
Eurachem 



CITAC
Cooperation on International
Traceability in Analytical Chemistry

Příručka EURACHEM/CITAC

**Hodnocení výkonnosti
a nejistota v kvalitativní
chemické analýze**

Záměrně ponecháno prázdné

Hodnocení výkonnosti a nejistota v kvalitativní chemické analýze

První vydání (2021)

Editoři

Ricardo Bettencourt da Silva (Faculdade de Ciências da Univ. de Lisboa, Portugalsko),
Stephen L R Ellison (LGC, Spojené království)

Složení pracovní skupiny *

Eurachem members

R. Bettencourt da Silva (předseda)	<i>Univ. Lisboa, Portugalsko</i>
S. Ellison (tajemník)	<i>LGC, Spojené království</i>
A. Togola	<i>BRGM, Francie</i>
D. Ivanova	<i>Eurachem Bulharsko</i>
E. Theodorsson	<i>LIU, Švédsko</i>
E. Totu	<i>University Politehnica of Bucharest, Rumunsko</i>
H. Emons	<i>Evropská komise, Evropská unie</i>
I. Leito	<i>Univ Tartu, Estonsko</i>
M. Segal	<i>INRIM, Itálie</i>
O. Levgar	<i>Ukrmetrteststandart, Ukrajina</i>
O. Pellegrino	<i>IPQ/DMET, Portugalsko</i>
P. Pereira	<i>IPST, Portugalsko</i>
R. Kaus	<i>Eurachem Německo</i>
S. Lardy-Fontan	<i>LNE, Francie</i>
W. Wegscheider	<i>Montanuniversitaet Leoben, Rakousko</i>

CITAC members

A. Botha	<i>NMISA, Jihoafrická republika</i>
F. Lourenço	<i>Univ. São Paulo, Brazílie</i>

*V čase schválení dokumentu

Poděkování:

Tento dokument byl vytvořen pracovní skupinou Eurachem/CITAC ve složení, které je uvedeno vpravo. Editoři jsou zavázáni všem jejím členům a organizacím a dalším, kteří přispěli připomínkami, návrhy a pomocí.

Příprava této příručky byla částečně podpořena Fundação para a Ciência e a Tecnologia, Portugalsko.

Tato publikace by měla být citována následovně *:

“R Bettencourt da Silva and S L R Ellison (eds.) Eurachem/CITAC Guide: Assessment of performance and uncertainty in qualitative chemical analysis. First Edition, Eurachem (2021). ISBN 978-0-948926-39-6. Dostupné z <https://www.eurachem.org>”

* Závísí na požadavcích časopisů

Obsah

Obsah	3
Předmluva	5
Účel	6
1 Úvod	9
2 Typy kvalitativní analýzy	11
3 Hodnocení výkonnosti kvalitativní analýzy	12
3.1 Obecné úvahy	12
3.2 Kvantifikace výkonu kvalitativní analýzy	12
3.3 Hodnocení falešně pozitivních a falešně negativních poměrů	15
3.4 Mez detekce a selektivita	18
4 Vyjádření spolehlivosti v kvalitativní analýze	20
4.1 Obecné úvahy	20
4.2 Věrohodnostní poměr	20
4.3 Posteriorní pravděpodobnost	22
4.4 Spolehlivost metrik	23
4.5 Nejistota poměrů	23
5 Uvádění kvalitativního analytického výsledku	25
6 Závěry a doporučení	26
7 Příklady	27
7.1 E1: Identifikace sloučenin hmotnostní spektrometrií s nízkým rozlišením pomocí vyhledávání v databázi nebo přítomnosti charakteristických iontů	27
7.2 E2: Identifikace čištěných sloučenin infračervenou spektrometrií	31
7.3 E3: Identifikace zneužívaných drog v moči technikou enzymové multiplikované imunoanalýzy (EMIT) a alternativní technikou	33
7.4 E4: Identifikace lidského genu SRY v biologickém materiálu pomocí qPCR	35
7.5 E5: Identifikace reziduí pesticidů v potravinách pomocí GC-MS/MS na základě retenčního času a zastoupení iontů	37
7.6 E6: Identifikace SARS-CoV-2 RNA testováním amplifikace nukleové kyseliny	40
Příloha A – Bayesova věta, šance a věrohodnostní poměr	43
A.1 Bayesova věta	43
A.2 Pravděpodobnost a šance	43
A.3 Forma Bayesovy věty pro šance a věrohodnostní poměr	44
Příloha B – Kvalitativní analýza spojená s posouzením shody a mezí stanovitelnosti	45
B.1 Posouzení shody jako kvalitativní analýza	45
B.2 Pozitivní a negativní výsledky posouzení shody	46
B.3 Příklad – Posouzení shody pro rezidua pesticidů v ovoci	46
B.4 Vzorce v tabulkovém procesoru pro posouzení shody pravděpodobností	47
Literatura	48

Záměrně ponecháno prázdné

Předmluva

Problému vyhodnocení a vyjádření nejistoty v kvalitativní chemické analýze je v literatuře věnováno mnohem méně pozornosti než nejistotě v kvantitativní analýze (tj. u měření) [1]. Zatímco někteří autoři se touto oblastí zabývali [2] – [11], obecné pokyny pro hodnocení výkonnosti v kvalitativní analýze nebo hodnocení a uvádění nejistoty kvalitativní analýzy jsou vzácné.

Neočekává se, že akreditované laboratoře v současné době vyhodnocují nebo uvádí nejistoty spojené s výsledky kvalitativní analýzy [12]. Normy ISO/IEC 17025 [13] a ISO 15189 [14] však vyžadují, aby laboratoře zajistily, že mohou dosáhnout platných výsledků kvalitativních a kvantitativních analýz. Je také důležité, aby si laboratoře byly vědomy spolehlivosti výsledků kvalitativní analýzy, což jim umožňuje v případě potřeby varovat před omezeními v interpretaci výsledků a přesně reagovat na dotazy zákazníků ohledně spolehlivosti. Kvantitativní posouzení spolehlivosti výsledků kvalitativní analýzy je zvláště užitečné, jsou-li chybné výsledky pravděpodobnější. Tato příručka je určena k použití v případech, kdy je žádoucí kvantitativní posouzení spolehlivosti výsledků kvalitativní analýzy.

Tato příručka se opírá o zkušenosti z několika analytických oblastí, kde je často používána kvalitativní analýza, např. ve forezních [15] a klinických oblastech [16] – [18], a o rozsáhlé obecné pokyny [7].

Účel

Tato příručka má laboratořím pomoci při stanovování a zavádění vhodných metodik pro hodnocení výkonnosti metod kvalitativní analýzy a vyhodnocování nejistot kvalitativní chemické analýzy.

V příručce je kvalitativní analýza definována jako „klasifikace podle stanovených kritérií“. Pro analytickou chemii a příbuzné obory se „kritérii“ obecně rozumí informace pro stanovení chemického složení, vlastností a/nebo struktury analyzovaných položek¹. V této příručce se počítá s následujícími typy kritérií:

- Kvantitativní kritéria, při nichž se číselný výsledek používá k zařazení zkoušeného předmětu do předem stanovené třídy;
- Kvalitativní kritéria, jako je přítomnost nebo nepřítomnost určitého znaku, změna barvy při zkoušce atd.

Tato příručka není vyčerpávající při popisu dostupných nástrojů pro hodnocení výkonnosti kvalitativních analytických metod a nejistoty kvalitativních analytických výsledků. Charakteristiky výkonnosti uvedené v příručce vycházejí z naměřených nebo odhadovaných podílů chybných výsledků a nezohledňují například míry shody mezi kvalitativními metodami nebo zpracování klasifikace na pořadových stupnicích² jinak než jako správnou nebo nesprávnou klasifikaci.

¹ Pozn. překladatele: v oblasti laboratorní medicíny lze tuto příručku aplikovat na parametry a hodnoty obsahu či množství.

² Pořadová stupnice je stupnice přirozeně uspořádaných kategorií, kde vzdálenost mezi kategoriemi není známa. Mohsova stupnice je pořadová stupnice tvrdosti minerálů.

Zkratky a symboly

V této příručce se vyskytují následující zkratky a symboly. Symboly použité v tomto dokumentu nejsou harmonizovány ve všech oblastech vědy, kde se používají. Například ve zdravotnické laboratoři jsou FN a FNR zkratky pro „počet falešně negativních výsledků“ a „poměr falešně negativních výsledků“.

A	Zastoupení iontů hmotnostního spektra	$LR(+)$	Věrohodnostní poměr pozitivních výsledků
\bar{A}	Průměrné zastoupení iontů hmotnostního spektra	$LR(-)$	Věrohodnostní poměr negativních výsledků
AR	Poměr zastoupení iontů hmotnostního spektra	n	Počet negativních výsledků
c	Naměřená koncentrace (nebo jakékoli jiné množství) analyzované položky	nc	Počet negativních případů
CI	Konfidenční Interval	NPV	Negativní prediktivní hodnota
c_{max}	Maximální přípustná koncentrace	$O(\cdot)$	Rozdíly ve prospěch události, např. $O(A)$ označuje rozdíly pro událost A
c_{min}	Minimální přípustná koncentrace	p	Počet pozitivních výsledků
DOR	Diagnostický poměr šancí	$P(\cdot)$	Pravděpodobnost události; např. $P(A)$ je pravděpodobnost události A
E	Účinnost	$P(+)$	Apriorní pravděpodobnost pozitivního případu
fn	Počet falešně negativních výsledků	$P(-)$	Apriorní pravděpodobnost negativního případu
FN	Poměr falešně negativních výsledků vztažený k pozitivním případům	pc	Počet pozitivních případů
fp	Počet falešně negativních výsledků	PN	Posteriorní pravděpodobnost negativního případu (viz příloha A)
FP	Poměr falešně pozitivních výsledků vztažený k negativním případům	PP	Posteriorní pravděpodobnost pozitivního případu (viz příloha A)
GC-MS	Plynová chromatografie s hmotnostní spektrometrií	PPV	Pozitivní prediktivní hodnota
GC-MS/MS	Plynová chromatografie s tandemovou hmotnostní spektrometrií	qPCR	Kvantitativní polymerázová řetězová reakce
GUM	Pokyn pro vyjadřování nejistoty měření	RT-PCR	Polymerázová řetězová reakce s reverzní transkripcí
$HL_{RR.95}$	Horní mez 95% konfidenčního intervalu pro poměr výsledků RR (např. SS)	RA	Relativní zastoupení
LC-MS	Kapalinová chromatografie s hmotnostní spektrometrií	Rn	Normalizovaná reportovaná hodnota
$LL_{RR.95}$	Horní mez 95% konfidenčního intervalu pro poměr výsledků RR (např. SS)	RR	Poměr výsledku
$LL_{RR.95}^{tg}$	Cílová nebo minimální hodnota pro $LL_{RR.95}$	S_A	Směrodatná odchylka zastoupení iontů
LOD	Mez detekce	SP	Specifičnost
LOQ	Mez stanovitelnosti	SS	Citlivost (klinická citlivost)
LR	Věrohodnostní poměr	S_{tRi}	Směrodatná odchylka retenčního času
		tn	Počet pravdivě negativních výsledků

TN	Skutečný negativní poměr vztažený k negativním případům
tp	Počet pozitivních výsledků
TP	Skutečný pozitivní poměr vztažený k pozitivním případům
t_R	Retenční čas
\bar{t}_{Ri}	Průměrný retenční čas
$u(c)$	Standardní nejistota c
w	Hmotnostní zlomek
Y	Youdenův index
ΔRn	Normalizovaná uváděná hodnota minus odezva základní linie
ρ	Spearmanův korelační koeficient mezi páry zastoupení iontů

1 Úvod

Mnoho relevantních socioekonomických nebo individuálních zájmů, jako je průmyslová produktivita a zdravotní stav, závisí na chemické analýze. Některé z těchto analýz jsou výhradně kvalitativní nebo zahrnují následnou kvantifikaci identifikované chemické entity. Zájmy, které mají být těmito analýzami chráněny, jsou zachovány pouze tehdy, je-li analytická kvalita vhodná pro zamýšlené použití.

V některých publikacích se pojem „vyšetření“ [1], „vyšetření nominální vlastnosti“ [19] nebo „zkouška“ a „zkoušení“ používá pro „kvalitativní analýzu“. Mezinárodní norma pro akreditaci zdravotnických laboratoří používá pojem „vyšetření“ pro kvantitativní i kvalitativní analýzu [14]. Vzhledem k tomu, že mezi různými relevantními mezinárodními společnostmi nebylo dosaženo konsensu ohledně těchto pojmů, používá tato příručka pro stanovení nominálních (kvalitativních) vlastností v chemické analýze termín „kvalitativní analýza“.

Obecně řečeno, kvalitativní analytický výsledek je jednoduchý výrok nebo kategorizace zkoušené položky nebo materiálu, tj. klasifikace. Rozhodnutí jsou vždy přijímána na základě kategorizace; například zda propustit šarži hnojiva, zda je voda vhodná k pití, zda osoba vlastní regulovanou látku či nikoli nebo zda má nově syntetizovaná látka správnou strukturu na základě požadavků. Nesprávná klasifikace – jako například „příjetí“ produktu, když je nevhodný k použití – představuje riziko pro všechny strany. Aby byla tato rizika pod kontrolou, odborníci zapojení do analýzy intenzivně pracují, aby zajistili, že jejich postupy vedou k přijatelně nízkým rizikům nesprávné klasifikace.

Z toho vyplývá, že v určité fázi vývoje takového zkušební postupu musí být provedeno hodnocení rizika nesprávné klasifikace. Proto je u většiny takových postupů rozumné očekávat, že laboratoř vytvoří informace o rizicích nesprávných výsledků nebo k nim bude mít přístup. Důležitou výjimkou je použití standardizovaných zkušebních postupů stanovených skupinami mimo laboratoř jako vhodných pro zamýšlený účel [20] [21] [22] [23]. Laboratoř může mít omezený nebo dokonce žádný přístup k údajům o výkonu takových zkušebních postupů. Tyto postupy však vždy specifikují zkoušku s příslušnými podrobnostmi

a od laboratoře se obecně očekává, že prokáže, že relevantní faktory pod její kontrolou skutečně splňují požadavky zkušební postupu. To zase může zahrnovat prokázání, že nejistota kontrolovaných parametrů a výkonnost zkoušky je adekvátní ve vztahu k účelu zkoušky.

Vyhodnocování nejistot spojených s kvantitativními parametry nebo výsledky analýz bylo předmětem značného úsilí od vydání „Pokynu pro vyjadřování nejistoty měření“ (GUM), který je dostupný jako ISO Guide 98 [24] a také jako dokument JCGM [25]. Na druhou stranu nejistotám v kvalitativní analýze je věnována mnohem menší pozornost. Po zveřejnění prvního vydání ISO/IEC 17025 [26] vzrostl zájem o nejistoty kvalitativní analýzy. Výzvy při stanovení nejistoty spojené s kvalitativní analýzou, jako je „vyhověl/neyhověl“, analýzy identity nebo srovnávací analýzy identity proto získaly více pozornosti, zejména v oblastech, kde je dopad falešných výsledků kvalitativní analýzy extrémně důležitý, např. ve forenzní nebo dopingové analýze.

Existuje široká škála metrik pro vyjádření nejistoty v kvalitativních výsledcích [7]. Existuje však omezená shoda ohledně toho, které metriky použít. Výjimkou jsou oblasti epidemiologie a klinické laboratoře, kde se jako parametry klinické přesnosti důsledně používají pojmy „klinická citlivost (senzitivita)“ a „klinická specifická“ [27].

Kvantitativní a kvalitativní analýzy se podstatně liší v tom, jak jsou výsledky a související nejistoty uváděny. Zatímco kvantitativní výsledky jsou uváděny jako interval, který zahrnuje „skutečnou hodnotu“ měřené veličiny s definovanou konfidenční úrovní, nominální vlastnosti jsou uváděny jako klasifikace s metrikami, které vyjadřují možnost správné nebo nesprávné klasifikace. Tato „možnost“ může být popsána pravděpodobností (v angličtině probability, likelihood, odds) nebo jinými metrikami odhadovanými z interpretace vstupních informací. Kvalita vykazovaných metrik závisí na počtu a rozmanitosti studovaných případů. Stanovení těchto metrik umožňuje identifikaci případů, kdy by měly být postupy vylepšeny, aby se snížila pravděpodobnost chybných výsledků.

Příručka popisuje obecné principy pro hodnocení výkonnosti kvalitativní analýzy pro vykazování nejistoty kvalitativních analytických výsledků a uvádí příklady použití popsané teorie. Příručka nepojednává o schopnosti položky zkoušení reprezentovat skupinu identických položek nebo větší předmět; to znamená, že nepojednává o dopadu odběru vzorků v těchto hodnoceních.

Pořadové výsledky mohou být redukovány na binární (ano/ne) výsledky a zpracovány pomocí metod uvedených v této příručce tak, že se výsledky pořadové klasifikace přiřadí jako „správné“ nebo „nesprávné“. Jiné metody zpracování pořadových škál jsou mimo rozsah této příručky.

2 Typy kvalitativní analýzy

Jak je zmíněno v účelu této příručky „kvalitativní analýza“³ je definována jako „klasifikace podle specifikovaných kritérií“ [28]. Tabulka 1 uvádí některé příklady. Ačkoli se všechny tyto případy jeví jako velmi odlišné, sdílejí jednu společnou vlastnost; jakmile jsou kritéria specifikována, výkonnost klasifikační metodiky se dá relativně jednoduše popsat z hlediska měř její úspěšnosti nebo neúspěšnosti. Tyto míry úspěšnosti a neúspěšnosti tvoří základ většiny výkonnostních metrik.

Kvalitativní analýzy obsažené v hlavním textu jsou rozděleny do dvou kategorií na základě různých typů klasifikačních kritérií, tj. kvalitativních nebo kvantitativních. Tabulka 1 uvádí příklady každého z nich. Oddíl 3 popisuje strategie hodnocení výkonnosti pro různé typy klasifikačních kritérií.

U kvalitativní analýzy, kde míra pravdivé nebo chybné odpovědi závisí na kvantitativní vlastnosti, jako je přítomnost zakázané látky, jejíž detekce závisí na přítomném množství, je rovněž zvažována mez detekce (viz oddíl 3.4).

Posouzení shody hodnoty kvantitativní vlastnosti položky s limitní hodnotou nebo intervalem lze někdy považovat za převod výsledku měření na výsledek kvalitativní („shodný“ nebo „neshodný“). Použití naměřených hodnot a jejich nejistoty měření pro posuzování shody je podrobně popsáno v jiné příručce Eurachem/CITAC [29], a proto se tím tato příručka podrobně nezabývá. Příloha B však pojednává o tom, jak lze určit některé metriky používané k posouzení výkonnosti nebo nejistoty kvalitativní analýzy pro kvantitativní posouzení shody.

Tabulka 1. Typy kvalitativní analýzy založené na různých typech klasifikačních kritérií.

Klasifikační kritérium	Příklad kvalitativní analýzy
Kvalitativní	1 Detekce alifatických aldehydů v roztoku změnou barvy po přidání Schiffova činidla.
	2 Identifikace krystalické formy materiálu pozorováním.
	3 Identifikace značky a ročníku vína senzorickou analýzou.
	4 Identifikace biologického druhu určením nebo detekcí konkrétní sekvence DNA.
	5 Identifikace lidské krevní skupiny pozorováním aglutinace.
Kvantitativní	1 Identifikace rezidua pesticidu v ovoci pomocí naměřených hmotností fragmentů a relativních zastoupení fragmentů v GC-MS.
	2 Stanovení infračervené spektrální ekvivalence mezi novou a dříve přijatou průmyslovou surovinou pomocí kritérií vlnové délky a intenzity.
	3 Identifikace diuretika v moči od sportovce pomocí retenčního času a naměřených hmotností fragmentů v GC-MS.
	4 Identifikace léku v krvi pomocí retenčního času a naměřených hmotností fragmentů v LC-MS.
	5 Detekce viru v klinickém vzorku na základě intenzity fluorescence v kvantitativní real-time polymerázové řetězové reakci (qPCR).

³ „Kvalitativní zkoušení“ je v zásadě širší obor než „kvalitativní analýza“, jednoduše proto, že chemická analýza, často označovaná jako analytická práce, je jednou konkrétní činností z mnoha oblastí zkoušení. V této příručce pro analytické chemiky a odborníky z příbuzných oborů se však pojmy používají jako synonyma.

3 Hodnocení výkonnosti kvalitativní analýzy

3.1 Obecné úvahy

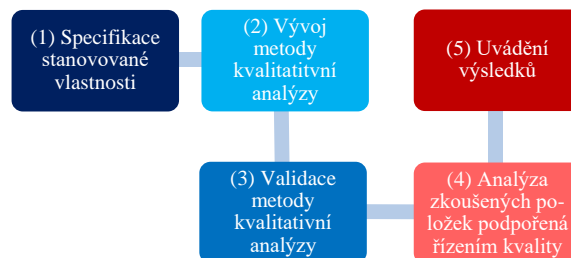
Tento oddíl poskytuje návod k posouzení a vyjádření výkonnosti pro postupy určené k jednoduché klasifikaci do dvou tříd („binární klasifikace“). „Třídy“ jsou zde označeny jako „pozitivní“ nebo „negativní“ k označení „člena zájmové třídy“ nebo „nečlena“. Tato klasifikace pokrývá většinu praktických situací, včetně „nad limit“, „přijatelné“, „nepřijatelné“, „identita jako“ nebo „přítomnost určitého druhu“.

