



Uncertainty factors and relative response factors: correcting detection and quantitation bias in extractables and leachables studies

Marco Giulio Rozio¹ · Davide Angelini¹ · Simone Carrara¹

Received: 11 March 2025 / Revised: 22 May 2025 / Accepted: 23 May 2025
© The Author(s) 2025

Abstract

The transfer of chemicals from packaging or medical devices to drug formulations, known as extractables and leachables (E&L) release, can affect drug strength and safety. These released substances must be monitored and assessed through toxicological evaluation. Identifying and quantifying analytes above a specific analytical evaluation threshold (AET) is crucial, but variability in response factors (RFs) complicates accurate detection, leading to potential errors in quantitation. An uncertainty factor (UF) can partially correct this, though it is limited by RF variability, and a multidetector approach improves characterization but does not fully resolve quantitation bias. The RRFlow model proposed in this study offers a solution by determining E&L concentrations without real-time reference standards analysis. It involves identity confirmation, RRF validation, and applies an average corrective factor (RRFi). A numerical simulation benchmark (NSB) is used to compare different scenarios, such as varying UF values, RRFlow application, and fixed rescaling factors. The benchmark assigns concentration values to model compounds with different response factors, iterating the process to evaluate the number of false positive and negative errors. The numerical simulations show that RRFlow reduces detection bias and outperforms UF-based methods, mitigating false positives and negatives.

Keywords Extractables · Leachables · Uncertainty factor (UF) · Analytical evaluation threshold (AET) · Relative response factor (RRF) · Numerical simulation benchmark (NSB)

Introduction

The contact between drug products and container closure systems (CCS) or single/multiple use systems (SUS/MUS) during manufacturing can release chemicals and polymers into the drug products [1–4]. Similarly, medical devices can release chemicals into patients through direct or indirect contact. These compounds, known as extractables and leachables, can affect the efficacy or safety of products, necessitating their assessment below specific toxicological thresholds.

Identifying relevant extractables often requires complex chromatographic analysis; nevertheless, the characterization of all detected analytes is not essential if it can be shown that a compound poses no toxicological risk below a certain concentration. The analytical evaluation threshold (AET) is used to determine which compounds should be included in safety assessments. This threshold assumes similar response factors (RF) for all compounds, but it is not what it is experimentally observed, and this situation can lead to inaccurate results. To address this, the relative response factor (RRF), which is a ratio between an analyte RF and a reference compound RF, can be used to adjust for these differences even if variability in RFs across detectors creates challenges in detection and quantification [5, 6].

✉ Simone Carrara
simone.carrara@bpt.eurofinseu.com

Marco Giulio Rozio
marcogiuliorz@gmail.com

Davide Angelini
davide.angelini@bpt.eurofinseu.com

¹ Department of Extractables & Leachables and Packaging Testing, Eurofins BPT, Via Buozzi 2, 20055 Vimodrone, MI, Italy

Correcting the detection bias

Detection bias in extractables and leachables (E&L) testing can cause the misdetection of the low-responding compounds. These compounds should be reported for the

toxicological evaluation, but due to their low RF, the results are below the AET.

To decrease the impact of detection bias, an AET adjustment has been proposed in recent years by introducing the uncertainty factor (UF). This is easily applied to take into account the analytical uncertainty of the screening methods used to estimate extractable concentrations in a test sample [7, 8]. ISO 10993–18:2020 is the only guideline which reports some indications about the use of statistical analysis on RRF data for the applied method, to establish the correct UF to apply. The UF should reflect the variability of the RRF of the extractables belonging to a specific dataset. Greater RRF variability should correspond to higher UF values, and the AET shall therefore be corrected by the ratio with the UF. If the RSD in RRF databases (DB) is high, it can be assumed that the adoption of a strategy of multiple complementary and orthogonal methods, also defined as a multidetector approach, increases the detection and identification capabilities. Considering this combination approach, LC/MS could compensate for the shortcomings of GC/MS and HS-GC/MS and vice versa [8–10].

Correcting the quantification bias

If the UF partially solves the problem of under-reporting extractable compounds, the problem of quantitative error remains [8–10]. The structural identification of the extractable compounds migrating from packaging, process components, or medical devices is crucial. This is because the actual toxicity of a molecule and its associated permitted daily exposure (PDE) can be evaluated only if its chemical structure is known. The quantitation of these compounds is critical since, in E&L studies, semi-quantitation is usually performed against a minimum number of reference standards. These could show different RF with respect to the compounds of interest. The result is an approximate means of quantitation that could lead to unwarranted concerns or, more dangerously, to false negatives. Misdetection bias, incorrect identification, and imprecise quantitation are in fact encountered during an E&L study, and these issues can lead to an incomplete or misleading toxicological assessment [11, 12]. They are mainly due to inadequate workflow or strategy and, specifically, to methods that produce imprecise responses. A more precise assessment is fundamental to overcome these drawbacks.

The following article describes a useful workflow, defined as “RRFlow” [13], as a new model for extractables quantitation. In this model, the analyses provide more precise and reliable data by applying the RRF of each compound in the concentration rescaling during the extractables assessment. The present work focuses on LC/MS and GC/MS analyses that represent the analytical techniques exhibiting the

greatest RRF variability. Additionally, in order to improve the understanding of the impact of a combined approach based on the RRF and UF, a numerical simulation benchmark (NSB) is built using a simplified model for extractables studies. Since the RRFlow method is applied with an experimentally determined UF, it is necessary to measure its impact in comparison to different UF-only approaches, i.e., with different UF levels or by applying fixed scaling of the results.

The benchmark is based on a set of 50 extractable compounds, covering a wide range of RRF. A random concentration value is assigned to each compound and compared to a selected AET. All false positive (type I) or false negative (type II) errors, that directly indicate the quality of the chosen approach, are counted and reported. The NSB is repeated multiple times to check numerous concentration values, thereby obtaining significant results and demonstrating the incidence of over-reporting and under-reporting.

A combined UF and RRFlow approach shows a lower incidence of type I and type II errors when compared to previous approaches in E&L studies.

Materials and methods

Numerical simulation benchmark

The numerical simulation benchmark (NSB) is implemented in a Microsoft Excel VBA macro.

Chemical reagents and materials

Standards used for the preparation of the HPLC–ESI–MS and GC/MS RF and RRF sessions are purchased from Sigma-Aldrich Co. (St. Louis, MO, USA) and TRC (Toronto, Canada). HPLC–ESI–MS grade methanol, ethanol, acetone, dichloromethane (DCM), hexane, dimethylsulfoxide, chloroform, isopropanol, acetonitrile, water, and ammonium formate are purchased from Sigma-Aldrich Co. (St. Louis, MO, USA), Honeywell Riedel-de Haën (Seelze, Germany), Merck (Darmstadt, Germany), and VWR (Radnor, PA, USA). DCM is analytical grade and is used without further purification from VWR (Radnor, PA, USA).

Standard preparation

Stock solutions of the standards for HPLC–ESI–MS and GC/MS RF checks and RRF sessions are prepared by individually dissolving them at 1 mg/mL. Portions of the stock solutions are then combined into working solutions, containing several standards that are subsequently diluted to concentrations in the range 0.05–10 µg/mL for analysis. Methanol is used to dissolve standards for QTOF-LC/MS, while DCM is

applied for GC/MS analysis. In rare instances, standards that show poor solubility in methanol or DCM are initially dissolved in another suitable solvent and subsequently diluted in methanol or DCM.

Instrumental analysis

HPLC-ESI-QTOF-MS HPLC-ESI-QTOF-MS analyses are performed to identify and quantify reference standards and extractable compounds using an Agilent 6530 HPLC/QTOF with an electrospray ionization (ESI) source coupled with an Agilent 1260 HPLC system (Agilent Technologies, Santa Clara, CA).

The instrumental parameters are gas temperature 350 °C; nebulizer gas (N₂) 60 psi; gas flow (N₂) 10 L/min; VCap 3500 V; and mass range 50–1500 m/z. Separations are performed on an Agilent Zorbax Eclipse XDB-C18 column, 2.1 mm × 50 mm, 1.8 μm (Agilent Technologies, Santa Clara, CA), maintained at 65 °C. The mobile phase consists of 5 mM ammonium acetate in water (solvent A) and 50:50 (v/v) acetonitrile/methanol (solvent B) delivered at 0.4 mL/min. Gradient elution has an initial condition of solvent A (90%), A:B 50:50 at 1.25 min, 100% B at 4 min, and 90% A at 12 min. The run time is 18 min. The HPLC eluent is introduced directly into the MS system. The high-resolution mass of the instrument is ensured by the continuous infusion of a calibration solution during the chromatographic run. The calibration solution consists of 0.3 mL purine + 1 mL HP-921 to 400 mL acetonitrile/water 95:5 (v/v), diluted 1:300 (ESI positive mode) or 4 mL purine + 0.2 mL HP-921 to 1100 mL of acetonitrile/water 95:5 (v/v) (ESI negative mode).

