# FY 2018 Research and Evaluation for the Testing and Interpretation of Physical Evidence in Publicly Funded Forensic Laboratories:

Final Research Report

| Award Information | |
|---|---|
| **Federal Awarding Agency:** | National Institute of Justice |
| **Federal Award Number:** | 2018-DU-BX-0184 |
| **Project Title:** | Establishing Sufficiency Thresholds for Assessing the Quality of Mass Spectral Data |
| **Award Amount:** | $282,703 |



| Project Director/Principal Investigator | |
|---|---|
| **Name:** | Preshious Rearden, Ph.D. |
| **Title:** | Director of Research and Development |
| **Email:** | prearden@houstonforensicscience.org |
| **Phone:** | 832-304-8978 |

| Award Recipient Organization | |
|---|---|
| **Name:** | Houston Forensic Science Center, Inc. |
| **Address:** | 500 Jefferson Street, 13th Floor<br>Houston, TX 77002 |



| Project Period | |
|---|---|
| **Start Date:** | 01/01/2019 |
| **End Date:** | 10/31/2020 |

# Table of Contents



*This project was supported by Award 2018-DU-BX-0184 through the National Institute of Justice, Office of Justice Programs, and U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.*

# 1. Executive Summary

## 1.1. Research Purpose

The 2009 National Academy of Science report, Strengthening Forensic Science in the United States: A Path Forward impacted the field of forensic science by detailing identifiable issues with pattern sciences.[1] The report however failed to identify a similar shortcoming with the analytical disciplines of forensic science, the importance of which has been overlooked. For example, the identification of controlled substances relies on the principles of analytical chemistry but possesses many characteristics of the pattern sciences in the interpretation of mass spectral results. At present, there is a major need for specific threshold guidelines that an analyst can use to determine the minimum mass spectral data quality necessary for confident and consistent comparison to the existing libraries. This problem becomes acute for novel substances such as synthetic opioids.

As forensic laboratories increasingly encounter novel or synthetic analog drugs, library matching or pattern recognition has become more challenging. The most widely accepted practice for mass spectral identification is the analyst's impression of concordance with existing library examples of established composition. However, samples with adulterants, or high noise-to-signal ratio pose greater difficulty for delineation of a sufficiency threshold for interpretable data.

Sample variation, whether due to concentration, adulteration or high noise-to-signal ratio, needs to be known to develop a quantitative threshold. The tolerance (allowable deviation) in the mass spectrum can be calculated from this information.

Determination of a quantitative threshold allows the analyst to determine whether the mass spectral data is of sufficient quality or the library basis for comparison is adequate for identification of compounds.

The premise for this research is that:

1. Like forensic pattern sciences, forensic drug analysts rely on subjective pattern recognition methodologies to identify samples.

2. Difficulty in identification via pattern recognition and matching depends on the quality of the test spectrum and the concentration of the compound along with the amount of variation in spectra encountered with known (or previously matched) standards.

3. Minimum acceptance criteria, plus safe-guarding practices for borderline samples should increase consistency (across time, across analysts, even across laboratories) and reliability of substance identifications.

### 1.2. Research Design and Methods

The Houston Forensic Science Center (HFSC) in collaboration with the Center for Intelligent Chemical Instrumentation at Ohio University (OHIO) proposed to implement a study to develop and test a sufficiency standard (quality value) for mass spectral data by building a spectral library of opioids across a wide range of concentrations, contaminant interference, and substance mixtures. From this (experimental) database, this project took advantage of statistical/mathematical methodology for pattern data to create and validate an Information Quality Model for determining analytical thresholds for sufficiency of information. The project was implemented as proposed with no changes to the original scope.

The project was executed as follows:

***Objective I*** – The overall strategy was to create a library of controlled substances to include natural opiates and synthetic opioids at a wide range of dilutions, including varying degrees of

contamination and/or background noise to use as a test bed for pattern modeling, and as a tool for analyst interpretation.

***Objective II*** – Analysis and Characterization of Variation of Samples: Develop mathematical models and statistical data-mining techniques to analyze the spectra in the Spectral Library. Examinations will focus on the actual inherent variation in spectral pattern resulting from two different concentration levels and to determine the thresholds at which identification becomes unreliable either due to excessive variation in spectral pattern sample-to-sample or to obscuring key spectral features.

***Objective III*** – Develop and Validate Information Quality Model: Develop and optimize a model to determine a function representing change in spectrum pattern as a function of dilution, adulterants, and interferences coupled with identification confidence levels, utilizing the results of objective I and II. Validate the model via experiments using reference samples and gas chromatography/mass spectrometry (GC/MS) data from forensic casefiles.

### 1.3. Research Applicability Expectations

The quantitative reliability metric (QRM) developed in this study was used to determine the probability of accuracy for a given mass spectrum searched against the library database. Mass spectra generated from opioids were the focus of this study. However, this process could be extended to other controlled substances and compounds, but more importantly, the methodology can be directly applied or adapted by other forensic laboratories. In addition, both the results and methodology from this project should have a direct extension to other forensic disciplines that utilize mass spectral data, such as Toxicology and Trace Analysis. The QRM is not only applicable to spectral database searching but can also be applied with any query of a collection of

reference data. It provides a uniform statistical probability of the reliability of the search results and can be applied to database searching in other industries.

### 1.4. Research Participants

**A. Houston Forensic Science Center**

Preshious Rearden, Ph.D.
Vivekananda Shetty, Ph.D.

**B. Ohio University**

Peter Harrington, Ph.D.
Zewei Chen

## 2. Background

GC/MS is the gold standard method for identification of seized drugs. Forensic drug samples often contain adulterants, and GC/MS analysis enables the separation of sample mixtures and subsequent identification of unknown compounds through mass spectral library searches. Generally, the mass spectrum of an unknown compound is compared to a collection of reference spectra during the library search. The closest matching spectra can then be used to identify the compound.

The advantage of a mass spectral library search is that it provides empirical results. However, the misidentification of compounds or the inability to identify a compound resulting from the hit list, could present challenges for the analysts. These challenges are often due to poor mass spectral quality arising from low concentration samples or impurities and/or background noise. Identification difficulty may also arise from the sparsity of reference standards in the resulting library data not contained in the mass spectral library.

Other factors complicating identification are peaks arising from column and septum bleed which can significantly alter the mass spectra obtained from low concentration samples. Column bleed

results from decomposition of the stationary phase or the septum and often contain silicon because they decompose from polysiloxane polymers. Typical *m/z* ions 207 and 283 are indicative of column bleed ions.

Complication may also occur when the chromatographic peak elution window is relatively short compared to the mass spectral scan rate. As a result, the mass spectra may be distorted because the concentration of the analyte changes while the mass spectrum is scanned. This complication is referred to as peak skewing and is manifested by relative change in the intensity of peaks at either high or low mass regions of the mass spectrum.

Conventional mass spectral library searches utilize similarity metrics to generate a list of potential compounds, commonly referred to as the hit list.[2-5] This type of search algorithm yields a list of candidates based on similarities in fragmentation patterns but does not provide information about the quality of the search. An ideal spectral search would not only look at the similarities between the reference and comparison spectra but also provide a probability measure of the library search reliability.

Probability measurements of the library searches can be achieved with a quantitative reliability metric (QRM).[6] This measure would give a score of 100% for each returned matching spectrum for ideal matches and give low scores when the search result is unreliable such as when the mass spectrum is distorted by peak skewing or is contaminated with column bleed components. In addition, if the compound's spectrum is missing from the reference library, a low reliability score would be obtained.

The QRM was originally used to evaluate the performance of infrared library searches but with modifications, it can be applied to mass spectral library searches. The QRM in its original form did not have a uniform scale but measured a relative difference. It also only worked on similarity

but not dissimilarity metrics, such as distance measures. To overcome the limitations, the QRM was improved and modified to work with both dissimilarity and similarity metrics and applied to a custom mass spectral library search for opioids.

