

---

**Eurachem** 

Řada příruček pro laboratoře

---

# **KVALIMETRIE**

## **27**

**Použití informací o nejistotě k posuzování shody**

**Výběr, použití a interpretace programů zkoušení  
způsobilosti**

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

EDITOR: David Milde

První část:

**Použití informací o nejistotě k posuzování shody**

Přeloženo z: A. Williams, B. Magnusson (eds.) Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2<sup>nd</sup> ed. 2021). ISBN 978-0-948926-38-9.

Překlad: David Milde a Zbyněk Plzák

Druhá část:

**Výběr, použití a interpretace programů zkoušení způsobilosti (PT)**

Přeloženo z: B. Brookman, I. Mann (eds.) Eurachem Guide: Selection, Use and Interpretation of Proficiency Testing (PT) Schemes (3<sup>rd</sup> ed. 2021).

Překlad: Zbyněk Plzák

Třetí část:

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

Přeloženo z: R. Bettencourt da Silva, S. L. R. Ellison (eds.) Eurachem/CITAC Guide: Assessment of performance and uncertainty in qualitative chemical analysis (1<sup>st</sup> ed. 2021). ISBN 978-0-948926-39-6.

Překlad: David Milde a Zbyněk Plzák

Editor: David Milde

Poděkování za finanční podporu patří MŠMT v rámci projektu LTV20008 a ÚNMZ v rámci projektu PRM 2022 VII/6/22.

Vydal Eurachem-ČR, z.s., Pasteurova 3544/1, 400 01 Ústí nad Labem, jako 27. publikaci v řadě příruček KVALIMETRIE.

První vydání, Ústí nad Labem 2022

Copyright © Eurachem-ČR 2022

**ISBN 978-80-86322-16-2**

---

# KVALIMETRIE

27

1. část

## Použití informací o nejistotě k posuzování shody

Pokyn Eurachem/CITAC

---



# Použití informací o nejistotě k posuzování shody

Druhé vydání (2021)

## Editoři

Alex Williams (UK)  
Bertil Magnusson (Švédsko)

## Složení pracovní skupiny\*

### Členové Eurachem

S. Ellison <i>předseda</i>	LGC, United Kingdom
B. Magnusson <i>tajemník</i>	Trollboken AB, švédsko
R. Bettencourt da Silva	Univ. Lisboa, Portugalsko
R. Becker	BAM, Běmecko
A. Brzyski	Eurachem Polsko
E. Christie	Eurachem Irsko
K. Darbinyan	Eurachem Arménie
D. Ivanova	Eurachem Bulharsko
M. Inkret	Eurachem Slovinsko
I. Leito	Univ. Tartu, Estonsko
R. Kaus	Eurachem Německo
O. Levbarg	Ukrmetteststandart, Ukrajina
T. Näykki	SYKE, Finsko
P. Pablo Morillas	EUROLAB-España, Španělsko
O. Pellegrino	IPQ/DMET, Portugalsko
M. Rösslein	EMPA St. Gallen, Švýcarsko
E. Sahlin	RISE, Švédsko
M. Sega	INRIM, Itálie
E. Theodorsson	LIU, Švédsko
P. Thomas	SCK CEN, Belgie
A. van der Veen	VSL, Nizozemí
A. Williams	United Kingdom
R. Wood	United Kingdom
P. Yolcu Omeroglu	Eurachem Turecko, Turecko
V. Zonaras	BPI, Řecko

### Členové CITAC

F. Rebello Lourenço	Univ. São Paulo, Brazílie
F. Pennechi	INRIM, Itálie

\* v době schválení dokumentu

## Poděkování

Toto vydání bylo vytvořeno 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.