GC/MS GC/MS analyses are performed using an Agilent 5977 mass spectrometry detector (MSD) with an electron ionization (EI) source coupled with an Agilent 8890 (G3540A) GC system (Agilent Technologies, Santa Clara, CA). The instrumental parameters are GC program temperature from 40 °C (held for 1 min) to 280 °C (held for 2 min) with a rate of 10 °C/min, then to 310 °C (held for 10 min) with a rate of 15 °C/min; helium carrier gas constant flow at 1.0 mL/min; injection mode splitless; and acquisition mode scan 33–800 m/z.

Separations are performed on an Agilent HP-5MS column, 30 m × 0.25 mm, 0.25 μm film thickness, or equivalent (Agilent Technologies, Santa Clara CA). Run time is 39 min.

MS data processing

HPLC/QTOF HPLC/QTOF data analysis is performed using Agilent MassHunter Qualitative Analysis (B.10.00) in

combination with the Eurofins Extractables Database (EED) (interfaced to MassHunter Qualitative Analysis through Agilent PCDL manager software), consisting of over 1500 compounds. Compound identities are confirmed using purchased reference standards when available. The Agilent MassHunter Molecular Structural Correlator (MSC) (B.07.00) interfaced with an online database (e.g., ChempSpider, mzCloud) is used for MS/MS analysis and structural characterization. The RRF determination is based on the sum of Qualifier (QLF) and Quantifier (QTF) mass signals [13] related to each investigative analyte at all concentrations and is performed by applying Agilent MassHunter Quantitative Analysis (B.08.00).

GC/MS GC/MS data analysis is performed using Agilent MassHunter Workstation Software, Qualitative Analysis Navigator (B.08.00), in combination with the Wiley Registry® 12th 7 Edition/NIST 2020 Mass Spectral Library.

The RRF determination is based on the extraction of the mass spectrum of each peak signal detected in the total ion current (TIC) chromatogram with a signal-to-noise (s/n) ratio greater than 3. These are then compared to the mass spectra stored in a personal compound and database library (PCDL) to identify the investigated analytes. The related peak areas (by software integration) are considered for quantitation. The evaluation is performed by applying Agilent MassHunter Workstation Software—Quantitative Analysis—Unknowns Analysis (B.08.00).

The RRFlow approach for HPLC-ESI-MS and GC/MS

The RRFlow approach is a quantitation model to determine the actual concentration of E&L compounds detected during a study in the absence of individual reference standards. The RRF of a compound *i* (RRF_{*i*}) is defined as the ratio between the RF of that compound (RF_{*i*}) and the RF of the reference compound for the E&L GC/MS and HPLC-ESI-QTOF-MS study (RF_{ref}).

$$\text{RRF}_i = \text{RF}_i / \text{RF}_{\text{ref}}$$

The RF of compound *i* (RF_{*i*}) is the ratio between the concentration of the standard solution (C_{*i*}) and the experimental peak area (Area_{*i*})

$$\text{RF}_i = C_i / \text{Area}_i$$

RRF_{*i*} is a corrective factor applied for practical reasons in RRFlow, which is equivalent to 1/RRF as normally reported in the majority of literature documentation (RRF = RF_{ref}/RF_{*i*}). It allows for the rescaling of the experimental concentration without the necessity of standard material analysis to determine the compound instrumental response.

The application of the rescaling factor is applied by GC/MS and HPLC-ESI-QTOF-MS to the extractables where data are available and that exhibit RRF_i less than 0.5 or RRF_i greater than 2. For those extractables that fall between these values, the approximation obtained by the traditional approach, based on a semi-quantification assessment, is considered acceptable.

RRFlow key points

The first step in the determination of a verified extractables RRF_i that can be applied to the extractables study data assessment is to eliminate analytical variability by setting specific key points as shown in Fig. 1.

Step 1—Extractable identity confirmation

The extractable identity confirmation, due to the different instrumental resolution (HR for QTOF-MS and LR for GC/MS), is performed in the following manner:

GC/MS Phenanthrene- d_{10} is chosen by the Eurofins E&L group as the most suitable GC/MS reference compound for semi-quantification of all migrating compounds found in E&L studies. A working standard solution of Phenanthrene d_{10} and the reference standards of the extractables to be investigated are purchased and analyzed at a concentration of 1 $\mu\text{g/mL}$. This analysis determines qualitative and quantitative data, such as relative retention time (RRT) and RRF_i values, to confirm identification of the extractables and the instrumental response level. All these useful data are reported in the in-house database for GC/MS.

The RRF obtained from the 1 $\mu\text{g/mL}$ solution of the reference standard shows the overestimation or underestimation bias of the extractable amount as determined during the semi-quantification stage of an extractable study. Only if the acceptance criteria reported in Table 1 are met, the second step of the RRFlow is executed.

HPLC-ESI-MS Reserpine and Irganox 1098 are chosen by the Eurofins E&L group as the LC/MS reference compounds (ESI+ and ESI- respectively) for the semi-quantification of all migrating compounds found in E&L studies. A working

standard solution of reserpine/Irganox 1098 (1 $\mu\text{g/mL}$) and the reference standards of the extractables to be investigated are purchased and analyzed at a concentration of 1 $\mu\text{g/mL}$. This analysis will determine qualitative and quantitative data, such as RRT and RRF_i values, to confirm the extractable identification and instrumental response level. After the identification step, information regarding RRT and RRF evaluated at a single concentration level and other useful data are reported in the EED for LC/MS.

The RRF_i determined from the 1 $\mu\text{g/mL}$ solution of the reference standard shows the overestimation or underestimation bias of the extractable amount as determined during the semi-quantification stage of the extractable study. Only if the acceptance criteria reported in Table 2 are met, step 2 of the RRFlow is executed.

Step 2— RRF_i session and method validation

The RRF_i evaluation is based on the most intense signals in the mass spectrum that significantly contribute to the instrumental response of each compound. For this reason, their detection at the specific RRT is used as the trigger prior to concentration rescaling.

GC/MS Mass signals are evaluated by the fragmentation of the molecular mass in the TIC, which is the same detection mode for extractables screening. For each compound subjected to the RRF_i assessment, the average RRF_i is determined through an eight-point calibration curve (0.05 to 10 $\mu\text{g/mL}$). DCM solutions spiked with the internal standard phenanthrene- d_{10} (1 $\mu\text{g/mL}$) are used to evaluate specificity, linearity, and quantitation limit (QL) of the method. For each compound subjected to RRFlow, the following method validation parameters are tested in order to verify data reliability.

Method validation parameters Specificity: RRF_i determination is performed by evaluating the chromatographic peak attributed to the compound in the TIC, by considering the entire mass spectrum profile. Specificity is assessed by checking that the blank solution (DCM) does not show any specific mass signals at the same retention time as the reference standard.



Fig. 1 RRFlow key points

Table 1 System suitability acceptance criteria GC/MS

Parameter	Acceptance criteria
System suitability	<ul style="list-style-type: none"> • The RSD of the response (peak area) of three injections of phenanthrene-d₁₀ solution (1 µg/mL) must be not more than 20% • The DEV%^a of the response (peak area) of the phenanthrene-d₁₀ solution (1 µg/mL) analyzed at the end of the analytical batch with respect to the average peak area of the three injections of phenanthrene-d₁₀ solution (1 µg/mL) at the beginning of the run, must be not more than 30%
RRF value (1 µg/mL)	• RRF < 0.5 or RRF > 2

^aDEV% between values x and y is calculated as $|x-y|/[(x+y)/2]$

Linearity: determination of a concentration range where the analytical response of the compound is linear is a crucial step for the application of a unique RRF_i for concentration compound rescaling. The range of the linearity assessment is from 0.05 to 10 µg/mL (0.05, 0.1, 0.2, 0.5, 1, 2.5, 5, 10 µg/mL). This represents the range of concentrations covering the majority of compounds determined in extractables studies.

Quantitation limit (QL): the assessment of method sensitivity is important to determine the lowest concentration at which the rescaling factor can be applied. For those compounds that show a low analytical response, the QL is the lowest point of the range for linearity determination. The QL is evaluated by the analysis of a diluted standard solution at the concentration producing a peak with a s/n ratio greater than or equal to 10.

RRF_i (average): this is the mean value between the RRF_i determined at each reference standard concentration level of the tested linearity range. The RRF_i is evaluated as the ratio of the RF of the reference standard solution determined by the evaluation of the chromatographic peak area in the TIC (triplicate analysis) compared to the RF of the internal standard phenanthrene-d₁₀ solution (1 µg/mL) (triplicate analysis).

If all acceptance criteria reported in Table 3 are met, the mean RRF_i for each extractable is used to rescale the amount detected in the extractables studies.