Předpokládá se, že třídy jsou ucelené a exkluzivní, aby bylo možné vypočítat jednoznačné poměry falešných odpovědí. To znamená, že žádná položka zkoušení nesmí být klasifikována jako člen třetí třídy. Toho lze obecně dosáhnout pečlivou specifikací klasifikačních kritérií. Je však možné, že výsledek nemusí poskytovat dostatečnou spolehlivost klasifikace. Za těchto okolností je zcela rozumné, aby analytik uvedl výsledek zkoušky (testu) jako „neprůkazný“ ve smyslu nedostatečně jistý. Neprůkazné výsledky vyžadují další studii, aby byly výsledky uvedeny jako „průkazné“. Tyto výsledky jsou ve zdravotnických laboratořích známy jako „šedá zóna“ nebo „nejednoznačná zóna“ (hraniční výsledek).

Některé z popsanych pojmů lze v zásadě rozšířit na více tříd, např. v klasifikaci na pořadové stupnici, a to posouzením správných a nesprávných poměrů klasifikace pro všechny třídy. Užitečným rozšířením je považovat identifikaci struktury nebo identity (formálně vícetřídní problém) buď za „správnou“ nebo „nesprávnou“, a tento přístup se zde předpokládá. Podrobné zpracování problému s více třídami, který může zahrnovat více současných úkolů nebo přiřazení několika třídám, však přesahuje rámec této příručky.

Kvalitativní analýza zahrnuje různé fáze, jmenovitě (1) popis problému, (2) vývoj metody a (3) validaci, (4) zkoušky na neznámých položkách kontrolovaných řízením kvality a (5) uvádění výsledků (obrázek 1). Rozhodující je jednoznačná specifikace vlastnosti, která má být stanovena, a posouzení vhodnosti analýzy pro zamýšlené použití. Uvádění kvalitativního analytického výsledku musí být podpořeno platnými postupy a odpovídajícím řízením kvality zkoušky. Způsob uvádění výsledků závisí na účelu analýzy

a příjemci zprávy. Tato příručka podrobně neuvádí, jak by měla být metoda vyvinuta nebo jak by mělo být navrženo řízení kvality.



Obrázek 1. Proces kvalitativní analýzy od popisu problému až po uvádění výsledků.

3.2 Kvantifikace výkonu kvalitativní analýzy

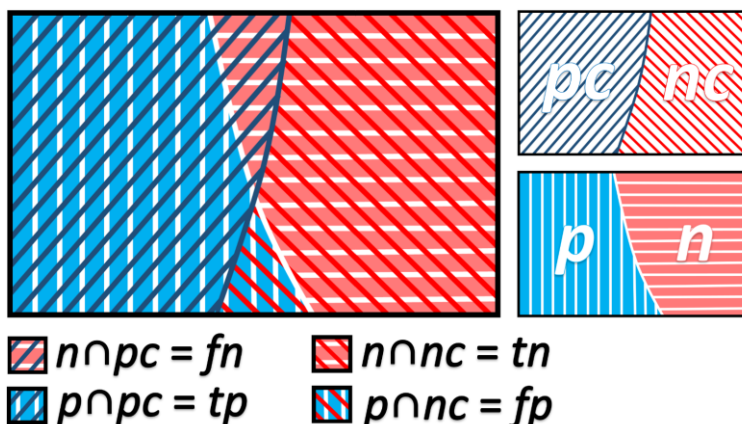
3.2.1 Definování základu pro hodnocení výkonnosti

Základním způsobem kvantifikace výkonnosti metody kvalitativní analýzy je výpočet četnosti chybných výsledků. U „pozitivních“ nebo „negativních“ výsledků je užitečné hlásit poměr „skutečně pozitivní“ a „falešně pozitivní“ nebo „skutečně negativní“ a „falešně negativní“. Tyto poměry však mohou být vztaženy buď k celkovému počtu konkrétního typu případu nebo výsledku, nebo k celkovému počtu možných případů nebo výsledků.

Například falešně pozitivní poměr může být definován jako:

- i) Podíl negativních případů, které jsou falešně hlášeny jako pozitivní (fp/nc), kde fp a nc jsou počty falešně pozitivních výsledků a negativních případů. Obrázek 2 graficky znázorňuje překrývání různých typů případů a výsledků. Zlomek (fp/nc) je reprezentován poměrem oblastí průniku (\cap) pozitivních výsledků, p , s , nc ($p \cap nc = fp$) a oblasti 'nc'. Toto stanovení uvádí FP z tabulky 2.
- ii) Podíl pozitivních výsledků, které jsou falešně vykázané jako pozitivní (fp/p), kde p je počet pozitivních výsledků. Na obrázku 2

Obrázek 2. Grafické znázornění příkladu překrytí počtu kladných pc nebo záporných nc případů s počtem kladných p nebo záporných n výsledků. Symbol „ \cap “ představuje průnik skupin; například $n \cap pc$, zde označuje sadu negativních výsledků z pozitivních případů, $n \cap pc$, $p \cap pc$, $n \cap nc$ a $p \cap nc$ definují fn , tp , tn a fp .



je tento poměr reprezentován poměrem mezi plochami ($p \cap nc = fp$) a ' p '.

- iii) Podíl z celkového počtu případů nebo výsledků, které jsou falešně hlášeny jako pozitivní ($fp/(pc + nc) = fp/(p + n)$), kde pc a n představují počet pozitivních případů a negativních výsledků. Na obrázku 2 je tato míra reprezentována poměrem mezi označenou plochou ($p \cap nc = fp$) a celkovou plochou obrázku.

Rozdíl mezi těmito definicemi je zásadní. V případě i) se poměr neliší s podílem „negativních případů“ v populaci, tj. $nc/(nc + pc)$, protože $FP = fp/nc$. Avšak pro případy ii) a iii) míra falešně pozitivních závisí na $nc/(nc + pc)$, protože je pozorováno více fp v populaci obsahující více nc .

Proto tyto definice charakterizují provádění kvalitativní analýzy různými způsoby, a proto zahrnují různé interpretace jejich hodnot.

Skutečně pozitivní TP (tp/pc) a skutečně negativní TN (tn/nc) poměry odkazují na relevantní počet případů je v klinické chemii znám jako „citlivost (senzitivita)“ a „specifičnost“ kvalitativní analýzy [7] (tabulka 2). Stanovení klinické senzitivity a specificity vyžaduje správné stanovení studovaných případů průkaznou klinickou diagnózou. Pro kvantitativní analýzu má termín „citlivost“ [1] nebo „analytická citlivost“ [30] jiný význam.⁴

Skutečně pozitivní poměr vztažený k pozitivním případům (tp/p) je také znám jako „preciznost“ nebo „pozitivní prediktivní hodnota“ kvalitativní

Tabulka 2. Alternativní výkonnostní charakteristiky pro vyjádření kvality kvalitativních analytických výsledků.

Výkonnostní charakteristiky	Výraz
Skutečně pozitivní poměr, TP (citlivost, SS)	$tp/pc = tp/(tp + fn) = 1 - FN$
Poměr falešně pozitivivity, FP	$fp/nc = fp/(tn + fp) = 1 - TN$
Skutečně negativní poměr, TN (specifičnost, SP)	$tn/nc = tn/(tn + fp) = 1 - FP$
Falešně negativní poměr, FN	$fn/pc = fn/(tp + fn) = 1 - TP$
'Preciznost' nebo 'Pozitivní prediktivní hodnota', PPV	$tp/p = tp/(tp + fp)$
Negativní prediktivní hodnota	$tn/n = tn/(tn + fn)$
Účinnost	$(tp + tn)/(p + n)$
Youdenův index	$SS(\%) + SP(\%) - 100$
Míra pravděpodobnosti pozitivních výsledků	TP/FP
Míra pravděpodobnosti negativních výsledků	TN/FN
Posteriorní pravděpodobnost	Viz příloha A

tp – počet skutečně pozitivních výsledků; fp – počet falešně pozitivních výsledků; tn – počet skutečně negativních výsledků; fn – počet falešně negativních výsledků; p – počet pozitivních výsledků ($tp+fp$); n – počet negativních výsledků ($tn+fn$); ks – počet pozitivních případů a nc – počet negativních případů.

⁴ Podle Mezinárodního metrologického slovníku „citlivost měřicího systému“ je „podíl změny indikace měřicího systému a odpovídající změny hodnoty veličiny, která má být měřena“ [1].

analýzy, PPV [30]. Termín „negativní prediktivní hodnota“ NPV , se používá pro skutečně negativní poměr vztažený k celkovému počtu negativních výsledků (tj. tn/n). Účinnost kvalitativní analýzy je definována jako podíl jakéhokoli typu správných výsledků daných všemi výsledky (tj. $(tp+tn)/(p+n)$). Youdenův index je alternativním způsobem kvantifikace úspěšnosti kvalitativní analýzy (Tabulka 2) [31].

Ačkoli metriky odkazující na počet pozitivních nebo negativních případů nezávisí na prevalenci typů případů, tato čísla samotná nemohou poskytnout pravděpodobnost, že konkrétní výsledek je správný. K odhadu pravděpodobnosti, že výsledek bude správný, je třeba vzít v úvahu také relevantní míru výsledků a prevalenci případů. Tato a další metriky pro spolehlivost kvalitativních výsledků jsou popsány v oddíle 4.

3.2.2 Definování reference hodnocení výkonnosti

Metriky používané ke kvantifikaci výkonnosti kvalitativní analýzy mohou mít další zvláštnosti. Pozitivní a negativní případy mohou být stanoveny různými způsoby. O některých případech nebo vzorcích použitých jako reference může být známo, že jsou „pozitivní“ pro charakteristiku kvůli jejich původu nebo formulaci. Jiné mohou být, jak je definuje AOAC International, případy, kdy výsledky „konfirmační techniky i jiné analytické techniky jsou oba pozitivní“ [32]. Některé příklady adekvátního původu pozitivních případů mohou být pacienti s diagnózou konkrétního onemocnění nebo půda, o které je známo, že je kontaminována. Pozitivní zkoušené položky mohou být připraveny přidáním specií, které mají být identifikovány, do matrice ekvivalentní analyzovaným položkám, jako je pesticid v potravinářském produktu, u kterého je potvrzeno nebo není potvrzeno, že má přirozené hladiny pesticidu. Pokud se identifikační výkonnost výrazně liší s kvantitativní vlastností (např. koncentrace látky, která má být identifikována nebo detekována), formulace (příprava) by měla umožňovat stanovení této

úrovně. Negativní případ lze také interpretovat jako případ, o kterém je známo, že je negativní svým původem, formulací nebo je definován jako negativní, protože „konfirmační technika i jiná analytická technika jsou obě negativní“. Mezinárodní definice pozitivních a negativních případů od AOAC mají komplexnější aplikaci, protože je to jediný přístup použitelný pro analýzu složitých položek, které je obtížné reprodukovat z formulace. Spoléhá se však na kvalitu výstupu použitých analytických technik. V některých oborech je obtížné uměle připravit položky se studovaným analytem a možnými interferencemi pro testování výkonnosti analýzy, protože matrice položek jsou neznámé a nepředvídatelné.

Pozitivní a negativní případy mohou být také poskytnuty jako referenční data, jako jsou spektra, o nichž je známo, že pocházejí ze specifické sloučeniny. Po definování identifikačních kritérií lze určit pravděpodobnost nahlášení správné nebo nesprávné shody složení ve srovnání s těmito kritérii. Například v hmotnostní spektrometrii může být identifikace založena na hodnocení přítomnosti nebo přítomnosti a zastoupení charakteristických iontů. Šance na spektroskopickou shodu lze předpovědět binomickou nebo hypergeometrickou statistikou, jak je diskutováno v příkladech E1 a E2.

3.2.3 Uvádění výkonnosti metody

3.2.3.1 Kontingenční tabulky

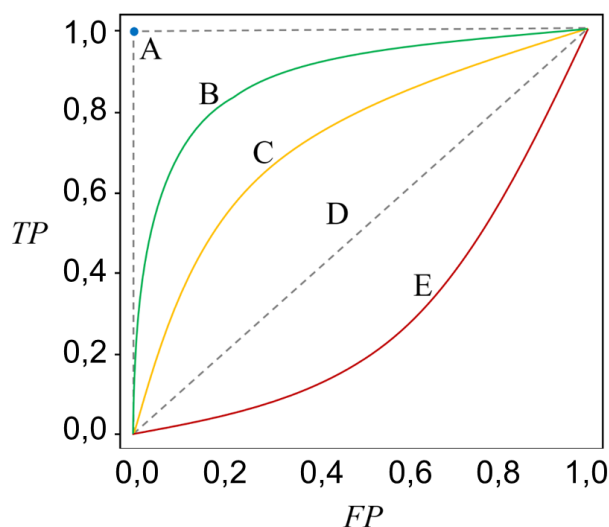
Velmi jednoduchý způsob uvádění výkonnosti kvalitativní metody analýzy, která se významně neliší v rámci analytického rozsahu, je pomocí kontingenční tabulky. Tabulka 3 představuje příklad takové tabulky. V tomto příkladu TP , FP , TN a FN jsou 97,8 % (228/233), 0,33 % (1/301), 99,7 % (300/301) a 2,1 % (5/233), v uvedeném pořadí.

Typicky může analytický rozsah zahrnovat různé úrovně studovaného druhu nebo vlastnosti a různé matrice analyzované položky. To může vyžadovat samostatné kontingenční tabulky pro různé oblasti analytického rozsahu.

Tabulka 3. Konkrétní příklad kontingenční tabulky, která popisuje výkonnost kvalitativní analytické metody, která by měla být v rámci analytického rozsahu přibližně konstantní.

		Případ		
		Pozitivní (pc)	Negativní (nc)	Celkové výsledky
Výsledek	Pozitivní (pc)	$tp = 228$	$fp = 1$	$p = 229$
	Negativní (nc)	$fn = 5$	$tn = 300$	$n = 305$
Celkový počet případů		$ks = 233$	$nc = 301$	

Obrázek 3. Pět příkladů ROC křivek, kde je znázorněna změna TP a FP se změnou kvantitativních identifikačních kritérií. Křivka A (modrý bod) představuje dokonalý test, kde identifikační kritéria neovlivňují výsledky, a TP a FP jsou 100 % a 0 %, v tomto pořadí. Křivka B a C představují vhodné metody, kde $TP \geq FP$. A z těchto tří je metoda B výhodnější než metoda C. Křivka D představuje úhlopříčku, kde $TP = FP$ pro všechny prahy rozhodování; to by nebyl užitečný klasifikátor. Křivka E se na první pohled zdá být velmi špatným klasifikátorem a trvale vykazuje četnost falešně pozitivních výsledků vyšší než skutečnou míru pozitivních výsledků. Avšak jednoduchým přepnutím hlášeného výsledku by vznikla křivka ROC blízko křivky C; klasifikátor se pak může ukázat jako užitečný.



3.2.3.2 Křivka operační charakteristiky přijímače (ROC)

Pro kvalitativní analýzu založenou na posouzení kvantitativní charakteristiky lze výběr klasifikačních kritérií, která vyvažují poměr pravdivých a nepravdivých výsledků, obvykle TP a FP , provést pomocí křivek operační charakteristiky přijímače (ROC), které vykreslují pár (TP, FP) se klasifikačním kritériem (tj. diskriminační práh) liší. Tyto křivky lze také použít k porovnání různých postupů kvalitativní analýzy [31]. Přestože podrobný popis těchto křivek přesahuje rámec tohoto průvodce, obrázek 3 představuje pět schematických příkladů křivek ROC. Každá křivka ukazuje, jak se mění skutečný pozitivní a falešně pozitivní poměr, jak se identifikační kritérium mění od přísnější k méně přísné identifikaci pozitivních případů spojených s nízkou nebo vysokou TP . Kromě poskytnutí vizuální ilustrace výkonu lze oblast pod křivkou (často zkráceně „AUC“, area under the curve) použít jako souhrn výkonnosti klasifikátoru [31].

3.3 Hodnocení falešně pozitivních a falešně negativních poměrů

3.3.1 Rozsah metody a podrobnosti validace

Validace metody kvalitativní analýzy zahrnuje definování požadavků na výkonnost a také kontrolu, zda jsou splněny [30].

Před tímto posouzením výkonnosti by měl být jasně definován rozsah analýzy z hlediska typu klasifikace (např. přítomnost pentachlorofenolu nad

1 mg kg⁻¹) a analyzovaných položek (např. kožené výrobky). Měla by být také specifikována metoda klasifikace, jmenovitě analytická technika (např. GC-MS/MS), způsob použití této techniky (např. příprava vzorku a instrumentální podmínky) a klasifikační kritéria. Kritéria klasifikace musí být jasně popsána, aby bylo zaručeno, že shromážděné údaje o výkonnosti budou použity pro následné analýzy.

V některých kvalitativních analýzách, řízených úvahami o účinnosti, je analytická metoda rozdělena do dvou fází: předběžná rychlejší a levnější screeningová metoda, po které, kdykoli je to požadováno, následuje časově náročnější a nákladnější konfirmační metoda. Potvrzení se provádí, když první zhodnocení přinese výsledky, které jsou v rozporu s očekávanými výsledky nebo mohou mít relevantní dopad na individuální nebo kolektivní zájem. Je však nezbytné vyhodnotit poměr falešně negativních a pozitivních výsledků pro celý postup, který zahrnuje screeningové a konfirmační zkoušky. Pokud například podléhají potvrzení pouze pozitivní výsledky, je nezbytné zkontrolovat, zda je poměr falešně negativních výsledků ve fázi screeningu přiměřeně nízký.

Pokud jde o podrobnosti o validaci metody, u metod použitelných na různé položky (např. různé potravinářské produkty) výkonnost by měla být testována pro reprezentativní soubor typů položek. Typy a počet testovaných položek závisí na vlivu analyzované matrice na výkonnost. V některých případech může pochopení principů klasifikace umožnit předvídat skupiny položek souvisejících s ekvivalentní výkonností kvalitativní analýzy, z nichž lze vybrat a studovat

reprezentativní položku. Výkonnost analýzy reprezentativní položky lze poté extrapolovat na skupinu položek souvisejících s ekvivalentní výkonností kvalitativní analýzy. Pokud to výkonnost klasifikační techniky umožňuje, lze se rozhodnout prostudovat výkonnost analýzy položek a/nebo hodnot vlastností, kde míry falešných výsledků dosahují nejvyšších hodnot. Laboratoř by měla kontrolovat důkladnost hodnocení výkonnosti a zároveň mít na paměti dostupný čas a zdroje pro toto hodnocení. V některých případech může být přijatelné provádět strategii průběžné validace, kdy pokaždé, když je zkoušena položka, která je v laboratoři nová, je prováděna dodatečná a specifická kontrola kvality analýzy.

3.3.2 Použití informací z literatury

U běžně používaných postupů kvalitativní analýzy lze očekávat, že informace o výkonnosti budou veřejně dostupné. Než se pustíte do zkoumání výkonnosti dobře zavedeného analytického postupu, měla by být provedena vhodná rešerše příslušné literatury, aby se shromáždily nezávislé informace o jeho vhodnosti pro zamýšlené použití. Publikované poměry falešných odpovědí by však měly být používány s opatrností; mohly být získány pomocí specifického vybavení, činidel a personálu a odkazovaly by na specifické matrice vzorků a charakteristické úrovně, takže analytik musí zvážit, zda je jeho situace odpovídající. Pokud například položky studované v literatuře mají charakteristické úrovně daleko od prahových hodnot používaných k rozlišení mezi třídami a pokud jsou jejich matice relativně bez interferencí, může být zjištěná identifikace příliš optimistická ve srovnání se „skutečnými“ analytickými problémy, se kterými se laboratoř setkává. Proto poměr pravdivých a nepravdivých výsledků silně závisí na dostupných datech.

V některých případech je možné předvídat, zda výkonnost pozorovaná v literatuře bude lepší nebo horší než výkonnost pozorovaná při kvalitativní analýze v laboratoři. Pokud se dojde k závěru, že

postup kvalitativní analýzy je platný pro nejhorší scénáře, tj. může poskytnout výsledky vhodné pro zamýšlený účel, lze tento postup použít k analýze neznámých položek bez omezení.

Oddíl 4 pojednává o tom, jak lze nastavit kritéria pro rozhodování, zda je analýza vhodná pro zamýšlené použití.

3.3.3 Hodnocení výhradně z experimentu

Bez ohledu na typ kvalitativní analýzy uvedené v oddíle 2 lze *FP* a *FN* odhadnout přímo z počtu falešných výsledků ze sady analýz. V kvalitativní analýze založené výhradně na kvalitativních vstupech (tabulka 1) je to jediný způsob, jak odhadnout nejistotu kvalitativní analýzy. Pokud jsou však falešné odpovědi nepravděpodobné, vyžaduje tento přístup velké množství testů (zkoušek).