## Citování

Tato publikace má být citována následovně\*: "A. Williams and B. Magnusson (eds.) Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2<sup>nd</sup> ed. 2021). ISBN 978-0-948926-38-9. Dostupné z [www.eurachem.org](http://www.eurachem.org)."  
\*Závisí na požadavcích časopisů



## Obsah

<b>Předmluva</b>	<b>6</b>
<b>Zkratky a symboly</b>	<b>7</b>
<b>1 Úvod</b>	<b>8</b>
<b>2 Účel</b>	<b>9</b>
<b>3 Definice</b>	<b>10</b>
<b>4 Rozhodovací pravidla</b>	<b>11</b>
<b>5 Výběr mezí pro oblast přijetí a odmítnutí</b>	<b>15</b>
<b>6 Určení přijatelné hodnoty standardní nejistoty</b>	<b>16</b>
<b>7 Doporučení</b>	<b>17</b>
<b>Příloha A – Určení velikosti ochranného pásma a meze přijetí</b>	<b>18</b>
<b>Příloha B – Příklady</b>	<b>22</b>
<b>Příloha C – Riziko výrobce a odběratele</b>	<b>26</b>
<b>Příloha D – Definice</b>	<b>29</b>
<b>Literatura</b>	<b>30</b>

## Předmluva

V době prvního vydání byly práce na posuzování shody prováděny v jiných oblastech, zejména ve strojírenství, a to pro zkoušení elektrických a mechanických výrobků, příručka se proto řídila zásadami stanovenými v ASME B89.7.3.1-2001.

Slova používaná v angličtině pro „shodu“ (compliance, conformity) jsou úzce spojené termíny. ISO často používá termín „posouzení shody“, ASME uvádí „shodu se specifikacemi“. Posuzování shody však může zahrnovat širokou škálu aktivit od zkoušení výrobků až po inspekce a licencování. Tato příručka Eurachem o posuzování shody se primárně zabývá tím, zda je výsledek měření v souladu s povolenými mezemi, např. specifikacemi, tolerancemi, regulačními nebo zákonnými mezemi. Příručka používá pojmy „shoda“ nebo „posuzování shody“ v souvislosti s rozhodováním o shodě se stanovenými mezemi. V normě ČSN EN ISO/IEC 17025 je shoda výsledku měření se stanovenými mezemi často používaná jako základ pro „výrok o shodě“.

Toto vydání bylo aktualizováno tak, aby zahrnovalo změny v např. *Pokynech pro používání rozhodovacích pravidel a uvádění výroků o shodě* (ILAC G8) a *Vyhodnocování naměřených dat – role nejistoty měření při posuzování shody* (JCGM 106).

Hlavní změny ve druhém vydání zahrnují:

- přidání seznamu zkratk a symbolů,
- zavedení pojmu mez přijetí,
- zavedení rozhodovacích pravidel vhodných pro podmíněné nebo neúplné výsledky (někdy nazývaná nebinární rozhodovací pravidla),
- zavedení logaritmicke normálního rozdělení pro asymetrické případy,
- přidání přílohy C, která zavádí globální a specifická rizika.