## 3. Methodology

### 3.1. Standards and Reagents

All reference standards were purchased from Cayman Chemicals (Ann Arbor, MI). Valeryl fentanyl, 4-ANPP, 2-furanyl fentanyl hydrochloride, cis-Tramadol hydrochloride, oxycodone, hydrocodone, and U-49900, were each purchased in methanol as 1 mg/ml solutions. The GC/MS Opioid Mixture 1 (U-47700, fentanyl hydrochloride, acetyl fentanyl hydrochloride, acrylfentanyl hydrochloride, butyryl fentanyl hydrochloride, cyclopropyl fentanyl hydrochloride, and furanyl fentanyl (hydrochloride) contained 1 mg/ml of each compound in methanol. Heroin hydrochloride, meta-Fluoroisobutyryl fentanyl (FIBF), phenyl fentanyl, and 3-methyl furanyl fentanyl (hydrochloride) were purchased as solids. The adulterants used in this study were cocaine purchased from Cayman Chemicals, acetaminophen from Spectrum Chemical (Brunswick, NJ), caffeine from Alfa Aesar (Ward Hill, MA), and diphenhydramine from Alfa Aesar (Ward Hill, MA). Methanol was obtained from Honeywell (Muskegon, MI).

### 3.2. Training Dataset

Sixteen opioids and four adulterants were used to develop the QRM and were obtained from experimental and archival data. The complete list of opioid substances and adulterants and their structures are shown in Table 1 and Table 2, respectively.

Table 1. Reference Standards

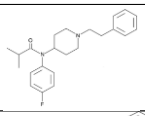
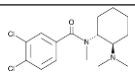
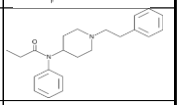
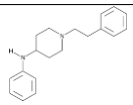
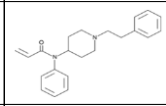
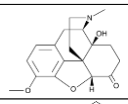
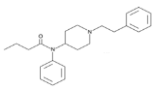
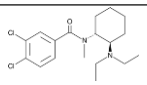
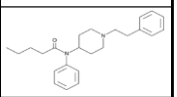
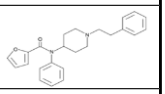
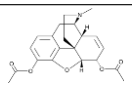
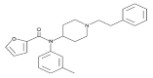
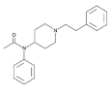
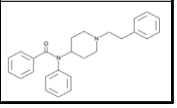
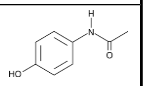
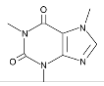
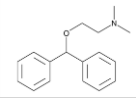
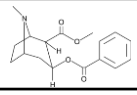
| | Opioid | Structure | | Opioid | Structure |
|---|---|---|---|---|---|
| 1 | Tramadol |  | 9 | meta-Fluoroisobutyryl Fentanyl |  |
| 2 | U-47700 |  | 10 | Fentanyl |  |
| 3 | 4-ANPP |  | 11 | Acryl Fentanyl |  |
| 4 | Oxycodone |  | 12 | Butyryl Fentanyl |  |
| 5 | U-49900 |  | 13 | Valeryl Fentanyl |  |
| 6 | Hydrocodone |  | 14 | 2-Furanyl Fentanyl |  |
| 7 | Heroin |  | 15 | 3-methyl Furanyl Fentanyl |  |
| 8 | Acetyl Fentanyl |  | 16 | Phenyl Fentanyl |  |

Table 2. Adulterants

| | Adulterants | Structure |
|---|---|---|
| 1 | Acetaminophen |  |
| 2 | Caffeine |  |
| 3 | Diphenhydramine |  |
| 4 | Cocaine |  |

The archival GC/MS datasets, generated from 2016 to 2018, were obtained from the seized drug case files at HFSC. For the experimental dataset, individual reference standards and opioid mixture stock solutions (with and without the adulterant mixtures) were prepared in methanol at 500 µg/mL. Working range concentrations from 1 to 250 µg/mL were prepared by serial dilution of the stock solution in methanol. In total, 2,219 GC/MS data files were collected from the archival experimental datasets.

### 3.3. Validation Dataset

GC/MS datasets obtained from archival seized drug casefiles at HFSC were used as validation data. The randomly selected case files contained GC/MS data in which opioids were identified

through forensic analysis from 2016 to 2019. These files included opioids which were not used in the training data. Experimental datasets were produced from mixtures of the sixteen reference opioids and four adulterants. Stock solutions of the opioid mixture was prepared in methanol at 2.5 µg/mL. Working standard solutions ranging from 1 to 250 µg/mL were prepared through serial dilutions in methanol.

### 3.4. Instrumentation

GC/MS experiments were performed using an Agilent Technologies 7890 GC equipped with a 5975C VL MSD. Separation was achieved using an HP-5MS capillary column (30 m × 0.25 mm i.d. × 0.25 µm). The helium carrier gas was set to a flow rate of 1 mL/min. The injection was made in split mode using a 20:1 split ratio. The injection volume was 1 µL and injection temperature set at 220 °C. The initial oven temperature was set at 100°C for 0.5 min and increased to 300°C at a rate of 25°C/min and then held for 5.5 min. The total runtime was 14 min. The mass spectrometer was operated in full scan mode using electron impact ionization (70 eV). The mass was scanned from *m/z* 41 to 400 at a scan rate of 4 sec/scan. The temperatures of the transfer line, ion source, and detector were set at 290 °C, 230°C and 150°C, respectively. Mass spectral identification of the compounds was performed using either the NIST library (Version 2.3) or a composite library constructed in-house. The in-house library records came from several sources, which include the Wiley Mass Spectral Library (6th ed), the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Mass Spectral Library (Version 3.8), and in-house reference standard library. The combined library contained 232,293 records.

### 3.5. Experimental Design

To minimize sampling variability, all samples were run using a random block design. This means that the sample concentration varied randomly with run order. The samples were partitioned into

three blocks. Each block contained randomized triplicates of each concentration, and a solvent blank was run between each sample.

### 3.6. Data Analysis

The preprocessing of the GC/MS data involved two steps. First, a basis set correction was used to correct the baseline of the chromatogram and then multivariate curve resolution (MCR) was applied to remove the background components from the mass spectra

#### 3.6.1. *Basis Set Correction*

In the basis set correction, an orthogonal basis set is built from the mass spectra collected from the preceding blank run.[7-8] The number of orthogonal components is determined using an algorithm for robust singular value decomposition (RSVD).[9] This algorithm ensures that the orthogonal basis constructed from the blank mass spectra can reconstruct each mass spectrum in the blank run with 99% accuracy. The robust property is helpful for removing septum bleed spectra that may occur as very sharp peaks.

The mass spectra for the sample is then split into the background and sample spectra using the RSVD orthogonal basis set. The total ion current (TIC) is calculated from the sample spectra and then the TIC is transformed to the second derivative. Second derivative values below a threshold define the number of peaks and the peak location. Windows are added to each peak at –0.1 and 0.2 min with respect to the peak maximum retention time.

#### 3.6.2. *Multivariate Curve Resolution*

The localized MCR correction is applied after the basis set correction. MCR is a powerful method that can provide pure response profiles from unresolved mixtures when no prior information is available. MCR can be applied to mass spectral data. It works on the principle of

linearity of the mass spectrum in that all the peaks in a mass spectrum of the analyte should vary proportionally. The mass chromatogram can be decomposed to:

$$\boldsymbol{X}=\boldsymbol{C}\cdot\boldsymbol{MS}+\boldsymbol{E} \quad (1)$$

where $\boldsymbol{X}$ is the mass chromatogram for which each row corresponds to a retention time and each column to a mass measurement at a given *m/z* value. The columns of $\boldsymbol{C}$ correspond to relative amount of a specific component (e.g., analytes or column bleed) and the rows correspond to the retention time. The pure mass spectra comprise the rows of $\boldsymbol{MS}$ for each component and the residual error $\boldsymbol{E}$ corresponds to data excluded from the model. Uppercase bold font indicates matrices and lowercase bold font indicates vectors. Adding more components to the model (i.e., columns of $\boldsymbol{C}$ and rows of $\boldsymbol{MS}$) will reduce the error $\boldsymbol{E}$. The goal is not to reduce the error in the model but obtain the purest analyte mass spectra.