Vzhledem k tomu, že počet falešných odpovědí by měl být v ideálním případě nízký, vyvstává problém, kolik vzorků otestovat, abychom si byli přiměřeně jisti, že nalezneme nenulový počet falešných odpovědí.

Z publikovaných informací (viz např. Ferrara et al. [33]) je zřejmé, že poměr falešně pozitivních nebo negativních výsledků může být jen 0,5 % a v některých případech dokonce nižší [6, 8, 9]. Pro rozsah pravděpodobnosti falešných výsledků ukazuje tabulka 4 počet vzorků, které by bylo třeba analyzovat, aby bylo jisté, alespoň v rámci uvedených konfidenčních úrovní, že nalezneme jeden nebo více falešných výsledků. Tabulka 4 používá binomické rozdělení a ukazuje, že pro 95% pravděpodobnost detekce jednoho nebo více falešných výsledků bude počet testů, které je třeba provést, třikrát vyšší než počet testů, které poskytnou průměrně jeden falešný výsledek. Například u metody s 1% poměrem falešně pozitivních výsledků je zjištěno, že (v průměru) na každých 100 analýz negativních případů je pozorován jeden pozitivní výsledek. Abychom si však byli „na 95 % jisti“, že je pozorován falešně

Tabulka 4. Minimální počet analýz k nalezení jednoho nebo více falešných (pozitivních nebo negativních) výsledku(ů).

Poměr chybných výsledků	Konfidenční úroveň	
	95 %	99 %
0,5 %	598	919
1 %	299	459
5 %	59	90

pozitivní výsledek, je třeba provést 299 (asi 3×100) testů na negativní případy.

Hodnoty v tabulce 4 nejsou dostatečné pro dobrý odhad četnosti chybných výsledků nebo pro srovnání různých metod. I přibližný odhad by obvykle vyžadoval pěti až desetinásobek minimálního počtu pozorování uvedeného v tabulce 4. Tuto tabulku lze také interpretovat jako minimální počet analýz nezbytných ke kontrole shody s různými přijatelnými poměry chybných výsledků, jak je uvedeno níže.

Analytik se ve snaze přímo určit poměr falešných výsledků pro nový postup z experimentování často potýká s dilematem. Na jedné straně je pro daný postup neznámý hodnocený poměr falešné odezvy, a proto mohou být jakékoli provedené klasifikace nespolehlivé. Na druhou stranu pouhá analýza, dokud se neobjeví první falešná odpověď, nemusí nutně poskytnout pravdivý obraz o poměru falešných odpovědí. K vyřešení tohoto problému se doporučuje, aby analytik předem rozhodl o přípustných úrovních pro dva poměry falešných odpovědí. Pro zvolenou konfidenční úroveň lze binomické rozdělení použít k odhadu počtu experimentů potřebných k nalezení jedné nebo více chybných odpovědí s dostatečnou spolehlivostí. Není zaručeno, že tento přístup poskytne přesné číslo pro poměr falešných odpovědí, ale omezí ho. Předpokládejme například, že analytik rozhodne, že *FP* 5 % je přijatelné, a po provedení 59 experimentů (tabulka 4), pokrývajících pravděpodobný rozsah matric, nebyly nalezeny žádné falešně pozitivní výsledky. V takovém případě lze dojít k závěru, že *FP* není větší než 5 %. Jako opatření řízení kvality validovaného postupu se dále doporučuje, aby vzorky byly proloženy slepými vzorky (tzv. blanky) a referenčními materiály obsahujícími cílovou charakteristiku (např. analyt) na příslušných úrovních charakteristik. Vždy je třeba mít na paměti, že poměr falešných výsledků velmi závisí na rozmanitosti a/nebo specifčnosti vzorkované populace a na strategii vzorkování této populace.

Tabulka 4 ukazuje, že pro nízký poměr falešných odpovědí může být nepraktické analyzovat dostatečný počet vzorků k detekci falešné odpovědi. V souladu s tím, pokud je test levný a/nebo je zamýšlen pro široké použití, např. jako screeningový test na drogy, může být přijatelné nejprve experimentem stanovit, že poměr falešné odpovědi nepřesahuje horní mez, řekněme 5 % a poté tento údaj upřesnit ve světle zkušeností s dalšími vzorky.

Tam, kde je pravděpodobné, že počet vzorků bude relativně nízký a/nebo použití testu je drahé, měly by být všechny testy prováděny souběžně s konfirmační zkouškou a čas od času by se měl přepočítat poměr falešných odpovědí.

K překonání některých omezení experimentálního stanovení poměru chybných odpovědí lze použít matematické zpracování dostupných informací (viz oddíly 3.3.4 a 3.3.5).

3.3.4 Hodnocení z databáze

Alternativou k určení poměru falešných výsledků z experimentování jsou studie náhodného nesouladu v referenčních databázích, jako jsou databáze hmotnostních spekter nebo infračervených spekter. V některých případech to umožňuje ekvivalent mnoha tisíc experimentů. I když jsou tyto databáze informativní a výkonné, současným omezením je, že takové databáze často zcela nereprezentují testovací populaci; například zatímco se výskyt různých materiálů obecně velmi liší, typická referenční databáze bude obsahovat pouze jeden z každého. To může vést k výrazně zkresleným odhadům pravděpodobnosti; opět je nepravděpodobné, že získané hodnoty budou lepší než řádové odhady. Příklady E1 a E2 ilustrují použití této metodologie pro hodnocení výkonnosti kvalitativní analýzy.

3.3.5 Vyhodnocení z kvantitativního datového modelování

Hodnocení výkonnosti vysoce selektivních a časově náročných a/nebo nákladných kvalitativních analýz výhradně z experimentů prováděných v jediné laboratoři není proveditelné.

V kvalitativní analýze založené na kvantitativním klasifikačním kritériu pro kvantitativní výsledky (jako je instrumentální metoda analýzy) lze modely rozptýlení výsledků použít k odhadu poměru pravdivých a nepravdivých výsledků. V příloze B jsou uvedeny další podrobnosti. Pokud je například relevantní instrumentální signál, jako je retenční čas analytu v chromatografické metodě, normálně rozdělen, lze předpovědět pravděpodobnost, že interferující složka bude mít retenční čas v rámci intervalu přijatelného retenčního času pro analyt (viz Rychlý odkaz 1).

Modelování se však opírá o platnost předpokladu modelu a hodnoty vstupních proměnných. Například protože relativní retenční časy nemohou být normálně rozděleny, předpoklad normality může podcenit poměr falešných výsledků. Simulace přístrojových signálů metodou Monte

Rychlý odkaz 1 – Příklad modelování signálu

Pokud je pro identifikaci deltametrinu v olivovém oleji pomocí GC-MS odhadovaná směrodatná odchylka opakovatelnosti retenčního času 0,022 min s $v = 32$ stupňů volnosti, může být tolerance retenčního času pro identifikaci této sloučeniny ve vzorku: $(t_R \pm t \cdot s_{IRI}) = (t_R \pm 2,04 \cdot 0,022) = (t_R \pm 0,045)$ min, kde t_R je retenční čas pozorovaný pro jeden denní nástřik standardního roztoku a t oboustranná 95% kritická hodnota pro t rozdělení s 32 stupni volnosti. Proto pro t_R 36,055 min by interval přijetí pro pík vzorku byl $(36,055 \pm 0,045)$ min. Za předpokladu, že interference má retenční čas o 0,05 min kratší než pro deltametrin a preciznost obou retenčních časů je ekvivalentní, pravděpodobnost, že interference bude mít retenční čas v rámci akceptačního intervalu, bude 1,5 %. Tato hodnota je odhadnuta pomocí kumulativního rozdělení t pro hodnotu $t(-0,05/0,022)$ a v (vzorec MS-Excelu: T.DIST(-0,05/0,022;32;TRUE)).

Carlo je pohodlný způsob odhadu FP a FN z nenormálně rozdělených parametrů [8, 9]. Příklad E5 ilustruje modelování rozptýlení instrumentálního signálu pro odhadování poměru falešných výsledků vysoce selektivních GC-MS/MS identifikací.

3.3.6 Hodnocení výkonnosti kvalitativního testu v závislosti na spojité proměnné

Mnoho konfirmačních nebo detekčních zkoušek ukazuje silnou závislost na pravděpodobnosti detekce nebo poměru falešné odpovědi na nějaké spojité proměnné. Například poměr detekce často závisí na koncentraci nebo počtu částic hledaného materiálu. Pak může být užitečné modelovat závislost poměru falešných odpovědí na spojitých (nebo jiných) proměnných.

Logistická regrese a probitová regrese [34, 35] se běžně používají na takové problémy a byly navrženy (s příklady) pro hodnocení výkonnosti kvalitativních metod analýzy [36]. Logistická regrese byla použita při detekci nízkého počtu kopií DNA [37]. Postup je dobře zdokumentován v učebnicích a dostupný v podstatě ve všech statistických softwarových balíčcích, proto zde není podrobně prezentován. Jednoduchá logistická regrese modeluje pravděpodobnost binární odezvy jako funkci nějaké spojité proměnné. Model je:

$$p = \frac{\exp(b_0 + b_1x)}{1 + \exp(b_0 + b_1x)} \quad (1a)$$

nebo ekvivalentně:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1x, \quad (1b)$$

kde p je pravděpodobnost zájmu (například pravděpodobnost pozitivního výsledku), x spojité proměnná (obvykle koncentrace analytu) a b_0 a b_1 regresní koeficienty. Většina statistických balíčků poskytne metodu prokládání buď z původních dat (páry koncentrace/kvalitativní výsledky) nebo z podílů vypočítaných z počtu výsledků. Všimněte si, že první vyžaduje pouze sekvenci hodnot ano/ne

(nebo 1/0); nevyžaduje proporce. To umožňuje aplikovat metodu na řadu zkušebních vzorků s různými (známými nebo nezávisle naměřenými) koncentracemi, které jsou podrobeny kvalitativnímu zkušebnímu postupu pouze jednou.

Jakmile je vztah stanoven, je možné odhadnout meze detekce (viz níže) z proloženého vztahu mezi koncentrací a pravděpodobností detekce, jednoduše výběrem vhodné meze pro pravděpodobnost detekce, který odpovídá definici používané detekční schopnosti.

Příklad E4 poskytuje praktický příklad logistické regrese.

3.3.7 Expertní posouzení

Pokud nejsou k dispozici žádné údaje o výkonu analytické metody od třetí strany a není možné posoudit výkonnost výhradně z experimentování (oddíl 3.3.3) nebo modelování (oddíl 3.3.4 a 3.3.5), může analytik využít své praktické zkušenosti s klasifikační technikou pro studované nebo podobné položky, aby rozhodl, zda je metoda vhodná pro zamýšlené použití.

Kdykoli je to možné, mělo by být rozhodnutí o vhodnosti metody pro zamýšlené použití podpořeno objektivními důkazy.

Proces formulování expertního posuzování je tématem několika studií. Posouzení jsou ovlivněna mnoha faktory vedoucími k odpovídajícím odhadům nejistoty výsledku [38].

3.4 Mez detekce a selektivita

3.4.1 Mez detekce

Mez detekce (LOD) typicky popisuje nejnížší koncentraci látky, která vede ke spolehlivé detekci. U zkoušek, kde klasifikace zahrnuje posouzení kvantitativní charakteristiky a hodnota této charakteristiky ovlivňuje kvalitativní výsledky, „mez detekce“ (LOD) a/nebo „mez

stanovitelnosti“ (LOQ) uvažovaná v kvalitativní a /nebo kvantitativní analýze by měla být kontrolována ve vztahu k výkonosti kvalitativní analýzy [30]. Výsledek kvalitativní analýzy by měl být vhodný pro zamýšlené použití na dané úrovni (úrovních).

POZNÁMKA: Nařízení Komise (EU) č. 589/2014 [39] a č. 152/2009 [40] definují LOQ jako „nejnižší obsah analytu, který lze měřit s přiměřenou statistickou jistotou, splňující identifikační kritéria popsaná v mezinárodně uznávaných normách“ [41].

Pro výhradně kvalitativní analýzu lze LOD určit aplikací postupu na položky obsahující postupně stále menší úrovně charakteristiky, dokud pravděpodobnost produkování falešných výsledků nedosáhne předem stanoveného kritéria. V tomto typu hodnocení lze také použít logistickou a probitovou regresi (oddíl 3.3.4).

3.4.2 Selektivita

Selektivita ve smyslu, ve kterém se tento termín obvykle používá v analytické chemii, se vztahuje k „rozsahu, v jakém lze konkrétní metodu použít ke stanovení analytů za daných podmínek v přítomnosti dalších složek podobného chování“ [42]. Mezinárodní metrologický slovník (VIM) definuje tento termín ekvivalentně jako vlastnost měřicího systému [1].

POZNÁMKA: Termín „specifičnost“ se v kontextu kvantitativní analýzy používá pro dokonale selektivní analýzu [42, 43], kterou lze v chemii potvrdit jen velmi zřídka. Existuje však jasné alternativní použití termínu „specifita“ v kontextu kvalitativní analýzy (viz tabulka 2). V této příručce se termín „selektivita“ používá v obecném smyslu a termín „specifita“ je vyhrazen pro účely uvedené v tabulce 2.

Selektivita může být hodnocena analýzou jedné nebo více zkušebních položek se známými nebo pravděpodobnými rušivými charakteristikami, tj. charakteristikami, které nejsou cílem analýzy, ale lze je považovat za pravděpodobné, že vyvolají testovací odezvu.

Někdy je možné identifikovat rušivé složky nebo možnosti, u kterých je obzvláště pravděpodobné, že generují falešně pozitivní výsledky. Například u zkoušek na amoniak lze rozumně očekávat, že budou reagovat na primární aminy, u testů na specifické bakteriální kmeny lze očekávat, že budou reagovat na jakékoli bakterie stejného obecného druhu.

Pokud kvalitativní analýza funguje relativně dobře i u nejtěžších případů, lze dojít k závěru, že postup je validní pro všechny typy položek.

Ačkoli lze míru falešné odezvy měřit pro každý jednotlivý materiál nebo každý přítomný rušivý prvek, je nepravděpodobné, že by studie selektivity poskytly jedinou definitivní hodnotu selektivity. Je tomu tak proto, že odezva závisí na potenciálních zkříženě reagujících složkách zahrnutých do studie a na úrovni těchto složek. Studie selektivity se proto nejlépe považují za studii, která poskytuje širokou indikaci adekvátnosti metody kvalitativní analýzy, pokud čelí různým výzvám.

4 Vyjádření spolehlivosti v kvalitativní analýze

4.1 Obecné úvahy

Zatímco prohlášení o nejistotě měření v kvantitativní analýze obvykle vedou k rozsahu hodnot, jako je interval rozšířené nejistoty nebo minimální čistota, prohlášení o klasifikaci obvykle nelze spojovat s rozsahem. Obecně nelze uvést, že materiál z 90 % „vyhovuje“, že analyt je na 99 % přítomen nebo že chemická látka je v nějaké souvislé sekvenci. Místo toho má typická forma informace o nejistotě pravděpodobnostní povahu. To znamená, že se udává pravděpodobnost, že daná klasifikace je správná, nebo typické pravděpodobnosti chybné klasifikace položek, jejichž správná třída je známa.

Hodnoty výkonnosti, které lze získat z validačních studií, lze uvést s výsledkem kvalitativního testu. Obecně však zřídka poskytují přímou informaci o (například) pravděpodobnosti, že je kvalitativní výsledek správný. V této části jsou popsány dvě metriky, které byly pro tento účel navrženy s cílem pomoci pochopit a zlepšit stav (v aj. *state of the art*) při vyjadřování nejistoty pro výsledky kvalitativní analýzy. Zde uvedené metriky používají varianty Bayesova pravidla [4] (viz příloha A). Ty mohou být použity k tomu, aby a) naznačily sílu důkazů, které poskytuje jeden nebo více kvalitativních výsledků ve prospěch jedné možné klasifikace ve srovnání s jinou; b) ve spojení se spolehlivými

informacemi o pravděpodobnostech výskytu různých (pravdivých) hodnot kvalitativních charakteristik v populaci naznačily pravděpodobnost, že určitá klasifikace je pravdivá pro určitý výsledek kvalitativní analýzy.

4.2 Věrohodnostní poměr

Nejznámější a nejrozšířenější formou vykazování výkonnosti kvalitativní analýzy je poměr falešných výsledků, zejména *FP* a *FN* nebo jejich komplementární poměry *TN* a *TP* (např. $TN = 1 - FP$). Dva z těchto poměrů lze pohodlně kombinovat do stejné výkonnostní charakteristiky: věrohodnostního poměru, *LR*.

Pokud je nahlášen pozitivní výsledek, *LR(+)* se odhadne pomocí rovnice (2):

$$LR(+) = TP/FP \quad (2)$$

LR(+) je poměr dvou pravděpodobností; pravděpodobnost nahlášení pozitivního výsledku, pokud je případ pozitivní, dělená pravděpodobností nahlášení pozitivního výsledku, pokud je případ negativní. Obecně platí, že věrohodnostní poměr udává míru změny v pravděpodobnosti, že vzorek je skutečně pozitivní po seznámení s pozitivním výsledkem testu. Matematicky je věrohodnostní poměr změnou pravděpodobnosti vyjádřenou jako

Rychlý odkaz 2 – Interpretace věrohodnostního poměru

Pokud je nahlášen pozitivní výsledek, pravděpodobnost, že případ bude ve skutečnosti pozitivní, *PP*, se vypočítá podle rovnice (Q2.1) (viz níže). Tato rovnice je dobře známá Bayesova věta (příloha A), která nahrazuje pravdivé a nepravdivé pozitivní poměry podmíněnými pravděpodobnostmi.

$$PP = \frac{P(+)TP}{P(+)TP + P(-)FP} \quad (Q2.1)$$

kde *P(+)* je pravděpodobnost, že případ bude pozitivní před zkouškou. To lze také vyjádřit ve formě „pravděpodobnosti“ (viz příloha A):

$$\frac{PP}{1 - PP} = \frac{P(+)TP}{P(-)FP} \quad (Q2.2)$$

V Q2.2 představuje poměr $P(+)/P(-)$ pravděpodobnost pozitivního případu před aplikací kvalitativního testu; tedy „apriorní pravděpodobnost“. Poměr TP/FP je vypočítaný věrohodnostní poměr *LR(+)*.

Pravděpodobnostní poměr tedy popisuje, jak se pravděpodobnost (reprezentovaná pravděpodobností) změní po pozitivním výsledku testu; lze jej považovat za měřítko doplňkových informací poskytnutých testem.

Ve speciálním případě, kdy $P(+)=P(-)=0,5$, takže apriorní šance je 1,0; *LR(+)* pak představuje poměr posteriorní pravděpodobnosti, že případ bude pozitivní nebo negativní. Například při stejné prevalenci (nebo předpokládané prevalenci) pozitivních a negativních případů by pozitivní výsledek s *LR(+)* 1000 znamenal, že posteriorní pravděpodobnost, že případ bude skutečně pozitivní, je 1000krát větší než pravděpodobnost, že případ bude negativní.

„šance“ (viz příloha A). Vysoký věrohodnostní poměr z testu naznačuje, že zkoušená položka je pravděpodobněji pozitivní, než by se dalo říci před provedením testu. Někdy je to interpretováno jako „váha důkazů“, k níž přispívá pozitivní výsledek testu ve prospěch toho, že zkoušená položka je skutečně pozitivní.

Ve zvláštním případě, kdy jsou pozitivní i negativní případy stejně pravděpodobné ($P(+)=P(-)=0,5$; kde $P(+)$ a $P(-)$ jsou prevalence pozitivních a negativních případů, v tomto pořadí), lze $LR(+)$ chápat tak, že ukazuje, o kolik více bude hlášený pozitivní výsledek pravděpodobně spíše pravdivý než nepravdivý (viz Rychlý odkaz 2).

Pokud jsou například pozitivní a negativní případy před testem považovány za stejně pravděpodobné, pozitivní výsledek spojený s $LR(+)$ ze 7300 znamená, že pozitivní výsledek je 7300krát pravděpodobnější a tudíž bude spíše pravdivý než nepravdivý.

Pokud je hlášen negativní výsledek, $LR(-)$ je:

$$LR(-) = \frac{TN}{FN} \quad (3)$$

Případ stejně pravděpodobných pozitivních nebo negativních případů před vyšetřením $LR(-)$ představuje, o kolik je pravděpodobnější, že negativní výsledek bude pravdivý než nepravdivý.

Někteří autoři kombinují oba věrohodnostní poměry v parametru „diagnostický poměr šancí“, DOR (Diagnostic odds ratio) ($DOR = LR(+)/LR(-)$) [30].

