Qualitative uncertainty (reliability) of chemical identification with High Resolution Mass Spectrometry

M. Kostakis, MSc, PhD Candidate

N Thomaidis, Associate Professor

Laboratory of Analytical Chemistry
Department of Chemistry
University of Athens

Introduction & Scope

- HRMS becoming more accessible in laboratories
- High mass accuracy and resolving power is the main advantage but it doesn't assure that HRMS is an non-error technique.
- In literature has been mentioned cases with errors in identification from matrix effect.
- To explore the identification capabilities of HRMS
- To study the identification criteria of HRMS
- To calculate the uncertainty of Identification for HRMS

Fast forward in identification

- **Identification:** is a qualitative result from a method capable of providing structural information that meets acceptable criteria for the purpose of the analysis.
- Identification is a qualitative parameter-binary.
- Identification criteria that should be fulfilled for HRMS:
  - Retention Time RT ± 0.2 min
  - Mass Accuracy <5 ppm
  - Isotopic Fit Score < 200 mSigma
  - Area-Intensity & their ratio (peak score)
  - Mass Fragmentation

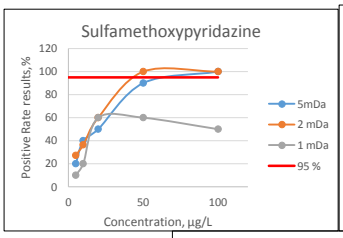
Our case

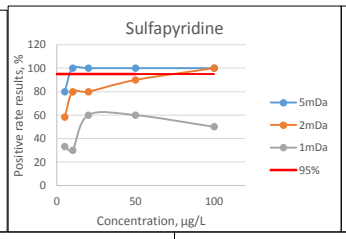
- Analytes: sulfonamides
  - Sulfaguanidine, sulfamethizole, Sulfachloropyridazine, Sulfaclozine, sulfadiazine, sulfamethoxazole, Sulfapyridine, sulfamerazine, sulfameter, sulfamethoxypyridazine, sulfamonomethoxine, sulfamoxole, sulfisoxazole, sulfadimidine, sulfadimethoxine, sulfadoxine, Sulfaquinoxaline
- Matrix material: fish tissue
- Matrix material was spiked in 5 different concentration levels, 5, 10, 20, 50, 100 μg/kg
- The samples analyzed in 5 replicates x 2 days = 10
- From data, it was studied the identification criteria of HRMS (mass accuracy, isotopic fit score, retention time.
- Instrumentation: Bruker Maxis Impact qTOF

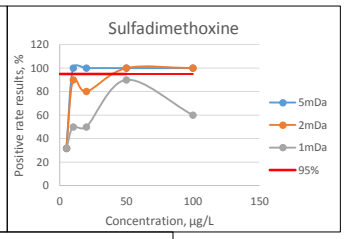
Mass Accuracy

- Mass accuracy was studied in 3 different areas, 5 mDa, 2 mDa, 1mDa
- **Performance curve**: Positive results rate (TP+FP) and concentration.
- Determined the Limit Of Identification (LOI) for 95% PRR.

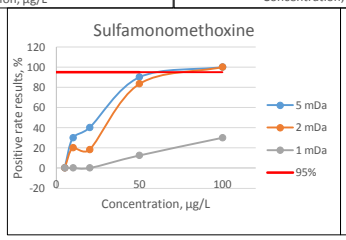
Performance Curve

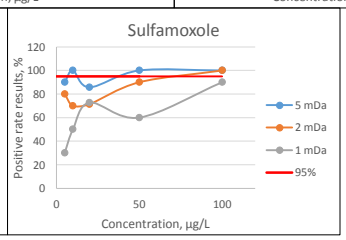






5 mDa: high number of positive results but larger probability for false positive detects
1 mDa: low number of positive results but larger probability for false negative detects





Performance Curve- Limit of Identification

ppb µg/L	5	10	20	50	100	mDa
Sulfaguanidine	0	10	**42**	100	100	**2**
sulfamethizole	10	27	82	**60**	100	**2**
Sulfachloropyridazine	**60**	100	100	100	100	***5***
Sulfaclozine	0	40	**44**	100	100	***5***
sulfadiazine	**90**	100	100	100	90	***5***
sulfamethoxazole	**90**	100	100	100	90	***5***
Sulfapyridine	58	80	80	**90**	100	**2**
sulfamerazine	63	90	**90**	100	100	***5***
sulfameter	0	0	40	**67**	100	**2**
sulfamethoxypyridazine	27	36	**60**	100	100	**2**
sulfamonomethoxine	0	20	18	**83**	100	**2**
sulfamoxole	80	70	71	**90**	100	**2**
sulfisoxazole	30	90	70	**90**	100	**2**
sulfadimidine	0	10	10	**80**	100	***5***
sulfadimethoxine	32	90	**80**	100	100	**2**
sulfadoxine	32	80	**90**	100	100	**2**
Sulfaquinoxaline	32	50	**80**	100	100	**2**

Isotopic Fit Score

- **Isotopic fit score** is a measure of the correlation between theoretical and measured isotopic pattern peak and expressed as mSigma value.
- Valid range 0-1000.
- The lower is mSigma, the better is the fitting.
- Organic compound with limited number of atoms (C,N,S,O) need more expanded limits of fitting
- But narrow window in isotopic fitting resulted higher number of false positives.
- It was measured the mean value and deviation of mSigma for every analyte in every concentration.