Using the MCR approach, the most abundant ion in each peak window of the sample is selected to model the analyte, and the most abundant ion in each peak window of the background is selected to model the baseline. This method relies on finding a pure variable that is a m/z value that is contained in the analyte or background but not in both. One approach when operating in dual-scan mode is to use the SIM to measure pure variables so that the MCR approach can be then used. However, most forensic chemists do not exploit this powerful GC/MS method and only use it for quantification. One disadvantage of using SIM mode is that it requires *a priori* knowledge of which ions to include in the target scan. Because most drugs are aromatic some general ions can be used, such as for phenylic m/z 91 or 77 or for indolic m/z 130, 116, or 77.

A new approach was devised to find pure ions automatically in the peak window without using the dual scan mode. First, in the baseline corrected chromatogram, the ion with the maximum

intensity is found and recorded with its m/z value. Then in the reconstructed baseline chromatogram (i.e., that which was subtracted to perform the baseline correction), the ion with the maximum intensity is found without having a maximum retention time close to the analyte. In other words, ions that have a peak shaped like the target analyte are excluded from the search. This is important because the mass spectrum of m-fluoroisobutyryl fentanyl has a common ion with the prominent column bleed mass spectrum of m/z 207; moreover, because both the analyte and baseline share the same ion, it cannot be used as a pure variable. Multiple background ions can be selected but one ion works as well.

Once the pure ions are selected, they are used to generate an ion chromatogram *c* in the peak window. A mass spectrum can be generated using a constrained least squares minimization of the fit of c into the GC/MS chromatogram window. It is constrained so that no peak in the obtained mass spectrum can be negative. The equation below describes the process and is implemented in MATLAB with the function *lsqlin*.

$$\boldsymbol{ms_j} = \frac{\boldsymbol{x_j}}{\boldsymbol{c}}, \boldsymbol{ms_j} \geq \mathbf{0} \quad (2)$$

for which $\boldsymbol{ms_j}$ is the mass spectra for *m*/z *j* for both the baseline and the analyte. The peak window $\boldsymbol{X}$ is a set of mass spectra with the spectra as rows and ordered by retention time in the window centered at the detected peak maximum; $\boldsymbol{x_j}$ is the column *j* of $\boldsymbol{X}$.

Once the mass spectrum $\boldsymbol{MS}$ is obtained it is regressed back into the peak window $\boldsymbol{X}$ to obtain a better estimate of the ion chromatograms $\boldsymbol{c}$.

$$\boldsymbol{c_i} = \frac{\boldsymbol{x_i}}{\boldsymbol{ms}}, \boldsymbol{c_i} \geq \mathbf{0} \quad (3)$$

for which $\boldsymbol{c_i}$ is the ion current for the signal and the background at retention time *i*, and $\boldsymbol{x_i}$ is the mass spectrum at retention time *i*. Note that the MCR operates on the uncorrected raw spectra.

### 3.6.3. *Quantitative Reliability Metric*

The refined QRM is given below.

$$\mathbf{QRM}(\boldsymbol{K}) = \mathbf{F}\left(\frac{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{K}}\left(\boldsymbol{r_i} - \boldsymbol{r_j}\right)^{\mathbf{2}}/\boldsymbol{K}\,,\, \boldsymbol{i} \longrightarrow \boldsymbol{j}}{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{r_i} - \bar{\boldsymbol{r}})^{\mathbf{2}}/(\boldsymbol{N}-\mathbf{1})}, \boldsymbol{K}, \boldsymbol{N}-\mathbf{1}\right) \qquad (2)$$

for which QRM(*K*) is the metric for the *K* closest matching spectra in the intra-database comparison, $r_j$ is the correlation of the intralibrary spectrum that is the *i*th spectrum in the intra-database match list, $i \longrightarrow j$ maps to the location of spectrum *i* of the intra-database list to its location in the query spectrum list of matches, and $r_j$ is the correlation of the corresponding spectrum from the intra-database list in the query spectrum list. F is the F cumulative distribution with *K* and *N*-1 degrees of freedom.

The QRM works with any comparison metric. In this study, the correlation $\boldsymbol{r}$ is measured by the dot product between every record in the database and the query mass spectrum, and provides $\boldsymbol{m}$ correlation coefficients $\boldsymbol{r}$. The records with the *K* largest correlation coefficients comprise a match list of the most similar reference spectra. The QRM uses an intra-database comparison of the spectra in the match list using the same metric. For each of the *K* closest matching spectra a list of *K* closest intra-database spectra and their corresponding match metrics $r_j$ in the interlibrary search are collected. The QRM compares the differences between the intra-database search order for the *K* with those from the query spectrum search by calculating the variances of the metrics.

## 4. Result and Discussion

The GC/MS data was processed in three steps: background correction, identification, and quantitative reliability metric calculation. The results from each step are discussed below.

### 4.1. Background Correction

Several background correction approaches were investigated in this study. The approach that was adopted used an orthogonal basis set that was built from the mass spectra collected from the preceding blank run and then used to estimate the background of the mass spectrum in the analytical measurement. This background subtraction method is compatible with most standard operating procedures (SOPs) for seized drug analysis because SWGDRUG protocols recommend running a negative control sample before an analytical run.[10]

Examples of basis set background correction are given in Figures Figure 1 Figure 3. In Figure 1, an RSVD-corrected TIC for oxycodone before and after the removal of septum degradation peaks and the column components is shown. Figure 2 and Figure 3 show the effects of the corrected mass spectrum on the library search for a mixture of oxycodone and caffeine, respectively. For each compound, the top library hit was observed for the MCR-corrected spectrum. This result shows that the corrected spectra did not degrade the library search. We were able to determine that the basis set method provided good background correction and library search results.

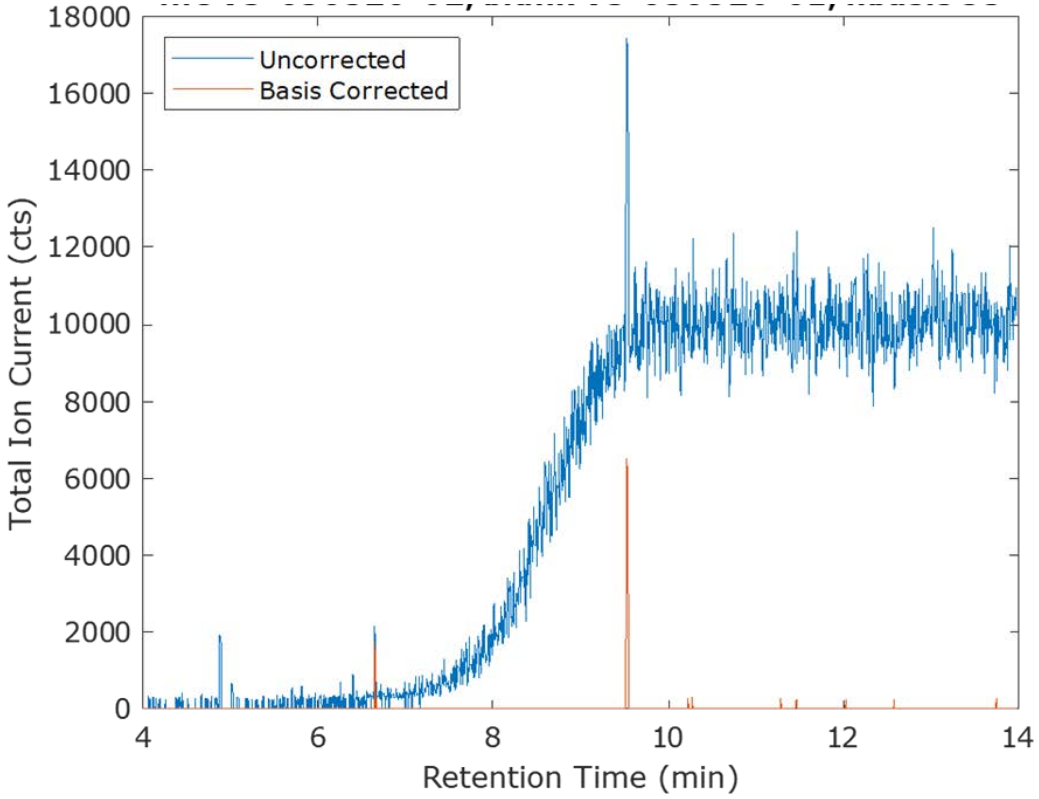



Figure 1. Comparison of the total ion chromatograms of oxycodone before and after the baseline correction using the basis set.