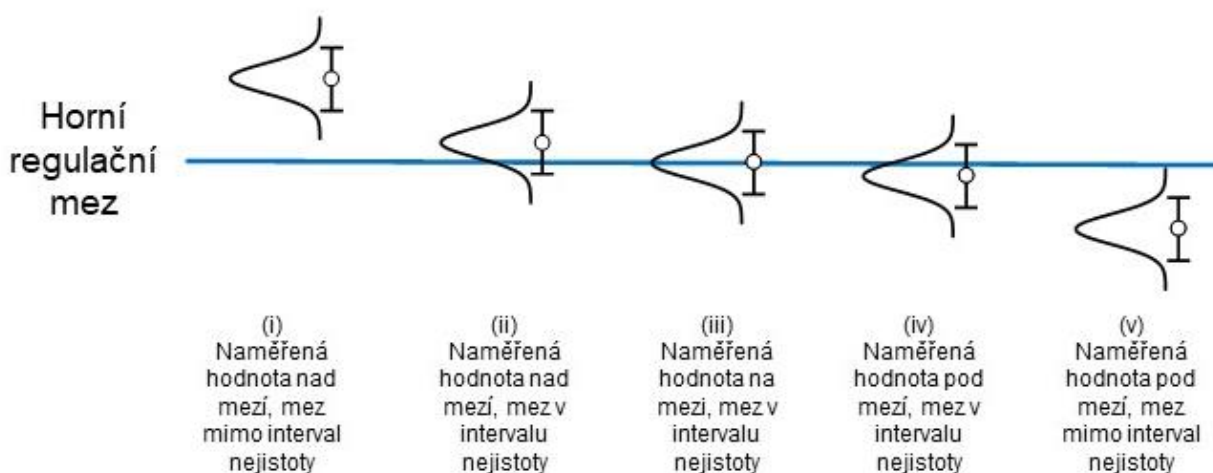
## Zkratky a symboly

V příručce se objevují následující zkratky, akronymy a symboly.

ASME	Americká společnost mechanických inženýrů	IEC	Mezinárodní elektrotechnická komise
BIPM	Mezinárodní úřad pro míry a váhy	ISBN	Mezinárodní standardní číslo knihy
CITAC	Mezinárodní spolupráce pro návaznost v analytické chemii	ISO	Mezinárodní organizace pro normalizaci
GUM	Pokyn k vyjádření nejistoty měření	IUPAC	Mezinárodní unie pro čistou a užitou chemii
JCGM	Společný výbor pro pokyny v metrologii	VIM	Mezinárodní metrologický slovník
$g$	ochranné pásmo	$s$	směrodatná odchylka
$k$	koeficient rozšíření	$s_G$	směrodatná odchylka $\log_e$ dat
$n$	počet měření	$x_i$	pozorovaná hodnota
$\exp$	exponenciální funkce; $\exp(x) = e^x$	$u$	standardní nejistota
$P$	pravděpodobnost (%) shody nebo neshody	$u_{\text{rel}}$	relativní standardní nejistota
$L$	povolená mez pro shodu	$^F U$	faktor nejistoty
$L_l$	dolní povolená mez	$U$	rozšířená nejistota
$L_u$	horní povolená mez		

## 1 Úvod

Aby se mohl výsledek použít k rozhodování, zda indikuje shodu či neshodu se specifikací, je nutné vzít v úvahu nejistotu měření. Obrázek 1 ukazuje typické situace, které vznikají, když se použijí výsledky měření, například koncentrace analytu, k posouzení shody s horní specifikační mezí. Vertikální linie označují u každé pozorované hodnoty rozšířenou nejistotu  $\pm U$  a přidružené křivky znázorňují odvozené funkce hustoty pravděpodobnosti hodnot měřené veličiny a ukazují, že je vyšší pravděpodobnost výskytu měřené veličiny v blízkosti středu intervalu rozšířené nejistoty než blízko okrajů. Případy (i) a (v) jsou dostatečně jasné; výsledky měření a jejich nejistoty poskytují spolehlivý důkaz, že hodnota měřené veličiny je zcela nad nebo zcela pod mezí. V případě (ii) však existuje vysoká pravděpodobnost, že hodnota měřené veličiny přesahuje mez, ale mez leží přesto uvnitř intervalu nejistoty. V závislosti na okolnostech, a zvláště na rizicích spojených s chybným rozhodnutím, může nebo nemusí být pravděpodobnost nesprávného rozhodnutí natolik malá, že neopravňuje rozhodnutí o shodě. Podobně může nebo nemusí být v případě (iv) pravděpodobnost, s níž je hodnota měřené veličiny pod mezí, dostatečná pro použití takového výsledku k potvrzení shody. V případě (iii) je pravděpodobnost nesprávného rozhodnutí 50 %. Bez dalších informací, které musí být založeny na rizicích souvisejících s chybným rozhodnutím, není možné použít tyto tři výsledky, tedy případy (ii), (iii) a (iv) k rozhodnutí o shodě.



