score" by a number of algorithms, with a higher match score corresponding, at least in principle, to the more probable identification. If a suitable match is obtained, then the identification is classified as tentative as one piece of supporting information has been secured (onedimensional identification).

If a match is not secured or in the complete absence of any database, the analyte is initially classified as unidentified (or unknown, as this term is commonly used) and remains unidentified until further actions are taken to establish its identity. For example, identification of an unidentified extractable can be accomplished via the process of structure elucidation, which involves the professional interpretation of the available mass spectral information. For instance, an expert could interpret the mass spectrum's fragmentation pattern to elucidate the analyte's probable functional groups and structural units. The individual groups and units can then be "assembled" to infer all or part of the analyte's structure. Generally, the elucidation process is inefficient and is prone to error and variation, as individual experts could easily come to different outcomes based on their experience and capabilities. Such an identification so secured is interpretive by its very nature and is properly classified as a tentative identification as it is based on one-dimensional data (hence the additional identification "grade" in Figure 1).

A third method for securing the identity of an unidentified analyte is to collect additional information about the analyte by subjecting the test sample to another method of analysis. For example, when it is possible to obtain and interpret a nuclear magnetic resonance (NMR) spectrum for the analyte, the interpretation could lead to the analyte's tentative (or perhaps confident) identification.

At this point in the identification process, we have either an unidentified analyte or an analyte that has been tentatively identified. Clearly, an unidentified analyte cannot be assessed for its impact (e.g., its effect on patient safety). Although an analyte that has been tentatively identified can be assessed for impact, it is understood that such an assessment is provisional as the certainty in the analyte's identity is lower. It is understood that a proper impact assessment can only be obtained when the tentative identity is "elevated" in grade by securing confirmatory information (two-dimensional identification). That is, one's confidence that an identity is correct is increased (a tentative identification becomes a confident identification) when a second dimension of confirmatory data is secured. For example, although a mass spectral match and an NMR interpretation separately produce tentative identities, the combination of a mass spectral match and an NMR interpretation would produce a confident identity as these are reinforcing pieces of information.

The term "confirmed" identity is typically reserved for an analyte whose tentative or confident identity has been investigated by analysis of a reference standard of the inferred compound (three-dimensional identification). The lower "grade" identification is confirmed if the key properties of the analyte have been matched to the key properties of the reference standard (e.g., mass spectral match and retention time match). However, one can envision a situation where the identification is supported by such a preponderance or "critical mass" of data (e.g., a three-dimensional identification) that it is virtually impossible that the identification can be in error. In such a case, surely the identity has been confirmed by the supporting data; thus, the possibility that a confirmed identity can be secured by either the traditional process of authentic standard matching (standard-based) or by the preponderance of mutually supporting data (data-based) is shown in Figure 1.

Given the "grading" scale, clearly an unidentified analyte reflects one type of error of inexact identification (i.e., the inability to secure an identity). Analytes that have been either partially or tentatively identified can be subject to the other error of inexact identification, which is misidentification. Given the additional confirmatory information that is required to secure a confident or confirmed identification, these identifications are not generally prone to inexact identification errors.

Lastly, it surely is the case that the distinction between a confident and a confirmed identity is vague and unclear, and that cases of high confidence (supported, for example, by three dimensions of compelling information) might, for all practical purposes, be considered to produce a confirmed identity. It such cases it is the high degree of confidence that is more important and the exact "grade" that is less important.

Error of Inexact Identification: A Fatal Error

Commission of an error of inexact identification is a fatal error because such an error precludes a proper assessment. If the error of inexact identification is that an identity cannot be secured, then clearly the substance's impact on the drug product's suitability cannot be assessed as the link between the extractable and its relevant related information cannot be established. If the error of inexact identification is that the wrong identity is secured (see Figures 2–7 for examples of this type of identification error), then clearly the substance's impact on the drug product's suitability cannot be correctly assessed. This is the case as the assessment is based on information relevant to what could be a completely unrelated compound.

The existence of an external database, that is, a database constructed and populated by a third party, addresses the aspect of errors of inexact identification to a certain extent. For example, as noted previously, the NIST/Wiley MS libraries are often utilized to secure identities in GC-MS analyses. In this case, the mass spectrum obtained for a compound via a screening method is compared to the mass spectra contained in the database, establishing those compounds in the database whose mass spectra closely match the spectrum of interest. Closely matching spectra are assigned a "match score" by a number of algorithms, with a higher match score corresponding, at least in principle, to the more probable identification.

"Simple" Identifications Using an External Database

Although spectral matching via an external database is a commonly employed and generally effective means of securing a compound's tentative identity, it is not without its problems. The first and foremost problem is that the more commonly used databases were not constructed with the intent of systematically, specifically, and completely addressing extractables or leachables. Rather, the external databases were constructed and populated to include compounds relevant to different situations encountered in a broad range of different industries (food, chemical, environmental, pharmaceutical, and so forth); thus, the presence of extractables/ leachables in these databases is incidental as opposed to intentional.

For example, many organizations involved in E&L testing refer to the Environmental Protection Agency

(EPA) Methods for Environmental Monitoring of Toxic or Hazardous Compounds (e.g., Method 8260 for Volatile Organic Compounds; Method 8270 for Semi-Volatile Organic Compounds) (7, 8) as a proper and valid methodology to discover, identify, and quantify extractables. Indeed, these methods will detect, identify, and quantify the many toxic and hazardous compounds that are targeted by each EPA method in an accurate and precise way. The selection of target compounds for these EPA methods was based on a careful evaluation of compounds that could be released into the environment in large enough quantities by a broad range of different industries that they could have a detrimental effect on the environment.

This aspect of the external database may confound its use in extractables or leachables screening. A careful evaluation of these lists of compounds suggests that only 10% of the volatile substances on the EPA list are relevant as volatile organic extractables and only 5% of the semivolatile compounds on the EPA list are relevant as semivolatile extractables or leachables. Thus, the EPA database linked to its methods does not include many of the commonly encountered extractables and includes a majority of compounds that are not extractables.

This situation produces two issues. The first issue is the content of the external database. Because the EPA database does not contain a large number of potential extractables, an extractable that needs to be identified is likely not in the database and thus a match will not be secured. If a match cannot be secured then clearly an identity cannot be proposed. Thus, the first issue is obtaining no identification, essentially concluding that the substance in question is unidentifiable.

The second issue is securing the wrong identity because of both the size and the content of the external database. Because the external databases contain so many compounds that are not extractables (e.g., the combined NIST/Wiley '17 Mass Spectral Library contains over 1 million mass spectra (9)), possibly a target spectrum will be closely matched to spectra from compounds that are not extractables, leading to false identifications. This problem is exacerbated to a certain extent by securing "simple identities", which, for example, is the process of accepting that a compound's identity is established by the match with either the highest match score or an acceptably high match score. Although the match score is an effective means of differentiating



Example of an exact identification via mass spectral fit with an authentic standard (1,1'-carbonothioyl-bispiperidine) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (incorrectly identified as 1-aminocyclopentane-carboxylic acid, N-hexyloxy-carbonyl, isohexyl ester).

potential identities from unlikely or impossible identities, the resolution between similar match scores is not always adequate to arbitrarily select the proper match based on generalizations such as "the highest match score always wins" and "any identity with a match score above 80 must be a good identity".

