



Reducing relative response factor variation using a multidetector approach for extractables and leachables (E&L) analysis to mitigate the need for uncertainty factors



Mark Anderson Jordi*, Kevin Rowland, Weixi Liu, Xingluan Cao, Jie Zong, Yuan Ren, Zhu Liang, Xiao Zhou, Michael Louis, Kaitlin Lerner

Jordi Labs, 200 Gilbert St., Mansfield, Ma, 02048, United States

ARTICLE INFO

Article history:

Received 20 February 2020

Received in revised form 17 April 2020

Accepted 22 April 2020

Available online 1 May 2020

Keywords:

Extractables

Leachables

E&L

Uncertainty factor

Analytical evaluation threshold

Response factor database

ABSTRACT

Characterization of Extractables and Leachables (E&Ls) is an important aspect of product quality in important fields such as pharmaceuticals, medical devices and food contact materials. The main goal of an E&L study is *identification and quantification* of those species which may leach from packaging materials used to contain pharmaceuticals or which may leach directly out of a medical device or food contact material and thus may result in patient exposure. It is common practice to perform relative quantitation of extractables and leachables using surrogate standards due to the large diversity of species observed and the lack of available reference standards. A key problem in obtaining accurate E&L results arises due to *response factor (RF) variation*. Different compounds at the same concentration give different signal intensities and thus have different RF values. Two key aspects of study quality are affected by this problem. First, the evaluation of the number of compounds which are above the toxicologically relevant threshold (*analytical evaluation threshold, (AET)*) can be affected (RF Problem 1: AET Underreporting). Second, quantitative accuracy is affected which can reduce the reliability of the margin of safety (MOS) calculations which serves as the basis of the toxicological evaluation (RF Problem 2: Quantitative Error). RF databases have been the main solution proposed for solving these problems but do not reduce the underlying RF variation and lack the scope required to address quantitative error for compounds not contained in the database. In the absence of other solutions, large uncertainty factors (UF) have been applied in the AET calculations to account for RF Problem 1: AET Underreporting. These UF factors have been assigned values of 4 for GCMS and up to 10 for LCMS. Large uncertainty factors have a number of unintended negative consequences including the need for large amounts of sample concentration (>10X) prior to analysis resulting in potential compound loss or degradation and increased matrix effects. To overcome these problems, this publication demonstrates a *multidetector approach* using an HPLC system coupled with a Quadrupole Time of Flight Liquid Chromatography Mass Spectrometer (QTOF-LCMS), Charged Aerosol Detector (CAD) and an Ultraviolet-Visible Detector (UV) and a dual detection Gas Chromatography Mass Spectrometry (GCMS) system using a Polyarc Reactor system with Flame Ionization Detection (FID). Herein, it is demonstrated that this combination of methods (the *multidetector approach*) allowed detection and accurate surrogate standard quantitation of 217 unique extractables spanning a wide range of chemical properties (Mw, logP, pKa and boiling point). The combination of optimized detector selection with appropriate standard selection was verified to provide positive detection for 94% of the compounds at the AET level and a high level of quantitative accuracy ($\pm 20\%$ for 85% of the compounds and $\pm 40\%$ for 91% of the compounds) while using only a UF of 2. Unlike the RF database approach, the *multidetector approach* is not limited to only those compounds contained in the database but is applicable to the majority of extractables.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail addresses: mjordi@jordilabs.com (M.A. Jordi), krowland@jordilabs.com (K. Rowland), wliu@jordilabs.com (W. Liu), xcao@jordilabs.com (X. Cao), jzong@jordilabs.com (J. Zong), dren@jordilabs.com (Y. Ren), zliang@jordilabs.com (Z. Liang), xzhou@jordilabs.com (X. Zhou), mlouis@jordilabs.com (M. Louis), klerner@jordilabs.com (K. Lerner).

1. Introduction

The characterization of extractables and leachables (E&Ls) is an important aspect of product quality. Leachables from pharmaceutical packaging, medical devices and food contact materials can lead to adverse effects on product stability, efficacy and patient safety. Leachables are those compounds or elemental impurities which migrate into the drug product from pharmaceutical packaging or into the patient directly or indirectly from medical devices under the conditions of use [5]. Extractables are those compounds which migrate under conditions which are generally harsher than the expected conditions of use and are thus meant to provide an exaggerated list of potential leachables [4]. In drug product analyses, leachables studies are typically conducted using the drug product. For medical devices, a simulated leachables condition (saline or other suitable simulant) is applied due to the difficulties in replicating the biological system encountered by implantable devices. Examples of the deleterious effects of leachables include the reaction of a therapeutic proteins with acrylic acid which leached from prefilled syringes or the leaching of a degradant of the common antioxidant Irgafos 168, (bis(2,4-di-tertbutylphenyl)phosphate (bDtBBP)) which was found to be highly detrimental to cell growth [15,16].

Key quality aspects in E&L study design include selection of appropriate extraction conditions, proper sample preparation procedures (prevent loss of E&Ls, enhance detection and remove matrix interferences) and the suitability of the analytical methods for identification and quantitation of all extracted E&Ls. Mass spectrometry (MS) detectors are commonly applied for both identification and quantification primarily because of their high sensitivity and broad applicability [12–14]. Identification of individual leachables and extractables is generally accomplished using a combination of gas chromatography–mass spectrometry (GCMS) for volatile and semi-volatile compounds and liquid chromatography–mass spectrometry (LCMS) for semi-volatile and non-volatile compounds. Secondary detectors for LC and GC have been applied in some cases but their role is typically not emphasized. Examples include the use of ultraviolet (UV) or evaporative light scattering (ELSD) detection for liquid chromatography and flame ionization detection (FID) for gas chromatography [14].

The purpose of this article is to demonstrate the suitability and advantages of the *multidetector approach* for confirming which extractables are above the *analytical evaluation threshold* (AET) and for providing accurate quantification. An analytical method is considered suitable for its intended purpose based on its performance characteristics as determined during method validation [1,2]. Methods should be selected with their intended purpose clearly defined. In an *E&L screening study*, the goal is to provide a complete list of all E&Ls present in the extract solution above the toxicologically relevant level and to accurately determine their associated concentrations for subsequent toxicological evaluation. The toxicological results from the screening E&L analysis can then be used to identify those leachables which pose significant risk and a determination can be made if additional analyses are warranted. For drug products, it is common to use this process to identify candidates for inclusion in a leachables assessment of stability study samples. For medical devices, the screening study results often serve as a part of a larger biocompatibility assessment. Based on these objectives, the suitability of E&L screening methods should be evaluated based on their ability to provide comprehensive identification and quantitation for all E&Ls above the relevant toxicological threshold. This is a difficult analytical challenge in pharmaceutical packaging studies due to the complexity of many drug product matrices and the high sensitivity required, (typically low ppb), but can be even more daunting for medical device extracts which often contain hundreds of individual extractables

at a wide range of concentrations for less well-studied polymer systems. The ideal analytical method would need to provide high selectivity, universal detectability and high sensitivity for identification while providing broad dynamic range and high accuracy for quantification (universal response factors). Unfortunately, no single detector technology currently exists which can meet these objectives for all potential E&L analytes. This has been noted in USP <1663>, “It is a reality that there is no analytical technique or combination of analytical techniques that is capable of the discovery, identification, and quantitation of any and all organic and inorganic extractable chemical entities known to science. [4]” It is however possible to reduce quantitative error and maximize identification potential through judicious application of detection methods best suited to their intended purpose (identification or quantification) and through proper surrogate standard selection.

To understand the requirements for the analytical methods, it is necessary to understand clearly the level of sensitivity needed. Guidance for characterization of extractables and leachables has been issued for pharmaceutical packaging [3–5] and medical devices [6–8]. Additional guidelines for control of DNA reactive impurities (mutagenic impurities) is provided in the ICH M7(R1) guidance [9]. This document defines a threshold of toxicological concern (TTC) (ranging from 1.5 µg/day for lifetime exposure to 120 µg/day for less than lifetime exposure) which is defined as an “acceptable intake for any unstudied chemical that poses a negligible risk of carcinogenicity or other toxic effects.” The TTC and other safety thresholds have been widely used in E&L studies to calculate an *analytical evaluation threshold* (AET value) which defines the toxicologically relevant identification threshold (concentration above which a compound should be identified and reported) during an E&L study [10]. The most current embodiment of the AET relevant to medical devices is defined in the ISO 10993-18:2020 guidance as follows [8].

$$\text{AET} (\mu\text{g/mL}) = \text{DBT} \times \frac{A}{B \times C \times D} \times \frac{1}{\text{UF}}$$

where: DBT is a dose based threshold such as the TTC

A is the number of devices extracted

B is the extract volume

C is the maximum number/mass of devices used per patient

D is the dilution or concentration factor

UF is an uncertainty factor applied to account for RF variation and has previously been defined as $\text{UF} = 1/(1-\text{RSD})$ where RSD is the relative standard deviation of a response factor database. A similar formula was detailed for pharmaceutical packaging as defined by PQRI [10]. The current embodiment of the ISO 10993-18 standard defines the UF as $\text{mean}/[1-(t \times \text{std})]$ where t is degrees of confidence, mean is the mean of the response factor database and std is the standard deviation. The authors would note that this definition is not equivalent to $1/(1-\text{RSD})$ and is in our opinion inappropriate as it can result in UF values < 1 in some cases and provides unreasonably large UF values for others cases where there is only modest RF variation.