Obrázek 1. Posouzení shody s horní mezí

## 2 Účel

Tuto příručku lze používat pro nastavení vhodných kritérií pro jednoznačná rozhodnutí o shodě založených na výsledcích a informacích o přidružené nejistotě. Klíčem k posouzení shody je koncept „rozhodovacích pravidel“. Tato pravidla poskytují předpis pro přijetí nebo odmítnutí položky založený na měřené veličině, její nejistotě a specifikační mezi nebo mezích, přičemž se bere v úvahu přijatelná úroveň pravděpodobnosti učinění špatného rozhodnutí.

Tento dokument se nezabývá případy, které zahrnují rozhodování založené na mnohočetných měřených veličinách. Příklady aplikací posuzování shody založené na mnohočetných měřených veličinách lze nalézt v literatuře [1, 2].

Když se rozhodování o shodě vztahuje na celou testovanou sérii nebo šarži látky nebo materiálu, může být příspěvek k nejistotě měření pocházející ze vzorkování významný. Tato příručka předpokládá, že když měřená veličina zahrnuje požadavek na vzorkování, zahrnuje nejistota i složky pocházející ze vzorkování [3].

### 3 Definice

Termíny používané v této příručce pocházejí hlavně z termínů Mezinárodního slovníku základních a obecných termínů v metrologii (VIM) [4] a z Pokynu ISO/IEC pro vyjádření nejistoty měření (GUM) [5] a ILAC G8 [6]. Další termíny jsou převzaty z ASME B89.7.3.1-2001 [7]). Souhrn nejdůležitějších definic použitých v tomto dokumentu uvádí příloha D: Definice.

## 4 Rozhodovací pravidla

### 4.1 Obecně

Klíčem k posuzování shody je koncept „rozhodovacích pravidel“. Tato pravidla poskytují předpis pro shodu nebo neshodu se specifikační mezí na základě přijatelné úrovně pravděpodobnosti, že se učiní špatné rozhodnutí. Norma ČSN EN ISO/IEC 17025 definuje rozhodovací pravidlo jako *pravidlo, které popisuje, jak se počítá s nejistotou měření při stanovování shody se specifickým požadavkem* [8]. ČSN EN ISO/IEC 17025 vyžaduje také, aby v případě relevantnosti bylo použití rozhodovacího pravidla dohodnuto se zákazníkem. Dokumenty ILAC G8 [6], JCGM106 [9], Zpráva Eurolab 1/2017 [10] a WADA TD2019DL [11] poskytují přehled rozhodovacích pravidel a shody s požadavky. Na základě takového rozhodovacího pravidla se určí „oblast přijetí“ a „oblast odmítnutí“ tak, aby pokud výsledek měření leží v oblasti přijetí, byl produkt deklarován jako vyhovující a pokud leží v oblasti odmítnutí, byl produkt prohlášen jako nevyhovující. Meze oblasti přijetí nazýváme „meze přijetí“.

Rozhodovací pravidlo má mít dobře dokumentovanou metodu určení umístění zón přijetí a odmítnutí, ideálně zahrnout i přijatelné hladiny pravděpodobnosti  $P$ , že hodnota měřené veličiny: 1. leží uvnitř meze specifikace nebo 2. leží mimo mez specifikace.

Případ 1 odpovídá vysoké spolehlivosti správného přijetí a nízké pravděpodobnosti falešného přijetí, zatímco případ 2 poskytuje vysokou spolehlivost správného odmítnutí a nízké pravděpodobnosti falešného odmítnutí.