Home Court Advantage of an Internally Developed Database

These issues notwithstanding, perhaps the most significant issue associated with use of an external database is the degree to which the experimental conditions



Example of an exact identification via mass spectral fit with an authentic standard (dicyclopentylsilanediol) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (incorrectly identified as 4-pentenoic acid, 2,2-diethyl-3-oxo-5-phenyl-, ethyl ester).

used to produce the information in the external database match the analytical conditions employed by the individual testing laboratory. The closer the match between the analytical conditions used to generate the information in the database and the analytical conditions that produced the information to be matched to the database, the better will be the outcome of the match (which is a potential identification). The poorer the match between analytical conditions, the more likely that either no identifications or false identifications will be generated when matching experimental data to the database.



Example of an exact identification via mass spectral fit with an authentic standard (1,4,7-trioxacyclotridecane-8,13-dione) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (incorrectly identified as succinic acid, 2-(2-chloro-phenoxy)ethyl ethyl ester).

For example, consider the case of mass spectral matching in GC. Because the industry has standardized the ionization conditions (electron impact at 70 eV), for a number of extractable and leachable compounds there is a generally good correlation between spectra in an external database and spectra obtained by independent laboratory analyses. However, lack of standardization in MS applied in liquid chromatography (LC) means that likely spectra contained in an external database were not collected under conditions that match the conditions used by an independent laboratory to collect its spectra, thus increasing the possibility that matching between the external database and the independent laboratory will produce aberrant identities.



Example of an exact identification via mass spectral fit with an authentic standard (PBT dimer) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (incorrectly identified as 2,6-bis [(5S)-5-phenyl-4,5-dihydrooxazol-2-yl]pyridine).

Most of the problems associated with the use of an external database are solved or ameliorated via application of an internally generated database. As the internal database contains only extractables and as the internal database is grown to include a significant number of extractables, the issues noted earlier in terms of producing no hits or producing aberrant hits are reduced.

Furthermore, the concept of "home court advantage" comes into play. That is, when a laboratory produces an internal database it certainly does so with the exact analytical



Example of an exact identification via mass spectral fit with an authentic standard (chlorinated rubber oligomer $C_{13}H_{23}Cl$) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (incorrectly identified as bicyclo[3.2.0]heptan-2-one, 6-hydroxy-5-methyl-6-vinyl-).

methods and conditions it uses to screen extracts for extractables (and drug products for leachables). Because the laboratory conditions for producing the reference and test spectra are closely matched, the test and database spectra will be more closely matched, resulting in better match scores that will more effectively differentiate the true identity from a smaller number of false potential identities. Home Court Advantage in Action: Examples of Inexact Identification Resulting from Spectral Matching via an External Database (Commercial Spectral Library)

Figures 2–7 provide examples that illustrate errors of identification propagated via spectral matching to an external database and corrected via matching to an



Example of an exact identification via mass spectral fit with an authentic standard (brominated rubber oligomer $C_{13}H_{23}Br$) in the Database (with retention time confirmation) versus an inexact "simple" identification, relying only on the mass spectral fit with the external NIST/Wiley mass, leading to a fatal identification error (wrongly identified as ethanone, 1-(3,3-dimethylbicyclo[2.2.1]hept-2-yl)-, exo-).

internal database. In each example, an electron impact (EI) mass spectrum was obtained for the compound and the compound's possible identity was secured by spectral matching to either an external commercial spectral database (NIST/Wiley) or an internal spectral database (generated by Nelson Labs, hereafter referred to as the

"Database") constructed from the analytical information obtained through analyzing authentic standards via the generic screening method for GC-MS.

Each example includes the analytically obtained data (retention time and mass spectrum) as well as data for

the "best hit" from each of these libraries, including library retention time in minutes (if available), spectral match score (%), spectrum, and structure. In all cases, the identity obtained with the internal database is correct whereas the identity obtained with the external database is incorrect.

The example in Figure 7 provides an opportunity to explore the difference between a partial and a tentative unknown. Let us, for a moment, imagine that the Database entry for this compound does not exist, meaning that either the investigator seeking to identify the compound does not have the database or that the database does not contain this compound. The investigator would have the match to the NIST/Wiley database to work with but the low match score and a visual comparison of the test spectrum to the match spectrum both suggest that the match is not the right identity for the test compound. Thus, without a match, an identification of "unidentified" seems proper. However, clearly the mass spectrum is interpretable, with the most obvious feature being the clear indication that the compound contains bromine (similar sized peaks at m/z 146, 147). Thus, even a cursory interpretation of the mass spectrum produces the partial identity of "bromine-containing compound". Further elucidation, and some familiarity with the E&L literature, might allow the investigator to conclude that the compound of interest is a "brominated rubber oligomer", which is a more detailed identification but still a partial one. That is, the spectral features of ions m/z 97, 57, 123, and 67 are also observed in the spectrum of the nonbrominated rubber oligomers. Spectral elucidation by either manual or software-assisted (e.g., MS interpreter, MassFrontier) interpretations strengthens this hypothesis by explaining the most abundant ions. For instance, the base ion m/z 97 is known to arise from ring cleavage. This information is sufficient to propose that the compound is a "brominated rubber oligomer".

Finally, identification of the molecular ion (the spectrum in Figure 7 shows clear molecular ions at m/z 258 [corresponding to $C_{13}H_{23}^{79}Br$] and 260 [corresponding to $C_{13}H_{23}^{81}Br$]) could allow the investigator to establish the compound of interest as the $C_{13}H_{23}Br$ rubber oligomer (molecular weight = 259), which is a tentative identification.

Taking this discussion one step further, if the extract were retested using a methodology that produces accurate mass information, such information could be used

Vol. 74, No. 1, January-February 2020

to provide a chemical formula that also provides a tentative identification of C13 brominated rubber oligomer. The two independent derivations of the same tentative identification, once by mass spectral interpretation and once by accurate mass MS to obtain the chemical formula, represents two-dimensional supporting data that taken together justify classifying the identification as confident.

The purpose of this discussion about identifications based on database matches is not to discourage the use of external databases, such as NIST/Wiley, which has the specific advantage of being curated by a government agency and reviewed by a wide scientific community. Rather, this discussion properly points out the identification error that can occur when identifications are only based on a match score and justifies the assertion that better matches can be obtained with a properly populated and maintained internal database.

Increasing Identification Confidence via Multidimensional Analysis with an Internal Database

To this point in the discussion, identification by comparing an analytical response to a database response has been one dimensional in the sense that the matching is based on one characteristic of the response. It is intuitively obvious that an identification secured on the basis of two (or more) independent characteristics is likely to be a more correct identification than one secured on the basis of a single characteristic. In fact, it is the ability to support an identification on the basis of multiple characteristics that differentiates between the commonly applied identification categories such as tentative, confident, and confirmed.