HPLC-ESI-MS

Qualifier and quantifier mass signals definitions The quantifier mass signals (QTFs) are the mass signals that

significantly contribute to the peak area of the compound, i.e., charged molecule, ion adducts, dimers, fragments, etc. The RRF_i evaluation is based on the sum of the quantifier mass signals (QTFs), which should represent at least 80% of the compound peak area in the TIC. There may be nevertheless a high number of these related ions for each candidate, and the analytes usually found in extractables studies can be very high; thus, target searching of these signals could result in a huge amount of data to be managed. A helpful technique to overcome the challenges of this evaluation is the utilization of the qualifier mass signal (QLF). The QLF is the most intense or representative mass signal found in the compound mass spectrum and is always present in the chromatogram if the extractable compound is detected in the analyzed samples.

QLF signals are searched by extraction ion currents (EIC) analysis. If there is a match at the expected retention time, QTFs are considered for quantification: the compound resulting area is the sum of the EIC peaks signal from each Quantifier Mass. This option greatly reduces the amount of data that needs to be utilized by the analytical software at the analyte detection stage. It permits the attention to be focused on those compounds that actually require deeper investigation. The QLF-QTFs approach, ensuring a specific and sensitive data assessment since done by focusing on selected and highly resolved m/z signals, allows moreover the detection of extractables exhibiting low response factors. These are often not detected by the standard TIC

Table 2 System suitability acceptance criteria LC/MS

Parameter	Acceptance criteria
System suitability	<ul style="list-style-type: none"> • The RSD of the response (peak area) of three injections of reserpine/Irganox 1098 solution (1 µg/mL) must be not more than 10% • The DEV% of the response (peak area) of reserpine/Irganox 1098 (1 µg/mL) analyzed at the end of the analytical batch with respect to the average peak area of the three injections of the reserpine/Irganox 1098 solution (1 µg/mL) at the beginning of the run, must be not more than 20%
RRF value (1 µg/mL)	• RRF < 0.5 or RRF > 2

extractables approach since they are below the detection limit (DL).

For each compound subjected to the RRFlow, the following parameters are tested to verify data reliability.

Method validation parameters Specificity: the RRF_i determination is performed by considering the sum of the most representative ions of the compound (QTF). Only non-interfering mass signals must be applied for compound quantification. Specificity is assessed by checking that the blank solution (ethanol/water, 50:50) does not show any specific mass signals at the same retention time window as the reference standard.

Linearity: the determination of a concentration range over which the analytical response of the compound is linear is a fundamental step in the application of a unique RRF_i for concentration compound rescaling. The range of the linearity evaluation is from 0.05 to 10 $\mu\text{g/mL}$ (0.05, 0.1, 0.25, 0.5, 1, 2.5, 5, 7.5, 10). This represents the range of concentrations covering most compounds detected in extractables studies. Triplicate analysis of standards at each concentration level is performed.

Quantitation limit (QL): the assessment of method sensitivity is important to determine the lowest concentration at which the rescaling factor can be applied. For those compounds that show a low analytical response, the QL is applied as the lowest point of the linearity range. The QL is evaluated by the analysis of a diluted standard solution with a concentration that produces a peak (as sum of the EIC) with a s/n ratio greater than or equal to 10.

RRF_i (average): this is the mean between the RRF_i determined at each reference standard concentration level of the linearity range. The RRF_i is evaluated as the ratio of the RF of the reference standard solution, determined by the evaluation of the chromatographic peak area in the EIC, with the RF of the internal standard reserpine (ESI positive detection) or Irganox 1098 (ESI negative detection).

If all acceptance criteria reported in Table 4 are met, the mean value of the RRF for each extractable is used to rescale the amount detected in the extractables study.

RRFlow application: from theory to practice

To test and evaluate the submitted workflow, two sets of compounds have been selected, 140 for HPLC–ESI–MS and 125 for GC/MS analysis, between the most common extractables found in E&L studies. They are analyzed by the HPLC–ESI–MS and GC/MS instrumental methods described in Sect. 3.3. An evaluation of the RRF_i has been conducted by comparing the responses of the target compounds and the reference standards used in the semi-quantitative approach. A total of 82 compounds from the 136 analyzed by HPLC–ESI–MS (60%) and 77 compounds from the 124 analyzed by GC/MS (62%) show RRF_i that fell in the rescaling factor range ($RRF_i < 0.5$ or $RRF_i > 2$). For this reason, they have been subjected to the RRFlow approach.

For LC/MS, a nine-point calibration curve is performed for each compound in the range 0.05–10 $\mu\text{g/mL}$ and an average RRF_i is defined. The same approach is applied to GC/MS analysis. An eight-point calibration curve is performed for each compound in the range 0.05–10 $\mu\text{g/mL}$. The acceptance criteria described in Tables 3 and 4 have been applied to the obtained analytical data. If the system suitability criteria are not achieved, the linearity range may be reduced (for compounds showing a low instrumental response) or split in two (for compounds that show instrumental response saturation) to meet the precision requirements. The average RRF_i determined by RRFlow is then applied to re-calculate the amount of the corresponding compound detected in the extractables study, using the following formula (Fig. 2).

Where:

- *Experimental concentration* is the extractable compound concentration calculated by the semi-quantitative screening against a fixed reference standard;
- *Average RRF_i* is the average value of the RRF_i calculated using the linearity range as a result of the RRFlow.

The above formula is applied whenever an extractable is detected, to rescale the concentration, which is then compared to the AET. If the concentration falls outside the linear range, the sample should be re-analyzed after analytical

Table 3 Method validation acceptance criteria

Parameter	Acceptance criteria
Specificity	No significant peak observed in blank solution at the retention time of reference standard
Quantification limit	S/N ratio ≥ 10
Linearity (peak area vs concentration)	$R^2 \geq 0.98$ (minimum of three concentration levels)
Precision	<ul style="list-style-type: none"> • The RSD of the response (peak area) in each set of the three injections of reference standard must be not more than 20% • Calculate the RRF at each concentration level over the tested linearity range. The RSD on the obtained RRF values must be not more than 30%

Table 4 Method validation acceptance criteria

Parameter	Acceptance criteria
Specificity	No significant peak observed in blank solutions 1 or 2 at the retention time of the reference standard
Quantification limit	S/N ratio ≥ 10
Linearity (response versus concentration)	$R^2 \geq 0.98$ (minimum of three concentration levels)
Precision	<ul style="list-style-type: none"> • The RSD of the response (peak area) in each set of three injections of reference standard at each concentration level must be not more than 10% • The RSD of the response (peak area) of three injections of reference standard at the QL must be not more than 20% • The RSD of the average RRF of the reference standard calculated at each concentration level must be not more than 25%
Stability evaluation	<ul style="list-style-type: none"> • The DEV% of the average concentration of three injections of reference standard (1 $\mu\text{g/mL}$) evaluated at the end of the RRFlow with respect to the concentration of the single injection of the corresponding reference standard evaluated at the beginning of the analysis must be not more than 20%

treatment (i.e., concentration or dilution) to permit a reliable rescaling.

Results and discussion

RRF and UF evaluations and their limitations

The chemical compounds present in the sample solution at the same concentration could exhibit different RF compared to the chosen internal standard.

One of the direct consequences of such discrepancy is misdetection of E&L compounds by over-reporting or under-reporting due to quantitative overestimation or underestimation, respectively. To mitigate the effect, uncertainty factors (UFs) are introduced in the AET calculation, usually expressed as integer values or percentages—for example, a UF of 2 corresponds to a reduction of the AET to 50% of the original value.

While the UF value is necessary to take into account measurement uncertainty (i.e., instrumental precision), its main effect is to lower the AET enough to be able to report low-responding compounds. UF, however, cannot be set arbitrarily high as it will produce AETs lower than the instrumental detection limit.

For this reason, an effective UF value should be:

- based on specific experimental data which takes into account the actual analytical method and procedures followed by a laboratory

- reliant on a statistical parameter

This approach is in line with ISO 10993–18:2020 which suggests in Annex E, a UF estimation method using the following formula:

$$\text{UF} = \frac{1}{1 - \text{RSD}}$$

where:

- RSD is the average relative standard deviation of the RRF for a set of compounds (GC/MS and LC/MS RSDs are calculated separately)

A RRF value database of 136 compounds for HPLC–ESI–MS and 125 compounds for GC/MS has been considered for UF calculation.

Nevertheless, the complete list of RRF values is not compatible with the formula shown above. RRF values can range over several orders of magnitude for a single technique, resulting in RSD values greater than 1—this renders the above formula unusable as it is; however, the exclusion of a few compounds from the evaluation, based on a rationale, proved to be effective in obtaining meaningful results.

Compounds are excluded considering the following:

1. High response factor

There is no risk of detection failure for compounds presenting RRF greater than 1. This is because the intensity

**Fig. 2** RRFlow application

of analytical response is higher than that of the reference standard. Therefore, as a criterion for RRF selection, only compounds with an RRF less than 1 are actually considered for UF calculation.