Isotopic Fit Score

mSigma	5 µg/kg		10 µg/kg		20 µg/kg		50 µg/kg		100 µg/kg	
	mean	deviation	mean	deviation	mean	deviation	mean	deviation	mean	deviation
Sulfaguanidine	613	30	591	55	586	90	512	62	409	139
sulfamethizole	123	46	75	28	76	32	45	13	14	6
Sulfachloropyridazine	205	133	172	166	119	57	75	77	46	57
Sulfaclozine	nm	nm	293	202	230	148	129	90	98	86
sulfadiazine	89	40	78	63	76	44	23	8	11	4
sulfamethoxazole	356	162	211	175	287	173	148	130	60	52
Sulfapyridine	114	13	54	29	45	24	9	5	6	3
sulfamerazine	463	125	264	112	271	141	67	29	42	16
sulfameter	nm	nm	67	34	48	13	26	11	10	3
sulfamethoxypyridazine	92	38	60	25	40	21	17	11	10	3
sulfamonomethoxine	nm	nm	69	29	44	20	25	10	10	4
sulfamoxole	315	142	165	69	186	162	42	75	8	4
sulfisoxazole	357	156	224	123	186	167	49	30	12	9
sulfadimidine	nm	nm	57	37	47	13	36	7	31	1
sulfadimethoxine	141	81	80	50	85	50	29	13	12	6
sulfadoxine	80	40	40	22	26	15	9	3	13	3
Sulfaquinoxaline	218	150	290	151	161	107	230	89	192	58

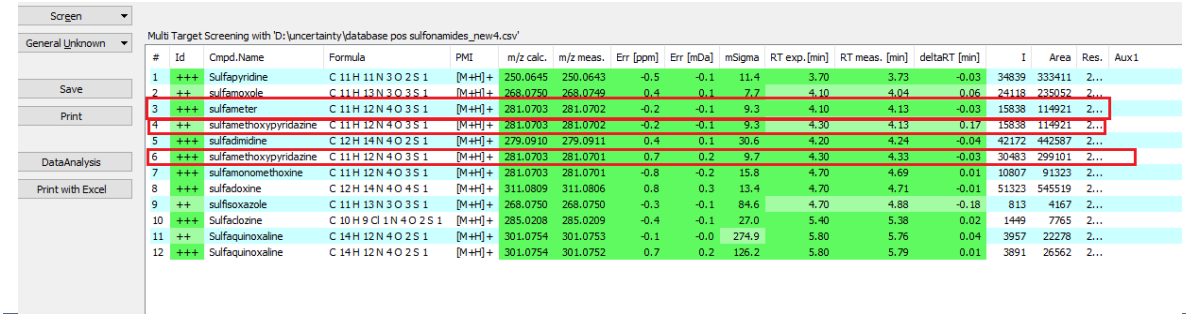
mSigma<200

300<mSigma<200

mSigma>300

Retention Time

- As RT tolerance was chosen 0.2 min
- In all compounds RT tolerance was <0.1 min.
- Exception Sulfameter and sulfamethoxypyridine, very close RT<0.2 min and software confuses the analytes.