When one generates one's own internal database, one standardizes not only the operating conditions of the mass spectrometer but also the analytical conditions preceding the mass spectrometer (e.g., the chromato-graphic conditions). So doing introduces a second, confirmatory characteristic to identification, retention time (absolute or relative). That is to say that an identity secured via a MS match can be confirmed by a retention time match, internal database versus laboratory test conditions. An example of an internal database that contains such complementary identifying information was presented as Table I in Part 1 of this series (6). Furthermore, two-dimensional mapping was illustrated in the examples contained in Figures 2–7, where in addition to

spectral matching, the retention times matched, internal database versus actual analysis.

The elevation of tentative identities derived from mass spectral matching to confident or confirmed identities based on supporting information is consistent with standard and recommended laboratory practices in extractables and leachables screening, which dictate that identities obtained by spectral matching alone are identified as tentative identities unless and until they are confirmed with additional supporting data. Such supporting information is rarely present in an external database, as reference spectra in that database may be obtained either (a) without a chromatographic separation (in which case there is no basis for a retention time match) or (b) by a chromatographic separation that is operationally different from the separation used in the independent laboratory (in which case the retention time match may be difficult to establish).

Furthermore, in certain situations it may be standard laboratory practice to collect additional information when performing extractables or leachables screening. For example, performing high-resolution mass spectrometry as the chromatographic detection approach is becoming more prevalent in extractables and leachables screening, thus providing accurate masses for extractables and leachables, which in turn can be used to specify potential empirical formulas. Although the empirical formula derived from accurate mass data might be adequate by itself to produce a tentative identity, it is more likely that the empirical formula is used with other analytical data to secure a confident identity.

It is logical and appropriate to note that an identification secured by mass spectra matching and supported by definitive and rigorous secondary information (such as retention time and empirical formula matches) is most likely the correct identification. However, even in the case of low information content detection methods such as ultraviolet (UV) absorbance, features of the detector response may be instrumental to and adequate for differentiating between candidate speculative or tentative identities.

Errors of Inaccurate and Imprecise Quantitation

General Discussion

Once an extractable has been discovered in an extract (or a leachable in a drug product) above a reporting

threshold, it becomes a candidate for impact assessment; for example, establishing its potential adverse effect on patient safety. The impact assessment considers two factors, the leachable's intrinsic ability to produce an effect (e.g., the leachable's safety hazard) and the amount of the leachable available to cause the effect. (e.g., a patient's exposure to the leachable). Specifically, it is the substance's identity that links it to its relevant effect-indicating information (e.g., toxicity data), thus facilitating its proper impact evaluation. However, the impact assessment cannot be completed without quantifying the patient's or drug product's exposure to the substance, which is established only by accurately determining the concentration of the substance in either the extract or the drug product. This is the case for safety assessment as a patient's exposure to a leachable (in terms, for example, of mass per day) is determined as the product of a leachable's concentration in a drug product and the daily dose volume of the drug product.

As a consequence, the most relevant and meaningful impact assessment is based on a sufficiently accurate and precise quantitation of the extractable in the extract (or the leachable in the drug product). Screening methods, however, do not always achieve these desired quantitation goals as the most commonly employed approaches to quantitation are prone to errors of inaccurate and imprecise quantitation. These errors, and their resolution by proper use of an internal database, are considered in greater detail as follows.

It is intuitive that the quality of reported extractable's or leachable's concentrations is also strongly influenced by method characteristics other than quantitation method; for example, extraction yields during sample preparation. Moreover, proper design and execution of the extraction itself is critical in terms of ensuring that the extractable's concentrations are relevant and meaningful. However, these topics are outside the scope of this series.

Errors of Inaccurate and Imprecise Quantitation

As was established in both Part 1 of this series and earlier in this Part 2, errors of omission or misidentification are considered fatal to the impact assessment as such errors irrevocably compromise the assessment's validity and applicability. An extractable that is not surfaced during screening (error of omission) escapes assessment, whereas an extractable that has not been identified correctly (error of identification) is inappropriately and incorrectly evaluated. Conversely, errors in quantitation are not strictly fatal as they do not preclude the impact assessment; rather, quantitation errors skew the assessment by either exaggerating or underestimating the true impact.

In this Part 2 of the series, two errors of quantitation are considered; the error of estimation (so-called "simple" quantitation) and the error of extrapolation.

"Simple" Quantitation-Error of Estimation

It goes without saying that the most accurate estimate of an analyte's concentration in a sample is obtained when the test method's response to the analyte has been calibrated via the generation of a calibration curve obtained by the analysis of standards prepared to contain the analyte at known concentrations. However, given the large and diverse population of potential extractables and leachables and the circumstance that most extractables or leachables profiles consist of numerous and largely unpredictable substances, generation of a response curve for each individual extractable (or leachable) is impractical, if not impossible, and hence rarely performed during screening. Rather, the quantitation of extractables and leachables in screening is most commonly performed using an alternative approach that can be described as "simple" quantitation.

In this approach, concentrations of extractables and leachables are estimated via a single reference compound, an internal standard, which is added to the extract or drug product in a known quantity before analysis. For quantitation, a response factor for the internal standard (RF_{IS}) is determined as the ratio of its analytical response (R_{IS}) versus its concentration in the sample (C_{IS}) (eq 1). The assumption that all compounds detected in screening mode exhibit the same concentration–response relation as the internal standard allows the concentration of each extractable or leachable in a sample to be estimated using eq 2, with R_{sample} and C_{sample} being the compound's observed analytical response and estimated concentration, respectively.

$$RF_{IS} = \frac{R_{IS}}{C_{IS}} \tag{1}$$

$$C_{sample} = \frac{R_{sample}}{RF_{IS}} \tag{2}$$

This assumption, that all extractables and leachables exhibit the same proportional response as a single reference compound, is the very root of estimation errors. When both the analyte and the internal standard respond similarly, concentration estimates obtained using the RF_{IS} can be highly accurate; as the responses of the analyte and the internal standard diverge, the concentration estimates become less accurate with the degree of inaccuracy increasing as the divergence increases. The degree of divergence between analyte and internal standard varies greatly across analytical methods and detection techniques. Gas chromatographic methods, targeting semivolatile compounds, typically use mass spectrometric or flame-ionization detection (FID). It is well established that many extractables and leachables exhibit a variation in GC-MS and GC/FID absolute response factors of a factor of ~ 4 (10–12). This means that if the response factor for an internal standard is arbitrarily assigned a value of 1, the majority of the substances' individual response factors will vary from 0.5 to 2.0. For example, if an extractable's concentration in an extract is calculated to be 1.0 mg/L via an internal standard, then the true concentration of the extractable will fall somewhere in the range of 0.5-2.0 mg/L. Of course, this generalization is not applicable to all extractables and there are many extractables whose absolute response factors fall well outside the range of 0.5-2.0 mg/L.