When performing E&L screening studies, the AET value must be related to the detector response by analyzing a reference standard or standards at the AET concentration. The signal from the reference standard is then compared to the analytical responses of the E&Ls to determine if they exceed the threshold (reference standard response at the AET concentration). It is important to keep in mind that at this stage in the analysis, the identity of the E&L compounds is not yet known. Hence, it is not practical to use the actual E&L compounds under study as standards since 1) their identity is not yet known, 2) many of the compounds found to extract are oligomers or other side products of the polymer system and hence are not commercially available and 3) the frequently large number of E&Ls would make this time and cost prohibitive. For these reasons, the

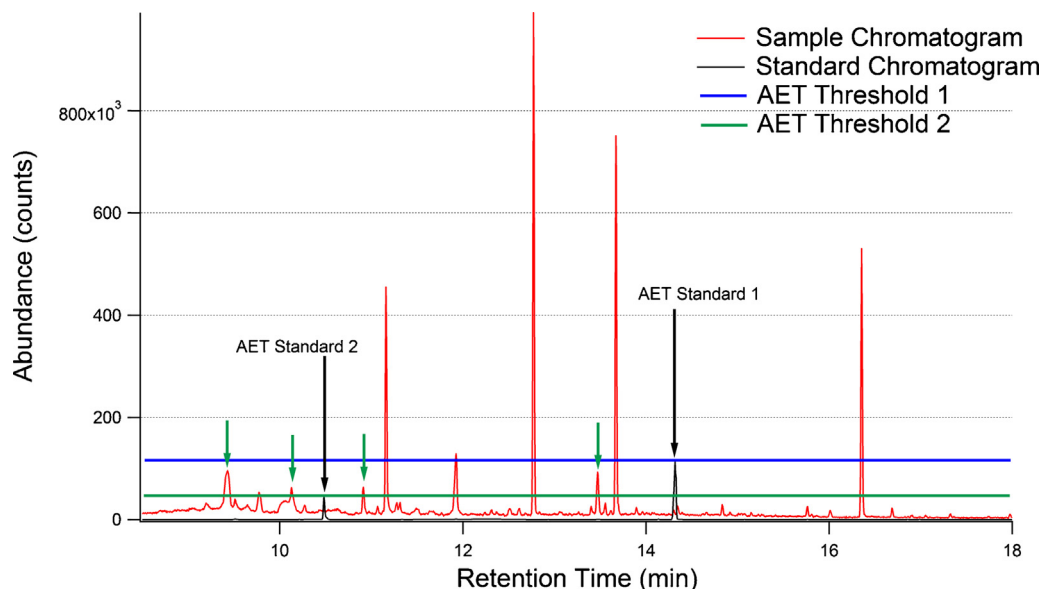


Fig. 1. Effect of RF variation on the AET threshold. Note the difference in peak magnitude for two standards at equal concentration (black arrows) resulting in different threshold levels and additional compounds which must be identified (green arrows). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

common practice is to use a surrogate standard to determine the response associated with the AET level. This practice introduces the potential for significant error into the E&L study due to response factor variation [11,17]. The response factor (RF) for a given compound is the amount of signal per unit concentration. The RF can be further defined as the slope of the calibration curve for the analyte under a given set of analytical conditions and using a specific detector system [18]. Additionally, it is often analytically expedient to define the relative response factor (RRF) as the slope of the calibration curve for a target compound divided by the slope of the calibration curve for a surrogate compound [18,21]. RRF values are frequently preferred over RF values as they account for variability over time including detector drift and sensitivity changes. Finally, it should be noted that linearity of the detector is assumed when using the slope as a measure of the RF as non-linearity would constitute variation in the response factor as a function of concentration.

When comparing a target compound with a surrogate standard, RRF values which approach a value of one indicate equal responses for equal concentrations. If the RRF is below one then the concentration of the target compound will be underestimated. This could result in the exclusion of compounds from the E&L study that are actually above the AET level (RF Problem 1: AET Underreporting). Fig. 1 shows an example of the determination of the AET threshold for a Gas Chromatography Mass Spectrometry (GCMS) analysis and demonstrates how surrogate standard selection can negatively affect the determination of which compounds exceed the threshold. From this demonstration, it is clear that the peak magnitude associated with the AET is not a single value but is instead a distribution based on the RF values for all E&L compounds. The degree of RF variation observed is a function of both the *method utilized (analytical conditions)* and the *detector system applied (choice of instrumentation)*.

A second, significant problem also arises due to RF variation. After it is determined that a compound is at or above the toxicologically relevant level, it is then necessary to determine the compound's identity and quantity for subsequent toxicological evaluation. Differences in the response factor for the E&L compound as compared to the surrogate standard can adversely affect quantitative accuracy (RF Problem 2: Quantitative Error) [18,21]. If the RF value for a compound is less than that for the surrogate stan-

dard then in the absence of any adjustment for RF differences, the concentration of the target compound will be underestimated. The concentration of each target compound is used to calculate a margin of safety (MOS) value. MOS values are used to assess the level of risk to the patient from exposure to that particular extractable or leachable at the reported concentration. An underestimated concentration could result in the calculation of an inflated MOS value potentially adversely affecting the accuracy of the biocompatibility assessment and underestimating the toxicological risk. On the other hand, if the RF value is greater than that of the surrogate standard, the compounds concentration will be overestimated. This could result in an underestimated MOS value and thus a potentially safe device could be considered to not be biocompatible.

To mitigate RF Problem 1: AET Underreporting, a UF was introduced into the AET equation. Some authors have suggested a UF of 4 for GCMS and a UF of 10 has been proposed for LCMS [17]. For studies utilizing the threshold of toxicological concern (TTC), the associated AET values typically require a detection limit in the mid to low ppb (ng/mL) range. As an example, consider a theoretical long-term contacting device for which the surface area is 30 cm², one device is used and the concentration factor is 1. Using the most conservative TTC value of 1.5 μg/day and a surface area to extraction volume ratio of 3 cm²/mL (derived from ISO-10993-12), an AET value of 150 ng/mL is calculated prior to applying a UF factor. If a UF of 10 is applied to account for RF variation then this drops to just 15 ng/mL.

$$\begin{aligned} \text{AET} \left(\frac{\mu\text{g}}{\text{mL}} \right) &= \text{DBT} \times \frac{\text{A}}{\text{B} \times \text{C} \times \text{D}} \times \frac{1}{\text{UF}} \\ &= 1.5 \mu\text{g/day} \times \frac{1}{30/3 \times 1 \times 1} \times \frac{1}{10} = 15 \text{ ng/mL} \end{aligned}$$

For analyses requiring large extraction volumes, (100 mL–5000 mL for large volume parenterals, or devices using high solvent volumes such as dialyzers) this value can be sub part per billion. This level of sensitivity approaches or exceeds the limits of what current analytical technology can achieve for screening analysis (non-targeted). It should also be noted that including a UF in the AET calculation does nothing to resolve RF Problem 2: Quantitative error. It is therefore strongly desirable to mitigate response factor

variation and with it the need for large UF values. This would provide enhanced reliability for the toxicological evaluation and would increase the viability of the AET value.

The amount of error introduced into an E&L study due to the RF problems is directly proportional to the amount of RF variation observed for the detectors used for quantitation. As noted above, mass spectrometry detection has been widely used in E&L analysis for both identification and quantitation. It is well known in the literature that the response factors for different compounds can differ substantially when using MS detection [19–21,23]. This is particularly significant for liquid chromatography mass spectrometry (LCMS) but is also observed to a lesser extent for gas chromatography mass spectrometry (GCMS) due to the different means used to ionize the samples. In a study reported by Blanz et al., which included the LCMS analysis of 132 structurally diverse drug candidates and 233 corresponding metabolites, 27 metabolites showed a response factor less than 0.2 and 7 had a response factor greater than 2 [21]. In LCMS, the response can further be affected by sample matrix effects (ion suppression or ion enhancement) which result in different response factors for a single compound as a function of the sample matrix or mobile phase composition [19,29].

Ultraviolet (UV) detection at low wavelength is currently the overwhelming method of choice for relative quantitation when screening for drug impurities [20,22,23]. UV response factors can change for different compounds due to changes in the molar extinction coefficient (type of chromophore) but this approach is generally preferred to MS relative quantitation for compounds with a consistent chromophore as it has the advantage of not being susceptible to sample matrix effects. Other detection methods have been applied to provide “varying degrees of orthogonality” and increase the “likelihood of separating and detecting all” impurities including Chemiluminescent nitrogen detectors (CLND), flame ionization detection (FID), refractive index (RI) and $^1\text{H NMR}$ [23,27,28]. Mass based detectors including charged aerosol detection (CAD) and evaporative light scattering (ELSD) were applied to supplement UV because they do not require a chromophore and they provide signals which are less dependent on the “properties of the individual analytes” [26,28]. While both mass-based detectors have limitations, including moderately less sensitivity than other methods, the CAD detector showed particular promise as reduced variation in response was observed across a wide range of chemicals.

The purpose of this publication is to demonstrate the suitability of the *multidetector approach* to obtain comprehensive and accurate quantification for an extensive collection of extractables (217 extractables) without the use of large uncertainty factors (UF) by surrogate standard quantitation. The extractables analyzed were selected to represent the universe of potential extractables and covered a very wide range of properties (Mw, log P, pKa and boiling point) and contained a significant number of compounds reported to be difficult to analyze (specifically 42 organic acids and 38 amines). The *multidetector approach* consists of an HPLC system using a Quadrupole Time of Flight Liquid Chromatography Mass Spectrometer (QTOF-LCMS), Charged Aerosol Detector (CAD) and an Ultraviolet-Visible Detector (UV) and a dual detection Gas Chromatography Mass Spectrometry (GCMS) system coupled with a Polarc Reactor system with Flame Ionization Detection (FID). In this approach, the QTOF-LCMS detector is used solely for identification of compounds above the AET threshold while the CAD and UV are applied for quantitation in HPLC. GCMS and FID are applied for identification and quantification with the primary advantage of the FID being its ruggedness for quantitation and the GCMS its identification capabilities. The *multidetector approach* described herein relies on the combination of HPLC (QTOF-LCMS-UV-CAD) and GC (GCMS-FID) to provide a high rate of coverage for the full range of extractables and leachables studied. The combination of high sen-

sitivity and specificity of the mass spectrometry detectors coupled with the more consistent responses of the CAD, UV and GCMS-FID detectors was shown to significantly improve quantitative accuracy and provide comprehensive detection for all species analyzed.

2. Materials and methods

2.1. Chemical reagents and materials

The standards used for relative quantitation listed in Table 2 and extractables standards of the 217 compounds used for the determination of the response factors were obtained from commercial sources in high purity. Purity of each standard was further verified during the analysis by confirming the absence of significant impurities peaks during QTOF-LCMS-UV-CAD and GCMS-FID analysis. The majority of the standards were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA), TCI America (Portland, OR, USA) or Alfa Aesar (Ward Hill, MA, USA). LC/MS grade methanol, isopropanol, acetonitrile (ACN), ammonium hydroxide and formic acid were purchased from Honeywell (Houston, TX, USA). Dichloromethane (DCM) and Tetrahydrofuran (THF) were analytical grade from Pharmco-Aaper and were used without further purification. Ultrapure water (18 M Ω cm, EMD Millipore Billerica, MA) was used. Chemical properties for the extractables (Mw, boiling point, log P values and pKa) were obtained from reputable websites such as PubChem, ChemSpider, ChemSRC and ChemAxon.