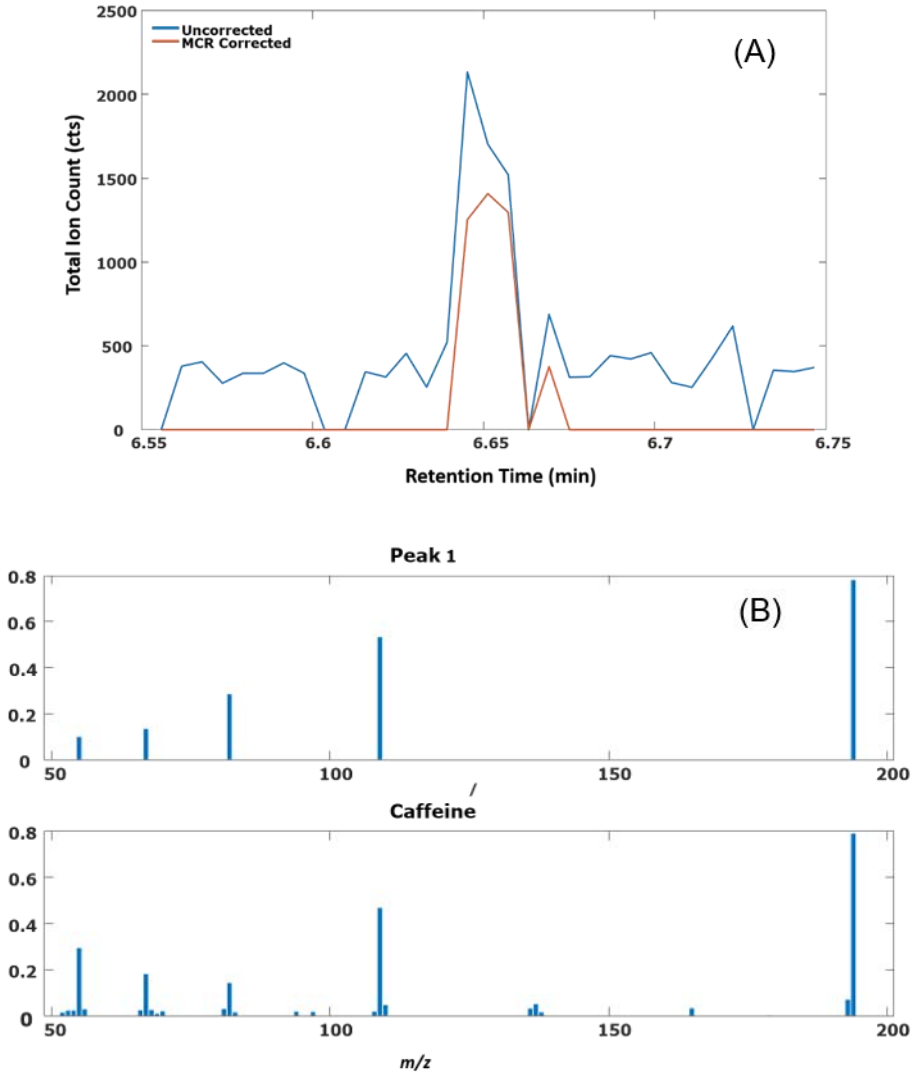



Figure 2. (A) Comparison of the local chromatograms of trace caffeine contaminant, found in the first oxycodone run, before and after the localized correction by MCR. (B) Top, the pure mass spectrum estimated by MCR for the caffeine peak. Bottom, the reference spectrum of the closest match from the library search using the uncorrected mass spectrum

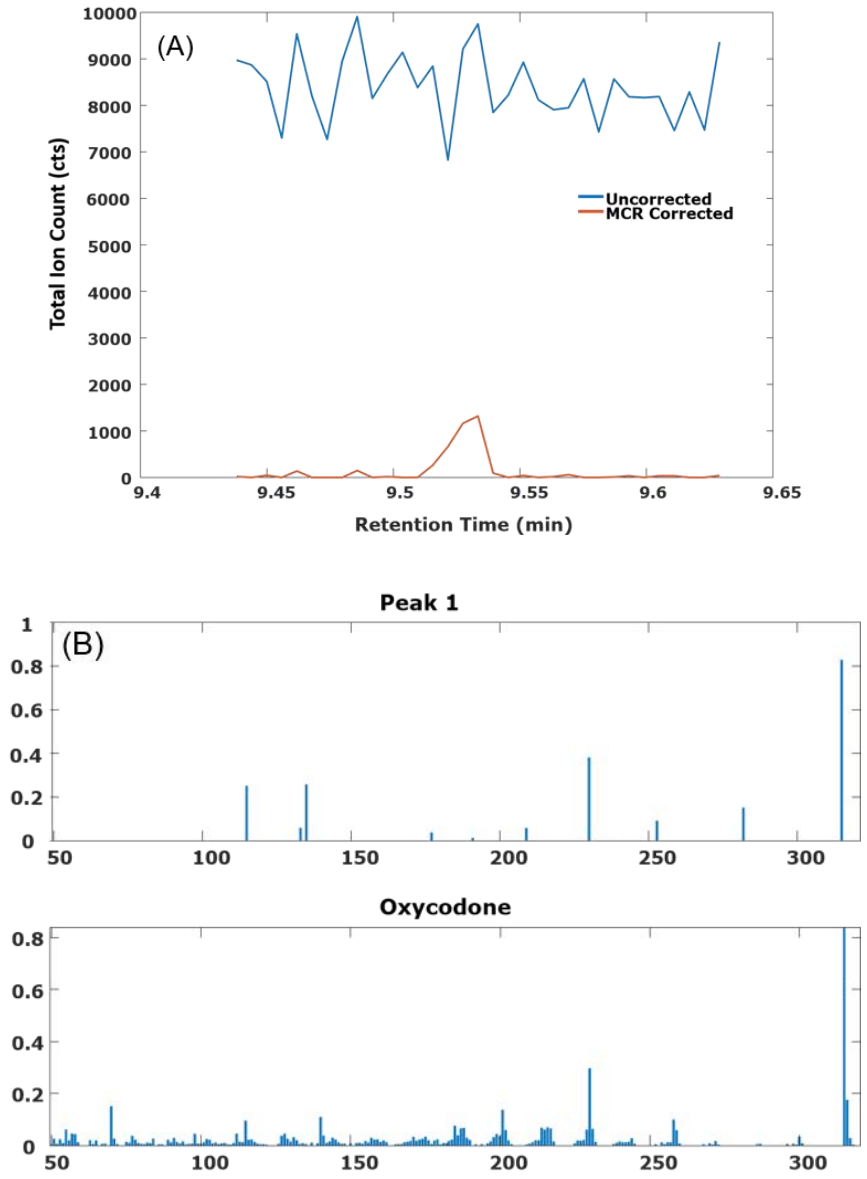



Figure 3. . (A) Comparison of the local chromatograms of oxycodone before and after the localized correction by MCR. (B) Top, the pure mass spectrum estimated by MCR for the oxycodone peak. Bottom, the reference spectrum of the closest match from the library search using the uncorrected mass spectrum.