The case for LC is even more disproportionate, as it is well established that absolute response factors for the commonly applied LC detection methods (MS and UV absorbance) are, to use a scientifically rigorous term, "all over the place" (13, 14). Although there is less published data to quantify this statement for LC (versus GC), the range of individual response factors for LC analyses will likely vary from 0.1 to 10 (given an internal standard with a response factor of 1). Moreover, there likely are more extractables that fall outside this LC range than there are extractables that fall outside the smaller GC range. For example, most organic extractables will produce a GC/FID response because they are carbon-containing. However, an extractable without a UV chromophore would have a low (if any) UV response and would have its concentration substantially underestimated if it were "quantified" against an internal standard with a strong UV chromophore.

Home Court Advantage of an Internally Developed Database: Relative Response Factors

Previously, we established how a database containing relevant compound properties obtained through ana-

lysis of reference standards could tackle E&L screening errors related to identification. To address and correct errors of estimation, such a database must be expanded to include analytical information establishing each specific compound's concentration–response relationship.

In this context, analysis of authentic standards of compounds at known concentration in the presence of an internal standard (*IS*) during database development enables determination of each compound's response factor (*RF*) and relative response factor (*RRF*) against the internal standard (eqs 3 and 4),

$$RF_{compound} = \frac{R_{compound}}{C_{compound}}$$
(3)

$$RRF_{compound} = \frac{RF_{compound}}{RF_{IS}} \tag{4}$$

with C the known concentrations and R the observed analytical responses. Availability of the RRF values in the database allows a more accurate estimation of the concentration of an extractable (or leachable) according to eq 5.

$$C_{sample} = \frac{R_{sample}}{RRF_{compound}} \times \frac{C_{IS}}{R_{IS}}$$
(5)

In essence, the RRF serves as an "adjustment" factor. If an internal database includes RRF values for the relevant population of extractables, then the concentration obtained for each extractable can be adjusted with its corresponding RRF.

For example, suppose that an identified extractable in a sample produces a response of 2 units and that the sample, spiked to contain an internal standard at a concentration of 1 mg/L, has an internal standard response of 1 unit. Calculating the concentration following the "simple" quantitation strategy (eq 2) will result in the following value:

$$C_{sample} = 2 \times \frac{1 \text{ mg/L}}{1} = 2 \text{ mg/L}$$
(6)

Further suppose that this extractable is registered in an internally developed database with an RRF of 0.4. This means that it was experimentally established that when

a sample containing equal concentrations of extractable and internal standard was analyzed, the extractable produced a response of 2 units while the internal standard's response was 5 units. In this case, the estimated concentration (eq 5) becomes:

$$C_{sample} = \frac{2}{0.4} \times \frac{1 \,\mathrm{mg/L}}{1} = 5 \,\mathrm{mg/L}$$
 (7)

This example demonstrates that a more accurate, truly semiquantitative concentration can be obtained for every substance present in an extractable profile using each compound's experimentally determined RRF. It should be noted, however, that this approach cannot be applied to unidentified substances, as the RRF for an unidentified compound is not known. For such unidentified substances, either "simple" quantitation can be performed, or, preferably, one could apply RRF correction using the mean (or median) RRF obtained for the population of all the identified extractables present in the database for the relevant analytical method.

Example of a Database of RRFs

To illustrate the contents and use of an RRF database, a Database compiled for volatile, semivolatile, polar semivolatile, and nonvolatile compounds (analysis by, respectively, headspace (HS)-GC-MS, GC-MS, derivatization GC-MS and LC-MS) can be considered. This Database contains, in total, data for more than 5000 unique entries, more than 4000 of which have RRF values. The information in the database was collected across different techniques and reflecting experimentally encountered volatile, semivolatile, and nonvolatile organic extractables/leachables. Although the concept and reality of informational databases is not unique to Nelson (meaning that other organizations may have, to some extent, similar databases), clearly the authors are able to illustrate their points with the Database as they have access to that Database.

An overview of the distribution and general statistics of the RRF data present for the different techniques in the database is given in Table I and is depicted in Figure 8. In general, RRF values for each technique cover a large range of up to three orders of magnitude, although most RRF values fall within a range of 0.5– 2.0. The distribution of RRF values is such that the mean is larger than the median for all analytical techniques, which implies that the RRF data is skewed

FABLE I
Distribution of RRFs Related to Compounds in the Database for HS-GC-MS, GC-MS, Derivatization GC-MS, and
LC-MS (APCI & ESI)

Parameter	HS-GC-MS ^a	GC-MS	Deriv. GC-MS	LC-MS (APCI & ESI)
Total number of entries	987	3076	408	1985
Total number of entries with RRFs	246	2247	408	1129
RRFs < 0.1	22%	7%	3%	33% ^b
Median RRF	$0.640(0.909)^{c}$	0.590 (0.628)	0.949 (0.960)	$0.251 (0.541)^b$
Mean RRF	0.799 (0.997)	0.645 (0.687)	0.989 (1.017)	$0.543 (0.749)^b$
Standard Deviation	0.811 (0.728)	0.425 (0.407)	0.530 (0.512)	$0.921 (0.651)^b$
Minimum RRF	< 0.001	< 0.001	0.008	< 0.001 ^b
Maximum RRF	5.390	4.611	3.547	16.500 ^b

^a Data obtained by analysis of standard solutions prepared in ultrapure water (UPW).

^b LC-MS (APCI) RRF Data.

^c Values in parentheses correspond to the subset of "quantifiable" data with RRF values between 0.1 and 5.0.

toward lower response values (meaning there are more compounds that respond much more poorly than the internal standard than there are compounds that respond much more strongly than the internal standard). Thus, Table I and Figure 8 establish that each technique contains a number of compounds with exceptionally low RRFs (ranging from 3% to up to 34% of the total number of detectable compounds) and have fewer compounds with RRF values considerably greater than 1. This distribution of RRF values suggests that for each technique there are compounds that are insensitive by that technique, meaning that they are not amenable to quantitation using the specified analytical method. Although these poorly responding analytes are identifiable by the corresponding method, they cannot be reliably quantified by the



Figure 8

Cumulative distribution of RRFs (range 0–5) of compounds present in the Database for HS-GC-MS, GC-MS, derivatization GC-MS, and LC-MS (APCI) (see Table I for total number of RRFs per technique).

method because their estimated concentrations would be much lower than their actual value. In other words, although the method would be appropriate for establishing the identity and presence of the corresponding compound in the extract, it would be considered unacceptable for generating concentration estimates. The fact that a concentration estimate could be calculated is irrelevant as it is surely the case that if the calculated concentration were reported, any assessment performed using the reported number would be seriously flawed. In general, if a compound has an RRF value much less than 0.5 or much greater than 2, then its "simple" concentration is sufficiently erroneous that it is not proper to report such a concentration for assessment purposes. This range establishes whether an analyte is quantifiable (within the range) or unquantifiable (outside the range).