2.2. Standard preparation

Stock solutions of the standards for RF determination were prepared by individually dissolving them at 1 mg/mL. Portions of the stock solutions were then combined into working solutions containing groups of standards which were subsequently diluted to concentrations of 5, 10 and 25 $\mu\text{g/mL}$ for analysis. Methanol was used to dissolve standards for QTOF-LCMS-UV-CAD while DCM was applied for GCMS-FID analysis. In rare instances, standards which were insoluble in methanol or DCM were initially dissolved into another suitable solvent (acetonitrile or tetrahydrofuran) and then subsequently diluted into methanol or DCM.

2.3. GCMS-FID analysis

The slope of the calibration curves (RF values) of each extractable were determined using the mass selective detector (MSD) and flame ionization detector (FID). All GCMS-FID analyses were performed on an Agilent 7890B gas chromatograph with dual detection combining both a 5975 MSD and an FID (Agilent Technologies, Santa Clara CA). The two detectors were connected post column using a flow splitter for simultaneous detection. A polarc catalytic reactor system (Activated Research Company, Eden Prairie, MN) was connected after the column splitter and prior to the FID to convert all organic molecules to methane for FID detection. MS analyses were performed in electron ionization (EI) mode. The mass spectrometer was set to monitor a mass range of 29–500 amu. Separations were performed on an HP-5MS 30 m x 0.25 mm x 0.25 μm film thickness (Agilent Technologies, Santa Clara CA). The GC oven was ramped from 28–270 $^{\circ}\text{C}$ at a ramp rate of 15 $^{\circ}\text{C}$ per minute and held for 10 min at 270 $^{\circ}\text{C}$.

2.4. QTOF-LCMS-UV-CAD analysis

The slopes of the calibration curves (RF values) for each extractable were determined using a Agilent 6545 quadrupole time of flight mass spectrometer (QTOF) using a Dual Agilent Jet Stream Electrospray Ionization source in both positive and negative ion

modes, an Agilent 1260 Diode Array Ultraviolet Visible (UV) detector and a Corona Veo RS charged aerosol detector (CAD) (Thermo Fisher Scientific, Waltham MA) combined with an Agilent 1290 HPLC system (Agilent Technologies, Santa Clara CA). The QTOF analysis parameters included: Gas temperature 350 °C; VCap 3500 V; Nebulizer (N2) 15 psi; Drying gas (N2) 12 L/min; Sheath gas 8 L/min; Sheath Gas Temperature 400 °C.; mass range m/z 80–3200. The UV detector was set to monitor at a wavelength of 230 nm using bandwidth of 4 nm. The CAD detector used a gas pressure of 58 PSI, evaporation temperature of 35 °C and a charge voltage of 2.37 kV. Separations were performed on an Agilent Zorbax Eclipse Plus C-8, 1.8 μm , 2.1 \times 50 mm (Agilent Technologies, Santa Clara CA) maintained at 45 °C. The mobile phase consisted of 0.05% formic acid/0.03% ammonium hydroxide/5% methanol in water (solvent A) and 0.05% formic acid in methanol (solvent B) and Isopropanol (solvent C) delivered at 0.65 mL/min. Gradient elution had an initial condition of solvent A (100%) for .3 min, at 5.3 min (B:C 80%/20%), at 8.3 min (B:C 50%/50%), at 10.54 min (B:C 50%/50%). Run time was 10.54 min with a 1.5-minute post time to allow for equilibration. The HPLC effluent was introduced directly into the MS system.

2.5. Data processing

Data reduction was performed in Mass Hunter Qualitative Analysis Version B.10.0.10305.0 (Agilent Technologies, Santa Clara CA). Integration of the total ion chromatogram (TIC) was used for determination of the peak areas for GCMS. Integration of the compound chromatogram (summation of all compound related ions) was used for determination of the peak area for QTOF-LCMS. The UV signal at 230 nm and the CAD and FID chromatograms were integrated without further data processing. A linear regression analysis was used to determine the slope of the calibration curve (RF value) for each extractable using duplicate injections at 5, 10 and 25 $\mu\text{g}/\text{mL}$. Irganox 245 was used as the surrogate standard for determination of the RRF values for QTOF-LCMS-UV-CAD. Decane was used as the surrogate standard for determining RRF values in GCMS-FID. Following the determination of all RRF values for each detector, the RRF values were further normalized using the average RRF value for each distribution thus making the choice of surrogate standard immaterial to the final values.

3. Results and discussion

3.1. Relative response factor distribution results

In order to understand the magnitude of RF Problem 1: AET Underreporting and RF Problem 2: Quantitative Error, it is necessary to understand the RF distribution for the universe of extractables for each detector system for a particular E&L screening method. To estimate the universe of extractables, 217 extractable compounds were analyzed containing a wide range of chemical properties including volatiles, semi-volatiles and non-volatile compounds. Highly volatile organic compounds (HVOCs) (compounds that exert a vapor pressure greater than 80 mm Hg when measured at 20 °C) have been excluded as these compounds are appropriate only for headspace GCMS and other specialized HVOC methods and should be analyzed separately. The properties examined for inclusivity included molecular weight (93–1177 amu), logP values (–2.38–25.156), boiling point (102–>600 °C) and pKa values (–0.83–18.9). The extractables standards had a wide range of functional groups including amines, acids, alcohols, phenols as well as sulfur, silicon and phosphorous containing species and many others. Table 2 shows a partial listing of some of the extractables analyzed to provide a sense of the diversity of compounds analyzed. Fig. 2 shows the distribution of chemical properties (molecular

weights, logP values, boiling point, pKa and pKb) for the extractables used to probe the RF variation of the method. The *multidetector approach* utilizes a triple detection system combining QTOF-LCMS-UV-CAD detectors and a dual detection system with GCMS-FID. QTOF-LCMS was operated in both positive and negative ion modes. Thus, a total of 6 different signals were acquired for each extractable compound (LCMS Positive ion, LCMS Negative ion, UV, CAD, GCMS and FID). The RRF values were normalized using the average RRF value for the complete distribution for each detector. This has several advantages. The primary advantage being that it greatly simplifies data interpretation such that a value of 1 is equal to the average RRF in each distribution. An RRF value greater than 1 then indicates a strongly responding extractable and a value less than 1 indicates a poorly responding extractable in comparison to the average RRF for the distribution. Table 1 summarizes the RRF distribution results for each detector. Fig. 3 shows the resulting RRF distribution plots for all 6 detectors. Only the compounds which showed a response on a given detector system are included in that detector's RRF distribution. This is appropriate because only compounds which show a response would be used for quantitative assessment (non-responding compounds are not measured at all and must be detected by another approach). All 217 extractables analyzed in this study showed a response by one or more detectors demonstrating the wide applicability of this strategy.

The distribution plots for the 6 detectors were found to differ substantially in terms of RRF variability and in terms of the breadth and type of compounds covered. This is a reflection of the different principles upon which each detector is based. LCMS positive and negative mode detected 72% and 34% of the overall extractables, respectively, making LCMS positive mode the single most comprehensive detector in terms of compound coverage. Unfortunately, the LCMS positive and negative mode distributions showed a very wide range of RRF values with a maximum number of compounds between 0 and 0.2 and the second most abundant region at > 1.8. This is reflected in the very large% RSD values for the distributions (LCMS Positive% RSD, 108%, Negative Mode LCMS%RSD, 119%). This indicates that a disproportionate number of compounds observed by LCMS give either very weak or very strong responses in comparison with the average (RRF = 1). This has very significant negative implications for both RF Problem 1: AET Underreporting and RF Problem 2: Quantitative Error when LCMS is applied for relative quantitation placing great importance on appropriate surrogate standard selection. Only 33% of the compounds detected by positive mode LCMS and 22% of the compounds detected by negative mode LCMS had an RRF value between 0.6–1.4 ($\pm 40\%$ of the actual concentration) without surrogate standard optimization. This further demonstrates that the approach of using a single internal standard for relative LCMS quantitation will frequently result in very substantial error in the quantitative values. LCMS detects compounds based on their ability to ionize. The ionization process is well known to vary widely based on sample chemistry with ionizability being favored for those compounds containing heteroatoms with acidic or basic character and disfavored for those without these groups [21]. Compounds which do not have heteroatoms or other ionizable groups show very weak response. This is the fundamental reason for the increased number of compounds at the extremes of the LCMS RRF distributions. Studies in our laboratories have shown that this problem is not unique to QTOF instruments but is also observed for triple quadrupole systems due to the fact that they rely on the same ionization processes.