2. Low compatibility with the analytical technique

It is clear that compounds with an extremely low RRFs are not adequately detected by that specific technique. It is likely that these chemicals can be suitably evaluated by a different technique. Thus, a lower RRF limit can be applied for both GC/MS and LC/MS. This serves as a lower limit for compounds that can be effectively detected by the specific technique. Compounds with RRF lower than 0.05 are excluded from UF calculation (20 times lower than the internal standard). The choice is in line with the range considered for RRF evaluation since the lowest concentration is 0.05 µg/mL or 20 times lower than the reference concentration of 1 µg/mL.

In addition to the categories above, in the case of compounds detectable by both GC/MS and LC/MS, only the RRF value related to the technique that shows the higher value (higher sensitivity) is considered. This is a direct result of applying the strategy of multiple complementary and orthogonal methods [8, 10].

The effectiveness of the application of such rationale can be shown below for each technique by plotting the calculated UF against the percentage of compounds considered for calculation. The final value is obtained by gradually reducing the set of compounds (shown from left to right) until all compounds that fall into either category 1 or 2 are removed from the pool (as indicated above).

GC/MS For GC/MS (Fig. 3, see supplementary material), at least ≈7% of total values have to be removed from the calculation pool to obtain a positive UF value. After 10% of values have been removed, the UF value has a slower rate of change and the final obtained value is 3. Compounds belonging to categories 1 and 2 account for ≈30% of total values (Table 5).

HPLC–ESI–MS For LC/MS (Fig. 4, see supplementary material), at least ≈25% of total values have to be removed from the calculation pool to obtain a positive UF value. After 30% of values have been removed, the UF value has a slower rate of change and the final obtained value is 3.9. The values consider both ESI+ and ESI- data. Compounds belonging to categories 1 and 2 account for ≈50% of total values (Table 6).

In both cases, applying the rationale significantly reduces the pool of values that are used for UF calculation; however, this is necessary to ensure valid data can be entered in ISO

Table 5 GC/MS data for UF calculation

Total no. of compounds considered	110
No. of compounds after applying rationale	74 (67%)
Experimental UF (corrected set)	3
Final suggested UF value	4

Table 6 LC/MS data for UF calculation

Total no. of compounds considered	128
No. of compounds after applying rationale	56 (44%)
Experimental UF (corrected set)	3.9
Final suggested UF value	4

10993–18:2020 Annex E formula. The process showcases the high variability of response factors for extractables compounds as it spans several orders of magnitude and further demonstrates the limitations of a UF-centered approach.

For this reason, the suggested value for UF is set to 4, which is a reasonable value to apply in E&L studies. UF=4 is in line with the theoretical value obtained for LCMS (3.9) and is a slightly worst case for GC/MS (3) remaining, however, acceptable.

Numerical simulation for evaluation of the combined UF/RRF approach

Concept

A simplified numerical simulation benchmark (NSB) has been developed to evaluate strengths and weaknesses of different combined UF/RRF approaches. The metric used to judge the efficacy and quality of each approach is to count false positive (type I) and false negative (type II) errors. Type I and type II errors have practical consequences on E&L studies since false positives lead to unfounded concerns, requiring additional effort and resources. False negatives represent a direct failure of the E&L evaluation, potentially delaying required action on compounds of concern. The scope of this benchmark is to demonstrate how the distribution and applied thresholds for UF and RRF values can impact results.

Because variability in E&L studies is high, a generic extractables study cannot be defined due to the variables involved. Results may differ depending on the material, formulation, geometry, or extraction conditions applied, i.e., solvents, pH range, temperature, and contact time. For this reason, a generic simulation of all physico-chemical processes of an extractables study is not feasible. Instead, a

simplified numerical simulation has been constructed with a heuristic approach due to intrinsic limitation.

The NSB is based on random concentration values, which are assigned to a selection of extractable-like compounds, reflecting an extended range of RRF (more than two orders of magnitude), and have the following characteristics:

- The random values are assigned from a predetermined distribution,
- The values act as the real concentration of the extractable impurity in the sample, defined as *base values* (expressed as $\mu\text{g/mL}$),
- Experimental (*exp*) values are calculated from the *base values* and act as the experimental value obtained through the semi-quantitation during the extractables screening process (expressed as $\mu\text{g/mL}$)

By comparing *base* and *exp values* to a certain AET, it is possible to determine if the outcome is correct or not, thereby distinguishing between different error types:

- If the measured concentration (*exp value*) is above the AET while the real concentration (*base value*) is below the AET, it is a type I error,
- If the measured concentration (*exp value*) is below the AET while the real concentration (*base value*) is higher than the AET, it is a type II error.

Since a single random value is not representative, the process is iterated multiple times to obtain a dataset that spans over the entire concentration range examined (six orders of magnitude). The intent of the numerical simulation is not to closely match a real case but to visualize the interaction between several parameters in a controlled process. For this reason, attention needs to be focused on trends and relative results and not on the obtained absolute values.

Benchmark execution

Building from a simple case An example can be used to understand the NSB process, considering a single extractable “Compound A” and its associated RRF value, RRF_a . The concentration of A in the extraction solution will be a certain *exp value*. However, as the amount is semi-quantified against an internal standard from the screening process, the *base value* of A in solution will be different. The RRF_a value is an expression of how close the *base* and *exp values* are.

Since a compound during the extractables studies will be reported only if its *exp value* is above the AET, it can be easily verified if Compound A is actually above or below AET. In fact, three outcomes are possible:

1. A false positive is reported: Compound A has a higher response factor than the internal standard. The base value is below the AET.
2. A false negative is reported: Compound A has a lower response factor than the internal standard. The base value is higher than the AET.
3. The extractable evaluation is correct. The same result is obtained using either the *exp value* or the *base value*.

The process above can be taken a step further by repeating the same consideration for multiple compounds in parallel—A, B, C... each one with a different RRF value, i.e., RRF_a , RRF_b , RRF_c .

To complete the picture, it is necessary to:

- Test multiple *base* and *exp values*
- Define the AET

The first point is solved by iterating the process. At each cycle, a random value is assigned to each compound, and then false positives and false negatives are tracked. The AET concentration is defined arbitrarily. The following paragraphs provide the rationale for this specific benchmark.

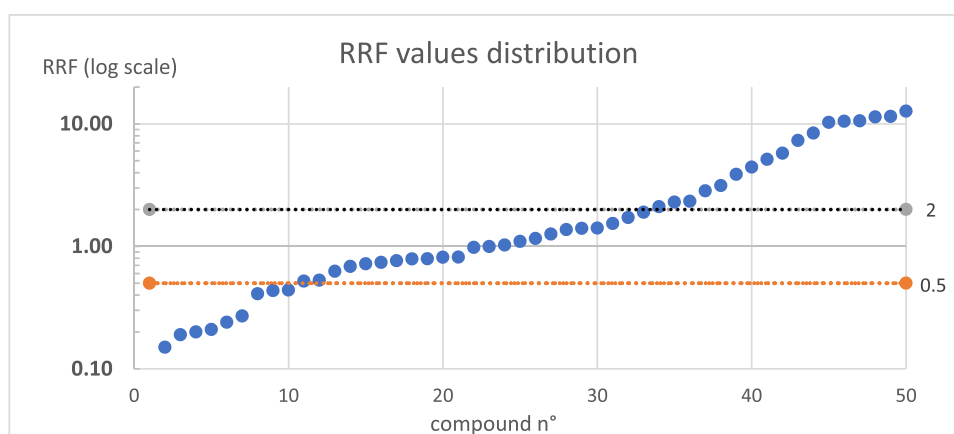
Compounds considered for the NSB (RRF distribution) To perform the evaluation, the following assumptions have been made regarding the compounds:

- The benchmark is performed by considering 50 compounds. Each compound has a different RRF value and is evenly distributed over a defined range (Fig. 3).
- Each compound is representative of an actual extractable compound with a known RRF value, which was determined experimentally.
- The response of each compound is linear over the tested range.

Each dot represents a different compound, with 50 compounds in total, ordered from lowest to highest value. The y-axis (RRF value) is shown as a log scale for readability. The 0.5–2 RRF value range is shown by dotted lines.

As shown in Table 7, almost half of the values are within the 0.5–2 range, which represents a no-action range for the RRFlow approach (the RRF calculation is not applied, refer to section “[The RRFlow approach for HPLC–ESI–MS and GC/MS](#)”). The remaining compounds ($\text{RRF} < 0.5$ and > 2) are rescaled by the RRF value when considering the RRFlow approach.

Distribution of *base values* The first step of the NSB is the assignment of *base values* to each one of the 50 compounds. To work, the NSB requires new base values at each iteration.