Jednou z nejužitečnějších vlastností věrohodnostního poměru ($LR(+)$ nebo $LR(-)$) je to, že pokud klasifikace závisí na dvou nezávislých důkazech (tj. výsledek je hlášen pouze tehdy, když jej potvrzují dvě nezávislé analýzy z nezávislých postupů), $LR_{(1\&2)}$ výsledek obou analýz se

odhaduje vynásobením LR , který kvantifikuje nejistotu každého důkazu ($LR_{(1)}$ a $LR_{(2)}$):

$$LR_{(1\&2)} = LR_{(1)} \cdot LR_{(2)} \quad (4)$$

Pokud je například přítomnost kontaminantu v potravinářském produktu, stanovená pomocí GC-MS, založena na retenčním čase s $LR(+)$ 99,9 a na datech hmotnostního spektra s $LR(+)$ 490, $LR(+)$ identifikací založených na obou těchto nástrojích se stane $4,9 \times 10^4$ (tj. $99,9 \times 490$). Rovnice (4) vyplývá ze skutečnosti, že pravděpodobnost konvergence dvou nezávislých výsledků se odhaduje vynásobením příslušných jednotlivých pravděpodobností.

Pokud je zvažováno m nezávislých důkazů ($i = 1$ až m) k nahlášení pozitivního nebo negativního výsledku, tj. výsledek je nahlášen pouze při indikaci m důkazů, LR z kombinovaných důkazů odhaduje rovnice (5).

$$LR = \prod_{i=1}^m LR_{(i)}, \quad (5)$$

kde Π označuje součin posloupnosti proměnných a $LR_{(i)}$ je věrohodnostní poměr z i -té kvalitativní analýzy ($LR_{(i)}(+)$ nebo $LR_{(i)}(-)$).

Pokud důkazy nejsou nezávislé, rov. (5) podcení nebo nadhodnotí společnou pravděpodobnost. Rychlý odkaz 3 ukazuje, jak lze kombinovat nezávislé pravděpodobnosti.

Interpretace věrohodnostních poměrů může být náročná, zejména pro laiky. Pro forenzní aplikace byla doporučena stupnice v tabulce 5 [44], která poskytuje slovní vyjádření síly důkazů. Podle této tabulky jsou shromážděné důkazy považovány za „extrémně silné“ pouze v případě, že LR je větší než 10^6 . V zásadě lze tento druh přístupu přizpůsobit i jiným okolnostem, je-li požadován

Rychlý odkaz 3 – Pravděpodobnost pro nezávislé důkazy

Pravděpodobnost výskytu dvou nezávislých jevů A a B, $P(A \cap B)$ je odhadnuta rovnicí. (Q3.1).

$$P(A \cap B) = P(A)P(B), \quad (Q3.1)$$

kde $P(A)$ a $P(B)$ jsou pravděpodobnosti událostí A a B, které nastanou, například produkcí pozitivního výsledku z analýzy pozitivního případu (tj. TP).

Pokud jsou však $P(A)$ a $P(B)$ spojeny, pravděpodobnost koincidence obou událostí je určena rovnicí (Q3.2), která zahrnuje podmíněnou pravděpodobnost výskytu události B za předpokladu, že nastala událost A:

$$P(A \cap B) = P(A)P(B|A) \quad (Q3.2a)$$

Nebo ekvivalentně:

$$P(A \cap B) = P(B)P(A|B) \quad (Q3.2a)$$

Pro přímé nebo inverzní korelace spojené s $r_{AB} > 0$ nebo $r_{AB} < 0$ v tomto pořadí, bude $P(A \cap B)$ v tomto pořadí větší nebo menší než pro případy, kdy A a B jsou nezávislé.

Tabulka 5. Interpretace věrohodnostních poměrů navržených pro forenzní vědy Evropskou sítí forenzních ústavů [44]

Hodnota věrohodnostního poměru	Slovní ekvivalent
1	Zjištění nepodporují jeden návrh před druhým
2 – 10	Slabá preference pro první návrh ve srovnání s alternativou
10 – 100	Mírná preference pro první návrh spíše než pro alternativu
100 – 1000	Mírně silná preference pro první návrh spíše než pro alternativu
1000 – 10 000	Silná preference pro první návrh spíše než pro alternativu
10 000 – 1 000 000	Velmi silná preference pro první návrh spíše než pro alternativu
> 1 000 000	Extrémně silná preference pro první návrh spíše než pro alternativu

obecný údaj o síle důkazu. Například pro identifikaci polymerního typu mikroplastů odebraných ze sedimentů při monitorování životního prostředí jsou kritéria uvedená v tabulce 5 příliš přísná. Pro tyto analýzy by měly být adekvátní výsledky spojené s $LR(+)$ větším než 19 (tj. s $TP \geq 95\%$ a $FP \leq 5\%$), protože kontaminace je stanovena po identifikaci mnoha částic z několika vzorků [45].

Přestože stanovení binární vlastnosti může poskytnout pouze jeden ze dvou výsledků, pokud je nejpravděpodobnější výsledek (např. ano nebo ne) spojen s nízkou hodnotou LR , lze rozhodnout, že bude výsledek vykazován jako neprůkazný místo uvedení slovního ekvivalentu z tabulky 5. Například lze rozhodnout o nahlášení pozitivního nebo negativního výsledku, pokud je příslušný LR větší než (například) 10^5 , přičemž nižší LR je hlášeno jako neprůkazné. Tuto „šedou zónu“ pro hodnotu LR lze nastavit pod 10^5 , 19 nebo jakoukoli jinou hodnotu v závislosti na účelu analýzy. Hlášení výsledku jako neprůkazného je užitečné, pokud mají falešně pozitivní i falešně negativní výsledky relevantní dopad. Při testování na dopingové látky v moči sportovce jsou falešně pozitivní výsledky mnohem závažnější než falešně negativní výsledky, což naznačuje, že pokud není pozorován žádný důkaz dopingů, může být výsledek hlášen jako negativní (tj. žádný důkaz o dopingů) [9, 46]. Falešně pozitivní i falešně negativní výsledky však mohou být problémem pro identifikaci mateřství nebo otcovství, což naznačuje, že pozitivní shoda s nízkou hodnotou LR by neměla být hlášena jako průkazná „ne-shoda“ [47].

4.3 Posterioční pravděpodobnost

Pokud existují spolehlivé informace o prevalenci konkrétní charakteristiky (např. populace s dobře zdokumentovanou prevalencí určitého onemocnění), lze $LR(+)$ spojené s výsledkem testu

převést na pravděpodobnost PP , že zkoušená položka je s ohledem na pozitivní výsledek testu pozitivní. Toto je známo jako posterioční pravděpodobnost a odhaduje se pomocí Bayesovy věty (příloha A). Jedna z forem výpočtu PP , s použitím věrohodnostního poměru, je:

$$PP = \frac{\frac{P(+)}{P(-)} LR(+)}{\frac{P(+)}{P(-)} LR(+)} + 1 \quad (6)$$

Zde $P(+)$ a $P(-)$ jsou apriorní pravděpodobnosti, tj. informace dostupné před testem, a PP a PN jsou posterioční pravděpodobnosti.

Vezmeme-li předchozí příklad analýzy kontaminantu v potravinářském produktu pomocí GC-MS, kde je pozitivní výsledek spojen s $LR(+)$ $4,9 \times 10^4$, za předpokladu, že $P(+)$ = $P(-)$ dosáhne PP 99,998 % ($PP = 4,9 \times 10^4 / (4,9 \times 10^4 + 1)$).

Pokud je hlášen negativní výsledek, posterioční pravděpodobnost, že vzorek bude skutečně negativní, PN , se odhaduje takto:

$$PN = \frac{\frac{P(-)}{P(+)} LR(-)}{\frac{P(-)}{P(+)} LR(-)} + 1 \quad (7)$$

Tato rovnice je podobná rovnici (6). $P(+)$ a $P(-)$ vyjadřují prevalenci pozitivních nebo negativních případů.

Protože posterioční pravděpodobnost se vztahuje k hlášené klasifikaci, lze posterioční pravděpodobnost považovat za míru pravděpodobnosti, že hlášená hodnota je správná.

Posterioční pravděpodobnosti může být obtížné aplikovat v praxi. Někdy nejsou k dispozici dostatečně relevantní a spolehlivé apriorní pravděpodobnosti. Ačkoli někteří autoři navrhli, že tento problém lze překonat předpokladem, že pozitivní a negativní výsledky jsou stejně pravděpodobné, takže $P(+)/P(-) = 1$, toto není vždy smysluplné. Někdy, zejména ve forenzní oblasti, může být nevhodné vyvozovat apriorní

pravděpodobnosti pro konkrétní případ ze znalosti nesouvisejících případů. V takových případech může věrohodnostní poměr (oddíl 4.2) poskytnout užitečné shrnutí spolehlivosti poskytované výsledkem testu, aniž by bylo nutné předem určovat pravděpodobnosti.

V některých oborech, jako jsou např. lékařské vědy, může být při rozhodování o kvalitativních analytických výsledcích pro diagnózu zásadní zohlednění prevalence stavu nebo charakteristiky. Diagnóza onemocnění nebo klinické situace na základě klinických analytických výsledků se bude opírat také o další informace, jako je barva sliznice, lokalizace a intenzita bolestí, věk a pohlaví, rizikové chování atd. Způsob, jakým tyto informace přispívají ke konečnému rozhodnutí o pozorovaném výsledku, lze ilustrovat výpočtem *PP* nebo *PN*, ačkoli lékaři tyto výpočty běžně neprovádějí; spíše se od nich očekává, že si budou vědomi obecného významu prevalence při stanovení diagnózy na základě výsledku zkoušky (testu).

Další podrobnosti o těchto metrikách jsou uvedeny v bibliografii [4, 8, 9].

4.4 Spolehlivost metrik

Spolehlivost vypočtených *LR*, *PP* nebo *PN* závisí na spolehlivosti poměru výsledků a posteriorní pravděpodobnosti, na spolehlivosti uvažované apriorní pravděpodobnosti. Tabulka 4 uvádí počet testů potřebných pro spolehlivou detekci jedné nebo více falešných odpovědí při různých pravděpodobnostech falešné odpovědi. Spolehlivé stanovení takového poměru obecně vyžaduje mnohem více (viz oddíl 3.3.3). Počet studovaných případů může být nutné dále zvýšit, aby pokryl celý rozsah zkušební metody; například při testování potravin může být nutné zkoumat více různých matric potravin. Modelování instrumentálního signálu uvažovaného v kvalitativní analýze může umožnit kvantifikaci nízkých chybných výsledků, ale závisí na adekvátnosti vstupních dat a modelovacím algoritmu.

Pro stanovení *PP* nebo *PN* z velmi neshodných $P(+)$ a $P(-)$ lze použít tabulku 4 k definování počtu případů (z cílové populace), které by měly být studovány.

Kvalita vstupních dat pro odhad těchto metrik je ještě kritičtější, když jsou kombinovány různé důkazy a počítány metriky kvantifikující sílu kombinovaných informací.

Prezentované metriky by proto měly být používány obezřetně, přičemž je třeba mít na paměti relevantní podrobnosti o vstupních datech, způsob použití hlášeného výsledku a příslušné důsledky. Přehnaná interpretace údajů o výkonu kvalitativní analýzy může být stejně škodlivá jako ignorování omezení konkrétní kvalitativní analýzy.

4.5 Nejistota poměrů

Statistická kvalita odhadovaného poměru výsledků, která závisí na počtu testů (zkoušek) použitých pro jejich stanovení, může být vyjádřena jako konfidenční interval, CI, pro vypočtený poměr. Tento konfidenční interval je také znám jako „podmínková nejistota“ (4.4.6 z [18]), obvykle se vypočítává pro 95% konfidenční hladinu (95% CI).

Například široký 95% CI pro citlivost *SS* naznačuje, že „skutečná“ hodnota *SS* se může velmi lišit od odhadu. Stejnou logiku lze použít i na jiné poměry výsledků, například *SP*. Vzhledem k tomu, že poměry výsledků nejsou odhadnuty z žádných předchozích znalostí o populaci případů, tyto intervaly charakterizují pouze odhadovanou kvalitu analytické výkonnosti.

Interpretace 95% CI je do určité míry podobná tomu, co se děje s rozšířenou nejistotou měření [1]. Pro 95% CI existuje 5 % pravděpodobnost, že „skutečná“ hodnota míry výsledků je mimo meze CI. Podobně 95% CI pro experimentálně stanovený poměr poskytuje statistickou nejistotu pro vypočítaný poměr.

Pokud je například schopnost metody kvalitativní analýzy správně identifikovat pozitivní případy testována na základě analýzy 400 takových případů a všech 400 výsledků je pozitivních, odhadovaný *SS* 100 % je spojen s 95 % CI ohraničené mezi 99 % a 100 %; tj. skutečná hodnota *SS* se pohybuje mezi 99 % a 100 % s 95% spolehlivostí. Pokud je metoda testována pouze s 5 pozitivními případy, bude 95 % CI *SS* omezen 57 % a 100 %. 95% CI umožňuje vyjádřit kvalitu výkonnostních parametrů analytické metody, která je nutná pro jejich správnou interpretaci. Ve výše uvedených příkladech jsou uvedeny obě hodnoty *SS* na 100 %, ale odhad *SS* je mnohem spolehlivější v prvním případě. Výpočet 95 % CI pro *SS* a *SP* je standardní praxí v klinické laboratoři (10.1.3 z [27]).

Bylo publikováno několik modelů pro výpočet CI [48] [49, 50, 51, 52, 53, 54]. Pro jednoduchost a použitelnost na malé počty byl použit interval Wilsonova skóre [54]. Rovnice 8 a 9 lze použít

k výpočtu dolní $LL_{SS,95}$ a horní $HL_{SS,95}$, meze 95 % CI pro SS nebo TP .

$$LL_{SS,95} = \frac{A_1 - A_2}{A_3} 100 \quad (8)$$

$$HL_{SS,95} = \frac{A_1 + A_2}{A_3} 100 \quad (9)$$

kde:

$$A_1 = 2 tp + 1,96^2$$

$$A_2 = 1,96 (1,96^2 + 4 tp \cdot fn / (tp + fn))^{1/2}$$

$$A_3 = 2 (tp + fn + 1,96^2).$$

Rovnice 10 a 11 lze použít k výpočtu dolní a horní meze 95 % CI pro SS nebo TP .

$$LL_{SP,95} = \frac{B_1 - B_2}{B_3} 100 \quad (10)$$

$$HL_{SP,95} = \frac{B_1 + B_2}{B_3} 100 \quad (11)$$

kde:

$$B_1 = 2 tn + 1,96^2$$

$$B_2 = 1,96 (1,96^2 + 4 fp \cdot tn / (fp + tn))^{1/2}$$

$$B_3 = 2 (fp + tn + 1,96^2).$$

Cílová nebo minimální hodnota $LL_{SS,95}$ nebo $LL_{SP,95}$ (tj. $LL_{SS,95}^{tg}$ nebo $LL_{SP,95}^{tg}$) by měla být definována podle účelu analýzy. Cíl je obzvláště kritický, když je kritický dopad falešných výsledků. Například u krevních složek používaných při transfuzích by měl být screening na infekční onemocnění proveden pomocí testů souvisejících s $LL_{SS,95}$ blízkým 100 %, což lze potvrdit pouze tehdy, bude-li během validace testováno mnoho pozitivních případů.

Když se poměr výsledků porovnává s cílovou minimální hodnotou nebo když se zkoumá zvýšení nebo snížení parametru, mělo by být provedeno jednostranné posouzení. Pro 95% test spolehlivosti by měl být faktor 1,96 změněn na 1,64.

5 Uvádění kvalitativního analytického výsledku

Neočekává se, že akreditované laboratoře v současné době vyhodnocují nebo uvádí nejistoty spojené s výsledky kvalitativní analýzy. Příklady v tomto oddílu jsou proto určeny k navržení možných přístupů k uvádění výsledků, pokud se tak laboratoř rozhodne udělat, aby pomohla zákazníkovi.

Pozitivní výsledek může být uveden s *TP* a *FP*, *LR(+)* nebo *PP* a negativní výsledek s *TN* a *FN*, *LR(-)* nebo *PN*. Ostatní metriky uvedené v tabulce 2 lze také použít k vykazování spolehlivosti výsledku.

Tyto metriky obvykle poskytují informace o jednotlivých výsledcích zkoušky. Avšak v případech, kdy je hodnota metriky pro analytický rozsah konstantní, lze takové parametry interpretovat jako charakterizující analytickou metodu.

Následující čtyři příklady ilustrují, jak lze vykazovat kvalitativní výsledky s příslušnou výkonností nebo nejistotou.

Příklad 1 (text kurzívou zmiňuje nejistotu kvalitativní analýzy):

Paní AB je infikována virem SARS-CoV-2.

(test s klinickou citlivostí 90 % a specifičností 99 %)

Příklad 2 (text kurzívou zmiňuje nejistotu kvalitativní analýzy):

Moč pana CD obsahuje rezidua kanrenonu

(identifikace s věrohodnostním poměrem $4,9 \times 10^4$)

Příklad 3 (text kurzívou zmiňuje nejistotu kvalitativní analýzy):

Kokain je přítomen ve vzorku 123

(identifikace s věrohodnostním poměrem $4,9 \times 10^4$ a považována za „velmi silný“ důkaz přítomnosti analytu)

Příklad 4 (text kurzívou zmiňuje nejistotu kvalitativní analýzy):

V troskách požáru byly identifikovány zbytky benzínu s kódem vzorku 456

(identifikace s posteriorní pravděpodobností 99,998 %, odhadem ze simulace modelu signálu a za předpokladu, že přítomnost nebo nepřítomnost analytu jsou stejně pravděpodobné)

6 Závěry a doporučení

Je důležité, aby laboratoře zkontrolovaly alespoň nejkritičtější poměr falešných odpovědí. U některých metrik musí být stanoven poměr falešně pozitivních i falešně negativních výsledků.

Je reálné očekávat, že většina laboratoří má příslušné parametry svých postupů kvalitativní analýzy (tj. podmínky analýzy) pod odpovídající kontrolou. Důkazy o tom budou obvykle zahrnovat:

- jasný důkaz o přiměřené metrologické návaznosti hodnot parametrů podléhajících kontrole vzhledem k jejich relevanci pro zkoušku;
- důkaz, že nejistoty těchto parametrů jsou pro daný účel dostatečně malé.

Je rozumné očekávat, že laboratoře se budou řídit publikovanými kodexy osvědčených postupů v kvalitativní analýze, pokud jsou k dispozici, včetně použití vhodných referenčních údajů a materiálů.

Kvantitativní (tj. numerické) zprávy o nejistotách ve výsledcích kvalitativních zkoušek by měly být předloženy způsobem, který zabrání nesprávné interpretaci.

Kdykoli dojde k závěru, že získaný analytický výsledek je spojen s příliš nízkým, respektive příliš vysokým poměrem pravdivých nebo nepravdivých výsledků, je zcela rozumné oznámit výsledek testu jako „neprůkazný“ ve smyslu nedostatečně jistý.

7 Příklady

Příklady jsou popsány po uvedení jejich účelu.

7.1 E1: Identifikace sloučenin hmotnostní spektrometrií s nízkým rozlišením pomocí vyhledávání v databázi nebo přítomnosti charakteristických iontů

7.1.1 Úvod

Tento příklad je rozdělen na případ A nebo B, kde se používají různé postupy k identifikaci sloučenin v komplexních maticích pomocí hmotnostní spektrometrie s nízkým rozlišením. Paralelní prezentace těchto dvou případů zdůrazňuje alternativní povahu možností identifikace.

Všimněte si, že v praxi se při identifikaci obvykle používá více kritérií, jako je kombinace shody hmotnostního spektra a chromatografického retenčního času. Tento příklad se zaměřuje pouze na složku hmotnostní spektrometrie. Příklad E5 uvádí aplikaci více kritérií.

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích

Položka/matrice: A) masné výrobky a B) forenzní nebo environmentální vzorky

Parametr/analyt: A) diethylstilboestrol, DES (zakázaný růstový hormon pro hovězí a drůbeží maso) nebo B) heroin, DES a dichlordifenyltrichlorethan (DDT)

Typ klasifikačního kritéria: 1) Identifikace na základě tolerancí pro relativní zastoupení (*RA*) specifických iontů hmotnostního spektra; 2) Identifikace na základě přítomnosti specifických iontů hmotnostního spektra bez ohledu na hodnoty *RA*.

Technika/přístroj: Plynová chromatografie spojená s hmotnostní spektrometrií s nízkým rozlišením využívající elektronovou ionizaci (GC-MS)

Typ hlášení výsledků: Věrohodnostní poměr

Tento příklad popisuje vyhodnocení nejistoty identifikace sloučenin pomocí GC-MS za použití různých identifikačních kritérií (oddíly 7.1.1 a 7.1.2). Příklady prezentují výsledky pro identifikaci tří sloučenin (tj. DES, heroin a DDT) ve dvou typech vzorků (masné výrobky a forenzní nebo environmentální vzorky).

Hmotnostní spektrometrie, zejména v kombinaci s chromatografickým separačním stupněm, je účinným nástrojem, který může pomoci identifikovat neznámé sloučeniny. Pro většinu účelů je zvolenou metodou hmotnostní spektrometrie s nízkým rozlišením využívající elektronovou ionizaci (EI), když je požadována identifikace místo kvantifikace. Hmotnostní spektrum může obsahovat mnoho iontů, z nichž ne všechny jsou užitečné pro diagnostické účely. To vyvolává otázku, zda existuje minimální počet iontů, který by postačoval k zajištění jednoznačné identifikace.

POZNÁMKA: V některých analytických oblastech je pro identifikaci sloučenin definován minimální počet iontů [20][21][22][23].