## 4.2. QRM

The QRM uses an intra-library comparison of the spectra within the inter-library hit list to provide a quality score of 100% for reliable or ideal matches, and low scores for unreliable

results. Low scores would occur in a situation whereby the mass spectrum is distorted by peak skewing, contaminated with column bleed components or when reference spectra are missing from the library. An interlibrary search is a process in which an unknown spectrum is searched against the reference library and returns a hitlist of *p* closest matches. The spectra inside the reference library can also be used to search the same library to generate an intralibrary hitlist of the *p* closest matches. The intralibrary search utilizes the referencing spectrum in the interlibrary hit list to search against the library and retrieve a new hit. If the query mass spectrum is of high quality, it should be the closest match (i.e., top hit). Furthermore, its intralibrary hitlist should match with its interlibrary hitlist. QRM compares the match indices of the intralibrary hitlist spectra with those in the interlibrary hitlist and calculates a score for each spectrum in the search hitlist.

The QRM used in this study went through four stages of refinement with extensive testing of large datasets. The training dataset contained over 1,000 data files. The first QRM algorithm only worked with similarity metrics. In the next iteration, the formula was normalized with respect to the maximum value. The third model used the F-distribution to compare the variance of the QRM to the hypothesis that the closest matching compound is missing from the library. This version of the model was very sensitive and gave low scores for what appeared to be good matches but was better at excluding misidentified search results. The fourth and final QRM used the F-distribution to compare the variance of the QRM to the variance of the comparison metrics in the top *N* matches.

**QRM Search Examples**

The U-47700 results are shown in Table 3. It contains the top hits and QRM values for the triplicate runs at injected masses ranging from 0.2 ng to 50 ng. Misidentified samples are shown in red. Also included in the table is the similarity score or match statistic, which estimates how well the unknown mass spectrum matches the mass spectrum in the reference library. For the U-47700 results, the similarity score and QRM value both decreased with decreasing amounts of mass injected on the column. However, this is not always the case because the similarity score does not account for the quality of the mass spectrum.

The majority of the QRM values for U-47770 were 99% and above for the higher masses and then drastically decreased at the lower mass. The lowest QRM values were recorded for the 0.2 ng runs. This indicates the unreliability of the mass spectral match at low masses. At this mass level, two of the three replicates were misidentified. The reason for the misidentification is that the mass injected on the column was below the detection limit of the instrument. The other replicate at 0.2 ng was correctly identified, but the spectrum was distorted and the smaller abundant peaks in the spectrum were not detected, resulting in a low QRM value.

In general, the U-47700 results are representative of all the opioid data analyzed.

Table 3. Hit list for U-47700.

| QRM (%) | Mass (ng) | Hit Name | Similarity Score |
|---|---|---|---|
| 100 | 50 | u47700 p | 0.997 |
| 100 | 50 | u47700 p | 0.997 |
| 100 | 50 | u47700 p | 0.998 |
| 100 | 25 | u47700 p | 0.995 |
| 100 | 25 | u47700 p | 0.996 |
| 100 | 25 | u47700 p | 0.996 |
| 100 | 12.5 | u47700 p | 0.992 |
| 100 | 12.5 | u47700 p | 0.993 |
| 100 | 12.5 | u47700 p | 0.992 |
| 100 | 6.25 | u47700 p | 0.988 |
| 100 | 6.25 | u47700 p | 0.988 |
| 100 | 6.25 | u47700 p | 0.987 |
| 100 | 3.13 | u47700 p | 0.989 |
| 100 | 3.13 | u47700 p | 0.988 |
| 100 | 3.13 | u47700 p | 0.989 |
| 100 | 1.56 | u47700 | 0.984 |
| 100 | 1.56 | u47700 | 0.984 |
| 99 | 1.56 | u47700 p | 0.984 |
| 100 | 0.78 | u47700 | 0.981 |
| 100 | 0.78 | u47700 | 0.982 |
| 99 | 0.78 | u47700 p | 0.984 |
| 99 | 0.39 | u47700 | 0.973 |
| 100 | 0.39 | u47700 | 0.98 |
| 98 | 0.39 | u47700 | 0.98 |
| 100 | 0.2 | u47700 | 0.952 |
| 100 | 0.2 | u47700 | 0.969 |
| 100 | 0.2 | u47700 | 0.971 |
| 98 | 0.1 | u47700 | 0.951 |
| 94 | 0.1 | u47700 | 0.952 |
| 100 | 0.1 | u47700 | 0.919 |
| 63 | 0.05 | u47700 | 0.889 |
| 100 | 0.05 | u47700 | 0.896 |
| 95 | 0.05 | u47700 | 0.938 |
| 9 | 0.02 | u47700 | 0.774 |
| 32 | 0.02 | 1-(2-(n-pyrrolidyl)ethylthio)-1-methylthiopropene | 0.794 |
| 11 | 0.02 | 2(1h)-pyridinone 4-hydroxy-6-methyl- | 0.729 |

Table 4Table 5 contain examples of contrasting QRM and similarity scores. Table 4 demonstrates a case where the match statistics are similar for the top two-hits in the tramadol list, but the QRM is very different. A QRM value of 100 was recorded for Hit 1 and 14 for Hit 2.

This difference is due to the quality of the reference spectra. The custom-built library contains multiple records for each reference spectrum. The reference spectra for tramadol was acquired from two sources: the Houston Forensic Science Center and the Wiley MS Library. The reference spectra are compared in Figure 4 and the mass spectrum for Hit 2, from the Wiley Library, appears to be of lower quality than that of Hit 1. Some of the identifying mass fragments in the m/z range of 100 to 200 for tramadol are missing in Hit 2, which makes Hit 1 a more reliable match. The difference in QRM values obtained for Hit 1 and 2 illustrates the sensitivity of this metric.

The last example, shown in Table 5**,** demonstrates a case where the match for tramadol is good, but the QRM value indicates that the result is unreliable. The tramadol is not identified in the search because similar to the U-47700 results, the injected mass of 0.2 ng is below the detection limit of the instrument.

Table 4. Hit list for Tramadol with differing QRMs

| Hit Number | Similarity | QRM (10) | Compound Name |
|---|---|---|---|
| 1 | 0.9997 | 99.97 | tramadol |
| 2 | 0.9970 | 14.02 | tramadol |
| 3 | 0.9962 | 10.93 | phenol 4-[2-(dimethylamino)ethyl]- |
| 4 | 0.9958 | 9.27 | 6-hydroxynobiline |
| 5 | 0.9958 | 6.57 | phenol 4-[2-(dimethylamino)ethyl]- |
| 6 | 0.9958 | 10.02 | *N*, *N*-dimethyl-undecylamine |
| 7 | 0.9957 | 9.03 | *N*, *N*-dimethyl-tridecylamine |
| 8 | 0.9957 | 0.54 | ephedrine (-)- |
| 9 | 0.9956 | 7.74 | *N*, *N*-dimethyl-pentadecylamine |
| 10 | 0.9956 | 7.17 | *N*, *N*-dimethyl-heptadecylamine |

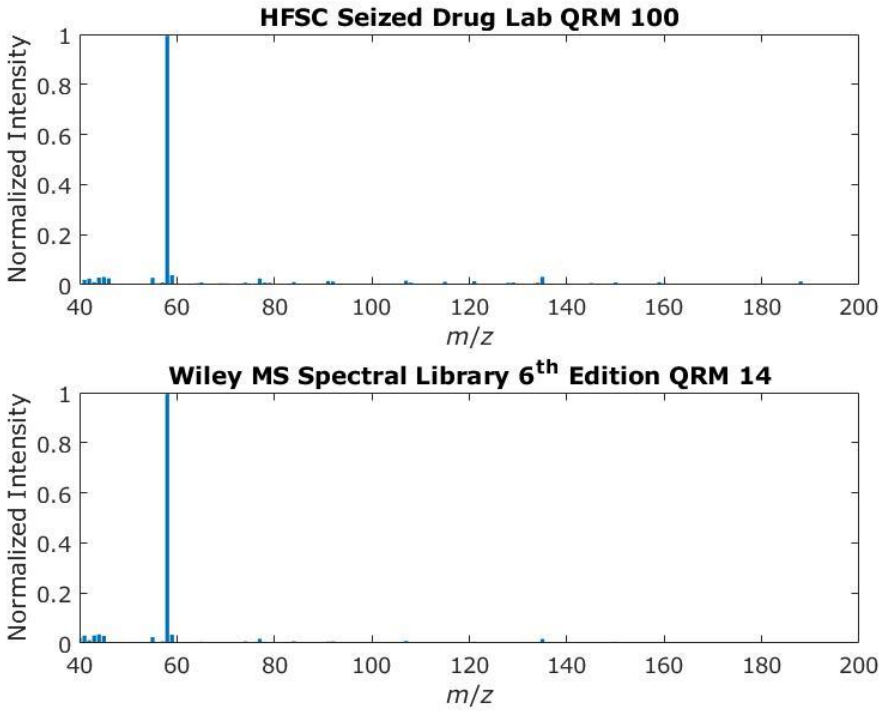



Figure 4. Tramadol Mass Spectrum acquired from two sources. Top - Houston Forensic Science Center. Bottom - Wiley MS library. Note that the smaller peak in the m/z range of 100 to 200 are missing in the Wiley spectra.