Fig. 3 RRF values chosen for the numerical simulation**Table 7** Summary table of the RRF values

Total	50 compounds
RRF within 0.5–2.0	23 compounds
Minimum RRF value	0.07
Maximum RRF value	12.70

Since the *base values* represent the actual concentrations in the sample, the following conditions are defined:

1. The base value is a random value;
2. A random value range is 0.001–100 (µg/mL or µg/mL as unit of measure);
3. Majority (> 50%) of random values generated are within the range 0.001–1 µg/mL.

Since the considered range covers five orders of magnitude, an additional requirement (point 3) is included to increase the statistical weight of values in the lower range—even though the scope of the NSB is not to simulate a real extractables profile, empiric evidence suggests that the majority of extractables are released at lower amounts, with few compounds being observed at higher concentrations.

This empiric evidence justifying the bias towards lower values is supported by the nature of the extractables. A molecular species could be an additive, impurity, degradation compound, or a fragment that crosses a phase separation (commonly the liquid–solid interface) in a process that maintains the structural integrity of the bulk material, i.e., the extractables source. This is also reflected by the different amounts of extractables usually reported when the AET is high, i.e., above 1 µg/mL, versus low AET values that are closer to the QL value of the analytical technique.

The required random numbers (point 1) are generated using the Microsoft Visual Basic Rnd() function. However, because the Rnd() function generates a number between 0

For each extractable compound (50)

r = random value (0 – 100)

If $r > Y$ and $< (Z + Y)$, then $x = \text{Rnd}^2$

If $r < Y$, then $x = \text{Rnd}^2 * 10$

If $r > (Z + Y)$, then $x = \text{Rnd}^2 * 100$

Base value = $x + m$

Fig. 4 Base values assignment function

and 1, a simple algorithm is developed in order to obtain the final *base values* under the requirements above.

Base value generation algorithm and resulting *exp values*. The algorithm structure is visible in Fig. 4:

1. A random r value is generated between 0 and 100.
2. Depending on the outcome, a *variable x* may be:

directly assigned a squared random value
 assigned a squared random value multiplied by 10
 assigned a squared random value multiplied by 100

3. The variable x is added to a fixed m value (0.001)

Y, Z parameters allow control over the generated random values:

Y = a value between 0 and 100, which represents the probability of multiplication by 10;

Z = a value between 0 and 100, which represents the probability of no multiplication (the value is unchanged);

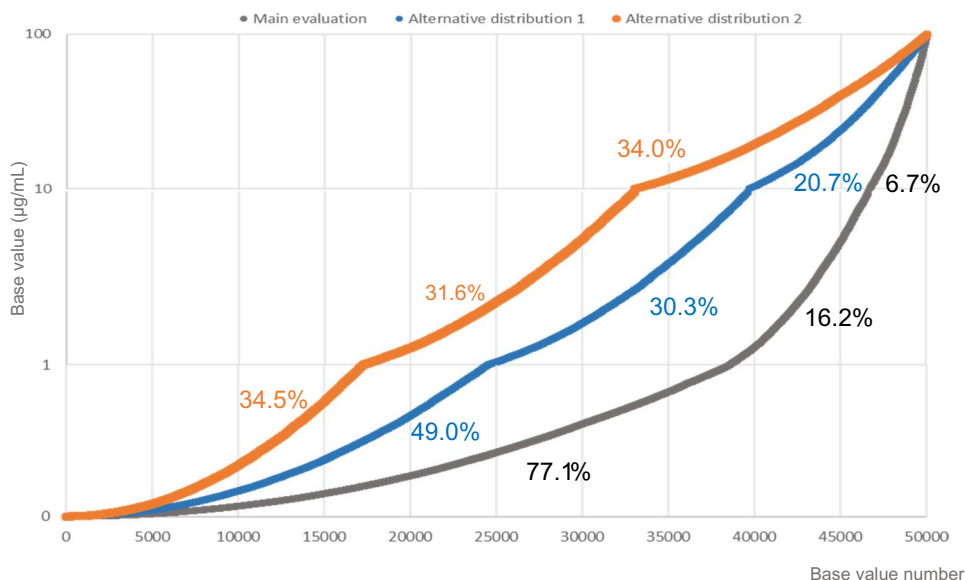
A residual K value can be defined as a value between 0 and 100, which represents the probability of multiplication by 100. However, if Y and Z are defined, then K is

Table 8 Distribution parameters

	Main evaluation	Alternative distribution 1	Alternative distribution 2
Y	20	35	30
Z	70	35	20
K^a	10	30	50

^aFraction of random values not multiplied by 10 or 100 (fixed by Y and Z values)

Fig. 5 Distribution of values. Plot of all the generated random *base values* (all distributions) ordered from lowest to highest (log scale for the y-axis). Each “line” is composed of 50,000 values (1000 values each for 50 compound)



fixed, since the sum of all probabilities must be equal to 100 ($Y + Z + K = 100$). The following examples show how the tuning parameters are used.

By setting $Y = 50$, $Z = 50$, and $K = 0$, 50% of the random values are multiplied by 10, while the remaining 50% of random values are not multiplied.

If $Y = 25$, $Z = 25$, and $K = 50$, then 25% of the random values are multiplied by 10, 25% are not multiplied, and the remaining 50% are multiplied by 100.

The random value is squared to increase the weighting of lower values within the range.

The fixed value m ensures that the lowest possible random value is always higher than 0.001.

The *exp values* are derived from the base value and using the RRF value of the specific compound. For example, if a compound has an RRF of 4, its *exp value* will be four times higher than the base value. If its base value is 2.5 µg/mL, the *exp value* will be 10 µg/mL.

The following Y, Z, and K values are selected for the numerical simulation benchmark.

Two alternative distributions have been tested to check the robustness of the main evaluation distribution to

demonstrate how the results may vary when considering higher *base values* (Table 8).

By increasing Y and K, the contribution of higher values on the total is increased since more random values are multiplied by 10 or 100. As a result, for alternative distributions 1 and 2, fewer *base values* fall between 0.001 and 1. The percentage decreases from 77.1 to 34.5% (Fig. 5).

Figure 5 is a plot of all the random values generated (*base values*) in order from the lowest to the highest over a vertical logarithmic scale to improve readability. Depending on

the chosen parameters for the main evaluation, alternative distribution 1, and alternative distribution 2, a different distribution is obtained.

The contribution over the total number of random generated values is indicated by a percentage value for each sub-range (0–1, 1–10, 10–100). The actual percentage for each range is lower than expected due to the use of the squared Rnd value. The alternative distributions have been tested only for the medium threshold category (see Table 13, supplementary material).

Thresholds and AET As indicated by ISO 10993–18:2020 Annex E [14], the AET is calculated as follows:

$$\text{AET } (\mu\text{g/mL}) = \text{TTC } (\mu\text{g/day}) \times (A/(B \times C \times D)) \div \text{UF}.$$

Where:

- TTC (threshold of toxicological concern) is selected according to ISO/TS 21726:2019
- A = number of medical devices extracted
- B = extract volume

Table 9 AET values grouped by five threshold levels

	Low	Medium–low	Medium	Medium–high	High
No UF ^a	0.200	0.600	1.80	5.40	15.0
UF=2 ^b	0.100	0.300	0.90	2.70	7.50
UF=4 ^c	0.050	0.150	0.45	1.35	3.75
UF=10 ^d	0.020	0.060	0.18	0.54	1.50

^aNo UF applied on the reporting threshold (AET), equivalent to UF=1

^b50% of the “no UF” value

^c25% of the “no UF” value

^d10% of the “no UF” value

- C = number of medical devices that are in contact with the body
- D = dilution factor
- UF = uncertainty factor

The UF directly reduces the AET value, i.e., a UF=2 is equivalent to a 50% reduction of the AET, a UF=4 reduces the AET to 25% of the original value.

AET values used in extractables studies are based on the UF and can range over several orders of magnitude. These are dependent on the TTC value and parameters A, B, C, and D. Several fixed values are used for the AET.

Five different values are considered as the starting point. For each one of those values, three UF are considered. This resulted in a total of 20 unique values. This permits subdivision of the AET range into five categories, i.e., high, medium–high, medium, medium–low, low.

The same AET value could be obtained through different starting parameters. For example, a threshold of 10 with a UF value of 4 would be equivalent to a threshold of 5 with a UF of 2 (2.5 µg/mL AET). For this reason, the grouping suggested in Table 13 (see supplementary material) is used for comparison purposes. The following values have been selected (Table 9).

Benchmark scenarios The scope of the numerical simulation is to evaluate the impact of UF values and the effect of the RRFlow approach. To gain sufficient insights, several scenarios are tested in the benchmark. A scenario is defined by the AET applied, itself dependent on the UF, and the processing of the experimental data.