7.1.2 Identifikace na základě relativního zastoupení charakteristických iontů

Sphon [55] zkoumal minimální počet iontů, které je třeba monitorovat, aby se dosáhlo jednoznačné identifikace diethylstilboestrolu⁵ (DES) v masných výrobcích. Údaje týkající se následné studie [56], založené na komerční knihovně hmotnostních spekter obsahující asi 270 000 záznamů, jsou uvedeny v tabulce E1.1.

Tabulka E1.1 ukazuje počet spekter v použité knihovně, která odpovídají specifikovaným kritériím pro relativní zastoupení *RA* jednoho nebo více iontů. *RA* se odhadne vydělením zastoupením studovaného iontu zastoupením nejhojnějšího iontu (tj. základním píkem). Tato normalizace má za cíl vytvořit identifikační parametr méně závislý na hladině analytu (např. koncentraci analytu). V tabulce E1.1 lze pozorovat, že když

⁵ DES byl používán jako růstový hormon pro skot a drůbež a následně zakázán poté, co byly prokázány jeho karcinogenní vlastnosti.

Tabulka E1.1. Počet spekter Wileyovy knihovny s 270 000 záznamy, které odpovídají specifickým kritériím pro relativní zastoupení některých iontů

#	Identifikační kritérium		Shody
	Ion (<i>m/z</i>)	<i>RA</i> (%) interval přijetí	
1	268	1 – 100	9 995
2	268	1 – 100	5 536
	239	1 – 100	
3	268	90 – 100	46
	239	10 – 90	
4	268	90 – 100	9
	239	50 – 70	
5	268	90 – 100	15
	239	50 – 90	
	145	5 – 90	
6	268	90 – 100	1 (DES) ^a
	239	50 – 70	
	145	45 – 65	

RA Relativní zastoupení (procento nejzastoupenějšího iontu, základní pík)

^a Jediná shoda odpovídá hmotnostnímu spektru DES

se vezme v úvahu více iontů a užší rozsahy zastoupení, počet shod vyskytujících se v databázi se dramaticky zmenší. Sada identifikačních kritérií č. 6 izoluje hmotnostní spektrum DES, což vede k jediné shodě.

Pro srovnání s alternativní knihovnou uvádí tabulka E1.2 počet shod z veřejně dostupné referenční knihovny, která pak obsahuje 62 235 spekter, přičemž se berou v úvahu tolerance pro relativní zastoupení (*RA*) jednoho nebo více iontů [57]. Jak se tolerance spojené s *RA* více iontů zužují, izolují se hmotnostní spektra méně sloučenin. Tabulka E1.2 uvádí počet spekter odpovídajících třem různým cílovým sloučeninám, jmenovitě DES, heroinu a DDT. Heroin a DDT jsou relevantní pro analýzu některých forenzních a environmentálních vzorků.

Porovnání izolace hmotnostního spektra DES v obou knihovnách, popsané v tabulkách E1.1 a E1.2, umožňuje dospět k závěru, že podle očekávání závisí počet shod na počtu spekter v knihovně (viz identifikační kritéria #1 a #2 v tabulkách E1.1 a E1.2). Pokud se počet shod vydělí celkovým počtem záznamů, rozdíly pozorované v tabulkách E1.1 a E1.2 se sníží.

Tabulka E1.3 shrnuje informace shromážděné v tabulkách E1.1 a E1.2 z nejselektivnějších identifikací. Tabulka E1.3 převádí shromážděné informace na *TP* a *FP*, dále kombinované do *LR(+)*, která odhaduje nejistotu pozitivního výsledku (tj. hlášení přítomnosti analytu).

Odhadovaná *TP* (tj. přibližně 100 %) předpokládá, že definované tolerance pro *RA* iontů zohledňují jejich variabilitu. V ideálním případě by tolerance měly být stanoveny z modelů variability signálu sestavených z replikovaných spekter roztoků vzorků s relevantními koncentracemi analytu [8, 9] (příklad E5).

FP uvedený v tabulce E1.3 předpokládá, že možnými interferencemi jsou všechny sloučeniny, jejichž hmotnostní spektrum je dostupné v použité knihovně. Mnoho sloučenin přítomných v roztocích vzorků nebude detekovatelných pomocí GC-MS nebo bude eliminováno při přípravě vzorku. Na druhou stranu mnoho sloučenin, jejichž spektrum je v knihovně, pravděpodobně nebude v analyzovaných vzorcích kvůli chemické nekompatibilitě nebo nezávislosti zdrojů nebo původu. Nejhorší případ *FP* ekvivalentní jednomu na celkový počet spekter N mínus 1 ($FP = 1/(N - 1)$) se odhaduje, protože je známo, že *FP* nebude nula. Při použití této hodnoty je třeba vzít v úvahu popsaná omezení způsobu vyhodnocení *FP*. *FP* lze alternativně odhadnout z modelů šumu signálu, jak je diskutováno v příkladu E5 [8, 9].

Z údajů v tabulce E1.3 je vidět, že při použití stejné knihovny mají všechny analyty s jednou shodou stejné *TP*, *FP* a *LR(+)* v závislosti pouze na počtu záznamů v knihovně. Podle kritérií definovaných Evropskou sítí

Tabulka E1.2. Počet spekter veřejně dostupné knihovny s 62 325 záznamy, které odpovídají specifickým kritériím pro relativní zastoupení některých iontů

#	Identifikační kritérium		Shody
	Ion (<i>m/z</i>)	RA (%) interval přijetí	
1	268	1 – 100	3597
2	268	1 – 100	1597
	239	1 – 100	
3	268	55 – 95	83
4	268	55 – 95	4
	239	30 – 70	
5	268	55 – 95	1 (DES)
	239	30 – 70	
	145	60 – 100	
6	369	1 – 100	1672
7	369	1 – 100	526
	327	1 – 100	
8	369	45 – 85	43
9	369	45 – 85	1 (heroin)
	327	60 – 100	
10	352	1 – 100	1242
11	352	1 – 100	234
	235	1 – 100	
12	352	1 – 40	1140
13	352	1 – 40	1 (DDT)
	235	60 – 100	
14	352	1 – 40	7
	235	1 – 100	
	237	48 – 88	
15	352	1 – 40	1 (DDT)
	235	60 – 100	
	237	48 – 88	

RA Relativní zastoupení (procento nejzastoupenějšího iontu, základní pík)

^a Jediná shoda odpovídá hmotnostnímu spektru DES, heroinu nebo DDT

forenzních vědeckých ústavů (tabulka 5) je identifikace analytu hmotnostní spektrometrie za použití popsaného identifikačního postupu poskytuje „velmi silný“ důkaz přítomnosti analytu (*LR*(+) mezi 10^4 a 10^6). Předpokládejme, že identifikaci podporuje také retenční čas analytu v chromatografickém systému (tj. v GC) a okno retenčního času je adekvátní. V takovém případě se může zvýšit *LR*(+) identifikace (příklad E5).

7.1.3 Identifikace na základě přítomnosti charakteristických iontů – odhadovaná pravděpodobnost shody

Předpokládejme, že namísto identifikace analytu pomocí tolerancí pro *RA* specifických iontů je identifikace založena na jednoduché přítomnosti tří selektivních iontů. V takovém případě lze pro hrubé vyhodnocení nejistoty identifikace použít následující výpočet. Předpokládejme, že je použit hmotnostní spektrometr s nízkým rozlišením, který dokáže rozlišit pouze jednotky *m/z* (tj. jednotky poměru hmotnosti k náboji) a ionty musí mít hodnoty *m/z* mezi 180 *m/z* a 480 *m/z*. V takovém případě lze ve spektru pozorovat přibližně 300 možných *m/z* hodnot. Vzhledem k tomu, že počet kombinací 300 objektů odebraných najednou po třech je $300!/[3! \cdot (300 - 3)!] = 4\,455\,100$, pro tři píky by pravděpodobnost tří náhodných píků odpovídajících třem vybraným iontům náhodou byla, za předpokladu, že všechny poměry *m/z* jsou stejně pravděpodobné, jsou $1/4,6 \times 10^6$ nebo přibližně $2,2 \times 10^{-7}$. To však neumožňuje skutečnost, že většina hmotnostních spekter má

Tabulka E1.3. Nejistota identifikace několika analytů v různých matricích pomocí GC-MS, odhadnutá z počtu shod spekter z použité knihovny, přičemž se berou v úvahu tolerance pro relativní zastoupení dvou nebo tří iontů

Analyt	Analyzovaná položka	Počet knihovny, N	spekter Počet shod ^a	TP (%) ^b	FP (%) ^c	LR TP / FP
DES	Masné výrobky	270 000	1 v N	~100	3.7×10^{-4}	2.7×10^5
DES	F&E	62 235	1 v N	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$
Heroin	F&E	62 235	1 v N	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$
DDT	F&E	62 235	1 v N	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$

DES: diethylstilboestrol; F&E: Forenzní a environmentální vzorky; TP : Skutečně pozitivní poměr; FP Poměr falešně pozitivních výsledků; $LR (+)$: Věrohodnostní poměr (TP/FP) (viz tabulka 2).

^a Počet shod s ohledem na definované tolerance pro RA specifických iontů.

^b Optimistický odhad u TP (cca 100 %).

^c Odhadováno jako $1/(N-1)$ (nejhorší scénář ze shromážděných informací).

v zájmové oblasti obvykle mnohem více než tři ionty; to zvyšuje pravděpodobnost, že hmotnostní spektrum obsahující n iontů může odpovídat vybraným třem faktorem $n!/[3! \cdot (n-3)!]$. Vezmeme-li deset jako typický počet píků, pravděpodobnost náhodné shody se zvýší o $10!/3! \cdot 7! = 120$. Odhadovaná pravděpodobnost shody náhody je tedy $120 \times 2,2 \times 10^{-7}$ nebo přibližně $2,6 \times 10^{-5}$. To poskytuje přibližný poměr falešně pozitivních výsledků, FP .

Za předpokladu, že hladiny kontaminantů jsou dostatečně vysoké, aby poskytly spolehlivá hmotnostní spektra, lze uvažovat, že TP je přibližně 1 nebo 100 %.

Odhadovaný TP a FP výše lze zkombinovat, aby se získal $LR(+)$ z $1/2,6 \times 10^{-5} = 3,8 \times 10^5$, což indikuje horní mez věrohodnostního poměru pro identifikace založené popsáního postupu.

POZNÁMKA: Tento odhad předpokládá, že všechny kombinace iontů jsou stejně pravděpodobné a možné; a že jejich výskyt je nezávislý; je známo, že jde o hrubou aproximaci (viz příklad 7.2). Pravděpodobnost shody efektivní šance je proto pravděpodobně mnohem vyšší než vypočítaná hodnota, což vede k nižšímu věrohodnostnímu poměru.

7.1.4 Závěrečné poznámky

Metodologie pro odhadování $LR(+)$ identifikací pomocí GC-MS uvedené v tomto příkladu mají tendenci být příliš optimistické, pokud jde o validitu kvalitativních analytických výsledků (viz příklad 7.2 pro srovnání se skutečnými pravděpodobnostmi shody). Tyto výpočty by proto měly být používány pouze jako počáteční posouzení platnosti identifikace. Příklad E5 popisuje alternativní a realističtější stanovení nejistoty identifikace prováděné pomocí GC-MS/MS.

Přestože zde prezentované metodiky pro odhad pravděpodobnosti shody a věrohodnostních poměrů mohou být optimistické, je stále možné dospět k závěru, že například identifikace heroínu ve vzorcích z místa činu pouze na základě přítomnosti iontů s m/z 369 a 327 by bylo neadekvátní, protože je spojeno s $LR(+)$ ze 118 ($118 = TP/\{FP\} = 100/\{[526/(62235-1)] \cdot 100\}$, (metodika oddílu 7.1.1: případ 7 tabulky E1.2) nebo $4,5 \times 10^4$ ($4,5 \times 10^4 = 100/\{[1/(44850-1)] \cdot 100\}$, kde $44850 = 300!/[2! \cdot (300-2)!]$), (metodika oddílu 7.1.2) v závislosti na přístupu použitým pro hodnocení spolehlivosti výsledků. V praxi by to znamenalo, že k zajištění dostatečné spolehlivosti by byla zapotřebí další kritéria nebo další potvrzující testy.

Podobně, ačkoli Sphon [55] a další navrhli, že rezidua veterinárních léčiv u skotu lze identifikovat pomocí tří iontů hmotnostního spektra, pro oficiální monitorování nepovolených látek Evropská unie (EU) vyžaduje shromažďování dalších důkazů o přítomnosti těchto sloučenin. [20]. Pokud je například možné monitorovat pouze dva charakteristické ionty pomocí GC-MS na vhodných úrovních analytu, je třeba zvážit dvě nezávislé chromatografické analýzy založené na elektronové nebo chemické ionizaci, aby se potvrdila přítomnost analytu [20].

7.2 E2: Identifikace čištěných sloučenin infračervenou spektrometrií

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích

Položka/matrice: Čištěná chemická sloučenina

Parametr/analyt: Sloučenina z dostupné databáze infračervených spekter

Typ klasifikačního kritéria: Shoda vlnočtu tří nebo šesti pásů infračerveného spektra v intervalu [500, 1800] cm^{-1}

Technika/přístroj: Infračervená spektrometrie

Typ hlášení výsledků: Věrohodnostní poměr

Několik autorů zkoumalo využití databázových statistik při hodnocení kritérií pro kvalitativní analýzu. De Ruig a kol. [58] navrhli kritéria, která mají být splněna pro identifikaci reziduí veterinárních léčiv v masných výrobcích (viz část 7.1.2). Autoři uvádějí orientační hodnoty pravděpodobnosti náhodné shody na základě jednoduchého binomického modelu. Ellison a kol. ukázali, že hypergeometrické rozložení bylo vhodnějším modelem pro náhodné porovnávání ve spektrech, protože umožňuje malý počet odpovídajících píků mezi dvěma spektry, z nichž obě obsahují větší počet píků [5]. Posledně jmenovaní autoři se zaměřili na náhodné shody, kdy se infračervené spektrum srovnává se spektrální knihovnou.

Ellison a kol. [5] studoval spolehlivost identifikace purifikovaných sloučenin porovnáním získaného infračerveného spektra se spektry z knihovny. Knihovna použitá Ellisonem a kol. byla Sadtlerova knihovna obsahující spektra z 59 626 různých materiálů. Z této knihovny byla vybrána náhodná podskupina třiceti sloučenin a pro každou sloučeninu byl zaznamenán počet pásů m v intervalu [500, 1800] cm^{-1} . Bylo zjištěno, že průměrný počet pásů na spektrum v intervalu [500, 1800] cm^{-1} , M , byl 16. Dostupné spektrální rozlišení bylo 4 cm^{-1} a to implikovalo existenci $1300/4 = 325$, N , diskretních poloh píku v intervalu [500, 1800] cm^{-1} . Pro každé odlišné spektrum ve zvolené podmnožině byla celá databáze prohledána dvakrát – nejprve pro minimálně $n = 3$ odpovídajících píku a podruhé pro minimum $n = 6$ odpovídajících píku.

Pro $n \geq 3$ byl počet pozorovaných shod přibližně dvojnásobkem počtu předpovězeného hypergeometrickým rozložením. Pro $n \geq 6$, ačkoli počet shod byl podstatně nižší, jak by se dalo očekávat, pozorované shody desetinásobně převýšily predikované shody. Část dat pro šest odpovídajících píku je uvedena v tabulce E2.1.

Tabulka E2.1. Možné shody vůči šesti pásům v infračervené databázi

Sloučenina	Počet pásů m v rozsahu	Pravděpodobnost ^a možné shody	Předpovězené shody ^b	Pozorované shody	LR
1-Chlor-3-(1-naftyloxy)-2-propanol	23	$3,19 \times 10^{-4}$	19	192	311
Methylester α -kyanoskořicové kyseliny	17	$5,03 \times 10^{-5}$	3	29	2056
Fenyl v-triazolo-[1,5- α]-pyridin-3-yl keton	24	$4,19 \times 10^{-4}$	25	190	314
Benzo- β -thiofen-6-akrylová kyselina	20	$1,34 \times 10^{-4}$	8	52	1147
3-((Dipropylamino)methyl)-5-nitroindol	17	$5,03 \times 10^{-5}$	3	29	2056
2-Mesityl-5-fenyloxazol	22	$2,52 \times 10^{-4}$	15	99	602
<i>p</i> -hydroxybenzoová kyselina	18	$6,71 \times 10^{-5}$	4	44	1355
Kapronová kyselina, isobutylester	8	$1,36 \times 10^{-7}$	0	1	59626
1-bromoadamantan	10	$9,64 \times 10^{-7}$	0	1	59626
Fenylpropylether	17	$5,03 \times 10^{-5}$	3	47	1269

^a Pravděpodobnost náhodné shody je pravděpodobnost, že $n=6$ píků se náhodně shoduje ve dvou spektrech m pásů, za předpokladu náhodného výskytu pásů napříč spektrem. Pravděpodobnost byla vypočtena pomocí hypergeometrického rozdělení (ref [5]).

^b Předpokládaný počet shod je pravděpodobnost shody vynásobená počtem spekter v databázi.

Vypočtené pravděpodobnosti náhodných shod pro shody v šesti píčích byly v intervalu $[10^{-7}, 10^{-5}]$. Pravděpodobnost shody pro sloučeninu, vynásobená počtem záznamů v databázi, odhaduje počet sloučenin vyhovujících kritériím vyhledávání. V případě dvou sloučenin v tabulce E2.1, viz isobutylester kapronové kyseliny a 1-bromadamantan, vyhledávací kritéria produkují jedinou shodu, a proto se zdají být adekvátní, pokud existuje podezření na tyto sloučeniny. Pro zbývající sloučeniny vzniká mnohem více shod, což naznačuje požadavek na přísnější kritéria.

Za předpokladu, že TP je přibližně 100 % (protože IR spektra čistých sloučenin se spolehlivě shodují s jejich vlastním referenčním spektrem) a vezmeme-li FP jako poměr mezi počtem pozorovaných shod a celkovým počtem spekter knihovny (tj. 59 626), je možné odhadnout $LR(+)$ identifikace. Poslední sloupec tabulky E2.1 představuje odhadované $LR(+)$ (vypočteno jako TP/FP s $TP = 1,0$). Hlášené $LR(+)$ jsou mnohem nižší než minimální hodnota 10^6 považovaná za klasifikaci důkazů jako „extrémně silné“ (tabulka 5), což naznačuje, že jednoduchá shoda šesti píků na vlnové délce sama o sobě nemusí poskytnout dostatečnou spolehlivost bez dodatečných kritérií. K zajištění dostatečně jednoznačné identifikace mohou být zapotřebí další kritéria (jako je další shoda intenzity píku, nepřítomnost píků, které nejsou přítomny v cílové sloučenině, odpovídající chromatografický retenční čas nebo požadavek na blízkou vizuální shodu pro kompletní spektrum).

Tento příklad zdůrazňuje, že referenční databáze, z nichž jedním typem jsou spektrální knihovny, mohou získat pouze orientační informace o poměru falešných odpovědí. Poměry odpovědí jsou striktně spolehlivé pouze pro populace podobné populaci očekávaných vzorků. Je tak odpovědností analytika rozhodnout, která, pokud nějaká, ze sady shod odpovídá neznámé.

7.3 E3: Identifikace zneužívaných drog v moči technikou enzymové multiplikované imunoanalýzy (EMIT) a alternativní technikou

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích (studovaná s použitím kvalitativních informací)

Položka/matrice: Moč

Parametr/analyt: Kokain, metadon nebo opiáty

Typ klasifikačního kritéria: Neuvedeno

Technika/přístroj: Technika enzymového multiplikovaného imunotestu, EMIT, a alternativní, patentovaná technika

Typ hlášení výsledků: Věrohodnostní poměr a pravděpodobnost, že pozitivní výsledek bude správný

Ačkoli stanovení EMIT zahrnuje zpracování instrumentálního signálu, tento příklad hodnotí výkonnost této kvalitativní analýzy na základě experimentálně stanovených poměrů falešně pozitivních a falešně negativních výsledků. Proto tento příklad ilustruje stanovení kvality kvalitativní analýzy na základě kvalitativních informací.

V literatuře bylo popsáno použití databází výsledků vzorků pro získání relevantních pravděpodobností pro Bayesovské hodnocení výkonnosti kvalitativní analýzy. Pro testování drog v moči Ferrara *a kol.* [33] vytvořili databázi obsahující informace o typech drog, analytických technikách, poměru falešných odpovědí pro techniky a prevalenci drog. Pro laboratoř citovaných autorů shrnuje tabulka E3.1 část těchto údajů pro techniku EMIT. Tabulka také uvádí odhadovanou posteriorní pravděpodobnost, že zkoušený vzorek bude skutečně pozitivní, PP , jak je popsáno v rovnici (6). Pro výpočet se bere prevalence negativních výsledků jako $(1 - P(+))$, $LR(+) = TP/FP$ a $TP = 1 - FN$.