Table 5. Hit list for Tramadol with misidentifications.

| Hit Number | Similarity Score | QRM | Compound Name |
|---|---|---|---|
| 1 | 0.9875 | 8.2 | 10, 11-dihydro-5-(1-hydroxy-3-dimethylaminopropylidenyl)dibenzo[b e]oxepin |
| 2 | 0.9871 | 12.3 | 1H-indole-3-ethanamine *N, N*-dimethyl- |
| 3 | 0.9868 | 0.6 | 1H-indole-3-ethanamine *N*, *N*-dimethyl- |
| 4 | 0.9854 | 44.5 | *N*, *N* 2, 7-tetramethyl-2, 7-octadien-1-amine |
| 5 | 0.9801 | 0.2 | 2-methyamino-*N*-heptane |
| 6 | 0.9775 | 0.0 | 5-amino-4-cyano-3-(4-ethylaminobutyl)pyrazole |
| 7 | 0.9769 | 0.0 | 2-penten-1-amine *N, N* 2-trimethyl- (e)- |
| 8 | 0.9691 | 0.0 | ethanedial monohydrate dimer |
| 9 | 0.9677 | 0.0 | 1-butanamine *N*-ethyl- |
| 10 | 0.9673 | 0.0 | 1-pentanamine *N*-ethyl- |

### 4.3. Validation Study

The QRM performance was validated with an external validation dataset which was not used to train the QRM. A total of 165 mass spectra obtained from eighteen opioids were run through the QRM model. The list of drugs used in the validation process are shown in Table 6**.** The dataset consisted of GC/MS data extracted from randomly selected forensic case files and lab generated mixtures containing the reference standards and adulterants. The raw data files corresponding to each drug and the blank preceding each run were extracted from each dataset.

To process the data in batch, an automated Matlab® pipeline was developed to locate gas chromatographic peaks and remove column bleed components from the mass spectrum prior to database searching. The output from the pipeline was analyzed to determine the percentage breakdown for when the drugs were identified as the top hit in the query, ranked among the top 10 hits, or misidentified. The QRM values were used to evaluate the quality of the mass spectrum.

Table 6. Validation Drugs

| | |
|---|---|
| Acetylmorphine | Hydromorphone |
| Benzyl Fentanyl | Methadone |
| Buprenorphine | Methoxyacetylfentanyl |
| Carfentanil | Morphine |
| Codeine | Oxycodone |
| Fentanyl | Oxymorphone |
| Heroin | Tramadol |
| Hydrocodone | U-47700 |
| Hydromorphine | 4-Aminophenyl-1-phenethylpiperidine (4-ANNP) |

**Forensic Case Files Results**

The results from the forensic case files data are shown in Table 7. Out of 165 drugs analyzed, all were ranked in the 10-hit list and as the top hit, except for one of the methadone datasets which did not identify as the top hit. As shown in Table 8, the top hit was identified as promethazine whereas methadone was the 9$^{th}$ hit in the list with a similarity score of 0.994 and a QRM value of 1.7%.

There are two factors that may have contributed to the low ranking of methadone. The first one is the close similarity in mass spectra fragmentation patterns between methadone and promethazine, as observed in Figure 5**.** The second one is the multiple promethazine records in the reference library and variation in the quality of the mass spectrum in each record. Promethazine was identified in 7 out of the 10 hits and each hit had a different record number as seen in Table 8. The issue arising from the multiple records could be mitigated by limiting the number of reference spectrum copies in the library.

Table 7. Results from the QRM performance evaluation.

| Compound | # of Files | % Top Hit per Query | % Ranked in the Top 10 Hit | Misidentification Rate (%) |
|---|---|---|---|---|
| Acetylmorphine | 1 | 100 | 100 | 0 |
| Benzyl fentanyl | 5 | 100 | 100 | 0 |
| Buprenorphine | 2 | 100 | 100 | 0 |
| Carfentanil | 9 | 100 | 100 | 0 |
| Codeine | 22 | 100 | 100 | 0 |
| Fentanyl | 14 | 100 | 100 | 0 |
| Heroin | 20 | 100 | 100 | 0 |
| Hydrocodone | 16 | 100 | 100 | 0 |
| Hydromorphine | 1 | 100 | 100 | 0 |
| Hydromorphone | 5 | 100 | 100 | 0 |
| Methadone | 9 | 89 | 89 | 11 |
| Methoxyacetyl Fentanyl | 7 | 100 | 100 | 0 |
| Monoacetylfentanyl | 1 | 100 | 100 | 0 |
| Morphine* | 11 | 100 | 100 | 0 |
| Oxycodone | 18 | 100 | 100 | 0 |
| Oxymorphone | 100 | 100 | 100 | 0 |
| Tramadol* | 8 | 100 | 100 | 0 |
| U-47700 | 15 | 100 | 100 | 0 |

**Morphine and Tramadol were identified twice within one chromatographic run, in each of their respective datasets.*

Table 8. Hit list for methadone.

| Hit Number | Similarity Score | QRM | Record Number | CAS | Compound Name |
|---|---|---|---|---|---|
| 1 | 0.998776 | 100 | 229699 | 60877 | Promethazine |
| 2 | 0.998012 | 92.58957 | 222137 | 60877 | promethazine |
| 3 | 0.997924 | 30.44633 | 222141 | 60877 | promethazine |
| 4 | 0.996946 | 48.20926 | 222139 | 60877 | promethazine |
| 5 | 0.996678 | 51.69157 | 222140 | 60877 | promethazine |
| 6 | 0.995256 | 2.485291 | 226566 | 7456248 | fonazine |
| 7 | 0.995219 | 3.776417 | 222142 | 60877 | promethazine |
| 8 | 0.995135 | 3.759733 | 222143 | 60877 | promethazine |
| 9 | 0.994931 | 1.701874 | 223517 | 76993 | 3-heptanone 6-(dimethylamino)-4 4-diphenyl- |
| 10 | 0.994206 | 2.011758 | 229015 | 0 | N N-dimethyl-4-tert-butylamphetamine |