Using the RRF Database to Establish Which Method's Result to Report

It is not uncommon that an extractable is detected by more than one screening method and that the estimated concentrations obtained from the various methods differ substantially. When this occurs, a choice must be made in terms of which concentration to assess. Although it may be standard practice to "assess the highest reported concentration as the worst case", the proper practice is to assess the most correct result, which may not be the highest amount.

An RRF database allows the evaluation of a compound's response across the different techniques and facilitates the selection of the appropriate quantification screening method. Examples of such complementing RRF entries among different screening methods in the Database are given in Table II. For each relevant combination of two complementing techniques, examples are given of compounds that are quantifiable in one technique and only identifiable in the other technique and vice versa. For example, consider the case of acetophenone. Although conceivably acetophenone could produce both a HS-GC-MS and LC-MS response if it were present in an extract at a sufficiently high concentration, such as response could only be used to identify acetophenone as an extractable as its "simple" concentration by either method would be in error by a large amount given its low RRF values. Alternatively, its GC-MS RRF is such that acetophenone can be both identified and more accurately quantified by this method.

Using an RRF database, the most appropriate technique to quantify compounds can be selected when taking the desired quantitation range into account. If a dynamic range of 4 is desired and it is assumed that the database is centered around an RRF of 1.0, then compounds eligible for quantitation correspond to those with RRF values between 0.5 and 2.0. The concentrations of these compounds can be estimated by application of the internal standard RF ("simple" quantification) or, preferably, by RRF correction. Compounds with RRF values outside this range will have less accurate concentration estimates using "simple" quantitation and if "simple" quantitation is used these compounds should probably only be reported for identification purposes.

Although more accurate concentrations can be obtained for these compounds using the RRF values, the fact that their RRF value is either lower than 0.5 or higher than 2.0 suggests that the response behavior of the analyte is suboptimal and that perhaps, in some cases, a different method might be more appropriate. This situation is even more evident when the RRF is particularly low or high (e.g., 0.1 and lower or 5 and higher). Although mathematically an accurate concentration estimate can be obtained using any RRF value, an RRF value less than 0.1 or greater than 5 suggests that the method is not appropriate for quantitation of the analyte in question. Thus, even if a concentration is obtained by RRF for compounds with an RRF value less than 0.1 or greater than 5, the analyst is welladvised to report the analyte's concentration obtained with a different method whose RRF is closer to 1.0. If such a superior method is unavailable, then the RRF concentration should be reported with a notation that indicates the suboptimal RRF value.

When this criterion is applied to each technique's specific dataset (meaning that compounds with RRF values less than 0.1 or greater than 5 are removed from the dataset), it reduces the skew in the data as the mean and the median value become much closer (Table I).

Error of Extrapolation

In constructing a database of RRF values relevant to thousands of compounds covering multiple analytical techniques, the most efficient and generally applied strategy is to record RRF values at a single concentration of both extractable and internal standard. Although concentration estimates corrected with RRF values

TABLE II Examples of Complementing RRFs Across Orthogonal Techniques in the Database for HS-GC-MS, GC-MS, and LC-MS (APCI)

		Technique Specific RRF Values		
Cas	Compound Name	HS-GC-MS	GC-MS	LC-MS
Complementing HS-GC-MS & GC-MS & LC-MS RRF Entries				
95-16-9	Benzothiazole	0.003	0.901 ^{<i>a</i>}	1.209
100-52-7	Benzaldehyde	0.042	0.699	0.178
98-86-2	Acetophenone	0.033	0.690	0.009
761-65-9	N,N-Dibutylformamide	0.001	0.559	1.206
1119-40-0	Pentanedioic acid, dimethyl ester	0.001	0.460	0.245
Complementing HS-GC-MS & GC-MS RRF Entries				
135-98-8	sec-Butylbenzene	2.990	0.928	$n.d^b$
95-47-6	o-Xylene	1.740	0.914	n.d
106-46-7	1,4-Dichlorobenzene	1.720	0.705	n.d
112-41-4	1-Dodecene	1.510 ^c	0.556 ^c	n.d
629-50-5	n-Tridecane	2.270 ^c	0.422 ^c	n.d
629-62-9	n-Pentadecane	0.533	1.025	n.d
104-76-7	2-Ethyl-1-hexanol	0.097	0.588	n.d
4265-25-2	2-Methylbenzofuran	0.468	0.894	n.d
111-70-6	1-Heptanol	0.146	0.631	n.d
112-73-2	Diethylene glycol di-n-butyl ether	0.071	0.669	n.d
Complementing GC-MS & LC-MS RRF Entries				
1568-83-8	Bisphenol A dimethyl ether	n.d	1.630	0.101
2943-75-1	Triethoxyethyl-n-octylsilane	n.d	1.210	0.013
80-46-6	4-tert-Pentylphenol	n.d	1.110	0.100
101-02-0	Triphenyl phosphite	n.d	0.922	0.279
80-07-9	Bis(4-Chlorophenyl) sulfone	n.d	0.893	0.050
149-30-4	2-Mercaptobenzothiazole	n.d	0.112	0.459
619-21-6	3-Formylbenzoic acid	n.d	0.078	1.081
1212-29-9	1,3-Dicyclohexythiourea	n.d	0.062	1.043
2306-33-4	Monoethyl phthalate	n.d	0.041	0.410
4559-70-0	Diphenylphosphine oxide	n.d	0.024	0.936

^{*a*} Bold entries reflect the method that would give the most accurate and reportable concentration estimate.

^b n.d. Compound outside scope of technique.

^c For these analytes, HS-GC-MS and GC-MS results are nearly equally acceptable.

obtained in such a manner will be more accurate than estimates obtained via "simple" quantitation, they may, nevertheless, have associated extrapolation errors. This type of error may be caused either by exceeding the dynamic range of the response curve or by dissimilar response curves between a compound and its internal standard.

Going beyond the Dynamic Range

This first type of extrapolation error has its roots in the assumption that RRF values remain constant as a function of the absolute and relative concentrations of the internal standard and the compound of interest. Over a certain range of concentration, the response functions for the internal standard and the analyte of interest will be well-defined and the RRF will be more or less constant. However, at some concentration either one or both of these compounds will exceed its dynamic range. Beyond this point, the assumed correlation between the responses of the internal standard and the target compound may become invalid. The accuracy of the estimated RRF corrected concentration consequently deteriorates outside the dynamic range of either compound.

This issue could occur when RRFs are applied to numerous extractables of varying concentration. Optimally, the internal standard concentration is selected to be within what is hopefully a narrow concentration range exhibited by all extractables in the extractables profile. The facts that (1) this range is not necessarily known during the design or acquisition of an internal database and (2) the concentration range exhibited by real extractables in a real extract is likely large can lead to a disconnect between the concentrations of the internal standard and the compounds that are present and need to be quantified in a sample. Such a disconnect could therefore cause extrapolation errors, especially for compounds present in a sample at a concentration that is substantially different from the value at which the RRF was determined.