Tabulka E3.1. Pravděpodobnosti detekce zneužívaných drog v moči pomocí EMIT [33]

Pravděpodobnost	Hodnoty výkonnostních charakteristik pro různé léky nebo třídy léků		
	Opiáty	Metadon	Kokain
$P(+)$	0,44	0,26	0,20
FP	0,028	0,004	0,009
FN	0,069	0,018	0,056
PP	0,963	0,988	0,963
$P(+)$ – prevalence pozitivních výsledků			

Například pro identifikaci metadonu se PP vyhodnocuje podle rovnice (E3.1) jako 0,988 (PP je posteriorní pravděpodobnost určená z Bayesovy věty). Jinými slovy, analytik si mohl být na více než 98 % jistý, že pozitivní odpověď na metadon skutečně ukazuje na přítomnost této drogy. Všimněte si však, že to částečně závisí na vysoké pozorované prevalenci v konkrétní vzorkované populaci. PP může být v běžné populaci mnohem nižší, tj. v populaci, kde užívání drog není tak časté.

$$PP = \frac{\frac{P(+)}{P(-)} LR(+)}{\frac{P(+)}{P(-)} LR(+) + 1} = \frac{\left(\frac{0,26}{1 - 0,26}\right) \left(\frac{1 - 0,018}{0,004}\right)}{\left(\frac{0,26}{1 - 0,26}\right) \left(\frac{1 - 0,018}{0,004}\right) + 1} = 0,988 \quad (E3.1)$$

Tabulka E3.2 ukazuje podobná data pro jinou, neimunochemickou techniku. Všimněte si, že FP pro kokain touto technikou je uváděn jako nula. Je však diskutabilní, zda poměr falešných odpovědí na takové screeningové testy může být skutečně nulový. V tomto případě nebyly nalezeny žádné falešné pozitivní, ale pokud by bylo analyzováno více vzorků, mohl se objevit jeden nebo více falešně pozitivních výsledků. Odhadovaný poměr falešné odpovědi byl proto použit při výpočtu posteriorní pravděpodobnosti.

Tabulka E3.2. Pravděpodobnosti detekce zneužívaných drog v moči patentovanou technikou [33]

Hodnoty výkonnostních charakteristik pro různé léky nebo třídy léků			
Pravděpodobnost	Opiáty	Metadon	Kokain
P (+)	0,44	0,26	0,20
FP	0,038	0,012	0,000
FN	0,276	0,179	0,247
PP	0,937	0,960	0,995 §

§ PP vypočtené pomocí odhadované hodnoty nejhoršího případu FP rovné 0,001 (jeden z 1000 testů).

Pokud vezmeme v úvahu opět metadon, PP je 0,960. To je přiměřeně vysoká pravděpodobnost, i když o něco méně přesvědčivá než ta, kterou poskytuje test EMIT.

Pro ilustraci účinku kombinace dat předpokládejme, že byly provedeny oba screeningové testy. Pokud je v každém případě získána pozitivní odezva, pak kombinované PP dosáhne 0,9998 (viz rovnice (E3.2) na základě kombinace rovnic (5) a (6).

$$PP = \frac{\frac{P(+)}{P(-)} LR(+)}{\frac{P(+)}{P(-)} LR(+) + 1} = \frac{\left(\frac{0,26}{1-0,26}\right) \left(\frac{1-0,018}{0,004}\right) \left(\frac{1-0,179}{0,012}\right)}{\left(\frac{0,26}{1-0,26}\right) \left(\frac{1-0,018}{0,004}\right) \left(\frac{1-0,179}{0,012}\right) + 1} = 0,9998 \quad (\text{E3.2})$$

V tomto příkladu jsou dostupné spolehlivé hodnoty prevalence (tj. apriorních pravděpodobností). Pokud by tyto nebyly po ruce, nebo kdyby je analytik raději nepoužil, mohly být místo nich použity věrohodnostní poměry; odpovídající hodnoty jsou 246 (EMIT) a ~ 68 (patentovaná), což dává kombinovaný věrohodnostní poměr přibližně 17 000 (rovnice (E3.3)), což podle tabulky 5 odpovídá „velmi silnému“ důkazu o přítomnosti metadonu.

$$LR(+) = \left(\frac{1-0,018}{0,004}\right) \left(\frac{1-0,179}{0,012}\right) = 1,7 \times 10^4 \quad (\text{E3.3})$$

Ve všech případech byla GC-MS použita jako referenční technika pro stanovení poměru falešných výsledků. Konkrétní databáze, na kterou se zde odkazuje, je poměrně obsáhlá pro studované analyty a byla navržena tak, aby poskytovala reprezentativní soubor, který umožňuje Bayesovskou analýzu dat. Nevyhnutelně některé hodnoty chybí, ale s přibývajícím údaji by se jejich počet měl snížit a přesnost předpovědi by se měla dále zlepšit.

Další výhodou databáze nastavené pro záznam reprezentativních dat pro několik různých technik jsou informace, které poskytuje, aby bylo možné optimalizovat analytickou výkonnost. Například výběr screeningové metody s nízkým poměrem falešně pozitivních výsledků by měl minimalizovat náklady na nákladné konfirmační analýzy. Je však třeba vzít v úvahu i další faktory, jako je mez detekce techniky, její poměr falešně negativních výsledků a rychlost a cena analýzy.

7.4 E4: Identifikace lidského genu SRY v biologickém materiálu pomocí qPCR

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích (fluorescence přesahující prahovou hodnotu)

Položka/matrice: Biologický materiál

Parametr/analyt: SRY gen (pohlaví určující oblast Y)

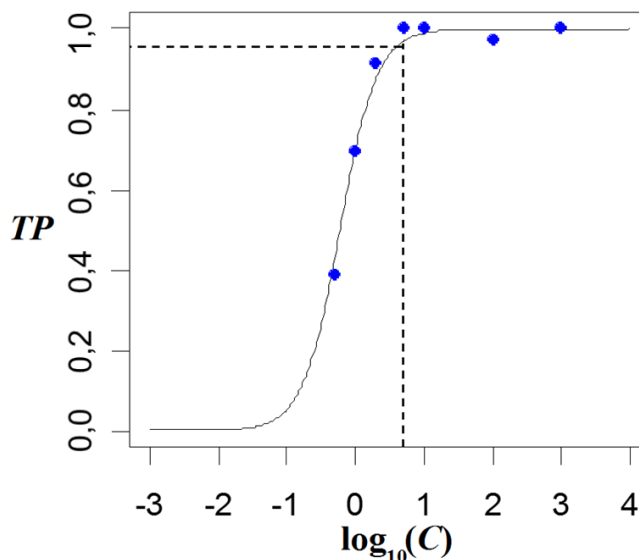
Typ klasifikačního kritéria: Neuvedeno

Technika/přístroj: Kvantitativní polymerázová řetězová reakce

Typ hlášení výsledků: Poměr pravdivě pozitivních výsledků.

Ačkoli interpretace qPCR zahrnuje zpracování instrumentálního signálu, v tomto příkladu je výkon hodnocen z určení TP při různých koncentracích DNA.

Obrázek E4.1 ukazuje některá experimentální data ze studie detekce lidského genu SRY pomocí kvantitativní PCR (qPCR) v biologickém materiálu [37]. Data byla generována z testu na 3 deskách s použitím 5'-nukleázového testu s duálně značenou fluorescenční sondou 'TaqMan' zaměřenou na oblast SRY specifické pro mužský lidský genom. Výsledek byl považován za „pozitivní“, pokud pozorovaná fluorescence překročila předem stanovený práh během 55 amplifikačních cyklů. Ačkoli byly jako součást validační studie použity různé preparativní metody, statistické testy neukázaly žádné významné rozdíly, takže data byla považována za jeden soubor dat.



Obrázek E4.1. Data detekce DNA zpracovaná logistickou regresí. Číslo detekce TP je vyneseno jako funkce $\log_{10}(C)$, kde C je číslo kopie genu ve studii schopnosti detekce a klasifikace DNA. Body ukazují podíl pozitivních výsledků z celkem 36 replikátů v každém čísle kopie. Plná čára ukazuje proložení logistické regrese s $b_0 = 0,85$ a $b_1 = 3,75$ (viz oddíl 3.3.6). Čárkovaná čára ukazuje $TP = 0,95$ a odpovídající log koncentraci $\log_{10}(C) = 0,56$ pro $C = 3,6$ kopií.

Všimněte si, že bod na $\log_{10}(C) = 2$ nevykazuje 100 % TP na obou stranách, přestože body na obou stranách ano, to vede k obtížnému správnému posouzení klasifikační schopnosti. Regresní křivka však umožňuje přiměřeně přesné umístění efektivní meze detekce, jak je znázorněno na obrázku; použití spojitého modelu účinně vyhladilo data náhodného počtu. To je značná výhoda; umožňuje analytikovi studovat velké množství

koncentrací s relativně malým počtem opakování na pouze jedné úrovni, místo aby vyžadoval mnoho opakování na několika úrovních, a přesto získat relativně spolehlivé odhady pravděpodobnosti.

Tento soubor dat ilustruje důležitou námitku při modelování. Údaje o koncentraci jsou vyneseny do grafu a modelovány v logaritmické doméně, což je běžná praxe při práci s experimentálními koncentracemi DNA nebo počty mikrobů. Pro závisle proměnnou je transformace často vyžadována rozdělením chyb. V tomto případě však neexistuje žádný pádný důvod pro volbu logaritmické transformace pro nezávisle proměnnou; výběr je v podstatě libovolný. Stejně tak i volba logistického modelu; jiné modely mohou také odpovídat údajům přiměřeně dobře. Tam, kde různé, ale hladké modely odpovídají datům podobně, není interpolace silně citlivá na výběr modelu. Extrémní pravděpodobnosti však mohou být skutečně velmi citlivé na výběr modelu. Z toho vyplývá, že i když model poskytuje dobrý popis dat a relativně spolehlivé pravděpodobnosti a meze detekce v tomto rozsahu, je velmi nebezpečné extrapolovat odhady pravděpodobnosti výrazně za studovaný rozsah bez podstatného důkazu o platnosti modelu.

7.5 E5: Identifikace reziduí pesticidů v potravinách pomocí GC-MS/MS na základě retenčního času a zastoupení iontů

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích

Položka/matrice: Potraviný rostlinného původu

Parametr/analyt: Methyl chlorpyrifos a malathion

Typ klasifikačního kritéria: Interval pro retenční čas a poměry intenzity charakteristických iontů

Technika/přístroj: GC-MS/MS

Typ hlášení výsledků: Věrohodnostní poměr a pravděpodobnost, že pozitivní výsledek bude správný

Tento příklad pojednává o odhadu *FP* vysoce selektivních stanovení reziduí pesticidů v potravinách pomocí GC-MS/MS s využitím modelování instrumentálního signálu a metody Monte Carlo. Simulace metody Monte Carlo byly provedeny v tabulkovém procesoru MS-Excel.

Identifikace analytů je založena na retenčním čase t_R v chromatografickém systému a na poměru zastoupení $AR = A_1/A_2$ dvou charakteristických iontů hmotnostního spektra analytu. Ačkoli má t_R přibližně normální rozdělení, AR se může výrazně odchylovat od normálního rozdělení. Je známo, že poměr korelovaných proměnných není normálně rozdělen, zejména pokud je proměnná s nižší precizností (tj. s větším rozptylem hodnot) ve jmenovateli [8].

Vývoj a validace postupu pro identifikaci stopových množství analytů v potravinách pomocí GC-MS/MS začíná definicí metody kvalitativní analýzy, včetně přípravy vzorku a podmínek GC-MS/MS. Specifikace podmínek GC-MS/MS zahrnuje výběr charakteristických iontů hmotnostního spektra analytu (methyl chlorpyrifosové ionty: 208 m/z a 271 m/z; malathionové ionty: 99 m/z a 127 m/z). Poté se provedou opakované nástřiky zásobních roztoků analytu a potravinových extraktů. Nástřiky zásobních roztoků analytu se používají ke studiu rozptylu a zastoupení iontů, A_1 a A_2 korelací každého analytu. Údaje o výkonnosti byly shromážděny při různých koncentracích analytu, protože hodnota a rozptýlení A_1 a A_2 se liší s koncentrací. Tabulka E5.1 uvádí souhrn výkonnostních parametrů t_R , A_1 a A_2 . Replikovaná analýza extraktů bez detekovatelných hladin analytu byla použita k definování modelů rozptýlení šumu signálu v okně retenčního času (tabulka E5.1). Byly studovány signály extraktů potravinářských výrobků reprezentujících nutriční vzorce potravin rostlinného původu s vysokým obsahem vody.

Z dat v tabulce E5.1 byly vyvinuty modely variability t_R a AR . Modely byly vytvořeny z konfidenčních intervalů založených na distribuci Studentova t rozdělení ($\bar{t}_{Ri} \pm t_{SiRi}$, kde \bar{t}_{Ri} a S_{iRi} představují průměr a směrodatnou odchylku t_{Ri} a t pochází z t -rozdělení pro definovanou konfidenční hladinu a stupně volnosti \bar{t}_{Ri} a S_{iRi}). Modely AR byly zkonstruovány ze simulací metody Monte Carlo. Z pozorovaných rozšíření různých A_1 a A_2 analytu byly simulovány poměry korelovaných A_1 a A_2 (tj. AR). Ze slepých extraktů byl simulován signál šumu a následně AR ve slepých pokusech. Šum signálu byl modelován jako normální rozdělení useknuté na nule, protože chromatografické píky nemají negativní oblasti. Tabulka E5.2 uvádí odhadovaný rozptyl t_R a AR . Tato tabulka také uvádí vzorce MS-Excelu použité při simulaci A_1 a A_2 . Konfidenční meze pro t_R a AR byly nastaveny na konfidenční úrovni 99,9 %, respektive 98 %, což odpovídá *TP*.

FP z identifikací založených na t_R byla stanovena na 10 % na základě zkušeností analytika. Toto *FP* představuje pravděpodobnost, že pík, u kterého není potvrzeno, že pochází z analytu, je pozorován v rámci definovaného okna retenčního času pro analyt. *FP* pro AR byl odhadnut ze simulací šumu signálu Monte Carlo a z určení, kolikrát tento simulovaný šum produkuje AR v rámci intervalu přijetí. Protože *FP* může být extrémně velký pro nízké hladiny analytu, byl stanoven při různých hmotnostních zlomcích analytu definováním minimálního zastoupení každého iontu. Tabulka E5.3 uvádí odhadované *TP*, *FP* a jejich kombinaci v *LR(+)*. V posledním sloupci tabulky je uvedena výkonnost identifikací na základě t_R a AR . Tabulka E5.3 také uvádí posteriorní pravděpodobnost *PP*, že zkoušená položka je pozitivní za předpokladu, že pozitivní nebo negativní výsledky jsou stejně pravděpodobné (tj. $P(+)=P(-)=0.5$).

Tabulka E5.1. Výkonnostní parametry relevantní pro identifikaci methyl chlorpyrifosu a malathionu v extraktech rostlinného původu pomocí GC-MS/MS. Všechny parametry byly odhadnuty s 11 stupni volnosti

Analyt: Methyl Chlorpyrifos								
Extrakt	w (mg kg ⁻¹)	Retenční čas		Zastoupení				ρ
		\bar{t}_{Ri} (min)	s_{tRi} (min)	Iont: 208 m/z		Iont: 271 m/z		
				\bar{A} (a.u.)	s_A (a.u.)	\bar{A} (a.u.)	s_A (a.u.)	
E §	3,33	17,24	0,024	10566	13,3	13867	6,14	0,9956
				8		8		
E §	0,33	17,24	0,024	10163	10,5	15025	8,10	0,6151
E §	0,083	17,24	0,024	4366	21,4	5790	15,1	0,3965
G	< 0,04	–	–	372	892	372	892	–
O	< 0,04	–	–	372	892	372	892	–
I	< 0,04	–	–	372	892	372	892	–

Analyt: Malathion								
Extrakt	w (mg kg ⁻¹)	Retenční čas		Zastoupení				P
		\bar{t}_{Ri} (min)	s_{tRi} (min)	Iont: 99 m/z		Iont: 127 m/z		
				\bar{A} (a.u.)	s_A (a.u.)	\bar{A} (a.u.)	s_A (a.u.)	
E §	3,33	19,45	0,070	22659	7,85	22676	10,3	0,9988
				2		5		
E §	0,33	19,45	0,070	22354	17,4	22969	15,6	0,9672
E §	0,083	19,45	0,070	5882	30,7	6345	28,0	0,7677
G	< 0,11	–	–	372	892	372	892	–
O	< 0,11	–	–	372	892	372	892	–
I	< 0,11	–	–	372	892	372	892	–

§ – rozptyl zastoupení iontů byl odhadnut kombinací signálů z analytu v čistém rozpouštědle se signály z rostlinných extraktů. Extrahované matrice: G – zázvor, O – jarní cibulka, I – řasa irský mech; E – nespecifikovaná matice.

w – hmotnostní zlomek analytu (mg kg⁻¹); \bar{t}_{Ri} – průměrný retenční čas (min) (tento parametr se může lišit podle dne nástřiku); s_{tRi} – směrodatná odchylka retenčního času odhadnutá za podmínek opakovatelnosti (min); \bar{A} – průměrné zastoupení iontů (libovolné jednotky, au); s_A – směrodatná odchylka zastoupení iontů; ρ – Spearmanův korelační koeficient.

Tabulka E5.2. Intervaly přijímané pro retenční čas a poměr zastoupení iontů

Analyt	Extrakt	Interval hmotnostních zlomků, w (mg kg ⁻¹)	Maximální rozdíl t_{Ri} (min) (c.l. 99,9 %) §	intervalAR (c.l. 98 %) †
Methyl chlorpyrifos	E	0,04 – 3,33	0,18	0,439 – 1,18
Malathion	E	0,11 – 3,33	0,54	0,467 – 1,54

E – nespecifikovaná matice, c.l. – konfidenční úroveň;

§ – Maximální rozdíl mezi retenčním časem analytu ve standardním roztoku a v analyzovaném vzorku ($\sqrt{2}ts_{tRi}$);

† – vzorec z MS-Excelu použitý v simulaci: První iont: $A_1 = \bar{A}_1 + s_{A1} * TINV(R1, v_1)$ a Druhý iont: $A_2 = \bar{A}_2 + s_{A2} * (TINV(R1, v_2) * \rho + TINV(R2, v_2) * (1 - \rho^2)^{0,5})$, kde v_i jsou stupně volnosti spojené s A_i a s_{A_i} a R1 a R2 dva nezávislé generátory náhodných hodnot U(0,1), vzorec v Excelu RAND()).

Podle tabulky E5.3, pouze pokud jsou identifikace prováděny na nebo nad mezí stanovitelnosti (0,14 mg kg⁻¹ nebo 0,38 mg kg⁻¹ pro methyl chlorpyrifos, respektive methyl malasparyrifos jsou podporovány „velmi silnými“ důkazy (tj. $10^5 < LR(+) < 10^6$). Všechny výkonnostní charakteristiky uvedené v Tabulce E5.3 (TP a FP a PP)

jsou validní alternativy pro uvádění výkonnosti nebo nejistoty kvalitativní analýzy při různých hmotnostních zlomcích analytu. Uvádění $LR(+)$ má tu výhodu, že kombinuje TP a FP v jedné metrice a nevyžaduje předpoklad prevalence pesticidu v analyzovaných vzorcích.

Tabulka E5.3. Výkonnostní charakteristiky identifikace methyl chlorpyrifosu a malathionu pomocí GC-MS/MS

Analyt	w (mg kg ⁻¹)	t_R	Výkonnostní charakteristiky na různých úrovních analytu, w			
			AR	AR		
Methyl chlorpyrifos	TP (%)	$\geq 0,04$	99,9	98	97,9	
	FP	0,04 §	10	30,2	3,02	
		0,08	10	0,2	0,02	
		0,14 †	10	0,001	0,0001	
		0,27	10	0,001	0,0001	
	$LR(+)$	0,04 §	9,99	3,24	32,4	
		0,08	9,99	490	4895	
		0,14 †	9,99	98000	979020	
		0,27	9,99	98000	929020	
	PP (%)	0,04 §	90,9	76,4	97,0	
		0,08	90,9	97,8	99,98	
		0,14 †	90,9	99,999	99,9999	
		0,27	90,9	99,999	99,9999	
	Malathion	TP (%)	$\geq 0,11$	99,9	98	97,9
		FP	0,11 §	10	29,8	2,98
			0,23	10	0,001	0,0001
0,38 †			10	0,001	0,0001	
0,77			10	0,001	0,0001	
$LR(+)$		0,11 §	9,99	3,29	32,9	
		0,23	9,99	98000	979020	
		0,38 †	9,99	98000	979020	
		0,77	9,99	98000	979020	
PP (%)		0,11 §	90,9	76,68	97,0	
		0,23	90,9	99,999	99,9999	
		0,38 †	90,9	99,999	99,9999	
		0,77	90,9	99,999	99,9999	

§ – mez detekce; † – mez stanovitelnosti; t_R – retenční čas analytu; AR – poměr četnosti dvou charakteristických iontů hmotnostního spektra analytu.

Tento příklad ilustruje, jak může simulace signálů Monte Carlo překonat obtížnost experimentálního stanovení falešně pozitivních poměrů vysoce selektivních identifikací.