In contrast, UV and CAD showed RRF distributions which are more symmetrical. While there was still a range of RRF values, the distributions tended to be more even and to have a maximum closer to the center of the distribution. This has significant positive implications for reducing RF Problem 2: Quantitative Error as error is minimized when the majority of compounds have an

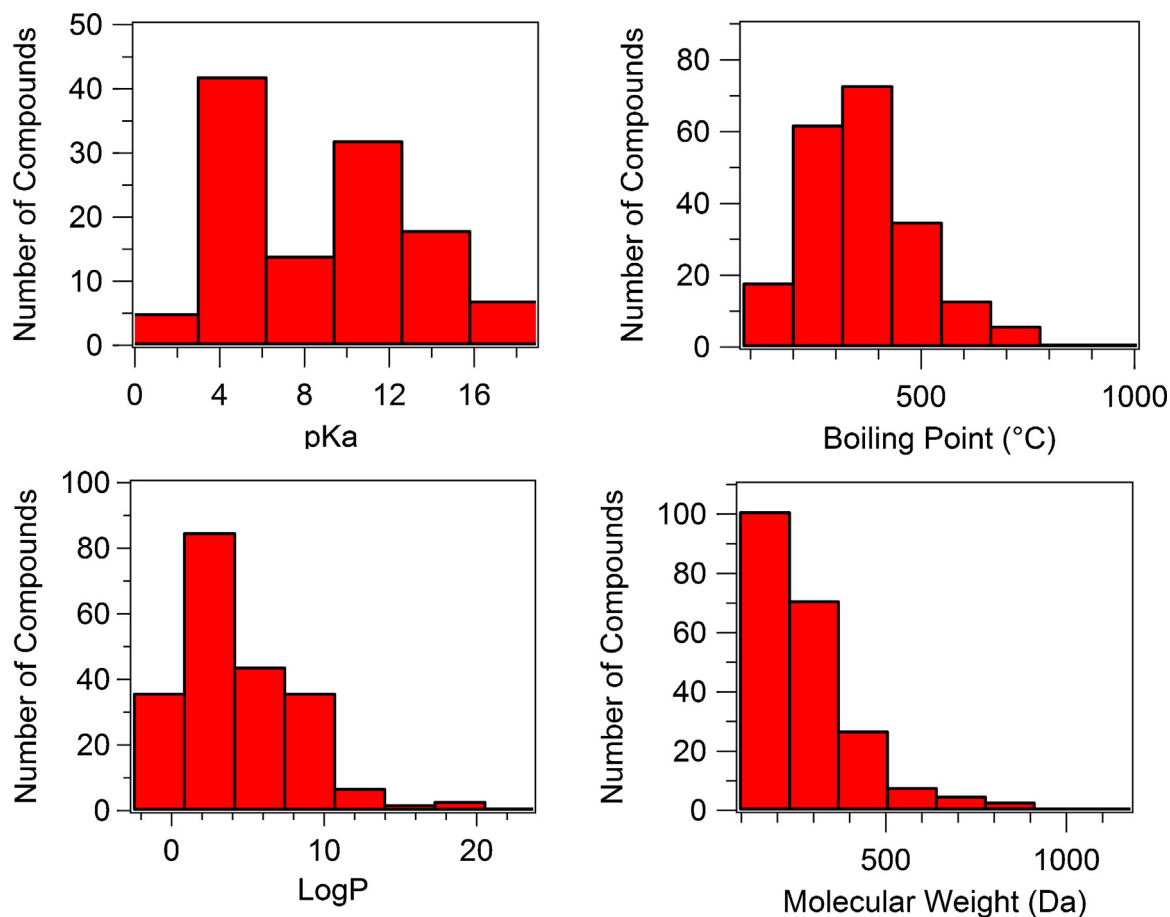


Fig. 2. Histograms showing the distribution of properties for the 217 extractables standards used to determine RF variation.

Table 1
Normalized relative response factors distribution statistics.

Detector	LCMS Positive	LCMS Negative	UV 230 nm	CAD	GCMS	FID
Average	1.0	1.0	1.0	1.0	1.0	1.0
RSD	108%	119%	60%	65%	52%	54%
Min	0.0013	0.0065	0.0364	0.0336	0.0318	0.0098
Max	6.1140	7.9816	2.8493	2.6467	2.8401	3.1218
Percent detected	72%	34%	52%	56%	59%	59%
% of compounds where RRF > .6 and < 1.4	33%	22%	44%	41%	61%	55%

RRF value closer to the center of the distribution (RRF = 1). In comparison to LCMS, UV and CAD showed %RSD values of 60% and 65% respectively, indicating that a much greater number of compounds fall within a more reasonable range of quantitative accuracy. Thus, a significant improvement in quantitative accuracy would be expected simply by replacing LCMS with UV and CAD for relative quantitation even without optimized standard or detector selection. UV and CAD detectors successfully detected 52% and 56% of the extractable compounds respectively, and a total of 74% of all compounds combined. This included 80% of the compounds detected by LCMS indicating a high degree of overlap. 44% of the compounds detected by UV at 230 nm and 41% of the compounds detected by CAD had an RRF value between 0.6–1.4 ($\pm 40\%$ of the actual concentration) without standard optimization. As previously reported, CAD detection was found to correlate strongly with compound boiling point, showing more consistent detection for compounds with boiling points above 400 °C and discrimination against low boiling compounds. This makes CAD an excellent

detector for many oligomeric and polymeric species. UV detection is based on the absorption of UV light (in this case 230 nm) and was found to be dependent upon the type and presence of chromophores. Examples of strongly absorbing compounds included those containing aromatic structures such as hindered phenolics (an important class of polymer antioxidants) while poorly responding compounds typically had a single weak chromophore such as acrylic monomers which are better suited to GC detection.

GCMS and FID response showed very analogous RRF distributions. While there was still a range of RRF values, the distributions were found to be the most Gaussian of all of the detectors with the majority of compounds showing values closer to the center of the distribution (RRF = 1). This indicates that the majority of the compounds would be expected to show more reasonable quantitative accuracy reducing RF Problem 2: Quantitative Error. This data supports that GCMS or FID relative quantitation would be expected to generally be more reliable than that for LCMS. GCMS and FID showed %RSD values of 51% and 53% respectively indicating that

Table 2

Partial list of the 217 extractables analyzed showing the diversity of compounds suitable for the method and the list of Surrogate Standard compounds. Normalized RRF values for each detector are also listed.

Name	Molecular formula	CAS	RRF LCMS Pos	RRF LCMS Neg	RRF UV 230 nm	RRF CAD	RRF GCMS	RRF FID
Example extractable compounds								
Bis(2-ethylhexyl) phthalate	C24H38O4	117-81-7	1.921	ND	0.622	0.802	1.706	1.855
4-Chlorobenzene sulfonic acid	C10H7ClO3S	98-66-8	ND	ND	ND	ND	0.081	0.034
Dibenzyl phosphate	C14H15O4P	1623-08-1	0.134	2.266	ND	1.411	ND	ND
Caprolactam	C6H11NO	105-60-2	1.045	ND	ND	ND	0.658	1.200
1-naphthylamine	C10H9N	134-32-7	0.580	ND	1.400	ND	0.918	0.248
Hexadecylamine	C16H35N	143-27-1	1.514	ND	ND	2.009	0.543	0.387
Tributylamine	C12H27N	102-82-9	1.872	ND	ND	ND	0.708	1.614
5-Amino-1-pentanol	C5H13NO	2508-29-4	0.126	ND	ND	2.597	0.095	0.663
Dodecamethylcyclhexasiloxane	C12H36O6Si6	540-97-6	ND	ND	ND	ND	0.630	0.182
Butyl methacrylate	C8H14O2	97-88-1	ND	ND	ND	ND	0.692	1.088
Irganox 245	C34H50O8	36443-68-2	1.705	1.744	0.550	1.570	ND	ND
Irganox 3114	C48H69N3O6	27676-62-6	0.513	ND	0.814	0.894	ND	ND
Irganox 1076	C35H62O3	2082-79-3	1.210	ND	0.334	0.634	ND	ND
Irganox 1035	C38H58O6S	41484-35-9	4.529	1.263	0.508	1.354	ND	ND
Irganox 1024	C34H52N2O4	32687-78-8	0.157	2.225	0.750	1.647	ND	ND
Tinuvin 320	C20H25N3O	3846-71-7	1.263	0.060	1.005	ND	2.126	1.287
Diphenyl phthalate	C20H14O4	84-62-8	0.120	0.034	1.177	0.938	1.667	1.243
Bisphenol A	C15H16O2	80-05-7	ND	ND	1.655	1.822	2.579	1.449
Bisphenol A diglycidyl ether	C21H24O4	1675-54-3	1.259	ND	1.371	1.451	ND	ND
Octaethylene glycol	C16H34O9	5117-19-1	1.763	ND	ND	2.139	ND	ND
Citric acid	C6H8O7	77-92-9	0.058	0.212	ND	1.463	ND	ND
1,3,5-Triphenylbenzene	C24H18	612-71-5	0.001	ND	1.976	0.819	ND	ND
LCMS surrogate standards								
1-octane sulfonic acid	C8H18O3S	5324-84-5	ND	3.164	ND	1.639	ND	ND
Irganox 245	C34H50O8	36443-68-2	1.695	1.744	0.551	1.570	ND	ND
Phthalic acid	C8H6O4	88-99-3	0.047	1.061	1.421	1.005	0.532	0.726
Glyceryl trioleate	C57H104O6	122-32-7	0.438	ND	ND	0.185	ND	ND
Oleamide	C18H35NO	301-02-0	1.414	ND	ND	2.680	1.504	0.779
Diisobutyl phthalate	C16H22O4	84-69-5	0.908	ND	0.853	ND	0.837	1.423
Diphenyl phthalate	C20H14O4	84-62-8	0.119	0.033	1.172	0.931	1.667	1.243
4,4'-Dichlorodiphenyl sulfone (DCDPS)	C12H8Cl2O2S	80-07-9	0.003	ND	1.402	0.149	1.398	1.158
Irgacure 184	C13H16O2	947-19-3	0.019	ND	1.163	ND	1.045	1.920
Diethylene glycol cyclic phthalate (Howflex Gbp)	C12H12O5	13988-26-6	0.458	ND	0.947	ND	1.173	1.311
GCMS surrogate standards								
Benzaldehyde	C7H6O	100-52-7	ND	ND	0.986	ND	0.815	1.758
Decane	C10H22	124-18-5	ND	ND	ND	ND	0.741	2.502
Decamethylcyclopentasiloxane (D5)	C10H30O5Si5	541-02-6	ND	ND	ND	ND	0.770	0.456
4-Methoxyphenol	C7H8O2	150-76-5	ND	ND	0.815	ND	1.092	1.802
Butylated hydroxytoluene (BHT)	C15H24O	128-37-0	ND	ND	0.584	ND	1.222	1.606
Hexadecane	C16H32	544-76-3	ND	ND	ND	ND	0.875	2.299
Irgacure 184	C13H16O2	947-19-3	0.019	ND	1.163	ND	1.045	1.920
Diisobutyl phthalate	C16H22O4	84-69-5	0.908	ND	0.853	ND	0.837	1.423
Eicosane	C20H42	112-95-8	ND	ND	ND	ND	1.227	3.154
Tetracosane	C24H50	646-31-1	ND	ND	ND	ND	1.249	1.812

ND = None Detected.

a much greater number of compounds fall within a more reasonable range of quantitative accuracy as compared to LCMS (%RSD of 107% and 119%). This is in line with the %RSD values determined in other studies (50.9% for GCMS and 44% for FID) [17]. Thus, when a compound is detected by both LCMS and GCMS, the GCMS values would generally be more reliable in terms of quantitative accuracy even without optimized standard selection. 61% of the compounds detected by GCMS and 56% of the compounds detected by FID had an RRF value between .6–1.4 ($\pm 40\%$ of the actual concentration) without standard optimization. The GCMS-FID distributions are modestly narrower than the UV and CAD distributions (%RSD 51–53% as compared to 61–66% for UV/CAD). GCMS and FID both successfully detected 60% of the extractable compounds. It was further observed that 50% of the extractables were detected by both QTOF-LCMS-UV-CAD and GCMS-FID indicating a high degree of overlap between methods. This allows useful opportunities for verification of compound identity and further protection against underreporting of compound concentration. Finally, an indication of the sufficiency of the data contained in this publication to estimate the universe of extractables (HVOCs excluded) was noted in the fact that the RSD of the distributions changed by less than 15% for the addition of the last 67 compounds.