The following list details the eight scenarios tested by the benchmark:

Scenario A:

- No UF applied (equivalent to UF=1)
- AET values: 0.200, 0.600, 1.80, 5.40, 15.0 µg/mL
- No further elaboration on exp values

Scenario B:

- UF=2
- AET values: 0.100, 0.300, 0.90, 2.70, 7.50 µg/mL
- No further elaboration on *exp values*

Scenario C:

- UF=4
- AET values: 0.050, 0.150, 0.45, 1.35, 3.75 µg/mL
- No further elaboration on *exp values*

Scenario D:

- UF=10
- AET values: 0.020, 0.060, 0.18, 0.54, 1.50 µg/mL
- No further elaboration on *exp values*

Scenario E:

- UF=4
- AET values: 0.050, 0.150, 0.45, 1.35, 3.75 µg/mL
- *Exp values* are multiplied by 4 (regardless of the actual RRF value)

Scenario F:

- UF=10
- AET values: 0.020, 0.060, 0.18, 0.54, 1.50 µg/mL
- *Exp values* are multiplied by 10 (regardless of the actual RRF value)

Scenario G:

- UF=4
- AET values: 0.050, 0.150, 0.45, 1.35, 3.75 µg/mL
- *Exp values* are multiplied by RRF value (RRFlow approach), if RRF < 0.5 or > 2 (if 0.5 < RRF < 2, no multiplication by RRF)

Scenario H:

- UF=10;
- AET values: 0.020, 0.060, 0.18, 0.54, 1.50 µg/mL
- *Exp values* are multiplied by RRF value (RRFlow approach), if RRF < 0.5 or > 2 (if 0.5 < RRF < 2, no multiplication by RRF).

Scenarios A, B, C, and D represent the classic E&L evaluation approach and have been previously considered in this section when defining the AET categories. Scenarios E and F represent an alternative approach to AET rescaling only. In these cases, the UF value is also

considered as a rescaling factor for the *exp values*. Two UF values (4 and 10) are considered to evaluate the impact of the increased effect of the UF value. Scenarios G and H represent the proposed RRFlow approach. Two UF values (4 and 10) are considered to verify the impact of the UF when the RRFlow approach is applied.

Numerical simulation benchmark Each scenario has been tested with the same *base values*, randomly generated as previously shown. For each case, the *exp value* is calculated and then compared to the original base value, with respect to the relevant threshold.

Type I and Type II errors have been evaluated for each set of *base values* and for a defined threshold. By combining all the data, a few statistic parameters can be calculated to compare the different cases. To obtain significant results, the process has been iterated 1000 times. Since 50 compounds are considered, individual scenarios are evaluated across a total of 50,000 values (50×1000) which are plotted in Fig. 5.

Results

Impact of increasing UF value The first evaluation is based on the impact of increasing the UF value (Table 14, see supplementary material). The considered scenarios are A, B, C, and D. These have been tested for each threshold group. The obtained results are summarized using the following statistics:

- Average error count and standard deviation,
- % of iterations with a perfect score (zero errors counted for all iterations),
- % of iterations with a score of at least 90% (≤ 5 errors counted for all iterations).

The average error count is the main evaluation parameter as it is indicative of the performance of a specific scenario. For example, the average number of type I errors for scenario A in the medium category (AET = 1.80 µg/mL) is 3.1. This value can be interpreted as follows: if 50 compounds are tested 1000 times randomly between 0.001 and 100 µg/mL versus an AET of 1.80 µg/mL, on *average*, 3.1 false positives errors will be committed. A similar statement can be formulated for any specific scenario and threshold.

An average below 1 indicates that most iterations resulted in zero errors. The standard deviation is shown to check how representative the average value is. For a normal distribution, a standard deviation of 1 indicates that 68.2% of error counts are within one standard deviation ($\pm 1\sigma$) of the average, or 95.4% of error counts vary between two standard deviations ($\pm 2\sigma$) of the average. The results obtained above are in line with typical normal distributions.

The main observations related to performance are as follows:

- An increased UF value does not equal improved results. In several scenarios, the trend is even reversed, i.e., a greater incidence of errors with a higher UF.
- Similar trends are observed for both types of errors at high or low AET. The main difference is observed in the medium AET range.
- Type II errors have more impact than Type I errors at lower thresholds.

Since each test case considers a different AET, a general trend can be visualized by plotting all the obtained average error count values, ignoring the UF value.

Figure 6 shows that for type I errors the average error count increases constantly. It reaches a maximum at 0.90 µg/mL before decreasing. For type II errors, a flex point is observed. For both false positives and false negatives, there is no monotonic variation; i.e., there is no direct correlation between UF and error count.

Tables 11 and 12 (see supplementary material) have been developed to provide more information regarding the distribution of errors at each AET (higher values are highlighted by a dark color).

For example, for type I errors at the 0.02 threshold, 47% of iterations resulted in a perfect score, 34% of iterations counted a single error, 15% of iterations had two errors, etc. By tracking the highest percentages for each threshold in Tables 15 and 16, we obtain a similar trend as that shown in Fig. 6. However, the spread around the maximum error count is visible and is limited and symmetric in each case (normal distribution).

These observations can be rationalized by considering the diagrams included in Fig. 7, 8, and 9.

The *exp value* and the *base value* are distinct concentration values due to the difference in response between the internal standard and the extractable compound. This difference is expressed by the RRF value for the target compound, which can be higher or lower than 1. The concentration range between the *exp* and *base values* represents a critical zone. If the AET falls between these values, depending on the RRF, two situations may arise:

- Type I (false positive) for compounds that have a higher instrumental response ($RRF > 1$) than the internal standard. The actual concentration is lower than the measured concentration;
- Type II (false negative) for compounds that have a lower instrumental response ($RRF < 1$) than the internal standard. The actual concentration is higher than the measured concentration.

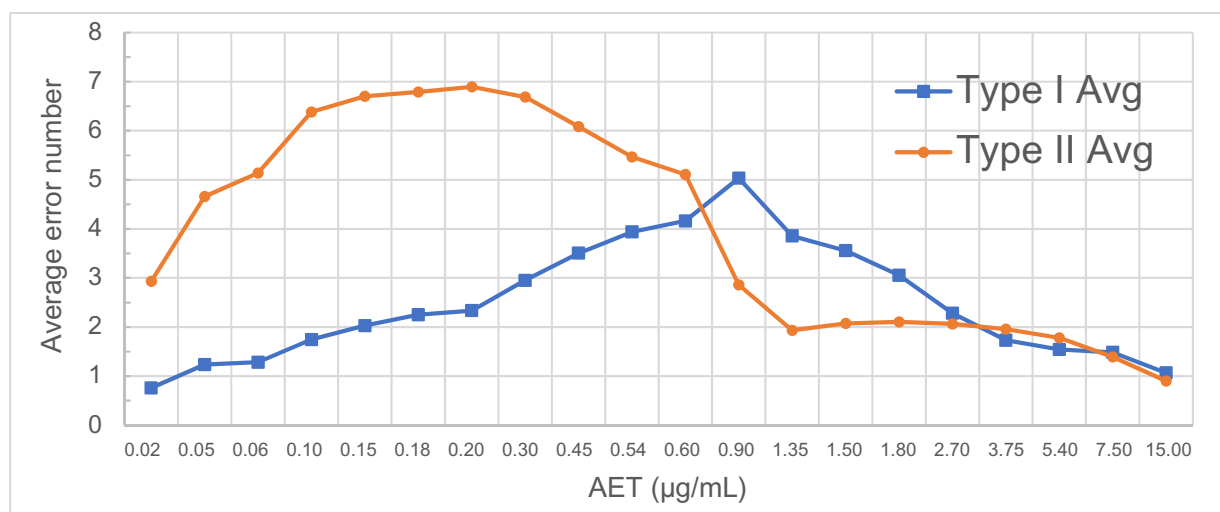
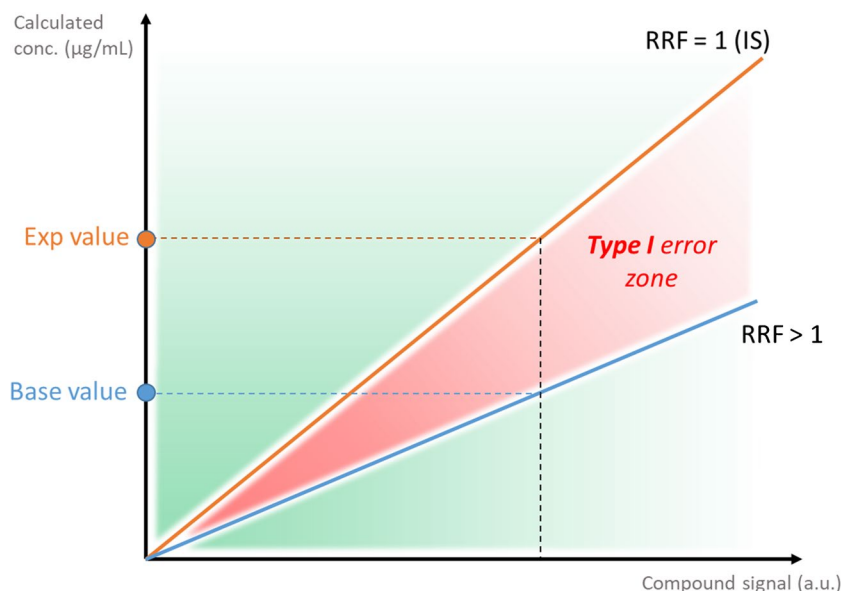


Fig. 6 Plot of all obtained average values. Plot of the average error (values from A, B, C, D scenarios in Table 10, see supplementary material) and the relative AET value applied. Blue line: false positives (type I). Orange line: false negatives (type II)

Fig. 7 Type I error zone (RRF > 1). Simplified plot for a high response compound (RRF > 1), exp value is always higher than the base value



Type I and type II errors are represented by the diagrams in Figs. 7 and 8, respectively. Since the requisite for both error types is for the AET to fall between the exp value and the base value, an error zone can be defined which is delimited by RRF lines. Target compounds with RRF values that deviate considerably from 1 result in wider error zones.