7.6 E6: Identifikace SARS-CoV-2 RNA testováním amplifikace nukleové kyseliny

Účel

Typ kvalitativní analýzy: Analýza založená na kvantitativních kritériích

Položka/matrice: Výtěry z nosu, výtěry z nosohltanu a horní části hltanu

Parametr/analyt: SARS-CoV-2 RNA

Typ klasifikačního kritéria: Prahová hodnota cyklu, Ct, hodnoty rovné nebo nižší než Ct mez jsou klasifikovány jako pozitivní; vyšší hodnoty jsou klasifikovány jako negativní.

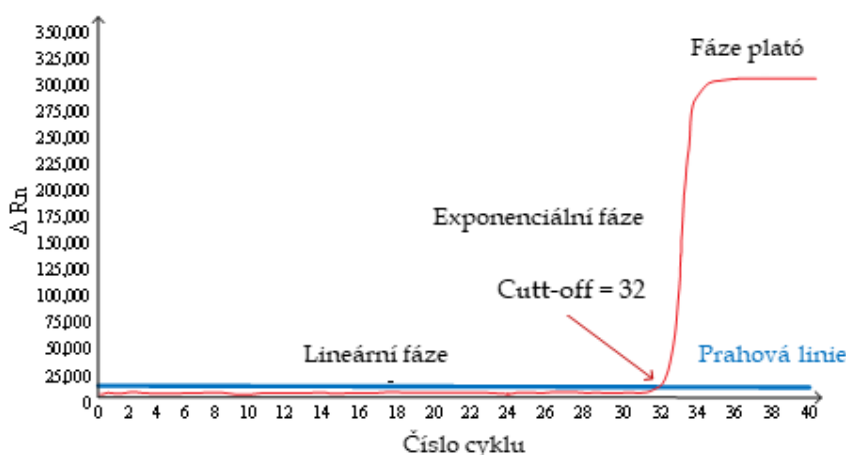
Technika/přístroj: Polymerázová řetězová reakce s reverzní transkripcí (RT-PCR)

Typ uvádění výsledků: Skutečný pozitivní poměr a skutečný negativní poměr.

Koronavirus těžkého akutního respiračního syndromu 2 (SARS-CoV-2) je virus, který způsobuje onemocnění COVID-19, respirační onemocnění odpovědné za pandemii COVID-19. Jeden ze screeningových testů na přítomnost tohoto viru v nosních výtěrech zahrnuje reverzní transkripční polymerázovou řetězovou reakci (RT-PCR), typ testování amplifikace nukleové kyseliny (NAAT). „Interní“ validace tohoto testu zahrnuje stanovení mezní hodnoty cyklu, LOD, (klinické) senzitivity (*SS*) a specifity (*SP*) s dalšími výkonnostními parametry (tabulka 2), (není uvedeno v tomto příkladu). V klinické analýze jsou *SS* a *SP* známé jako odhady klinické přesnosti. Klinická přesnost je omezena *FN* a *FP* a epidemiologickou prevalencí, typy a podtypy agens, mutacemi a dalšími biologickými faktory.

7.6.1 Prahová hodnota cyklu

„Počet cyklů amplifikace potřebných k tomu, aby se stal detekovatelným nad pozadím“ je definován jako prahová hodnota cyklu (*Ct*) [59] – počet cyklů potřebných k amplifikaci virové RNA k dosažení detekovatelné úrovně. Aby bylo možné porozumět aplikaci prahové hodnoty *Ct*, měly by být rozpoznány některé proměnné. *Rn* (normalizovaná reportovaná hodnota) je velikost signálu generovaného danou sadou podmínek PCR. ΔRn (obrázek E6.1) je normalizovaná reportovaná hodnota minus odpověď pozadí. Prahová hodnota je úroveň signálu, která odráží statisticky významný nárůst oproti vypočtenému signálu pozadí (viz obrázek E6.1). Tato rozhodovací čára je určena k rozlišení relevantních amplifikačních signálů od pozadí. V příkladu software nastaví práh na 10násobek standardní odchylky základní hodnoty fluorescence. Mez je definována v oblasti související s exponenciálním růstem produktu PCR. Obrázek E6.1 ilustruje pozitivní klasifikaci (menší nebo rovna *Ct* prahové hodnotě 32) lidského vzorku. Výsledky fluorescence vyšší než 32 jsou klasifikovány jako negativní. Všimněte si, že pouze sigmoidální amplifikační křivky vypovídají o skutečné amplifikaci (viz obrázek E6.1).



Obrázek E6.1. Detekce RNA SARS-CoV-2 ze skutečně pozitivního vzorku

7.6.2 Odhad LOD

Schopnost detekce blízko prahové hodnoty Ct se vyhodnocuje pomocí *LOD*. *LOD* je definována jako násobek koncentrace (například vyjádřený v počtu kopií/ml) spojený s TP z 95 % ($LOD_{95\%}$). Hodnota $LOD_{95\%}$ je odhadnutá modelováním rozptýlení *TP* v závislosti na koncentraci a odhadem koncentrace tam, kde *TP* je 95 % užitím probitové regrese [60] – [63]. V příkladu je $LOD_{95\%}$ odhadnutá v sérii sedmi ředění ze vzorku o koncentraci 500 kopií/ml je 114 kopií/ml. Více podrobností o předloženém postupu pro stanovení $LOD_{95\%}$ je k dispozici v bibliografii (5.5 z [63]).

7.6.3 Klinická přesnost

Posouzení klinické přesnosti zahrnuje definování cílových nebo minimálních hodnot pro spodní meze 95 % konfidenčního intervalu *SS* a *SP*, $LL_{SS,95}^{tg}$ a $LL_{SP,95}^{tg}$, zda jsou meze stejné nebo vyšší než příslušná cílová hodnota (tj. pokud $LL_{SS,95} \geq LL_{SS,95}^{tg}$ a $LL_{SP,95} \geq LL_{SP,95}^{tg}$). Pro tento test by měly být $LL_{SS,95}^{tg} = 95\%$ a $LL_{SP,95}^{tg} = 90\%$, což je spodní mez klinické citlivosti větší než 95 % a spodní mez specifčnosti větší než 90 %.

Pro validaci metody bylo analyzováno 200 vzorků výtěrů z nosu, nasofaryngeálních a orofaryngeálních výtěrů: 100 od jedinců, o nichž je známo, že jsou infikováni SARS-CoV-2, a 100 od jedinců, u kterých bylo potvrzeno, že nejsou tímto virem infikováni. Tabulka E6.1 ukazuje kontingenční tabulku získanou z 200 testů. Tato tabulka ukazuje, že nebyly zaznamenány žádné falešně negativní a tři falešně pozitivní výsledky, takže klinické *SS* a *SP* jsou 100 % resp. 97 %. Meze 95 % CI *SS* a *SP*, vypočtené z rovnic (8) až (11), jsou [96,3; 100] a [91,6, 99,0] (tabulka E6.2). Protože $LL_{SS,95}$ a $LL_{SP,95}$ jsou vyšší než 95 %, resp. 90 %, je analytická metoda považována za validní. *SS* je doplněna sérokonverzní senzitivitou [64].

Tabulka E6.1. Kontingenční tabulka, která shrnuje účinnost metody pro detekci SARS-Cov-2 RNA ve výtěrech z nosu, nasofaryngeálních a orofaryngeálních výtěrech

		Případ		Celkové výsledky
		Pozitivní (<i>pc</i>)	Negativní (<i>nc</i>)	
Výsledek	Pozitivní (<i>pc</i>)	$tp = 100$	$fp = 3$	$p = 103$
	Negativní (<i>nc</i>)	$fn = 0$	$tn = 97$	$n = 97$
Celkový počet případů		$pc = 100$	$nc = 100$	

Tabulka E.6.2. Klinická přesnost metody detekce SARS-Cov-2 RNA v lidském séru nebo plazmě

Klinická citlivost		
$SS = 100\%$	$LL_{SS,95} = 96,3\%$	$HL_{SS,95} = 100\%$
Klinická specifčnost		
$SP = 97\%$	$LL_{SP,95} = 91,6\%$	$HL_{SP,95} = 99,0\%$

Úmyslně ponecháno prázdné

Příloha A – Bayesova věta, šance a věrohodnostní poměr

A.1 Bayesova věta

Bayesova věta popisuje, jak se pravděpodobnost události A (jako zkoušená položka, která je skutečně pozitivní) mění s novou informací E , jako je pozitivní výsledek zkoušky (testu). Bayesova věta se nejčastěji uvádí pro dvě události, A a E , jako:

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)} \quad (\text{A.1})$$

Zde $P(A)$ a $P(E)$ jsou pravděpodobnosti událostí A a E , které se vyskytnou samostatně, $P(E|A)$ je pravděpodobnost události E , která nastala, $P(E|A)$ je pravděpodobnost události E , za předpokladu že nastala A . Ve statistice se $P(A|E)$ a $P(E|A)$ obvykle označují jako „podmíněné pravděpodobnosti“, například $P(E|A)$ může být označováno jako podmíněná pravděpodobnost události E dané A .

V kontextu kvalitativní analýzy, vezmeme-li jako příklad pozitivní případ, lze $P(A)$ chápat jako pravděpodobnost, že náhodně vybraná zkoušená položka je skutečně pozitivní před provedením jakýchkoli zkoušek. $P(E|A)$ je pravděpodobnost, že skutečně pozitivní testovaná položka vygeneruje pozitivní výsledek testu – skutečný pozitivní poměr TP v tabulce 2. $P(E)$ je pravděpodobnost pozitivního výsledku zkoušky bez ohledu na stav testovaného předmětu. A konečně, $P(A|E)$ je pravděpodobnost, že zkoušená položka je skutečně pozitivní, po zvážení informace přidané pozitivním výsledkem zkoušky. Protože se vypočítává poté, co budou k dispozici důkazy E , $P(A|E)$ se obvykle nazývá „posteriorní pravděpodobnost“ pro A . Odhadovaná posteriorní pravděpodobnost poskytuje přímou indikaci spolehlivosti, kterou lze klasifikovat.

Je důležité si pamatovat, že (pokračujeme pozitivním případem) $P(E)$ zahrnuje jak pravdivé pozitivní, tak falešně pozitivní výsledky, a také to, že $P(E)$ platí pro kompletní populaci zkoušených položek. To znamená, že $P(E)$ je citlivá jak na poměr pravdivé a falešné positivity, tak na poměry skutečně pozitivních a negativních zkoušených položek. Kvantitativně lze pro dva případy A a $\neg A$ (označující „Ne- A “, skutečně negativní zkoušenou položku), $P(E)$ zapsat jako vážený součet:

$$P(E) = P(A)P(E|A) + P(\neg A)P(E|\neg A) \quad (\text{A.2})$$

Pokud vezmeme v úvahu E jako pozitivní výsledek zkoušky, rovnice A.2 říká, že kombinovaná pravděpodobnost E je skutečný pozitivní poměr krát podíl skutečně pozitivních vzorků plus poměr falešně pozitivních krát podíl skutečně negativních vzorků. To je důvod, proč vysoký poměr falešně pozitivních výsledků snižuje důvěru v pozitivní výsledky zkoušky. S odkazem na rovnici A.1, pokud existuje vysoká pravděpodobnost pozitivních výsledků ze skutečně negativních zkoušených položek, $P(A|E)$ se snižuje, protože $P(E)$ se zvyšuje. To odpovídá intuici; jakkoli vysoká je skutečně pozitivní míra, pravděpodobnost velkého počtu falešně pozitivních výsledků by nám měla snížit jistotou, že pozitivní výsledek ukazuje na skutečně pozitivní testovanou položku.

A.2 Pravděpodobnost a šance

Pravděpodobnost P se obvykle vyjadřuje jako číslo mezi 0 a 1. Lze ji však vyjádřit také ve formě „šance“⁶ (angl. odds), což je termín snad nejnámější ve sportovním sázení. Pokud je pravděpodobnost události A $P(A)$ a alternativní možnost je jednoduše „Ne A “, lze šanci $O(A)$ ve prospěch A vypočítat pomocí:

$$O(A) = \frac{P(A)}{1 - P(A)} \quad (\text{A.3})$$

Na rozdíl od pravděpodobností mohou šance nabývat jakékoli nezáporné hodnoty; možná je šance 10^6 nebo „milion ku jedné“.

Šance lze převést zpět na pravděpodobnosti přeskupením A.3 tak, aby:

⁶ Výraz „odds“ (šance) je obecně považován za množné číslo.

$$P(A) = \frac{O(A)}{O(A) + 1} \quad (\text{A.4})$$

A.3 Forma Bayesovy věty pro šance a věrohodnostní poměr

Pokud existují pouze dvě alternativní a komplementární hypotézy, A a $\neg A$ (tj. „Ne A “), a nějaký důkaz E (jako je výsledek zkoušky, který je pozitivní pro A) použije se k aktualizaci pravděpodobností každé z nich, Bayesova věta dává posteriorní pravděpodobnosti jako:

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)} \quad (\text{A.5a})$$

$$P(\neg A|E) = \frac{P(E|\neg A)P(\neg A)}{P(E)} \quad (\text{A.5b})$$

Poměr jejich pravděpodobností je pak:

$$\frac{P(A|E)}{P(\neg A|E)} = \frac{P(E|A)P(A)}{P(E|\neg A)P(\neg A)} \quad (\text{A.6})$$

nebo pro srozumitelnost lze oddělit výrazy

$$\frac{P(A|E)}{P(\neg A|E)} = \frac{P(E|A)}{P(E|\neg A)} \times \frac{P(A)}{P(\neg A)}$$

Protože existují pouze dvě hypotézy, A a $\neg A$, apriorní pravděpodobnosti a posteriorní pravděpodobnosti musí být sečteny 1; tedy $P(\neg A) = 1 - P(A)$ a $P(\neg A|E) = 1 - P(A|E)$. To znamená, že levá strana A.6 je rovna $P(A|E)/[1 - P(A|E)]$. Ve srovnání s A.3 je to jen šance ve prospěch A , daná E , nebo $O(A|E)$, „posteriorní šance“ ve prospěch hypotézy A . Podobně se apriorní šance $O(A)$ objeví na pravé straně A.6 jako $P(A)/P(\neg A) = P(A)/[1 - P(A)] = O(A)$. Zbývající poměr, $P(E|A)/P(E|\neg A)$, je znám jako „věrohodnostní poměr“. Pro případ kvalitativní analýzy, kde $\neg A$ odpovídá skutečně negativní zkoušené položce, uvádí tabulka 5 (odhadovaný) věrohodnostní poměr jako TP/FP .

Forma Bayesovy věty pro šance tedy může být zapsána jako

$$O(A|E) = O(A) \times \frac{P(E|A)}{P(E|\neg A)} \quad (\text{A.7})$$

nebo schematicky:

$$\text{posteriorní šance} = \text{apriorní šance} \times \text{věrohodnostní poměr}$$

Věrohodnostní poměr lze tedy kvantitativně interpretovat jako změnu šancí ve prospěch konkrétní hypotézy.

Příloha B – Kvalitativní analýza spojená s posouzením shody a mezi stanovitelnosti

B.1 Posouzení shody jako kvalitativní analýza

Posouzení shody hodnoty kvantitativního parametru analyzované položky s mezní hodnotou nebo intervalem lze považovat za kvalitativní analýzu za použití jediného kvantitativního kritéria (oddíl 2) s výsledky „vyhovují“ nebo „nevyhovují“. Tabulka B.1 uvádí některé příklady těchto typů analýz.

Tabulka B.1. Příklady kvalitativní analýzy založené na posouzení shody hodnoty kvantitativního parametru analyzované položky s mezní hodnotou nebo intervalem

-
- 1 Hodnocení barvy suroviny porovnáním měření absorbance s hraniční hodnotou.
 - 2 Posouzení shody slitiny s minimálním obsahem její hlavní složky.
 - 3 Posouzení shody léku se specifikačním intervalem pro koncentraci účinné látky.
 - 4 Posouzení shody reziduí pesticidů v ovoci s ohledem na maximální meze reziduí.
 - 5 Posouzení zdravotního stavu jedince porovnáním naměřené krevní složky s intervalem hodnot zdravých jedinců.
-

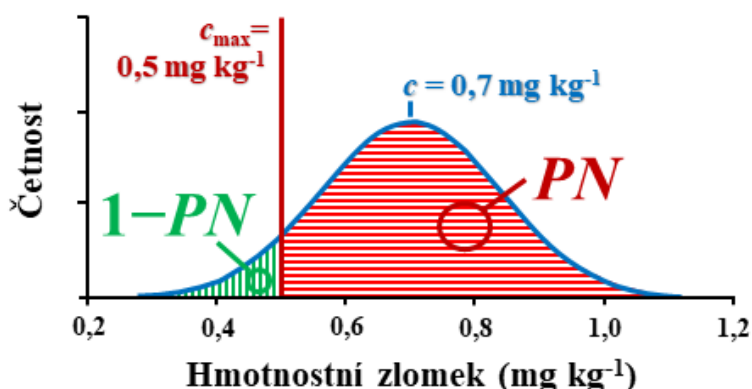
Použití rozhodovacích pravidel a nejistoty měření při posuzování shody je podrobně diskutováno v příručce Eurachem/CITAC „Použití informací o nejistotě při posuzování shody“ [29] („Příručka shody“). Pro úplnost však tato příloha pojednává o tom, jak lze informace o nejistotě nebo výkonnosti pro kvantitativní analýzu použít k odvození některých metrik v tabulce 2. Ty pak mohou být použity k charakterizaci výkonnosti postupů kvalitativní analýzy založených, zcela nebo částečně, na porovnání výsledků měření s mezí nebo specifikací.

Pokud analýza zahrnuje posouzení, zda je měřená vlastnost nad, pod nebo v rámci specifikační meze nebo intervalu, lze nejistotu měření použít ke kvantifikaci spolehlivosti posouzení shody.

POZNÁMKA: Tato příručka nepojednává o tom, jak by měla být vyhodnocena nejistota měření. Vyhodnocení nejistoty měření je podrobně popsáno v příručce Eurachem/CITAC „Stanovení nejistoty analytického měření“ [65].

Použití nejistoty měření pro rozhodování o shodě popsané v příručce shody Eurachem/CITAC [29] zahrnuje stanovení kritéria pro rozhodování, zda položka vyhovuje nebo nevyhovuje, s maximální pravděpodobností chybných rozhodnutí o shodě x %. Příručka shody rozlišuje „specifická“ a „globální“ rizika. „Specifické riziko“ kvantifikuje pravděpodobnost chybného rozhodnutí o shodě konkrétní položky; je založeno výhradně na rozdělení spojeném s výsledkem měření pro danou položku. Naproti tomu „globální riziko“ kvantifikuje pravděpodobnost chybných rozhodnutí o shodě náhodně vybrané budoucí položky [66]. Globální riziko bere v úvahu rozložení možných hodnot pro měřené položky, jako je rozložení hodnot položek z výrobní linky nebo oblasti životního prostředí. Například výpočet globálního rizika výrobce vyžaduje pravděpodobnost, že výrobní linka bude vyrábět produkty s hodnotou blízkou mezní hodnotě, takže je lze mylně považovat za nevyhovující. Proto pro stanovení globálního rizika musí být dobře charakterizováno rozdělení hodnot pro populaci položek.

Nejčastěji se analytici zajímají o posouzení shody konkrétní analyzované položky. Jak lze v takových případech určit metriky používané ke kvantifikaci spolehlivosti jiných typů kvalitativní analýzy? Případová studie a vzorce použité pro stanovení těchto parametrů jsou uvedeny níže.



Obrázek B.1. Grafické znázornění pravděpodobnosti, že analyzovaná položka je nevyhovující vůči maximální mezí c_{max} z měřené hodnoty c s přidruženou standardní nejistotou, $u(c) = 0,14 \text{ mg kg}^{-1}$ a odpovídající pravděpodobností $1 - PN$, že je nevyhovující.

B.2 Pozitivní a negativní výsledky posouzení shody

Je-li položka považována za vyhovující z „pozitivního výsledku“, lze rozdělení popisující nejistotu měření použít k poskytnutí buď věrohodnostního poměru $LR(+)$ ve prospěch shody, nebo za určitých okolností, posteriorní pravděpodobnosti PP položky skutečně shodné („pozitivní případ“) a pravděpodobnosti, že položka je ve skutečnosti neshodná („negativní případ“) ($1 - PP$). Stejně tak, pokud je výsledek hlášen jako „neshodný“ nebo „negativní“ ve srovnání s mezí nebo intervalem, lze nejistotu měření použít k získání odpovídajícího věrohodnostního poměru $LR(-)$ nebo posteriorních pravděpodobností PN a ($1 - PN$). Níže uvedený příklad uvažuje posteriorní pravděpodobnosti.

POZNÁMKA: Definice „pozitivních výsledků“ a „negativních výsledků“ jako „vyhovující“ a „nevyhovující“ je libovolná; lze použít opačnou konvenci.