Generally, modern analytical detectors applied in chromatographic screening exhibit a dynamic range that spans several orders of magnitude. The selection of the internal standard and its applied concentration should take this range into account. It is generally considered good practice that the internal standard results in a response that is linear within at least one order of magnitude centered on its applied concentration. This range is then considered to result in concentration estimates with an acceptable accuracy. For example, for an internal standard at 1 mg/L in a sample and with a dynamic range of 25, estimated analyte concentrations within the 0.2 to 5.0 mg/L range are consequently considered acceptably accurate whereas values outside this range can be compromised owing to dynamic range issues. To extend the quantitation range, two internal standards can be added at different concentrations, for example one compound at 0.1 mg/L and the other at 1 mg/L. This practice will increase the original quantitation range from a factor of 25 (0.2 mg/L to 5.0 mg/L) to a factor of 250 (0.02 mg/L to 5.0 mg/L). It should be noted, however, that response differences between the two internal standards must be properly accounted for.

Response Functions

The second type of extrapolation error is rooted in the assumption that all compounds detected in screening mode exhibit the same concentration–response relationship as the internal standard. RFs that are determined at a specific internal standard concentration do not account for differences in response functions. Such dissimilar relationships are usually caused by varying physicochemical properties between a compound and its corresponding internal standard. This is especially relevant for techniques with inherently complex response functions (e.g., LC-MS). In such situations, the error of extrapolation can be considerable because it can be anticipated that the concentration–response functions are nonuniform for all detected compounds.

Depending on a compound's concentration–response function, different situations can occur as illustrated in Figure 9. In Case 1, the concentration–response curve of the compound and the internal standard are identical. Consequently, the RRF is 1 within the entire applied range and no extrapolation errors are encountered.

In Case 2, the response functions for the analyte and the internal standard are both linear within the applied range but with different slopes. The average RRF can be calculated by dividing the slope of the response curve of the extractable by that of the internal standard (eq 4). The RRF accounts for the different slopes between the curves and prevents extrapolation errors within the applied range. This RRF value is applicable to the entire applied range and no extrapolation errors are encountered when the RRF is used.

From the third case on, differences in response functions for the analyte and the internal standard result in extrapolation errors within the applied range. This is the case when response functions have similar slopes but variable intercepts (Case 3), when both intercepts and slopes between curves differ (Case 4), and certainly when the response functions are dissimilar and nonlinear (Case 5).

It should be noted that the error of extrapolation is small within the limited concentration region where the analyte and the internal standard are present in a sample at concentrations close to the respective concentrations at which their RRF was estimated. Beyond this point, however, the magnitude of the extrapolation error will increase as the concentration difference between the analyte and the internal standard increases and as the analyte's and internal standard's response functions diverge.



Concentration-response function examples to illustrate different situations related to extrapolation errors.

Previously, a collection of RRF data was proposed as a means of addressing, to some extent, the error of estimation. Although this step improves the accuracy of "quantitation" compared to the "simple" quantitation approach, it is clear from the above discussion that the greatest accuracy is achieved when the database is constructed to contain response functions for the extractables residing in the database. Such response functions

TABLE III

Comparison of Library RRFs with Average RRFs Based on the Acquisition of Response Curves for a Relevant Subset of GC-MS Amenable Compounds in the Database across Different Solvents and Instruments in a 10–50 mg/L Range

		Library	Response Function RRF		Library RRF Value/ Response Model RRF
Cas	Compound Name	RRF	Estimate ± SE	Linear Range	%
583-39-1	2-Mercaptobenzothiazole ^a	0.112	0.535 ± 0.038	2 mg/L-50 mg/L	21
149-57-5	2-Ethylhexanoic acid	0.346	0.494 ± 0.016	500 µg/L-50 mg/L	70
5464-77-7	N,N-Dibenzylformamide	0.485	0.748 ± 0.019	500 µg/L-50 mg/L	87
112-52-7	1-Chlorododecane	0.568	0.856 ± 0.026	50 µg/L-50 mg/L	66
104-76-7	2-Ethyl-1-Hexanol	0.515	0.518 ± 0.020	50 µg/L-50 mg/L	99
112-12-9	2-Undecanone	0.627	0.692 ± 0.030	50 µg/L-50 mg/L	91
122-39-4	Diphenylamine	0.834	0.938 ± 0.021	50 µg/L-50 mg/L	89
593-49-7	n-Heptacosane	0.995	1.149 ± 0.038	50 µg/L-50 mg/L	87
128-37-0	BHT	1.010	0.923 ± 0.010	50 µg/L-50 mg/L	109
117-81-7	DEHP	1.010	1.104 ± 0.027	50 µg/L-50 mg/L	91
31570-04-4	Irgafos 168	1.298	0.964 ± 0.016	50 µg/L-50 mg/L	135
129-00-0	Pyrene	1.377	1.193 ± 0.017	50 µg/L-50 mg/L	115
			Average RRF accuracy (%):		$93 \pm 22\%$

^{*a*} Based on the RRF database, 2-mercaptobenzothiazole quantification should be performed by LC-MS, whereas estimated concentrations by GC-MS are expected to have a higher degree of uncertainty owing to the low RRF in this technique.

or calibration curves should be generated on the basis of response ratios against an internal standard at a constant level. In this way, the internal standard maintains its utility as a means of accounting for injection to injection variations but relinquishes its role as a means of quantitation. In this situation, the database would contain a response function in the following form for each extractable in the database:

$$C_{E\&L \ sample} = f \left(\frac{R_{E\&L}}{RF_{IS}} \right) \tag{8}$$

Given the large number of E&L compounds, it is practically untenable and next to impossible to determine all response functions for thousands of compounds. Moreover, although a database of response functions facilitates the quantitation of compounds contained therein, it does not address the quantitation of compounds that are not contained in the database such as "unknowns". As previously proposed, quantitation of such compounds could be performed using a mean (or median) RRF. Such a concept does not translate easily into the realm of response functions as the concept of an "average response function" likely has little practical meaning, especially considering the challenges in determining such a mean response function. If a mean response function could be calculated, then it could be applied to compounds that are not in the database in the same way RRFs were applied. If a mean response function cannot be established, then the use of a mean (or median) RRF remains the recommended way to estimate concentrations for compounds that are either unidentified or not in the database.

Case Study Comparing Relative Response Factors and Response Functions

The following experiment was performed to illustrate the magnitude of errors of extrapolation. Eleven extractables whose GC-MS RRFs had been previously determined were used to generate response functions (see Table III for a list of the analytes). Specifically, standards containing these extractables at levels spanning the concentration range of $10 \,\mu$ g/L to $100 \,$ mg/L were prepared in three commonly employed extraction solutions (hexane, isopropanol [IPA], and dichloromethane [DCM]). These



Relative response function for BHT (concentration range 0–50 mg/L) using an internal standard at a concentration 10 mg/L. The slope of the best-fit line, 0.923, is the value of the RRF and is valid over the entire concentration range.

standards, containing an internal standard at a concentration of 10 mg/L, were injected over the course of multiple experimental runs performed on three separate instrument systems. The resulting data were used to generate linear response functions (calibration curves) for each analyte, where the calibration function used was:

where the relative response is the ratio of the response of the analyte to the response of the internal standard $(R_{E\&L}/R_{IS})$ and the relative concentration is the ratio of the concentration of the analyte to the concentration of the internal standard $(C_{E\&L}/C_{IS})$. If the response function is truly linear, then the slope is equal to an RRF value that is valid over the entire linear range.