Figures 9 and 10 show how the interaction between the threshold, the *exp* and *base value* pair may lead to nonintuitive results—a lower AET (obtained from higher UF) could bring the threshold value in or out of the error zone. Meaning that a false positive or a false negative can appear or disappear depending on the considered variables.

The diagram in Fig. 9 shows a compound with RRF > 1 (higher response in comparison to the internal standard) while the diagram in Fig. 10 shows a compound with RRF < 1 (lower response in comparison to the internal standard).

Three AETs are shown, depending on the applied UF:

- UF = 1 (equivalent to no UF applied): in both cases the base value (actual sample concentrations) and *exp values* fall below AET, resulting in no type I or II errors;
- UF = 2 (50% of reporting threshold): the AET falls in the error zone in both diagrams—one of them has an

Fig. 8 Type II error zone (RRF < 1). Simplified plot for a low response compound (RRF < 1), exp value is always lower than the base value

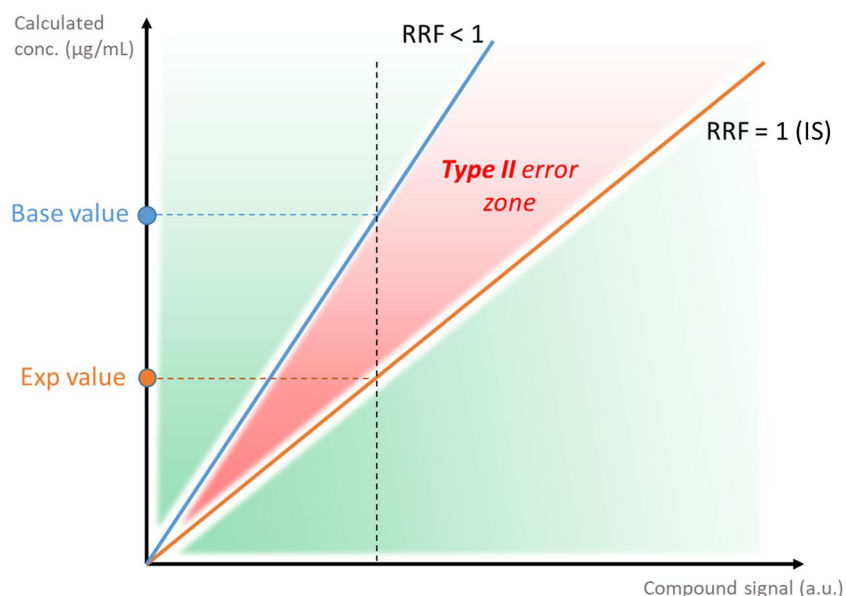
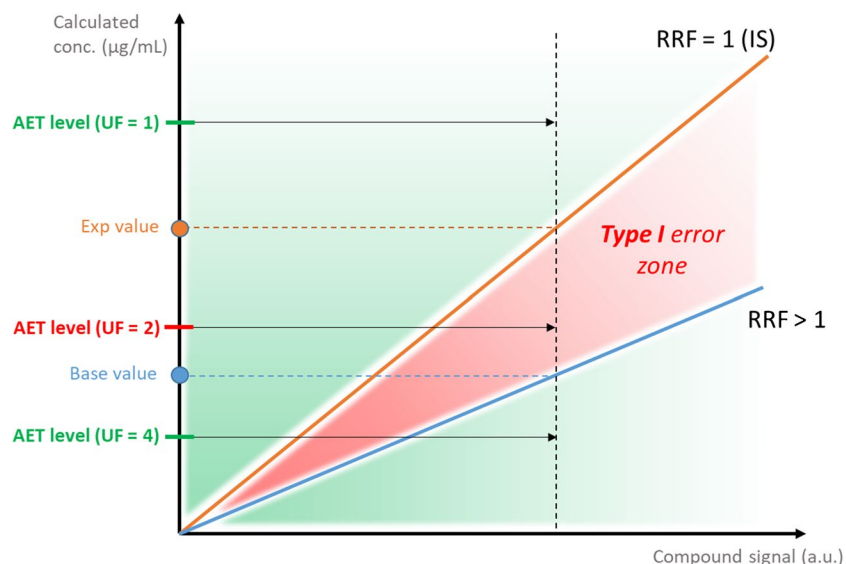


Fig. 9 Example of interactions of AET values, exp and base values for RRF > 1. Simplified plot for a high response compound (RRF > 1) intersecting with different AET values



RRF > 1 leading to a false positive, the other has an RRF < 1 leading to a false negative;

- UF = 4 (25% of reporting threshold): in both cases the AET is below both base and exp value, resulting in no type I or type II errors.

By including both *base value* and *exp value* in the analysis, it becomes apparent that it is possible to obtain worse results by lowering the AET value. Depending on the specific value UF and the response of the compound compared to the internal standard, type I and type II may arise even when less expected.

Further lowering the AET could resolve a single type I or type II error, but as demonstrated by the benchmark, the overall effect is not a positive gain.

Impact of RRFlow approach value at constant UF The impact of the RRFlow approach by error counting has been evaluated for two UF values: 4 and 10. The tested scenarios are C, E, G and D, F, H (Table 17).

- Scenarios C and D are previously considered when comparing UF values.
- Scenarios E and F are included to highlight how the RRFlow process is more refined than the application of

Fig. 10 Example of interactions of AET values, exp and base values for $RRF < 1$. Simplified plot for a low response compound ($RRF < 1$) intersecting with different AET values

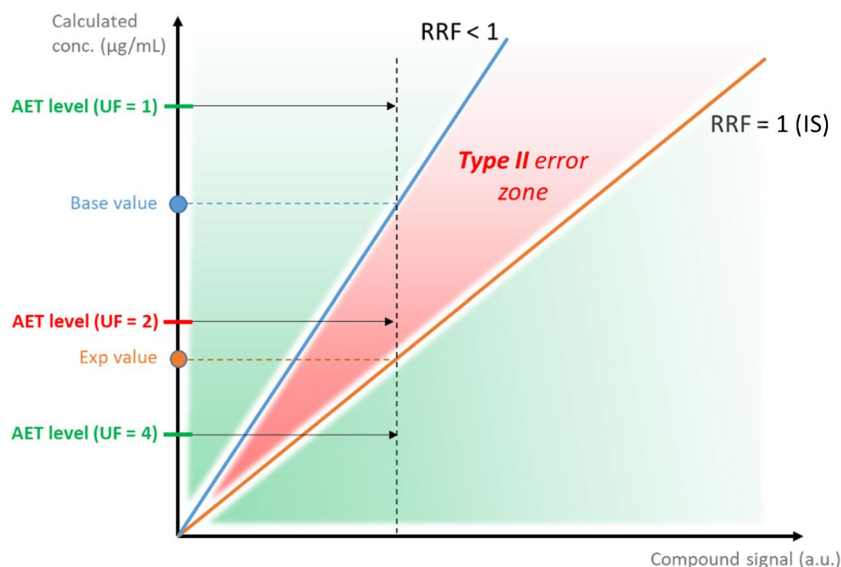
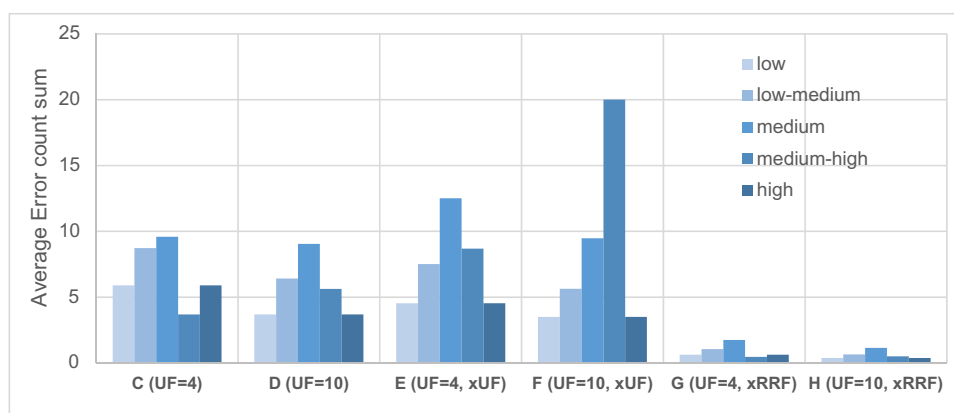


Fig. 11 Summary of results. Bar plot of the average values shown in Table 13 (see supplementary material)



a rescaling factor. In those cases, a fixed value equivalent to the UF is used to rescale the *exp values*.