B.3 Příklad – Posouzení shody pro rezidua pesticidů v ovoci

Předpokládejme, že shoda vzorku hroznů je posuzována vůči maximálnímu obsahu reziduí $c_{max} = 0,5 \text{ mg kg}^{-1}$ pro acetamiprid a naměřený hmotnostní zlomek ve vzorku je $c = 0,70 \text{ mg kg}^{-1}$. Výsledek měření má normální rozdělení se standardní nejistotou, $u(c) = 0,14 \text{ mg kg}^{-1}$, odhadnuté s velmi vysokým počtem stupňů volnosti⁷. Protože $c > c_{max}$ je nejpravděpodobnějším závěrem o shodě 'neshoda' hroznů ('negativní výsledek') [67].

Formálně posteriorní pravděpodobnost, jako je PP nebo PN vyžaduje apriorní pravděpodobnost. V tomto případě nejsou k dispozici žádné informace o obecném rozdělení acetamipridu v hroznech. V případech podezření na kontaminaci je však často rozumné předpokládat, že rozdělení je tak široké, že v oblasti výsledku měření je v podstatě neinformativní. V takovém případě lze nejistotu měření brát jako aproximaci posteriorního rozdělení. Aplikujeme-li zde tento přístup, posteriorní pravděpodobnost negativního (nevyhovujícího) zkušební vzorku PN , je plocha pod funkcí hustoty pravděpodobnosti výsledku měření napravo od c_{max} , znázorněná na obrázku B.1 červeně. Plocha je horní koncová pravděpodobnost normálního rozdělení se střední hodnotou c a směrodatnou odchylkou rovnou $u(c)$. To lze vypočítat v tabulkovém procesoru nebo statistickém balíčku, v aplikaci Microsoft Excel je například požadovaný vzorec $1 - \text{NORM.DIST}(c_{max} = 0,5, c = 0,7, u(c) = 0,14, \text{TRUE})$ (viz tabulka B.2). Pro tento příklad je plocha

⁷ Symbol c pro koncentraci se v této příručce používá pro případy použitelné pro různé typy veličin, jako je hmotnostní koncentrace, hmotnostní zlomek a pH.

0,923 nebo 92,3 %. Odpovídající pravděpodobnost PP , že vzorek je pozitivní (vyhovující), je $(1 - PN)$ nebo 7,6 %.

Pokud jsou k dispozici informativní předběžné informace a je vhodné je využít, zahrnují výpočty integraci přes apriorní rozdělení. Integrály pro normálně rozdělené apriorní nejistoty měření jsou uvedeny například v JCGM 106 [66] spolu s pokyny pro další rozdělení.

POZNÁMKA Může být překvapivě obtížné stanovit skutečně neinformativní apriorní rozdělení. Například pro jednoduché rovnoměrné rozdělení v rozsahu podle obrázku B.1 ($0,2 - 1,2 \text{ mg kg}^{-1}$) by odpovídala 30% apriorní pravděpodobnosti, že zkoušená položka je shodná s mezí, jednoduše proto, že pouze 30 % z toho rozmezí je pod mezí $0,5 \text{ mg kg}^{-1}$. Kompletní Bayesovská analýza proto bude zahrnovat kontrolu, obvykle zahrnující alternativní volby apriorní distribuce k ujištění, že závěr není příliš citlivý na předpokládané apriorní rozdělení.

B.4 Vzorce v tabulkovém procesoru pro posouzení shody pravděpodobností

Tabulka B.2 uvádí vzorce z MS-Excelu, které by se měly použít, když jsou uvažovány různé meze shody nebo intervaly a naměřená hodnota je pod, nad, uvnitř nebo mimo mez(e).

Pokud je standardní nejistota $u(c)$ odhadnuta s malým počtem stupňů volnosti, v , místo popisu rozptylu odhadu měřené veličiny normálním rozdělením, je třeba zvážit použití Studentova t rozdělení. V takovém případě by obecný vzorec použitý v tabulce B.2, $\text{NORM.DIST}(C,c,u(c),\text{TRUE})$, měl být nahrazen $\text{TDIST}(\text{ABS}(C - c)/u(c),v,\text{TRUE})$.

Tabulka B.2. Vzorce v MS Excelu používané k odhadu pravděpodobnosti shody, PP , nebo neshody, PN , rozhodnutí o konkrétní analyzované položce. Vzorce lze použít k výpočtu $LR(+)=PP/(1-PP)$ a $LR(-)=PN/(1-PN)$.

S	Mez	Shodnost výsledku Scénář	položky (typ	Spolehlivost shody	Vzorec MS-Excel (založený na kumulativním normálním rozdělení)
1	Max.	Vyhovující (pozitivní) $c \leq c_{\max}$		PP	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE})$
2	Max.	Nevyhovující (negativní) (%) c		PN	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE})$
3	Min.	Vyhovující (pozitivní) $c \geq c_{\min}$		PP	$\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$
4	Min.	Nevyhovující (negativní) (%) c		PN	$\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$
5	Inter val	Vyhovující (pozitivní) $c_{\min} \leq c \leq c_{\max}$		PP	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE}) -$ $\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$
6	Inter val	Nevyhovující (negativní) $c > c_{\max}$ nebo $c < c_{\min}$		PN	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE}) -$ $\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$

S – scénář; Inter. Max. nebo Min. – interval maximální nebo minimální meze; pozitivní nebo negativní výsledek – vyhovující nebo nevhovující výsledek c a $u(c)$ – měřená koncentrace a související standardní nejistota; c_{\max} nebo c_{\min} – maximální nebo minimální přípustná koncentrace

Literatura

- [1] JCGM, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (3rd edn.) (JCGM 200:2012), Sevres: BIPM, 2012.
- V češtině vydáno jako: TNI 010115 Mezinárodní metrologický slovník - Základní a všeobecné pojmy a přidružené termíny (VIM). ÚNMZ 2009. Zdarma dostupné na: <https://www.unmz.cz/metrologie/slovníky/>.
- [2] W. G. D. Ruig, G. Dijkstra and R. W. Stephany, "Chemometric criteria for assessing the certainty of qualitative analytical methods," *Anal. Chim. Acta*, pp. 277-282, 1989.
- [3] B. L. Milman and L. A. Konopelko, "Identification of Chemical Substances by Testing and Screening of Hypotheses. I. General," *Fresenius. J. Anal. Chem.*, vol. 367, pp. 621-628, 2000.
- [4] S. L. R. Ellison and S. Gregory, "Quantifying uncertainty in qualitative analysis," *Analyst*, vol. 123, pp. 1155-1161, 1998.
- [5] S. L. R. Ellison and S. L. Gregory, "Predicting chance infrared spectroscopic matching frequencies," *Anal. Chim. Acta*, vol. 370, pp. 181-190, 1998.
- [6] S. L. R. Ellison, "Uncertainties in qualitative testing and analysis," *Accred. Qual. Assur.*, vol. 5, pp. 346-348, 2000.
- [7] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhart, R. W. Stephany, A. Townshend, A. Zschunke and M. Valcárcel, "Quality assurance of qualitative analysis in the framework of the European project 'MEQUALAN'," *Accred. Qual. Assur.*, vol. 8, pp. 68-77, 2003.
- [8] R. B. Silva, "Evaluation of trace analyte identification in complex matrices by low-resolution gas chromatography - mass spectrometry through signal simulation," *Talanta*, vol. 150, pp. 553-567, 2016.
- [9] J. Narciso, C. Luz and R. B. d. Silva, "Assessment of the Quality of Doping Substances Identification in Urine by GC/MS/MS," *Anal. Chem.*, vol. 91, no. 10, pp. 6638-6644, 2019.
- [10] ISO, Reference materials - Examples of reference materials for qualitative properties (ISO/TR 79:2015), Geneva: ISO, 2015.
- [11] P. Pereira, B. Magnusson, E. Theodorsson, J. O. Westgard and P. Encarnaçao, "Measurement uncertainty as a tool for evaluating the 'grey zone' to reduce the false negatives in immunochemical screening of blood donors for infectious diseases," *Accred. Qual. Assur.*, vol. 21, pp. 25-32, 2016.
- [12] ILAC, ILAC Guidelines for Measurement Uncertainty in Testing (ILAC G17:01), Silverwater: ILAC, 2021.
- V češtině vydáno jako: ILAC-G17: 01/2021 Pokyny pro stanovení nejistoty měření ve zkoušení, ČIA 2021.
- [13] ISO, IEC, General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025), Geneva: ISO, 2017.
- V češtině vydáno jako: ČSN EN ISO/IEC 17025 Všeobecné požadavky na kompetenci zkušebních a kalibračních laboratoří. ÚNMZ 2018.
- [14] ISO, Medical laboratories – Requirements for quality and competence (ISO 15189), Geneva: ISO, 2012.
- V češtině vydáno jako: ČSN EN ISO 15189 Zdravotnické laboratoře - Požadavky na kvalitu a způsobilost. ÚNMZ 2013.

- [15] R. Bramley, A. Brown, S. Ellison, W. Hardcastle and A. Martin, "Qualitative analysis: A guide to best practice - forensic science extension," *Sci. Justice*, vol. 3, no. 40, pp. 163-170, 2000.
- [16] L. Wide and C. A. Gemzell, "An immunological pregnancy test," *Acta Endocrinol. (Copenh.)*, no. 35, pp. 261-267, 1960.
- [17] U. Forsum, H. O. Hallander, A. Kallner and D. Karlsson, "The impact of qualitative analysis in laboratory medicine," *Trends Anal. Chem.*, vol. 6, no. 24, pp. 546-555, 2005.
- [18] P. Pereira, Quality control of qualitative tests for medical laboratories, Lisbon: Author-edition, 2019.
- [19] G. Nordin, R. Dybkaer, U. Forsum, X. Fuentes-Arderiu and F. Pontet, "Vocabulary on nominal property, examination, and related concepts for clinical laboratory sciences (IFCC-IUPAC Recommendations 2017)," *Pure Appl. Chem.*, vol. 5, no. 90, pp. 913-935, 2018.
- [20] EU, Commission decision implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2002/657/EC), EU, 2002.
- [21] EU, Commission Regulation No 2017/644 of 5 April 2017 laying down methods of sampling and analysis for the control of levels of dioxins, dioxin-like PCBs and non-dioxin-like PCBs in certain foodstuffs and repealing Regulation (EU) No 589/2014, EU, 2017.
- [22] SANTE, Guidance document on analytical quality control and method validation procedures for pesticide residues and analysis in food and feed (SANTE/12682/2019), DG SANTE, 2019.
- [23] WADA, Technical Document – TD2010IDCR, Identification Criteria for qualitative assays incorporating column chromatography and mass spectrometry, WADA, 2010.
- [24] JCGM, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (JCGM 100:2008), Sèvres: BIPM, 2008.
- [25] ISO/IEC, Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)(ISO/IEC Guide 98-3), Geneva: ISO, 2008.
V češtině vydáno jako: TNI 014109-3 Nejistoty měření - Část 3: Pokyn pro vyjádření nejistoty měření (GUM 1195) (Pokyn ISO/IEC 98-3), ÚNMZ 2011.
- [26] ISO, IEC, General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025)(superseded), Geneva: ISO, 1999.
V češtině vydáno jako: ČSN EN ISO/IEC 17025 Všeobecné požadavky na kompetenci zkušebních a kalibračních laboratoří. ÚNMZ 2001.(Zrušena 2005)
- [27] Clinical Laboratory and Standards Institute, EP12-A2 User protocol for evaluation of qualitative test performance, 2nd ed., Wayne (PA): CLSI, 2008.
- [28] N. R. Campbell, Physics, the elements, Cambridge: Cambridge University Press, 1920.
- [29] A. Williams and B. Magnusson, (Eds.), Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment, 2nd ed., Eurachem, 2021.
- [30] B. Magnusson and U. Örnemark, (Eds.), Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics (2nd Edn.), Eurachem, 2014.
V češtině vydáno jako: Suchánek M., Milde D. (eds.): Kvalimetrie 20: Vhodnost analytických metod pro daný účel. Eurachem-ČR, Praha 2015.
- [31] A. Agresti, Categorical Data Analysis (3rd Ed.), New Jersey: Wiley, 2012.

- [32] AOAC International, Official Methods of Analysis of AOAC International (20th Edn.), Maryland: AOAC International, 2016.
- [33] S. D. Ferrara, L. Tedeschi, G. Frison, G. Brusini and F. Castagna, “Drugs-of-Abuse Testing in Urine: Statistical Approach and Experimental Comparison of Immunochemical and Chromatographic Techniques,” *J. Anal. Toxicol.*, vol. 18, no. 5, pp. 278-291, 1994.
- [34] R. J. Freund and W. J. Wilson, Regression Analysis, San Diego, CA: Academic Press, 1998.
- [35] J. Fox, An R and S-Plus companion to applied regression, Thousand Oaks, CA.: Sage Publications Inc., 2002.
- [36] S. L. R. Ellison and T. Fearn, “Characterising the performance of qualitative analytical methods: Statistics and terminology,” *Trends Anal. Chem.*, vol. 24, pp. 468-476, 2005.
- [37] S. L. R. Ellison, C. A. English, M. J. Burns and J. T. Keer, “Routes to improving the reliability of low level DNA analysis using real-time PCR,” *BMC Biotechnology*, vol. 6, pp. 33 (1-11), 2006.
- [38] I. Kuselman and F. Pennechi, “IUPAC/CITAC Guide: Classification, modelling and quantification of human errors in chemical analytical laboratory (IUPAC Technical Report),” *Pure Appl. Chem.*, vol. 88, pp. 477-515, 2016.
- [39] EU, Commission Regulation No 589/2014 of 2 June 2014 laying down methods of sampling and analysis for the control of levels of dioxins, dioxin-like PCBs and non-dioxin-like PCBs in certain foodstuffs and repealing Regulation No 252/2012, EU, 2014.
- [40] EU, Commission Regulation No 152/2009 of 27 January 2009 laying down the methods of sampling and analysis for the official control of feed, EU, 2009.
- [41] T. Wenzl, J. Haedrich, Schaechtele, Alexander, P. Robouch and J. Stroka, Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food, JRC, 2016.
- [42] J. Vessman, R. I. Stefan, J. F. Van Staden, K. Danzer, W. Lindner, D. T. Burns, A. Fajgelj and H. Müller, “Selectivity in Analytical Chemistry (IUPAC Recommendation 2001),” *Pure Appl. Chem.*, vol. 73, no. 8, p. 1381–1386, 2001.
- [43] ISO, In vitro diagnostic medical devices – Information supplied by the manufacturer (labelling) – Part 1: Terms, definitions and general requirements (ISO 18113-1), Geneva: ISO, 2009.
V češtině vydáno jako: ČSN EN ISO 18113-1. Diagnostické zdravotnické prostředky in vitro - Informace poskytované výrobcem (označování štítky) - Část 1: Termíny, definice a obecné požadavky (2012).
- [44] European Network of Forensic Science Institutes, ENFSI Guideline for evaluative reporting in forensic science, ENFSI, 2015.
- [45] V. Morgado, C. Palma and R. J. N. Bettencourt da Silva, “Microplastics identification by Infrared spectroscopy – Evaluation of identification criteria and uncertainty by the Bootstrap method,” *Talanta*, vol. 224, p. 121814, 2021.
- [46] A. J. Nunes, P. Paixão, J. Proença and R. J. N. Bettencourt da Silva, “Early warning of suspected doping from Biological Passport based on multivariate trends,” *Int. J. Sports Med.*, vol. 41, pp. 44-53, 2020.
- [47] N. Pinto, M. Magalhães, E. Conde-Sousa, C. Gomes, R. Pereira, C. Alves, L. Gusmão and A. Amorim, “Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers,” *Forensic Sci. Int. Genet.*, vol. 7, pp. 16-21, 2013.

- [48] B. Meijer, J. Thijs, J. Kleibeuker, A. van Zwet and R. Berrelkamp, "Evaluation of eight enzyme immunoassays for detection of immunoglobulin G against *Helicobacter pylori*," *J. Clin. Microbiol.*, vol. 35, no. 1, pp. 292-294, 1997.
- [49] B. G. Armitage P, *Statistical methods in medical research*, 3rd ed., Cambridge: Blackwell Science, 1994.
- [50] D. Zwillinger and S. Kokoska, *Standard probability and statistics tables and formulae*, Boca Raton (FL): Chapman & Hall/CRC, 2000.
- [51] C. B. Agresti A, "Approximate is better than "exact" for interval estimation of binomial proportions," *Am. Stat.*, vol. 52, no. 2, pp. 119-126, 1998.
- [52] D. Altman, D. Machin, T. Bryant and M. Gardner, *Statistics with confidence*, 2nd ed., M. D. B. T. G. M. Altman DA, Ed., London: BMJ Books, 2000.
- [53] R. Newcombe, "Two-sided confidence intervals for the single proportion: comparison of seven methods," *Stat. Med.*, vol. 17, no. 8, pp. 857-872, 1998.
- [54] E. Wilson, "Probable inference, the law of succession, and statistical inference," *JASA*, vol. 22, no. 158, pp. 209-212, 1927.
- [55] J. A. Sphon, "Use of Mass Spectrometry for Confirmation of Animal Drug Residues," *J. Assoc. Off. Anal. Chem.*, vol. 61, pp. 1247-1252, 1978.
- [56] R. Baldwin, R. Bethem, R. Boyd, W. Budde, T. Cairns, R. Gibbons, J. Henion, M. Kaiser, D. Lewis, J. Matusik, J. Sphon, R. Stephany and R. Trubey, "1996 ASMS FALL WORKSHOP: Limits to Confirmation, Quantitation, and Detection," *J. Am. Soc. Mass Spectrom.*, vol. 8, pp. 1180-1190, 1997.
- [57] K. S. Webb and D. Carter, "GC Report number LGC/VAM/1998/010," LGC Limited, London, 1998.
- [58] W. G. De Ruig, R. W. Stephany and G. Dijkstra, "Criteria for the detection of analytes in test samples," *J. Assoc. Off. Anal. Chem.*, no. 72, pp. 487-490, 1989.
- [59] Clinical Laboratory and Standards Institute, *MM17 - Validation and Verification of Multiplex Nucleic Acid Assays* (2nd ed.), Wayne (PA): CLSI, 2018.
- [60] J. Gaddum, "Medical Research Council, Special Report Series no. 183," *Br Med J*, 1933.
- [61] C. I. Bliss, "The method of probits," *Science*, vol. 79, no. 2037, pp. 38-39, 1934.
- [62] D. J. Finney, *Probit analysis*, Cambridge: Cambridge University Press, 1947.
- [63] Clinical Laboratory and Standards Institute, *EP-17A2 Evaluation of detection capability for clinical laboratory measurement procedures*, 2nd ed., Wayne (PA): CLSI, 2020.
- [64] Clinical Laboratory and Standards Institute, *MM53-A - Criteria for Laboratory Testing and Diagnosis of Human Immunodeficiency Virus Infection*, Wayne (PA): CLSI, 2011.
- [65] S. L. R. Ellison and A. Williams, (Eds); *Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement*, Third edition, 2012.
- V češtině vydáno jako: Suchánek M., Milde D. (eds.): *Kvalimetrie 19: Stanovení nejistoty analytického měření*. Eurachem-ČR, Praha 2014.
- [66] JCGM, *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment (JCGM 106:2012)*, Sèvres: BIPM, 2012.
- [67] I. Kuselman, F. Pennechi, R. J. N. Bettencourt da Silva and D. B. Hibbert, "IUPAC/CITAC Guide: Evaluation of risks of false decisions in conformity assessment of a multicomponent material or object

- due to measurement uncertainty (IUPAC Technical Report),” *Pure Appl. Chem.*, vol. 93, no. 1, pp. 113-154, 2021.
- [68] Association of Forensic Science Providers, “Standards for the formulation of evaluative forensic science expert opinion,” *Science and Justice*, vol. 2009, pp. 161-164, 2009.
- [69] L. A. Currie, “Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995),” *Pure Appl. Chem.*, vol. 67, no. 10, pp. 1699-1723, 1995.
- [70] European Food Safety Authority, The 2016 European Union report on pesticide residues in food, EFSA, 2018.
- [71] I. Kuselman, F. Pennechi, R. J. N. Bettencourt da Silva and D. B. Hibbert, “Risk of false decision on conformity of a multicomponent material when test results of the components’ content are correlated,” *Talanta*, no. 174, pp. 789-796, 2017.
- [72] R. B. Silva and A. Williams, (Eds.), *Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement*, Eurachem, 2015.
- Český překlad součástí: Suchánek M., Milde D. (eds.): *Kvalimetrie 21: Referenční materiály v chemické analýze. Nastavení a používání cílové nejistoty v chemických měřeních*. Eurachem-ČR, Praha 2016.
- [73] EU, Commission Regulation 2017/626 of 31 March 2017, EU, 2017.
- [74] D. R. Cox, “The regression analysis of binary sequences (with discussion),” *J R Stat Soc B*, vol. 20, no. 2, pp. 215-242, 1958.
- [75] D. J. Finney, *Probit analysis*, 2nd ed., Cambridge: Cambridge University Press, 1952.
- [76] S. L. R. Ellison and A. Williams, (Eds.), *Traceability in Chemical Measurement*, 2nd ed., UK: Eurachem, 2019.
- Český překlad součástí: Milde D. (ed.): *Kvalimetrie 24: Návaznost chemických měření. Metodický návod pro pořádání malých mezilaboratorních porovnání*. Eurachem-ČR, Ústí nad Labem 2019.