An example of a representative response function is shown in Figure 10 (for BHT). For BHT and nine of the other selected extractables, the response function was linear over the entire concentration range and exhibited a near zero intercept. For these analytes the slope of the best-fit linear regression line is the RRF value for the respective analyte.

The single analyte that did not exhibit a linear response function over the entire concentration range was 2mercaptobenzothiazole, which had the lowest response of the 11 selected analytes. For this compound, a linear range could only be established between 2 and 50 mg/L.

The magnitude of an error of extrapolation for the selected 11 analytes is illustrated by comparing the RRF value contained in the Database (based on a single analyte concentration/internal standard ratio) to the RRF value obtained as the slope of the response function. As shown in Table III, the agreement between the single point library RRF and the response function RRF is quite good for the 10 analytes that exhibited linear response functions over the entire concentration range studied. The only case where the agreement was poor was for the one analyte (2-mercaptobenzothiazole) that had the lowest library RRF and exhibited a truncated linear range. For these reasons, and because of the better LC-MS response of this analyte, GC-MS



Comparison of different quantitation approaches for three compounds at 10 mg/L in an artificial extract using an internal standard at a concentration 10 mg/L.

would not be the method of choice for reporting 2-mercaptobenzothiazole's concentration.

For GC-MS in general, it is concluded that errors of extrapolation would be relatively small for all analytes that have an RRF value in the range of 0.5 to 2.0 (consistent with the range of RRF for the compounds in Table III). Analytes with RRF values outside of this range would likely exhibit a significant extrapolation error if the analyte's concentration was greatly different from the internal standard's concentration.

Comparison of Concentrations Obtained by "Simple" Quantitation versus Application of Relative Response Factors

Ultimately, the decision to compile and routinely apply a database of RRF values is dictated by the answer to the question "to what extent does RRF correction improve the accuracy of estimated concentrations in extractables and leachables screening compared to "simple" quantitation?" If the answer is that the improvement in accuracy is small for use of RRF versus "simple" quantitation, then the value of using RRF is small versus the effort in compiling RRF values. On the other hand, if the improvement in accuracy is considerable, then the value of using RRF is significant versus the effort in compiling RRF values. To address this question, an "artificial" extract containing three GC-MS amenable extractables (with different RRFs) was prepared at a set concentration of 10 mg/L and spiked with an internal standard at the same concentration. Analysis of this extract and subsequent quantification of the target compounds (a) using only the internal standard, (b) using the library RRF for each compound, and (c) using a calibration curve obtained with analytical standards illustrates the differences among these quantitation approaches. As shown in Figure 11, the experimentally determined accuracy of RRF correction outperforms "simple" quantitation for the compound with the lower and the compound with the higher library RRF value.

It has been proposed and is generally accepted that "quantitation" for extractables during screening is acceptable if the concentration estimate is from 50% to 200% of the real value, as this provides sufficient accuracy to perform a credible safety assessment. Thus, concentration estimates obtained via "simple" quantitation would be "good enough" for extractables whose RRF values are between 0.5 and 2. Although use of the RRF values for such extractables would produce a more accurate concentration estimate than would "simple" quantitation, in most cases the increased accuracy would not necessarily translate into a more effective safety assessment. For the Database, 59% of the GC- MS amenable compounds and 32% of the LC-MS amenable compounds have RRF values that fall in this range. It can hence be concluded that RRF correction is necessary to obtain acceptable concentration estimates for the other 41% of GC-MS compounds and 68% (more than two-thirds) of the LC-MS compounds. Moreover, among these compounds that require RRF correction are compounds that are potentially toxic, that is classified as Cramer Class III substances. Hence, it is important to get these compound's concentration "right" to enable a proper safety assessment. Putting this into perspective, 82 of the 261 GC-MS amenable and 45 of the 92 LC-MS amenable compounds with Cramer Class III classification would be reported incorrectly by "simple" quantitation!

If one would consider more stringent acceptance criteria and require that the estimated concentration be within 70% to 130% of the real value (which is typically expected for target analysis), then the percentage of GC-MS and LC-MS amenable compounds that would benefit from RRF correction (versus "simple" quantitation) increases respectively to 67% and 85%. For Cramer Class III substances, this would imply that acceptable results would be obtained only with RRF correction for 143 of the 261 GC-MS amenable compounds and 67 of the 92 LC-MS amenable compounds.

Lastly, it is noted that use of the RRF makes the choice of the internal standard largely irrelevant, as the RRF value makes the internal standard essentially transparent in terms of accomplishing the actual quantitation. With "simple" quantitation, the choice of the internal standard remains highly relevant, especially in terms of any systematic quantitation bias. For example, if an internal standard has a RF that is the median of the RFs for the entire population of extractables and leachables, then an equal number of these substances will have their concentrations either overestimated or underestimated by use of the internal standard. However, if the internal standard has a RF that is either above or below the population median, then more concentrations will be either underestimated or overestimated. The further away the internal standard's RF is from the population median, the greater is the inherent quantitation bias.

Concluding Thoughts

As the identity of an extractable or leachable is a critical input to its impact assessment, errors of inexact identification (either an identity cannot be secured or



Figure 12

Means by which an internal database can improve the identification of extractables or leachables.

the identity that is secured is incorrect) are fatal in the sense that they irreversibly prejudice the assessment. Errors of inexact identification reflect fundamental issues associated with the processes by which identities are procured, including spectral matching.

The use of an internal database to reduce the number of identification errors by improving the quality of the match between a sample response to a standard response (via a database) is illustrated in Figure 12. For example, if a database consists of extractables and leachables only, this will reduce the occurrence of complicating incorrect matches to substances that are contained in external databases but are neither extractables nor leachables. If the sample spectrum and the match spectra are generated on the same instrumentation with the same operating conditions, then there will be fewer matches of higher quality, likely producing more definitively correct identities. Moreover, the use of orthogonal confirming data (such as MS spectral match plus retention time) will decrease the possibility of misidentifications and increase the "confidence level" in, and the "grade" of, tentative identities secured on the basis of spectral matching only.

Additionally, accurate quantitation of extractables and leachables is a critical input into an extractables/leachables impact assessment, as quantitation establishes exposure. For example, considering patient safety as a critical quality attribute, it is the mathematical product of the concentration of a leachable in the drug product and the drug product's daily concentration volume that establishes the patient's exposure to the leachable. Armed with patient exposure and an estimation of the allowable daily intake (driven by the leachable's identity), one can properly assess the impact of a leachable on a patient's health.