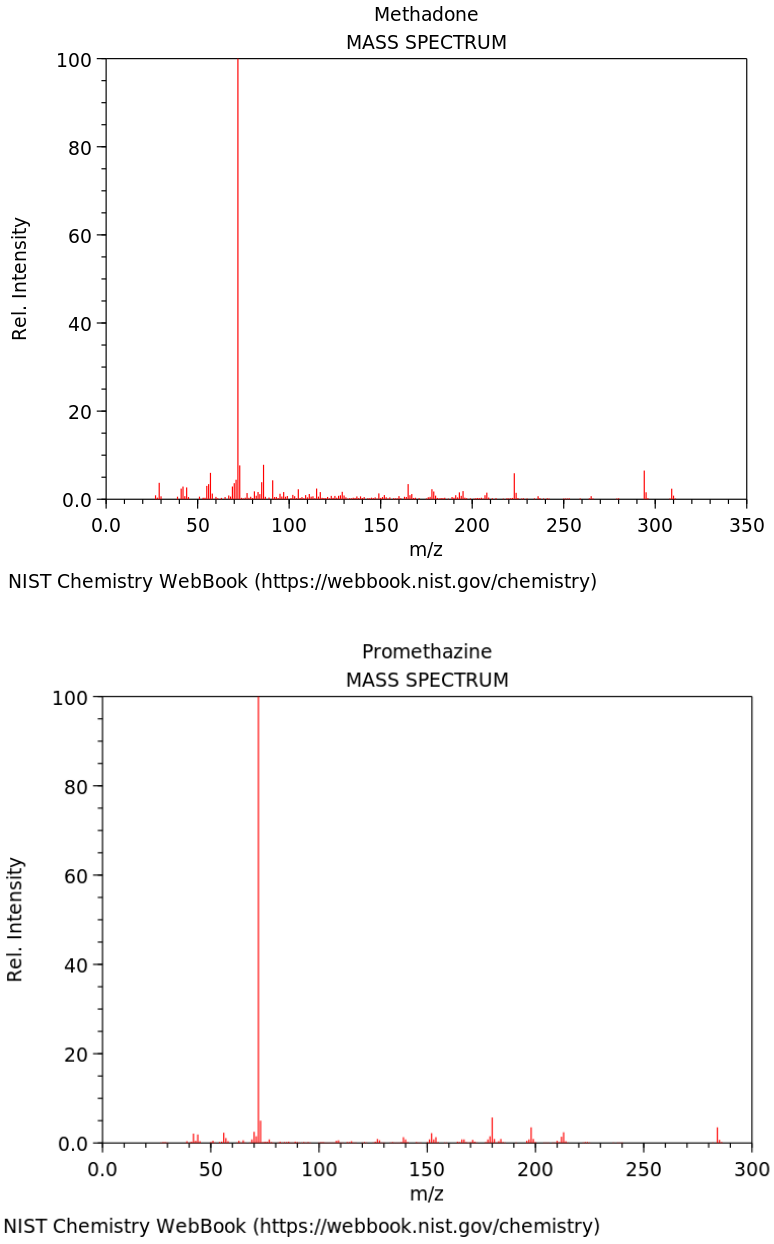



Figure 5. Mass Spectrum for (Top) Methadone and (Bottom) Promethazine. Note the similarity in mass fragment patterns. Spectra acquired from NIST Chemistry WebBook.

As for the morphine and tramadol samples, they were both identified as the top hit in their respective datasets. However, upon closer inspection of the data, we observed that morphine and tramadol, in each of their respective datasets, were identified twice within one chromatographic run. Unlike the methadone-promethazine example, this duplication is not due to multiple records in the reference library. In the case of morphine, sixteen peaks were detected and identified within the chromatographic run. Peak information can be found in Table 9. Usually each peak has a unique identity, but the two peaks at retention times 9.01 min and 9.10 min identified as morphine and have similar mass spectral fragmentation patterns. The peak at 9.10 min elutes on the tail of the 9.01 min peak and is much smaller in terms of area. These peaks may be isomeric compounds, but further analysis is required to confirm the peak identity. Similar results were observed for tramadol.

Table 9. Hit list for morphine.

| Peak # | QRM | Area | CAS | Hit Name | Similarity Score |
|---|---|---|---|---|---|
| **1** | 1 | 906900 | 3891983 | dodecane 2 6 10-trimethyl- | 0.87 |
| **2** | 3 | 546652 | 146805 | xanthosine | 0.931 |
| **3** | 0 | 2624302 | 142621 | hexanoic acid | 0.917 |
| **4** | 0 | 2292119 | 142621 | hexanoic acid | 0.917 |
| **5** | 2 | 3310978 | 57103 | hexadecanoic acid | 0.872 |
| **6** | 0 | 3019475 | 13432252 | 3-hexanol 2 4-dimethyl- | 0.831 |
| **7** | 52 | 713593 | 18593367 | cyclopropyl carbinyl-d2-methyl ether | 0.956 |
| **8** | 100 | 1410788 | 78693 | 3-octanol 3 7-dimethyl- | 0.922 |
| **9** | 0 | 1225388 | 112721 | 1-tetradecanol | 0.993 |
| **10** | 0 | 113126986 | 112721 | 1-tetradecanol | 0.979 |
| **11** | 0 | 340434 | 111900 | ethanol 2-(2-ethoxyethoxy)- | 0.851 |
| **12** | 100 | 862091 | 57103 | palmitic acid | 0.996 |
| **13** | 3 | 528922 | 112925 | 1-octadecanol | 0.988 |
| **14** | 72 | 116850436 | 112925 | 1-octadecanol | 0.977 |
| **15** | 96 | 1679644 | 57114 | stearic acid | 0.997 |
| **16** | 42 | 166556330 | 57272 | morphine | 0.962 |
| **17** | 98 | 827838 | 57272 | morphine | 0.925 |

## Opioid Mix Results

To determine how the QRM would perform when presented with a complex mixture, we processed samples containing the sixteen reference standards and four adulterants at varying concentrations. As expected, the mixture shown in Figure 6 resulted in poorly separated peaks. These un-resolved peaks led to misidentification or low QRM scores, thus identifying potential limitations in the QRM that need to be addressed in future studies. The QRM values are presented in Figure 7.

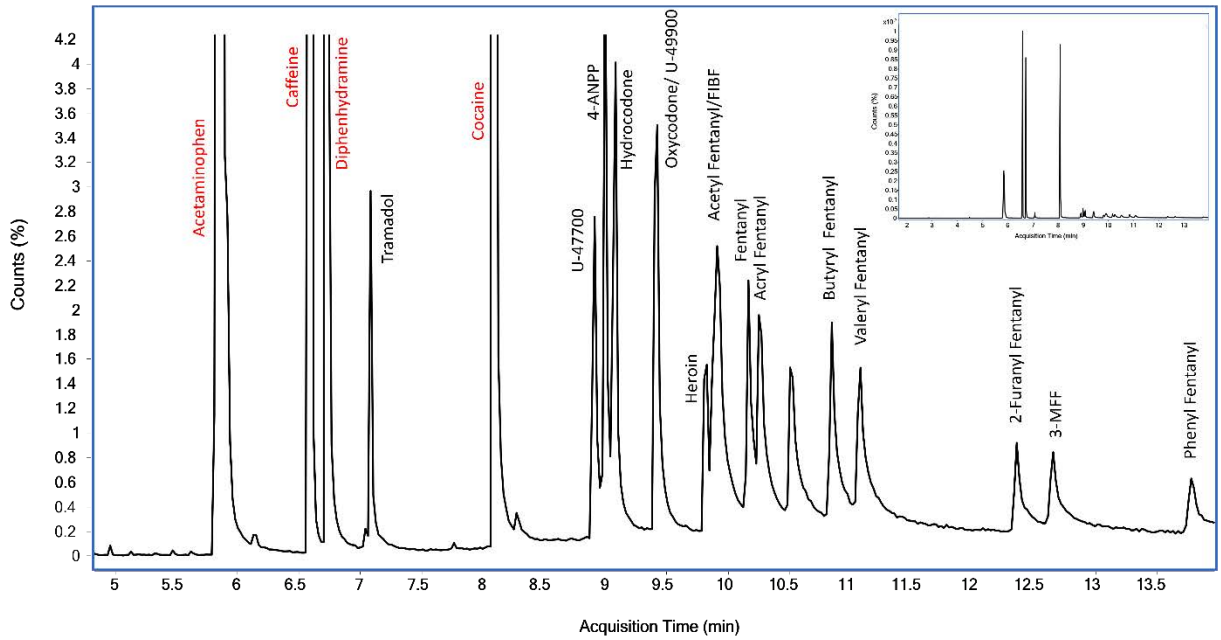



Figure 6. Zoomed view of the GC/MS chromatogram for the opioid/adulterant mixture at 45 µg/mL. The full chromatogram is shown in the inset.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