3.2. Multidetector as a solution to RF problem 1: AET underreporting

One of the most important aspects of E&L study quality is the decision as to which peaks should be identified and reported in the study because they are at or above the toxicologically relevant level. The *analytical evaluation threshold* (AET) approach is commonly applied to relate the toxicologically relevant threshold to the observed analytical responses. For a compound to be identified and reported in the study, two criteria must be met. First, the compound must be detected (give a measurable response on at least one detector) and second, the compounds must have an RRF factor sufficiently strong that it equals or exceeds the threshold set using the surrogate standard. In our previous publication, we noted that the number of compounds determined to be at the toxicologically relevant level could be dramatically affected by the choice of surrogate standard [11]. It is important to understand that the surrogate standard response defines the location of the AET threshold (see Fig. 1). If a surrogate standard with an RRF of 1 is selected, then any compound with an RRF equal or greater than 1 would need to be identified and reported. In setting the AET threshold value, it is preferred that the surrogate standard selected would be at the center

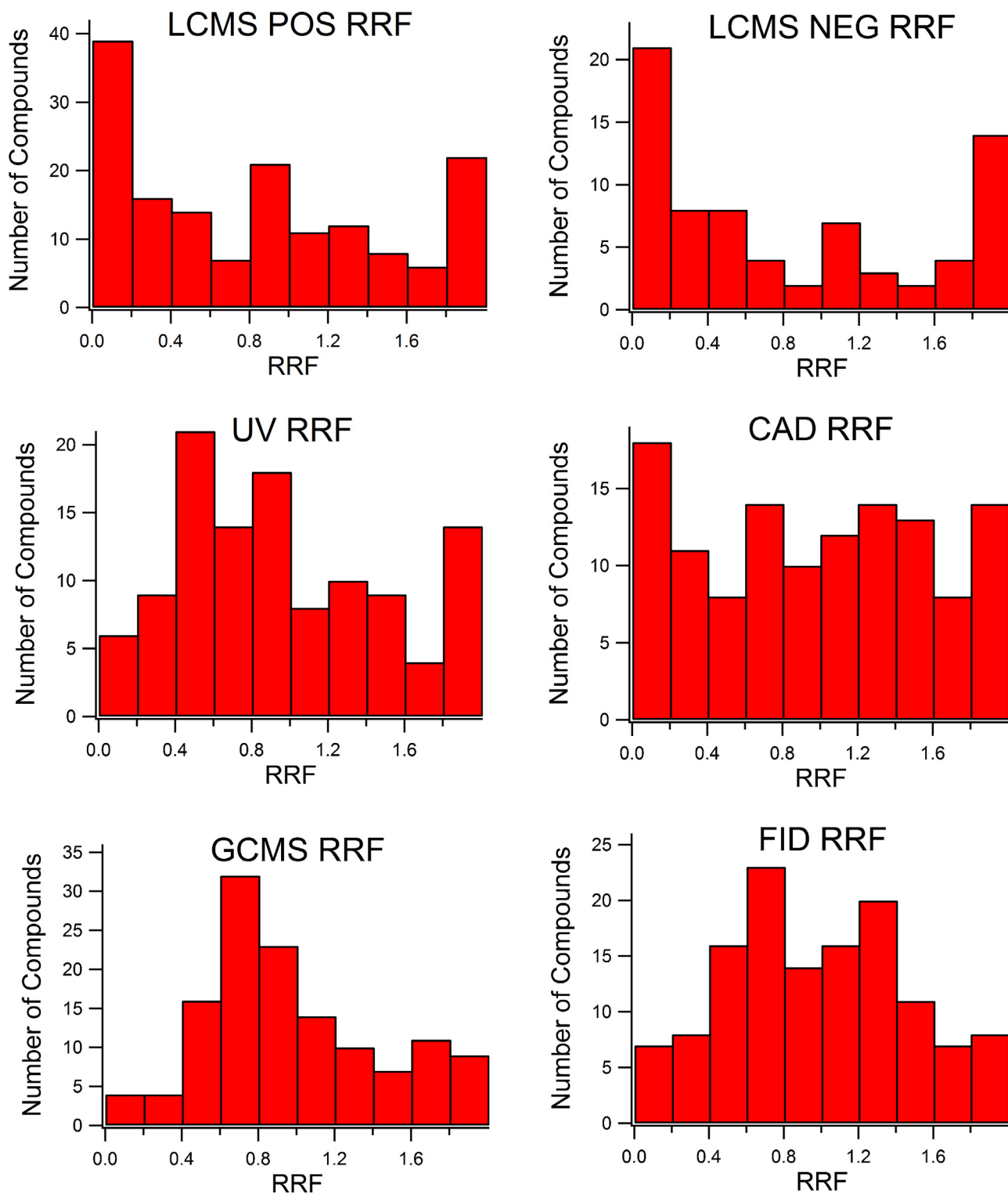


Fig. 3. Histograms showing the distribution of RRF Values for 6 different detectors for 217 extractables.

of the distribution ($RRF = 1$ for a normalized RRF distribution). This allows the application of any UF factors to be applied starting at the center of the distribution. If a standard was selected which had an RRF value less than 1, then an additional measure of conservatism would be added to the study (more inclusive) and conversely a value greater than one builds into the study design less conservatism (less protective of safety). Surrogate standard selection and the UF factor work together in tandem such that if a surrogate stan-

dard with an RRF of 0.5 is selected this is equivalent to using a UF of 2 with a surrogate standard with an RRF of 1 ($RRF/UF = \frac{1}{2} = .5$).

Even if an appropriate surrogate standard or standards are selected, it is an unfortunate fact that no single detector system can detect all compounds. However, it is not necessary that every compound is detected by every detector in order to obtain complete coverage of all extractables. When using the *multidetector approach*, it is only necessary that a compound show sufficient signal on at

least one detector. For instance, in a study conducted using only GCMS and LCMS, it is only necessary that the compound be detected by either GCMS or LCMS to be adequately accounted for in the toxicological evaluation. It is important to understand that for the *multidetector approach* described herein, the signal intensity on any one detector is not correlated to the signal observed by any other detector. This is because each detector operates based on different fundamental principles. As described above, the combination of QTOF-LCMS-UV-CAD with GCMS-FID (*multidetector approach*) provides five distinct mechanisms of detection (ionizability for LCMS, light absorption for UV, charged particle detection for CAD, electron ionization (EI) for GCMS and charged ion detection for Polycarb FID). To be included in the study, a compound only needs to 1) be detected by any one of these five mechanisms (one of the five detectors) and 2) give a signal which is greater than the AET threshold value on any one detector.

To confirm the sufficiency of the *multidetector approach* to mitigate RF Problem 1: AET Underreporting, Fig. 4 shows a plot comparing the maximum RRF value for each compound using either just GCMS and LCMS or using the multidetector strategy (QTOF-LCMS-UV-CAD and GCMS-FID). Each dot on the figure represents the highest RRF value for a given compound after comparing all detectors. All compounds represented by a value greater than 1 would be included in the study using a UF of 1. If a UF of 2 was applied, then all compounds with an RRF of greater than 0.5 would be included. Additional UF values can be considered by dividing the RRF of 1 by the desired UF value. An examination of the data indicates that 54% of the compounds would be determined to be above the toxicologically relevant threshold using a UF of 1 with only LCMS and GCMS while 76% would be included using the *multidetector approach*. If a UF of 2 is applied then this improves significantly to 85% for the combination of LCMS and GCMS and 94% for the *multidetector approach*. Applying a UF of 10 results in 95% coverage for LCMS and GCMS and 99% for the *multidetector approach*. Thus, using the *multidetector approach* with a UF of 2 was shown to be equally protective (1% difference) as a UF of 10 when using LCMS and GCMS only. While both strategies may be equally inclusive, the multidetector strategy has several advantages over the use of a larger UF factor ($UF > 4$). First, the *multidetector approach* can reduce underreporting due to signal overlap. This is due to the fact that it is much less likely that an interfering signal would cover the peak for a given extractable on multiple detectors simultaneously. Second, instrument sensitivity is not infinite. Thus, if the AET calculation uses a larger UF value, then it frequently becomes necessary to use extensive sample concentration ($>10X$). Extensive sample concentration has many potential negative consequences including potential compound loss or degradation, increased matrix interferences which could cause underreporting (the very thing we are attempting to avoid), the requirement for a larger amount of original extract (i.e. more devices or packaging must be extracted increasing cost and study time) and the potential for increased false positives due to contamination or concentration of background impurities. In some instances, it is also impossible to sufficiently concentrate the sample to reach the AET when using a large UF value ($UF > 4$). This is especially common for salt containing samples where concentration factors are limited due to salt solubility. While small amounts of sample concentration ($<10X$) are often necessary in order to increase method sensitivity, excessive concentration should be avoided where possible. The *multidetector approach* provides a means by which to obtain sufficient assurance of detection without resorting to large UF values and the corresponding large concentration factors.

It is important to note that for some analyses, the AET threshold may be so low (low ng/mL levels) that method sensitivity on an individual detector may not be sufficient and a full *multidetector approach* may not be feasible. The inclusiveness of the strategy

then depends on how many detectors can be applied at the AET level. In our laboratories, it has been observed that CAD detection is not as sensitive as LCMS or UV. The results of this study indicate that application of a QTOF-LCMS-UV and GCMS-FID utilizing a UF of 2 results in 91% coverage. This remains a high level of certainty especially when one considers the difficulty of detecting all unknown species at low or sub ppb levels and the fact that CAD detection at a limit of detection (LOD) greater than the AET may still catch some remaining compounds. In considering the negative potential consequences of extensive sample concentration ($>10X$), it seems advisable that 91% coverage may be the most protective option from a safety perspective even when the CAD LOD cannot reach the AET.