The size and diversity of the population of extractables and leachables presents two challenges with respect to E&L quantitation via screening. Firstly, the large number of potential substances precludes the generation and use of calibration curves for each and every potential substance and thus concentration estimates obtained in screening must practically be based on certain assumptions about responses and response functions across the entire population of substances. Secondly, shortcomings in these necessary assumptions cause concentration estimates obtained in screening to be of relatively poor and varying accuracy. However, internal databases, which contain information about response behavior for members of the E&L population, address these shortcomings and the use of this information increases the accuracy of, and reduces the compound to compound variability in, screening quantitation.

Moving Forward

Once all extractables have been discovered, confidently identified, and accurately quantified, then a rigorous assessment of the extractables' potential impact on critical product quality attributes (such as purity, efficacy, stability, and safety) can be performed. Properly leveraging an internal database mitigates the undesirable effects of errors in the screening processes of discovery, identification, and quantitation.

In the last installment (Part 3) of this series, we will consider the situation where a method is properly suited for its purpose but its implementation at time of use is flawed in some critical manner. Because of such errors of implementation, a perfectly capable method can produce data of unacceptably poor quality. The role of system suitability as a means of addressing implementation errors will be considered and the use of a database of system suitability data to anticipate and manage implementation errors will be discussed. Furthermore, we will consider the internal database as a means of enabling good and practical science, considering how an internal database can continue to advance the state of the art in organic extractables/leachables testing.

Conflict of Interest Declaration

The authors declare that they have no competing or conflicting interests, noting their relationship with Nelson Labs, a provider of extractables and leachables testing and consulting services.

References

- U.S. Food and Drug Administration. Guidance for Industry: Container Closure Systems for Packaging Human Drugs and Biologics. Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services: Rockville, MD, 1999.
- 2. European Medicines Agency. *Guideline on Plastic Immediate Packaging Materials*. EMA: London, 2005.
- Jenke, D. Material Screening and Characterization. In Compatibility of Pharmaceutical Products and Contact Materials: Safety Considerations Associated with Extractables and Leachables; John Wiley and Sons: Hoboken, NY, 2009; Chapter 4, pp 124– 131.
- Norwood, D. L.; Feinberg, T. N.; Mullis, J. O.; Pennino, C. J. Analytical Techniques for Identification and Quantitation of Extractables and Leachables. In *Leachables and Extractables Handbook: Safety Evaluation, Qualification and Best Practices Applied to Inhalation Drug Products*; Ball, D. J.; Norwood, D. L.; Stults, C. L. M.; Nagao, L. M., Eds.; John Wiley and Sons: Hoboken, NY, 2012; Chapter 13, pp 241–288.
- U.S. Pharmacopeial Convention. General Chapter <1663> Assessment of Extractables Associated with Pharmaceutical Packaging/Delivery Systems. In USP 41–NF 36; USP: Rockville, MD, 2018; p 7910.
- Christiaens, P.; Beusen, J. M.; Verlinde, P.; Baeten, J.; Jenke, D. Identifying and Mitigating Errors in Screening for Organic Extractables and Leachables: Part 1: Introduction to Errors in Chromatographic Screening for Organic Extractables & Leachables and Discussion of the Error of Omission. *PDA J. Pharm. Sci. Technol.* 2020, 74 (1), 90–107.

- U.S. Environmental Protection Agency. Method 8260D (SW-846): Volatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/MS). EPA: Washington, DC, 2018; Revision 4.
- U.S. Environmental Protection Agency. Method 8270E (SW-846): Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/ MS). EPA: Washington, DC, 2018; Revision 6.
- 9. Wiley Registry 11th Edition/NIST 2017 Mass Spectral Library; Wiley, NIST, 2017.
- Mullis, J. O.; Granger, A.; Quin, C.; Norwood, D. L. The Analytical Evaluation Threshold (AET) Concept, Sensitivity and Analytical Uncertainty. Presented at Leachables and Extractables 2008, Dublin, Ireland, Mar 5–6, 2008.
- 11. Jenke, D.; Odufu, A. Utilization of Internal Standard Response Factors to Estimate the Concentration of Organic Compounds Leached from Pharmaceutical Packaging Systems and Application of Such

Estimated Concentrations to Safety Assessment. J. Chromatogr. Sci. 2012, 50 (3), 206–212.

- Zdravkovic, S. A.; Duong, C. T.; Hellenbrand, A. A.; Duff, S. R.; Dreger, A. L. Establishment of a Reference Standard Database for Use in the Qualitative and Semi-Quantitative Analysis of Pharmaceutical Contact Materials Within a Extractables Survey by GC-MS. J. Pharm. Biomed. Anal. 2018, 151, 49–60.
- Jenke, D.; Liu, N. Chromatographic Considerations in the Standardization of Liquid Chromatographic Methods Used for Extractables Screening. *J. Liq. Chromatogr. Relat. Technol.* **2016**, *39* (13), 613– 619.
- Jordi, M. A.; Khera, S.; Roland, K.; Jiang, L.; Solomon, P.; Nelson, J.; Lateef, S. S.; Woods, J.; Martin, L.; Martin, S.; Aiello, F.; Chen, N. Qualitative Assessment of Extractables Form Single-Use Components and the Impact of Reference Standard Selection. *J. Pharm. Biomed. Anal.* 2018, *150*, 368–376.

PDA Journal of Pharmaceutical Science and Technology



An Authorized User of the electronic PDA Journal of Pharmaceutical Science and Technology (the PDA Journal) is a PDA Member in good standing. Authorized Users are permitted to do the following:

Search and view the content of the PDA Journal

Download a single article for the individual use of an Authorized User

Assemble and distribute links that point to the PDA Journal •Print individual articles from the PDA Journal for the individual use of an Authorized User Make a reasonable number of photocopies of a printed article for the individual use of an Authorized User or for the use by or distribution to other Authorized Users

Authorized Users are not permitted to do the following:

Except as mentioned above, allow anyone other than an Authorized User to use or access the PDA Journal

· Display or otherwise make any information from the PDA Journal available to anyone other than an Authorized User

Post articles from the PDA Journal on Web sites, either available on the Internet or an Intranet, or in any form of online publications

Transmit electronically, via e-mail or any other file transfer protocols, any portion of the PDA Journal

·Create a searchable archive of any portion of the PDA Journal

Use robots or intelligent agents to access, search and/or systematically download any portion of the PDA Journal

·Sell, re-sell, rent, lease, license, sublicense, assign or otherwise transfer the use of the PDA Journal or its content

Use or copy the PDA Journal for document delivery, fee-for-service use, or bulk reproduction or distribution of materials in any form, or any substantially similar commercial purpose Alter, modify, repackage or adapt any portion of the PDA Journal

Make any edits or derivative works with respect to any portion of the PDA Journal including any text or graphics

Delete or remove in any form or format, including on a printed article or photocopy, any copyright information or notice contained in the PDA Journal