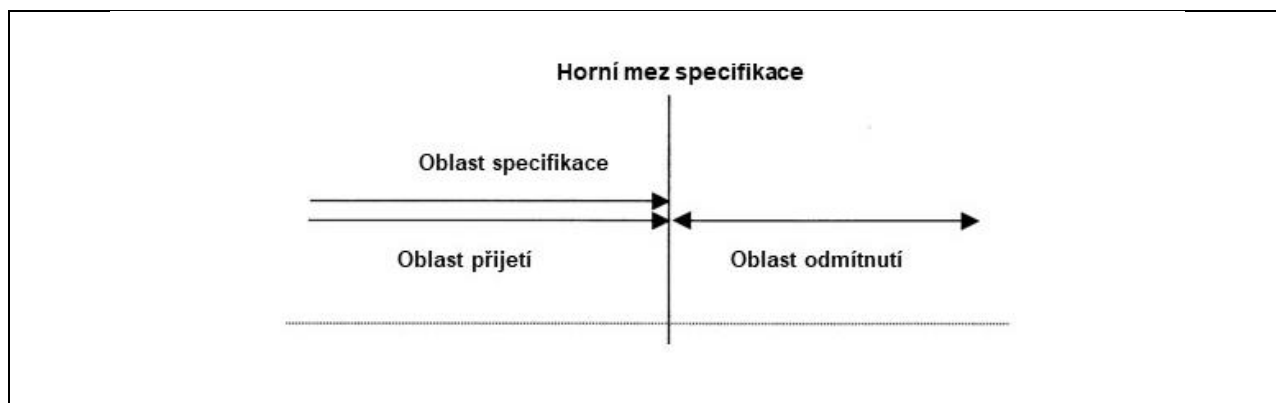
Stanovení oblasti přijetí/odmítnutí provádí běžně laboratoř při aplikaci rozhodovacího pravidla.

Rozhodovací pravidlo může také poskytnout:

- maximální povolenou nejistotu na mezi,
- předpokládané rozdělení, např. normální nebo lognormální (viz příloha A),
- pravidla pro zaokrouhlování nebo useknutí pozorovaných hodnot před hodnocením shody,
- požadovaný počet opakovaných měření (je-li takový) a postup pro použití replikátních měření včetně (například) toho, zda mají výsledky vyhovovat každý zvlášť nebo mají být před porovnáním s mezemi zprůměrovány,
- postupy pro řešení odlehlých hodnot,
- postupy pro další kroky v případě nebinárního rozhodnutí, když je rozhodnutí podmíněné (vyhovuje/nevyhovuje),
- postupy, které je třeba použít v případě podmíněného rozhodnutí, které vyžaduje další měření,
- doporučení, jak uvádět shodu/neshodu, např. vyhovuje/nevyhovuje, v toleranci/mimo toleranci, v rámci specifikace/mimo specifikaci,
- doporučení, jak použít rozhodovací pravidlo ve výročních o shodě.