One potential draw back of the *multidetector approach* is the potential for overreporting. It is important to note that inclusion of a compound in the study at this stage does not mean that the compound is actually at the toxicologically relevant level. It is possible that a compound has an RRF greater than 1 on a particular detector (compound is overestimated). If this occurs, the actual concentration may be below the toxicologically relevant level but the compound would be considered to be above the threshold (false positive). While this consequence may be less severe than that for underreporting, the increased cost and time to manufacturers can also have negative implications for patients. For this reason, we would recommend that a 2nd AET evaluation be performed following compound quantitation. Inclusion in the study at this early stage only means that a compound should be subjected to additional scrutiny as to its actual identity and concentration. Following quantitation as described below, a final decision will need to be made regarding if the compound should be included in the study based on the resulting quantitative values. This 2nd AET evaluation is made at a stage in the analysis when more information is available which allows for optimum surrogate standard and detector selection and thus is based on better information.

Regardless of the strategy applied for AET verification, it is important to stress the necessity of verifying the limit of detection (LOD) at the AET threshold at the time of sample analysis. The authors have frequently seen studies for which method sensitivity was not demonstrated but rather assumed. This is a crucial aspect of study quality without which there are no assurances that the toxicologically relevant compounds are being detected. Those skilled in the art know that method LOD values change daily based on transient conditions such as instrument cleanliness (LCMS ion sources being particularly vulnerable) and thus verification as a part of system suitability is essential. Ideally, this should be accomplished by analyzing a group of standards at the AET concentration with a range of RRF values whose normalized average approaches one and which are representative of the RRF distribution obtained using a method specific E&L database. This proves that the method has the needed level of sensitivity in order to detect those compounds at or above AET and to cover the use of UF factors. This brings us to one final advantage of the *multidetector approach* which is that UV, CAD and FID detectors are generally more rugged than MS and are less susceptible to loss of sensitivity throughout the analysis due to the high matrix backgrounds often observed in E&L analyses.

3.3. Multidetector and standard selection as solutions to RF problem 2: Quantitative error

A second key aspect of E&L study quality is quantitative accuracy. The accuracy of surrogate standard quantitation is directly related to the difference in the RRF values for the surrogate standards and the extractable compounds. Thus, if an extractable compound has an RRF of 0.2 and it is quantitated using a compound with an RRF of 1, the compounds reported concentration will only be 20% of the actual value. Similarly, if a compound with an RRF of

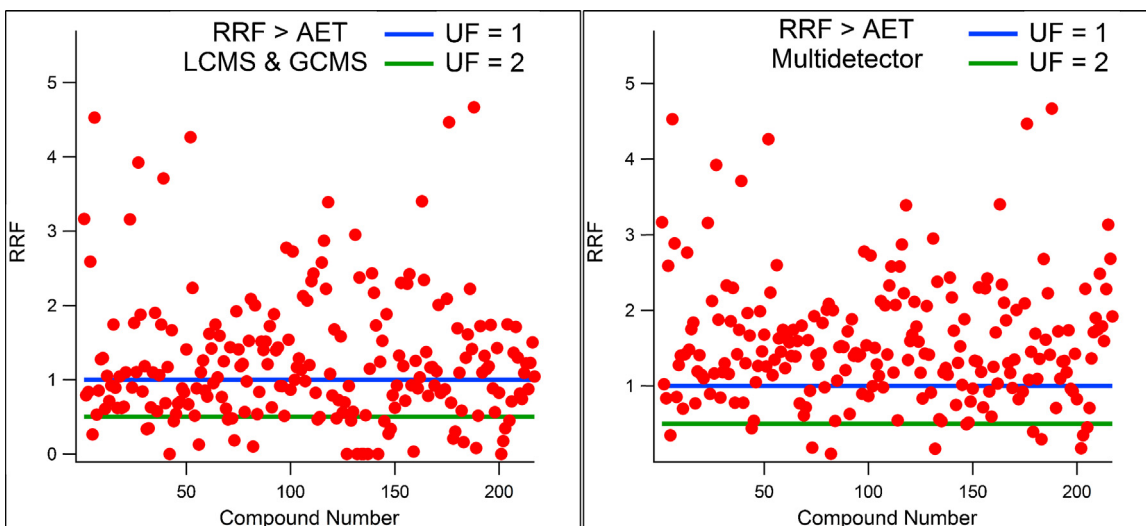


Fig. 4. Multidetector solution for RF Problem 1: AET Underreporting.

2 is quantitated using a surrogate standard with an RRF of 1 then its reported concentration will be 200% of the actual value. The magnitude of this error for the estimated universe of extractables can be seen by considering the data in Fig. 3 and in Table 1 where it is shown that errors exceeding 40% are the most common outcome on all but the GCMS detector. Even on the GCMS detector, 39% of the compounds are poorly estimated (>40% error). There is currently no mechanism in the toxicological risk assessment process (margin of safety (MOS) calculation) to account for this error and some discussion of the addition of an additional uncertainty factor designed to account for analytical uncertainty has occurred at national conferences. Those who are frequent practitioners in E&L analysis know that this would have serious negative consequences for the toxicological process as many studies have compounds with MOS values close to one and these products would potentially need to be unnecessarily rejected if an additional UF factor was added to account for this analytical uncertainty. It is therefore highly desirable that a method for improved quantitative accuracy be identified.

Based on the data in Fig. 3 and Table 1 it is clear that no single detector has a sufficiently universal response to provide highly accurate quantitation for the universe of extractables. Fortunately, two mechanisms are available to improve quantitative accuracy using the *multidetector approach*: 1) proper detector selection and 2) proper surrogate standard selection. Since the response on each detector is independent of all other detectors, proper selection of the detector (detector which produces an RRF closest to 1) can be used as a means to significantly improve quantitative accuracy. Fig. 5 shows a comparison of the quantitative accuracy which can be obtained using optimized detector selection for the combination of GCMS and LCMS as compared to the *multidetector approach* using (UV-CAD and GCMS-FID). We intentionally omit the LCMS detector in the *multidetector approach* as the RF variability in this detector is so large and the difficulty in determining proper surrogate standard selection so great (see discussion below) that we do not feel this detector can be applied confidently for relative quantitation. In the LCMS/GCMS approach, only 33% of the compounds are within $\pm 20\%$ of the actual value and only 55% are within $\pm 40\%$ even using the optimum detector. In contrast, using the *multidetector approach*, 40% of compounds are within $\pm 20\%$ of the accurate value and 72% are within $\pm 40\%$. In the LCMS/GCMS strategy, 12% of the compounds are significantly underreported (<40% of the true value) while only 7% are underreported using the *multidetector approach*. These improvements are accomplished without

any optimization of surrogate standard selection and excluding the LCMS detector.

In the LCMS/GCMS strategy, 6 compounds were found to have no signal from which to perform the quantitation. In the *multidetector approach* 5 compounds showed no signal on UV-CAD or GCMS-FID. However, there is one important distinction, in the case of the *multidetector approach*; the compounds which were missed on UV-CAD or GCMS-FID did respond in LCMS. Thus, the data at least exists from which an analyst could choose to perform LCMS formal quantitation for those compounds. In the case of the LCMS/GCMS strategy, no data exists to indicate that 6 compounds were missed. In examining the nature of these compounds, it was observed that for the LCMS/GCMS strategy these compounds were primarily higher Mw compounds (average Mw 527 amu) with few or poorly ionizing functional groups. It has frequently been observed in our laboratories that many E&L extracts contain substantial oligomer peaks (500–2000 amu) which show no signal at all in LCMS or GCMS but are detectable by CAD. In many cases, these oligomers are one of the most abundant species in the extract but they are missed using the LCMS/GCMS approach. In the case of the *multidetector approach*, 2 of the 5 compounds which showed no quantitative signal were highly volatile (Hexamethylcyclotrisiloxane and 3-(3-acetoxypropyl)heptamethyltrisiloxane) and may be better suited to HGCMS. In the remaining 3 of 5 cases, they were generally highly polar volatile acids (leucic acid and heptafluorobutyric acid). This is not to imply that the method is not good for acidic species ($pK_a < 5$) as 42 of 44 acidic species tested were quantifiable.

Another strategy which can be used to improve quantitative accuracy is appropriate surrogate standard selection. If a surrogate standard can be selected which has an RRF value close or equal to that of the extractable compound, then quantitative error can be minimized or eliminated. In the current study, 10 LC and GC surrogate standards were analyzed as shown in the bottom of Table 2. These standards were selected to provide a diversity of chemistry and because they elute at a range of retention times allowing the creation of calibration curves for each surrogate standard. The reason that 10 surrogate standards were selected was to provide a minimum of four compounds which respond by each detector. The dataset was then examined by choosing the standard with the RRF value closest to that of the extractable to serve as its surrogate standard (RRF of surrogate standard closest to RRF of extractable) and then further optimized using the detector which resulted in the RRF value closest to 1.0. The results of this analysis are shown in Fig. 6. It was found that this approach resulted in 85% of the compounds

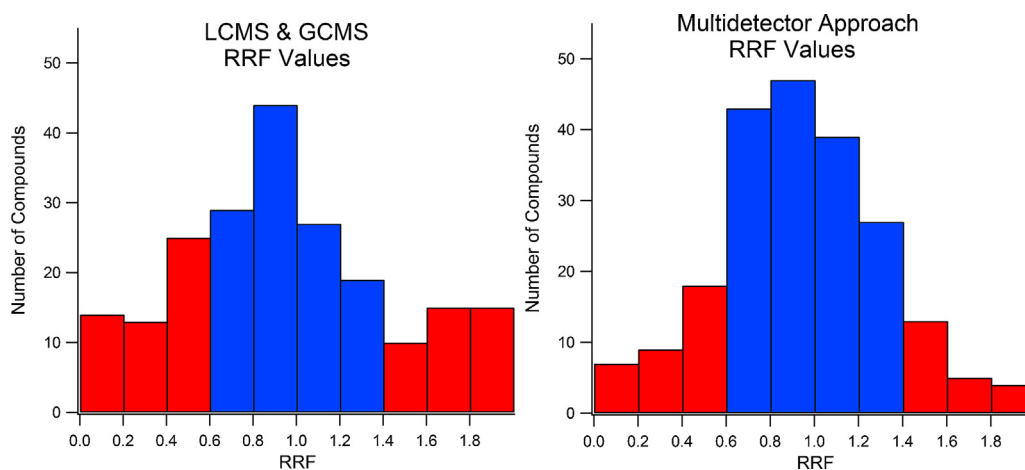


Fig. 5. Comparison of quantitative accuracy for GCMS/LCMS versus the *multidetector approach* using only optimized detector selection.