- Cases G and H apply the *RRFlow* approach as described previously. The obtained results are summarized by considering the following statistics: average error count and its standard deviation, percentage of iterations with a perfect score (0 errors, 1000 iterations in total), percentage of iterations with at least 90% score (≤ 5 errors, 100 iterations in total).

The main observations are as follows:

- The *RRFlow* approach (G, H) outperforms both the application of the UF value and rescaling with a fixed factor. The average error count is always less than 1 for both error types at each threshold.
- The fixed value rescaling (E, F) performs considerably worse for false positives, as expected. This is especially the case for the medium and medium-high thresholds. As the fixed value rescaling increases the experimental

value, several more compounds are pushed above the AET. For the same reason, performance is improved for false negatives, which have comparable results to the *RRFlow*.

- The *RRFlow* approach reduces the effect of applying the UF value. Similar results are obtained when considering UF = 4 or UF = 10.

Figure 11 shows the results grouped by scenario, with the average error count sum (type I and II errors combined) for each threshold level.

Cases G and H show improved performance compared with the C, D, E, and F test cases. For the average error sum, it is noted that increasing the UF from 4 to 10 (cases C and D) does not correlate with better results. Although the application of a higher UF (lower AET) increases the average amount of compounds detected, no logical improvement can be demonstrated regarding false negatives or positives. The *RRFlow* approach has instead a critical impact on the quality of the results. The approach



Fig. 12 Plot of results. Bar plot of the average values shown in Table 14 (supplementary material)

is effective independently of the UF factor applied. This is consistent with the RRFlow approach [13].

Assessment of different value distribution at the medium level To test two alternative random value distributions and to check the robustness of the benchmark and how the results are impacted, all scenarios from A to H are considered at the medium threshold (1.80 $\mu\text{g/mL}$ as AET for $\text{UF}=1$). The alternative distributions are generated using different Y and Z parameters (see Sect. "Benchmark execution") resulting in a greater contribution of *base values* between 1 and 100 than the main evaluation.

- Alternative distribution 1: 50% of the values generated are between 0.001 and 1;
- Alternative distribution 2: 34% of the values generated are between 0.001 and 1, each range (0.001–1, 1–10, and 10–100) contributes equally to the evaluation.

Testing is conducted using alternative distributions 1 and 2, and the obtained results are shown in Table 14 (supplementary material) and Fig. 12.

The same trend can be observed in each distribution for both type I and type II errors:

- Type I (false positive): The average errors for scenarios A, B, C, and D (different UF values) are slightly lower for the alternative distributions than for the main evaluation. This is consistent with an increased average base value. The *exp values* tend to be greater than the AET in more cases, thus reducing the count of type I errors. The same effect can be applied for cases E and F. As the results are only shifted by multiplication using the UF factor, the average number of errors remains high. For the RRFlow cases (G and H), there is no significant impact since the average error count remains below 1.
- Type II (false negative): The average errors for cases A, B, C, and D (different UF values) are slightly higher for the alternative distributions than the main evaluation. This is consistent with previous results. The average error counts are lower for cases E and F (UF rescaling factor). There is no significant impact observed for the RRFlow cases.

In both cases, the results can be justified. Shifting the *base values* towards higher values has the same effect as decreasing the AET.

- The average error count trend for alternative distribution 1 is similar to the results obtained for the medium–low threshold.
- The average error count trend for alternative distribution 2 is similar to the results obtained for the low threshold.

This indicates that the results depend on the relationship between the AET and *base values*, not their absolute values. Similar results should be obtained by increasing the *base values*, i.e., higher concentration values, or by reducing the AET (and vice versa).

Conclusions

The stages of identifying and quantifying extractables in E&L studies are critical, and inaccurate evaluations can compromise the safety and efficacy of pharmaceutical drugs. The use of semi-quantification methods for the E&L analysis, based on a few reference standards, can lead to incorrect estimations due to the variability of chromatographic detector responses (RF). The introduction of the uncertainty factors to correct the AET is partially effective, as it is influenced by the high variability of the RRF databases, carrying the risk of false positives and negatives.

The work described introduces a new analytical workflow (RRFlow) aimed at a more precise estimation of extractables, combining identification with specific validation and real-time correction of compounds previously estimated incorrectly. This approach seeks to resolve issues of quantitative overestimation or underestimation. The research also evaluated the impact of the intrinsic variability of RRFs and determined a reliable UF value for GC/MS and HPLC–ESI–MS techniques (UF = 4). Through numerical simulation, the effect of increasing the UF value and different scenarios on the quality of E&L studies was analyzed, measuring type I (false positives) and type II (false negatives) errors. The simulation demonstrated that increasing the UF value does not always improve data quality and can even worsen it depending on the applied thresholds and the discrepancy in RRF values. Applying a fixed rescaling factor (based on UF) produces asymmetric results, while RRF-based rescaling proves more effective in improving data quality and reducing the drawbacks associated with UF.

In conclusion, while improving RRF databases is crucial, the use of UF values remains necessary to reduce the risk of under-reporting. However, excessively increasing the UF can be counterproductive. The most effective strategy for improving data quality and preventing detection

and quantification bias is the rescaling of extractables amounts based on their relative response factors.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00216-025-05946-5>.

Author contribution M.G. Rozio: conceptualization, methodology, investigation, formal analysis, visualization, writing—original draft. D. Angelini: conceptualization, methodology, investigation, formal analysis, visualization, writing—original draft, writing—review and editing. S. Carrara: supervision, conceptualization, methodology, writing—review and editing.

Data availability Data are available from the authors on request.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. USP 43-NF 38 (1663) Assessment of Extractables Associated with Pharmaceutical Packaging/Delivery Systems
2. USP 43-NF 38 (1664) Assessment of Drug Product Leachables Associated with Pharmaceutical Packaging/Delivery Systems
3. USP 43-NF 38 <665> Plastic Components and Systems Used to Manufacture Pharmaceutical Drug Products and Biopharmaceutical Drug Substances and Products
4. FDA-1997-D-0145 Container closure systems for packaging human drugs and biologics. Guidance for industry, July 1999
5. Jenke D. Correcting the analytical evaluation threshold (AET) and reported extractable's concentrations for analytical response factor uncertainty associated with chromatographic screening for extractables/leachables. PDA J Pharm Sci Technol. 2020;74(3):348–58.
6. 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;50:206–12.
7. International Organization for Standardization, Biological Evaluation of Medical Devices, Part 18: Chemical characterisation 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).
8. Jenke D, Christiaens P, Beusen JM, Verlinde P, Baeten J. A practical derivation of the uncertainty factor applied to adjust the extractables/leachables analytical evaluation threshold (AET) for response factor variation. PDA J Pharm Sci Technol. 2022;76(3):178–99.

9. Jordi MA, Rowland K, Liu W, Cao X, Zong J, Ren Y, Liang Z, Zhou X, Louis M, Lerner K. Reducing relative response factor variation using a multidetector approach for extractables and leachables (E&L) analysis to mitigate the need for uncertainty factors. *J Pharm Biomed Anal.* 2020;186:113334.
10. Jordi MA, Heise T. An analytical strategy based on multiple complementary and orthogonal chromatographic and detection methods (multidetector approach) to effectively manage the analytical evaluation threshold (AET). *PDA J Pharm Sci Technol.* 2021;75:289.
11. Christiaens P, Beusen J-M, Verlinde P, Baeten J, Jenke D. Identifying and mitigating errors in screening for organic extractables and leachables: Part I; Introduction to errors in chromatographic screening for organic extractables & leachables and discussion of the error of omission. *PDA J Pharm Sci Technol.* 2020;74:90–107.
12. Christiaens P, Beusen J-M, Verlinde P, et al. Identifying and mitigating errors in screening for organic extractables and leachables: Part II; Errors of inexact identification and inaccurate quantitation. *PDA J Pharm Sci and Tech.* 2020;74:108–33.
13. Rozio M, Rosato A, Iadarola L, Carrara S. Compounds relative response factor, a reliable quantification within extractables testing. ResearchGate Publications; 2020.
14. Daniel L Norwood et al. Best practices for extractables and leachables in orally inhaled and nasal drug products: an overview of the PQRI recommendations. *Pharm Res.* 2008;25(4):727–39.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.