#	Id	Cmpd.Name	Formula	PMI	m/z calc.	m/z meas.	Err [ppm]	Err [mDa]	mSigma	RT exp.[min]	RT meas. [min]	deltaRT [min]	I	Area	Res.	Aux1
1	+++	Sulfapyridine	C 11 H 11 N 3 O 2 S 1	[M+H]+	250.0645	250.0643	-0.5	-0.1	11.4	3.70	3.73	-0.03	34839	333411	2...	
2	++	sulfamoxole	C 11 H 13 N 3 O 3 S 1	[M+H]+	268.0750	268.0749	0.4	0.1	7.7	4.10	4.04	0.06	24118	235052	2...	
3	+++	sulfameter	C 11 H 12 N 4 O 3 S 1	[M+H]+	281.0703	281.0702	-0.2	-0.1	9.3	4.10	4.13	-0.03	15838	114921	2...	
4	++	sulfamethoxypyridazine	C 11 H 12 N 4 O 3 S 1	[M+H]+	281.0703	281.0702	-0.2	-0.1	9.3	4.30	4.13	0.17	15838	114921	2...	
5	+++	sulfadimidine	C 12 H 14 N 4 O 2 S 1	[M+H]+	279.0910	279.0911	0.4	0.1	30.6	4.20	4.24	-0.04	42172	442587	2...	
6	+++	sulfamethoxypyridazine	C 11 H 12 N 4 O 3 S 1	[M+H]+	281.0703	281.0701	0.7	0.2	9.7	4.30	4.33	-0.03	30483	299101	2...	
7	+++	sulfamonomethoxine	C 11 H 12 N 4 O 3 S 1	[M+H]+	281.0703	281.0701	-0.8	-0.2	15.8	4.70	4.69	0.01	10807	91323	2...	
8	+++	sulfadoxine	C 12 H 14 N 4 O 4 S 1	[M+H]+	311.0809	311.0806	0.8	0.3	13.4	4.70	4.71	-0.01	51323	545519	2...	
9	++	sulfisoxazole	C 11 H 13 N 3 O 3 S 1	[M+H]+	268.0750	268.0750	-0.3	-0.1	84.6	4.70	4.88	-0.18	813	4167	2...	
10	+++	Sulfaclozine	C 10 H 9 Cl 1 N 4 O 2 S 1	[M+H]+	285.0208	285.0209	-0.4	-0.1	27.0	5.40	5.38	0.02	1449	7765	2...	
11	++	Sulfaquinoxaline	C 14 H 12 N 4 O 2 S 1	[M+H]+	301.0754	301.0753	-0.1	-0.0	274.9	5.80	5.76	0.04	3957	22278	2...	
12	+++	Sulfaquinoxaline	C 14 H 12 N 4 O 2 S 1	[M+H]+	301.0754	301.0752	0.7	0.2	126.2	5.80	5.79	0.01	3891	26562	2...	

Uncertainty of Identification

- ***The possibility of false identification, also called as "reliability/unreliability", "confidence".***
- ***Binary response***
- ***Its not assurance but the <u>probability</u> of correct or false detect.***
- ***Uncertainty - 2 approaches:***
  - ***Contingency table***
  - ***Bayesian Method***

Uncertainty of Identification

Contingency table

$$sensitivity = \frac{tp}{tp + fn} \times 100$$

$$specificity = \frac{tn}{tn + fp} \times 100$$

$$\text{PPV} = \frac{tp}{tp+fp} \times 100$$

$$\text{NPV} = \frac{tn}{tn+fn} \times 100$$

Bayesians

$$P(A|A) = \frac{PREV \times SENS}{PREV \times SENS + (1 - PREV) \times (1 - SPEC)}$$

$$P(nA|nA) = \frac{PREV \times (SPEC)}{(1 - PREV) \times SPEC + PREV \times (1 - SENS)}$$

$$PREV = \frac{number\ of\ real\ positive\ results\ (tp + fn)}{total\ number\ of\ samples}$$

Historical Data: Databases, previous results, validation dataset or 0.5 in cases without any prior information

Uncertainty of Identification
Results

Contingency Table

NPV: is near to 100% because the concentration of samples is near to LOI
PPV: large variation because of different sensitive of every compound in HRMS, different LOI

	contingency table approach		Bayes approach		
	PPV	NPV	PREV	P(A|A)	P(nA|nA)
Sulfaguanidine	38.5	92.3	52.0	73.5	78.3
sulfamethizole	100	100	58.0	84.7	100
Sulfachloropyridazine	40.0	87.0	92.0	96.9	100
Sulfaclozine	80.0	80.0	56.0	69.4	75.9
sulfadiazine	97.5	10.0	96.0	99.6	14.3
sulfamethoxazole	96.0	0	96.0	100	0
Sulfapyridine	31.1	81.8	82.0	95.2	30.6
sulfamerazine	97.4	81.8	80.0	85.6	88.9
sulfameter	45.8	84.4	46.0	64.1	74.2
sulfamethoxypyridazine	62.5	100	64.0	81.4	100
sulfamonomethoxine	52.6	100	38.0	68.9	100
sulfamoxole	23.3	100.0	86.0	97.6	100
sulfisoxazole	36.8	100	76.0	92.6	100
sulfadimidine	100	80.4	40.0	62.5	100
sulfadimethoxine	48.8	92.3	73.5	73.5	78.3
sulfadoxine	47.6	100	86.0	96.7	100
Sulfaquinoxaline	57.1	92.3	70.0	73.5	78.3

Bayesians

PREV: Important factor, depends on the dataset, and the correct estimation of uncertainty. Main drawback of Bayesians.

Conclusions & Perspectives

- The identification criteria for HRMS, namely mass accuracy, Isotopic fitting score and retention time, were investigated.
- A mass accuracy of 2 mDa (and in some few cases, at 5 mDa) is the most appropriate value in order to avoid false detects.
- Isotopic fitting need a caution on identification, because it is dependent on analyte concentration and the elemental structure in order to be reliable.
- Retention time is a very reliable and stable criterion for identification
- The uncertainty for identification was calculated with both approaches (contingency tables and Bayes theory)
- As next step is the study of the mass fragmentation

THANK YOU

Acknowledgements: Associate Professor N. Thomaidis