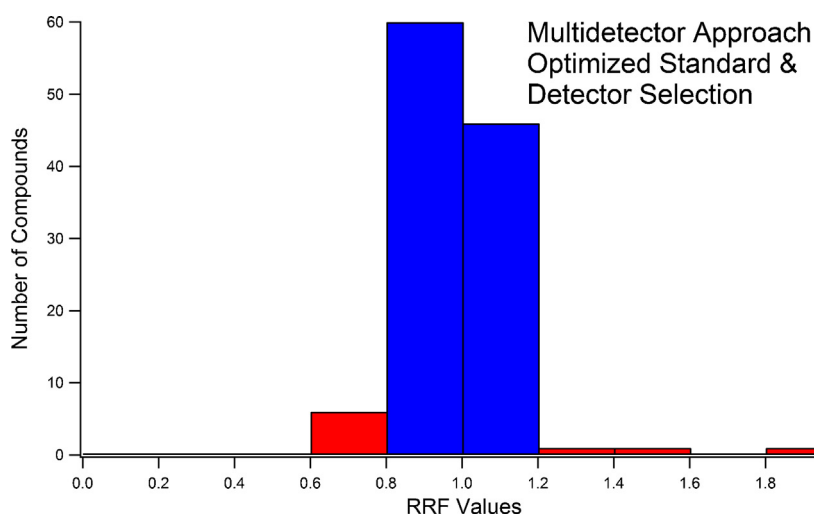


Fig. 6. Quantitative accuracy using the *multidetector approach* with optimized surrogate standard and detector selection.

being within $\pm 20\%$ of the accurate value and 91% being within 40%. This is a clear demonstration that the *multidetector approach* contains the data necessary to produce accurate quantitation with the addition of a reasonable number of surrogate standards. Accuracy which is generally within $\pm 40\%$ is in keeping with published regulatory expectations for food migration studies (a related field) using formal quantitation [30]. Given the inherent difficulties associated with surrogate standard quantitation and considering the RSD for even less variable methods such as GCMS, an accuracy result within 50–200% would be a reasonable expectation. The ability of the *multidetector approach* to exceed this performance standard supports its adoption as an effective analytical practice. In this analysis, the percentage of extractables for which each detector provided an optimum response (listed in the order UV, CAD, GCMS, FID) was 22.6, 28.3, 33.5 and 15.6% respectively demonstrating the need for all four detectors to obtain optimum quantitative accuracy in the *multidetector approach*.

One potential critique of this approach might be to question the ability of an analyst to choose the optimum detector or surrogate standard. While this critique is not entirely unmerited, it is important to remember that the RRF distributions for UV, CAD, GCMS and FID are narrower than that for LCMS. This reduces the burden on the analyst to make a perfect detector selection. The UV, CAD, GCMS and FID RRF distributions all showed sufficient quantitative accuracy ($\pm 40\%$ accuracy) for a higher percentage of the compounds

than the LCMS detector. Based on this one consideration alone, the use of any detector other than LCMS is strongly preferred for relative quantitation. Also, a comparison of different detector responses (UV vs CAD vs GCMS vs FID) can be used to assess if one detector might be strongly underestimating. Even more importantly, detector and surrogate standard selection can be reasonably guided by an assessment of the chemistry of the extractable following identification. Selection of an appropriate surrogate standard for UV and CAD can be guided by relatively simple principles which can be readily determined from the elemental formula of the extractable obtained from accurate mass data. For instance, the number of double bond equivalents (DBE) is generally an excellent predictor of the reliability of UV relative quantitation while molecular weight can be used as a reasonable predictor of CAD response. GCMS or FID values are usually calculated and reported separately from UV and CAD, providing a separate concentration value increasing the certainty that any safety assessment is not underestimating the concentration of an individual extractable. Thus, it is not necessary to achieve the optimum detector or standard selection in every case to substantially improve the quantitative accuracy of the extractables analysis.

Finally, it may be asked if LCMS can be applied effectively so long as sufficient surrogate standards are available. Unfortunately, it is our experience that LCMS produces widely different RRF values for seemingly very similar compound chemistry. For instance, consider

the case of Bisphenol A diglycidyl ether quantified against bisphenol A as a surrogate standard. Bisphenol A is observed to ionize only in negative mode while bisphenol A diglycidyl ether ionizes only in positive mode making relative quantitation impossible. As a second example, consider the relative quantitation of 5 very similarly structured hindered phenolic antioxidants (Irganox 245, Irganox 3114, Irganox 1076, Irganox 1035 and Irganox 1024) whose RRF data is shown in Table 2. All of these compounds are hindered phenols of relatively high monoisotopic mass (586, 783, 530, 642 and 552 amu), which elute at similar retention times (5.06, 5.96, 6.26, 5.78 and 5.31 min) and which have reasonably similar elemental formulas (C₃₄H₅₀O₈, C₄₈H₆₉N₃O₆, C₃₅H₆₂O₃, C₃₈H₅₈O₆S and C₃₄H₅₂N₂O₄). Based on their structure, it would seem logical to propose that LCMS negative mode should be used for relative quantitation and that any one of these compounds would be a reasonable surrogate standard for all the others. However, two of these compounds gave no response at all in LCMS negative mode. The RSD of the RRF values in LCMS positive and negative mode was 106% and 96% respectively. This is in spite of the fact that they have very similar chemistry and the same ionizable functional group (phenolic compounds). These seemingly random RRF values make intelligent and rational surrogate standard selection in LCMS very difficult and in our opinion, impractical. Comparing that to CAD and UV, the RSD of the RRF values was only 36% and 32%. This level of variability allows for reasonable levels of confidence when selecting surrogate standards.

3.4. RF Databases and the multidetector strategy

As a final point of discussion, a number of recent publications have proposed to use an RF database as a solution to both RF Problem 1: Underreporting and RF Problem 2: Quantitative Error [24,25]. In regards to RF Problem 1: Underreporting, it is acknowledged that having a comprehensive database would allow for determination of a UF value which would be sufficiently protective to avoid under reporting. However, this approach does nothing to reduce RF variation and hence the underlying need for a UF value or the uncertainty which it is designed to mitigate. UF values provide an added measure of conservatism by reducing the corresponding AET. For large UF values, the resulting AET may extend below the instrumental detection limit. In order to reach the reduced AET, it is frequently necessary to perform additional sample concentration (>10X) which increases the risk of compound loss and increases matrix effects which can potentially conceal extractables and leachables and complicates the process of identification. A larger UF also increases the risk of over reporting of compounds which are actually below the threshold but which have higher RF values. Some authors have suggested that the creation of an associative database could be used to “provide alerts to potential omissions [25]. “Presumably this would then allow for additional targeted screening to be performed specifically for the associated extractables. While the potential of this approach seems valuable, the authors also acknowledge that “the power of the database is derived from the number of substances in the database” and thus this approach is not helpful for catching those compounds which are not present in the database or which are not correctly associated. It further requires that additional targeted work be performed. In contrast, the *multidetector approach* requires no such prior knowledge and actually serves to reduce the need for a UF by reducing RF variation. This prevents under reporting without the potential for negative consequences associated with increased sample concentration. Similarly, the ability of an RRF database to improve Problem 2: Quantitative Error is based on the premise that the extractable under study is in the database. The largest commercial databases currently available include on the order of 1500 compounds (Agilent E&L Database) and some ven-

dors of E&L services have reported having databases greater than 2500 compounds [25]. While it is currently unknown how many compounds constitute the universe of all extractables, discussions at national meetings have frequently proposed numbers ranging from 10,000–100,000 or potentially more. In our own laboratories, we have identified and added more than 4000 extractables into a proprietary database but are daily detecting many new compounds which were as yet previously unencountered. This problem is further exacerbated by a lack of commercially available reference standards. Many of the most frequently encountered extractables are degradation products of antioxidants or oligomeric side products of the polymerization process. There are generally no commercial sources from which to purchase these standards. In addition, as new polymer systems and additives are being created the database approach would lag behind needing to catch up to current technology. For these reasons, the RRF database approach seems to leave very significant gaps given that the most extensive databases (including our own) likely contain far less than 25% of the extractables universe and do not include many frequently encountered extractables due to their not being commercially available. For these reasons, it is our opinion that the chief utility of an E&L RRF database is to define the RF distribution of the method to aid in standard selection, facilitate identification of compounds and to provide accurate quantitation for a subset of the most commonly observed commercially available extractables. Qualification of each E&L method should demonstrate that quantitative accuracy can be obtained not only for compounds contained in the database but also for compounds not in the database using relative quantitation with surrogate standards. This should be done using surrogate standard quantitation for a significant number of compounds not in the database with varied chemical properties with reasonable quantitative accuracy (the authors would suggest $\pm 40\%$ for >90% of compounds) and reasonably comprehensive compound coverage (the authors would suggest >90%). This proves that the method can successfully handle compounds which are not in the database. This is an essential aspect of insuring method quality since the majority of E&L compounds are not commercially available as standards. The *multidetector approach* can be applied broadly and without the requirement for a reference standard of each compound as it reduces RF variation.