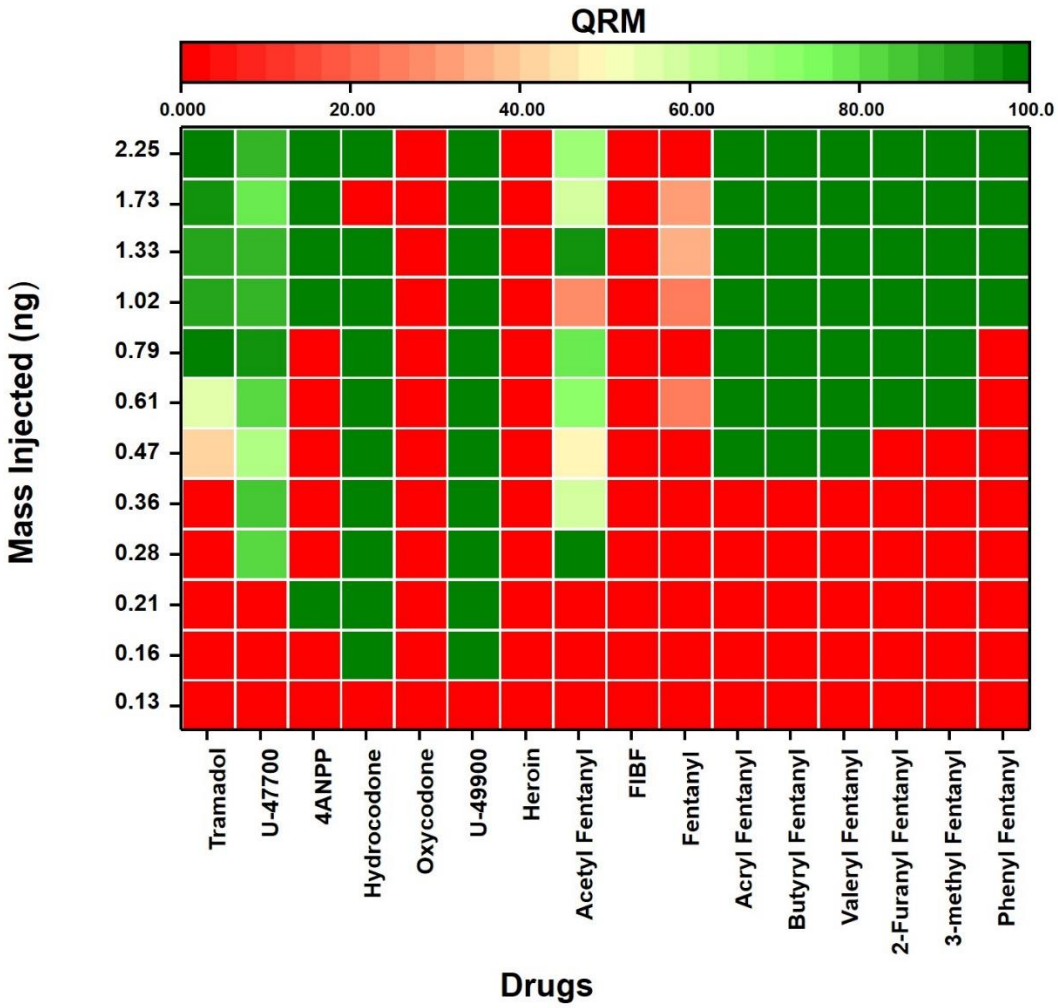



Figure 7. QRM values for opioid mixture analyzed in the validation study.

## 5. Conclusions

The goal of developing an Information Quality Model that could determine the probability of identity accuracy for a given mass spectrum searched against the library was accomplished in this project. The QRM, developed during this project, provides an independent measure of the quality of each library search result. The key advantages of the QRM is that it can be used with any kind of library and similarity metric, in addition to providing a statistical result of the match reliability. For example, if the similarity measure is high and the QRM is low, then there is not

much confidence in the result. However, when the similarity measure is low and the QRM is high, then the analyst can be confident in the identification. In addition to the QRM, the development of the basis set and MCR data processing techniques proved invaluable for chromatographic peak detection, background removal, and library search optimization.

### 5.1. Limitations and Implications for Future Research

The results of this study showed that the QRM could be used as a probability measure of reliability in a library search. The current iteration of the QRM was designed to work with chromatographic peaks that are baseline resolved. However, as shown in the opioid mix results, co-eluting peaks render it difficult to select identifying ions arising from each opioid and therefore result in misidentification or variations in the QRM value. The QRM may also give wrong estimates when the mass spectrum is distorted to the point that it resembles a spectrum in the library. These estimates would result as false positives when the compound is missing from the library. Additionally, in some cases, the QRM value can be influenced by concentration. For example, at low concentrations, the QRM value may be low but is also the highest for all the retrieved spectra in the hitlist, indicating a likely match. To expand and improve the capabilities of the QRM, further research is needed. There are many areas of development that can be pursued for metric refinement. This includes:

1. Development of a reverse search algorithm that is resistant to spurious peaks in the query mass spectrum.
2. Improvement of chromatographic peak detection for overlapping peaks.
3. Use of advance deep learning methods to recognize novel synthetic drugs that are not contained in the reference library by recognition of the active template of the drug molecule.

## 5.2. Implications for Policy and Practice

With creation of the Organization of Scientific Area Committees (OSAC) by the National Institute of Standards and Technology, the forensic community has moved toward standardization of disciplines with the goal of strengthening forensic science through development of nationalized standards and guidelines. As for seized drug analysis, the importance of a sufficiency standard for drug analysis has been noted; however, discussion of development and implementation has been limited. As the results from this study suggest, the QRM provides an empirical quality value that can be applied across multiple disciplines and serve as a model for the nationalization of quality standards in library searches. Implementation of a national quality scoring system can improve consistency across individuals, instruments, and forensic laboratories.

# 6. Dissemination of Research Findings

**Publications:**

A Quantitative Reliability Metric for Querying Large Databases—Manuscript in preparation Z. Chen, P.B. Harrington, V. Shetty, Angelica Noyola, and P. Rearden

**Presentations:**

Harrington, P.B., Chen, Z., Shetty, V., Noyola, A. and Rearden, P. (2020, October 14). *A Quantitative Reliability Metric for Querying Large Databases* [Oral Presentation}] SciX 2020 Meeting, Virtual Convention.

## 7. References


1. Council, N. R., *Strengthening forensic science in the United States: a path forward*. National Academies Press: 2009.

2. Stein, S. E.; Scott, D. R., Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry* **1994,** *5* (9), 859-66.

3. Garg, A.; Enright, C. G.; Madden, M. G., Improving spectral library search by redefining similarity measures. *Journal of chemical information and modeling* **2015,** *55* (5), 963-71.

4. Samokhin, A.; Sotnezova, K.; Lashin, V.; Revelsky, I., Evaluation of mass spectral library search algorithms implemented in commercial software. *Journal of mass spectrometry : JMS* **2015,** *50* (6), 820-5.

5. McLafferty, F. W.; Zhang, M. Y.; Stauffer, D. B.; Loh, S. Y., Comparison of algorithms and databases for matching unknown mass spectra. *Journal of the American Society for Mass Spectrometry* **1998,** *9* (1), 92-5.

6. Harrington, P. B.; Isenhour, T. L., A quantitative measure of the reliability of searches of spectral libraries. *Analytica Chimica Acta* **1987,** *197*, 105-119.

7. Xu, Z.; Sun, X.; Harrington, P. d. B., Baseline Correction Method Using an Orthogonal Basis for Gas Chromatography/Mass Spectrometry Data. *Analytical Chemistry* **2011,** *83* (19), 7464-7471.

8. Wang, Z.; Zhang, M.; Harrington, P. d. B., Comparison of Three Algorithms for the Baseline Correction of Hyphenated Data Objects. *Analytical Chemistry* **2014,** *86* (18), 9050-9057.

9. Ammann, L. P., Robust Singular Value Decompositions: A New Approach to Projection Pursuit. *Journal of the American Statistical Association* **1993,** *88* (422), 505-514.

10. SWGDRUG Recommendations - 2019. http://www.swgdrug.org (accessed October 20).