4. Conclusions

The accurate quantification of extractables and leachables presents a significant analytical challenge due to the large diversity of extractable compounds and the lack of a universal detector. The generally applied strategy of using LCMS and GCMS for relative quantitation has significant, difficult to resolve issues due to the wide RRF variation inherent primarily in the LCMS results. The magnitude of this error was estimated for the universe of extractables (HVOCs excluded) using 217 extractables and it was found that only 33% of the compounds were within $\pm 20\%$ of the actual value and only 55% were within $\pm 40\%$. Correction of this error for LCMS analysis using optimized surrogate standard selection also did not show readily predictable, logical trends based on the chemistry of the extractables. In contrast, the use of the *multidetector approach* (proper detector selection) with optimized surrogate standard selection (using 10 surrogate standards) was shown to produce accurate values within $\pm 20\%$ for 85% of the compounds and 91% of compounds were within $\pm 40\%$ substantially reducing RF Problem 2: Quantitative Error. It was further demonstrated that the *multidetector approach* significantly enhances the assurance that compounds are not missed during the initial discovery phase of the analysis thus reducing RF Problem 1: Underreporting. This is accomplished not through the use of UF values but through reduc-

ing RF variation and through the synergies between detectors. As a step towards improving quality assurance in E&L analyses, it is recommended that the sensitivity of E&L methods be verified at the time of the analysis utilizing a group of surrogate standards at the AET concentration whose average RRF approaches the center of the method's RRF distribution ($RRF \approx 1$ for normalized values) and which also contains a standard with an RRF value approximately equal to one divided by the UF ($1/UF$). This proves adequate method sensitivity. It is further recommended that the distribution of RRF values for the surrogate standards should be sufficiently broad to allow for correction of RF Problem 2: Quantitative Error. To identify appropriate standards, each laboratory should develop a method specific RF database, the primary purpose of which is to aid in identification, improve quantitative accuracy for frequently observed extractables and leachables and confirm that the RRF values for the standards are at appropriate locations in the distribution. The use of an RF database does not reduce the underlying problem of RF variation and as a solution for RF Problem 1; AET Underreporting and RF Problem 2; Quantitative Error has limited scope. RF databases allow calculation of a method specific UF but do nothing to reduce the underlying method uncertainty and the resulting lowered AET values can require additional sample concentration which can lead to compound loss and increased matrix effects. Similarly, quantitative error is reduced only for those compounds in the database. The RF database approach is in essence a targeted quantitative approach applied to non-targeted analysis. In contrast, the multidetector approach is at its core a non-targeted analytical approach for a non-targeted analytical problem (extractables and leachables screening). This study has demonstrated that the *multidetector approach* with a UF of 2 produced 94% coverage for all tested extractables and was equally protective (1% difference) as a UF of 10 when using LCMS and GCMS only.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors statement

The Relative response factor data used to generate this publication is part of a proprietary database developed and maintained by Jordi Labs. The authors are happy to answer questions related to this data but must maintain the proprietary nature of the database due to business interests.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

I would like to acknowledge the dedicated staff at Jordi Labs who tirelessly supported this study. We wish to recognize the contributions of the following people; Xiaoqing Kong, Jordan Tocher, Nisansala Hewage, Hongda Cao, Pravin Pathak, Xunzhe Cui and Lindsay Ross

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jpba.2020.113334>.

References

- [1] International Conference on Harmonisation, Q2(R1) Validation of Analytical Procedures: Text and Methodology, Retrieved from https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-2-r1-validation-analytical-procedures-text-methodology-step-5_en.pdf (last Accessed 1 January, 2020), 2020.
- [2] Food and Drug Administration, Guidance for Industry: Q2B Validation of Analytical Procedures: Methodology, Retrieved from <https://www.fda.gov/media/71725/download> (last Accessed 1 January, 2020), 1996.
- [3] United States Pharmacopeia, USP <661.2> Plastic Packaging Systems for Pharmaceutical Use, Retrieved from https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/revisions/661.2_rb_notice.pdf (last Accessed 1 January, 2020), 2020.
- [4] USP 40 General Chapter Assessment of Extractables Associated With Pharmaceutical Packaging/Delivery Systems <1663>, United States Pharmacopeial Convention, Rockville, MD, 2017, pp. 2020–2035.
- [5] USP 40 General Chapter Assessment of Drug Product Leachables Associated With Pharmaceutical Packaging/Delivery Systems. <1664>, United States Pharmacopeial Convention, Rockville, MD, 2017, pp. 2035–2047.
- [6] International Organization for Standardization, Biological Evaluation of Medical Devices, Part 12: Sample Preparation and Reference Materials, 10993-12: Retrieved from <https://www.iso.org/standard/53468.html> (last Accessed 1 January, 2020), 2012.
- [7] International Organization for Standardization, Biological Evaluation of Medical Devices, Part 17: Establishment of Allowable Limits for Leachable Substances, 10993-17: Retrieved from <https://www.iso.org/standard/23955.html> (last Accessed 1 January, 2020), 2002.
- [8] International Organization for Standardization, Biological Evaluation of Medical Devices, Part 18: Chemical Characterization of Medical Device Materials Within a Risk Management Process, 10993-18: Retrieved from <https://www.iso.org/standard/64750.html> (last Accessed 1 January, 2020), 2020.
- [9] International Conference on Harmonisation, M7(R1) Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals To Limit Potential Carcinogenic Risk: Guidance for Industry, Retrieved from <https://www.fda.gov/media/85885/download> (last Accessed 1 January, 2020), 2020.
- [10] D. Norwood, L. Nagao, The Analytical Evaluation Threshold (AET), Product Quality Research Institute, 2015, Retrieved from <http://pqri.org/wp-content/uploads/2015/08/pdf/PQRI.AET.Poster.pdf> (last Accessed 9 June, 2017).
- [11] M. Jordi, S. Khera, K. Rowland, L. Jiang, P. Solomon, J. Nelson, S. Lateef, J. Woods, S. Martin, F. Aiello, N. Chen, Qualitative assessment of extractables from single-use components and the impact of reference standard selection, *J. Pharm. Biomed. Anal.* 150 (2018) 368–376.
- [12] D. Jenke, Moving forward towards standardized analytical methods for extractables and leachables profiling studies, *PDA J. of Pharm. Sci. Technol.* 69 (2015) 471–476.
- [13] Y. Gao, N. Allison, Extractables and leachables issues with the application of single use technology in the biopharmaceutical industry, *J. Chem. Technol. Biotechnol.* 91 (2016) 289–295.
- [14] K. Li, G. Rogers, Y. Nashed-Samuel, H. Lee, A. Mire-Sluis, B. Cherney, R. Forster, P. Yeh, I. Markovic, Creating a holistic extractables and leachables (E&L) program for biotechnology products, *PDA J. Pharm. Sci. Technol.* 69 (2015) 590–619.
- [15] L. Marghitou, J. Liu, H. Lee, L. Perez, K. Fujimori, M. Ronk, M. Hammond, H. Nunn, A. Lower, G. Rogers, Extractables analysis of single-use flexible plastic biocontainers, *PDA J. Pharm. Sci. Technol.* 69 (2015) 49–58.
- [16] M. Hammond, H. Nunn, G. Rogers, H. Lee, A. Marghitou, L. Perez, Y. Nashed-Samuel, C. Anderson, M. Vandiver, S. Kline, Identification of a leachable compound detrimental to cell growth in single-use bioprocess containers, *PDA J. Pharm. Sci. and Technol.* 67 (2013) 123–134.
- [17] D. Jenke, A. Odufu, 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.* 50 (2012) 206–212.
- [18] M. Iqbal, K. Rao, G. Sridhar, P. Raju, G. Deshpande, J. Babu, Characterization and relative response factor determination of process related impurity in naproxen by nuclear magnetic resonance spectroscopy, *J. Pharm. Biomed. Anal.* 56 (2011) 484–490.
- [19] Guide to achieving reliable quantitative LC-MS measurements, in: M. Sargent, A. Contributors Sage, C. Wolff, C. Mussell, D. Neville, G. Lord, M. Saeed, R. Lad, R. Godfrey, S. Hird, V. Barwick (Eds.), *RSC Analytical Methods Committee*, first edition, 2013, ISBN 978-0-948926-27-3.
- [20] E. Adams, M. Raffie, E. Roets, J. Hoogmartens, Liquid chromatographic analysis of streptomycin sulfate, *J. Pharm. Biomed. Anal.* 24 (2000) 219–226.
- [21] J. Blanz, G. Williams, J. Dayer, T. Délmonté, W. Gertsch, P. Ramstein, R. Aichholz, M. Trunzer, D. Pearson, Evaluation of relative MS response factors of drug metabolites for semi-quantitative assessment of chemical liabilities in drug discovery, *J. Mass Spectrom.* 52 (2017) 210–217.
- [22] P. Sun, X. Wang, L. Alquier, C. Martanoff, Determination of relative response factors of impurities in paclitaxel with high performance liquid

- chromatography equipped with ultraviolet and charged aerosol detectors, *J. Chromatogr. A* 1177 (2008) 87–91.
- [23] S. Baertschi, Analytical methodologies for discovering and profiling degradation-related impurities, *Trends Anal. Chem.* 25 (2006) 758–767.
- [24] M. Rozio, A. Rosato, L. Iadarola, S. Carrara, Compounds Relative Response Factor, a Reliable Quantification Within Extractables Testing, White Paper, Eurofins Biopharma Product Testing Italy, 2019, December.
- [25] P. Christiaens, J. Beusen, P. Verlinde, J. Baeten, D. Jenke, Identifying and mitigating errors in screening for organic extractables and leachables: part 1; Introduction to errors in chromatographic screening for organic extractables and leachables and discussion of the error of omission, *PDA J. Pharm. Sci. Technol.* 74 (2020) 90–107.
- [26] O. Wahl, U. Holzgrabe, Impurity profiling of ibandronate sodium by HPLC–CAD, *J. Pharm. Biomed. Anal.* 114 (2015) 254–264.
- [27] M. Nussbaum, S. Baertschi, P. Jansen, Determination of relative UV response factors for HPLC by use of a chemiluminescent nitrogen-specific detector, *J. Pharm. Biomed. Anal.* 27 (2002) 983–993.
- [28] P. Hong, A. Phoebe, M. Jones, Study of relative response factors and mass balance in forced degradation studies with liquid chromatography/photo-diode array detector/evaporative light scattering detector/mass spectrometry system, *J. Chromatogr. A* 1512 (2017) 61–70.
- [29] S. Gustavsson, J. Samskog, B. Langstrong, K. Markides, B. Långström, Study of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents, *J. Chromatogr. A* 937 (2001) 41–47.
- [30] Guidance for Industry, Preparation of Premarket Submissions for Food Contact Substances: Chemistry Recommendations, Retrieved from <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-preparation-premarket-submissions-food-contact-substances-chemistry> (last Accessed 9 March, 2020), 2020.