## 4.2 Rozhodovací pravidlo vyhovuje/nehovuje s využitím jednoduchého přijetí

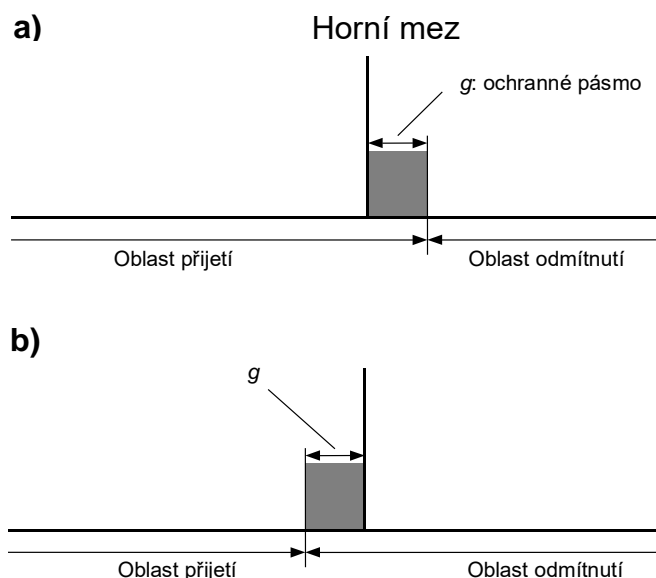
Rozhodovací pravidlo je v řadě situací navrženo tak, aby poskytlo přesvědčivé rozhodnutí o shodě: vyhovuje, nebo nevyhovuje. Nejjednodušší příklad používá mez specifikace jako mez přijetí a to tak, že výsledek uvnitř meze je považován za vyhovující. Toto nazýváme „jednoduché přijetí“ nebo „sdílené riziko“ [6]. Odpovídá to oblastem přijetí a odmítnutí na obrázku 2. Leží-li výsledek v oblasti přijetí, je produkt prohlášen za vyhovující (shodný) a pokud se nachází v oblasti zamítnutí, je prohlášen za nevyhovující (neshodný). Na obrázku 1 popisujícím shodu s horní mezí leží případy (iv) a (v) v oblasti přijetí a případy (i) a (ii) v oblasti odmítnutí. Příklad (iii) je ve většině případů považován za ležící v oblasti přijetí. Pro použití tohoto pravidla je obvykle požadováno, aby nejistota měření byla posouzena a shledána přijatelnou, aby riziko špatného rozhodnutí bylo přijatelné. Avšak pro pozorované hodnoty blízko mezi existuje v případě jednoduchého přijetí riziko (až 50 %) nesprávného rozhodnutí – viz přílohu C – riziko výrobce a odběratele.



Obrázek 2. Oblasti přijetí a odmítnutí pro jednoduché přijetí s horní mezí. Mez přijetí se rovná mezi specifikace

## 4.3 Rozhodovací pravidlo vyhovuje/nehovuje s využitím ochranného pásma

V případě pozorovaných hodnot velmi blízko mezi, nebo pokud je nejistota vysoká, může vést jednoduché přijetí k nepřijatelně vysokému riziku nesprávného rozhodnutí. Často si potřebujeme být více jisti přijetím nebo odmítnutím produktu. V těchto situacích lze oblasti přijetí a odmítnutí určit podle obrázku 3. Například pokud na obr. 3b leží pozorovaná hodnota v oblasti přijetí, je velmi nepravděpodobné, že by produkt zkoušky byl shledán nevyhovujícím. Interval  $g$  mezi mezí a koncem oblasti přijetí se nazývá „ochranné pásmo“, které snižuje riziko nesprávného rozhodnutí. Používání ochranných pásem poskytuje mimořádně jednoduchý způsob definování rozhodovacích pravidel. Volba velikosti ochrany definuje oblast přijetí, kterou lze použít pro provedení rozhodnutí. Ochranné pásmo  $g$  bude obecně násobkem standardní nejistoty  $u$ . V případě relativní nejistoty menší než 15 až 20 % a pokud pozorovaná hodnota je přesně na horní nebo dolní mezi přijetí, tak ochranné pásmo rovné  $1,64u$  poskytne pravděpodobnost nesprávného rozhodnutí  $\alpha$  rovnou 5 % a ochranné pásmo rovné  $2,33u$  vede k pravděpodobnosti  $\alpha = 1$  %. Násobek lze nastavit explicitně, např. ochranné pásmo může být nastaveno na  $2u$ , tedy rozšířená nejistota  $U$  se používá jako ochranné pásmo poskytující pravděpodobnost nesprávného rozhodnutí  $\alpha = 2,5$  %. Ochranné pásmo lze také nastavit na nulu,  $g = 0$ . To nazýváme jednoduchým přijetím nebo „sdíleným rizikem“, viz odstavec 4.2. Určení ochranného pásma je diskutováno v příloze A a rizikem špatného odmítnutí a přijetí se zabývá příloha C.



**Obrázek 3. Oblasti přijetí a odmítnutí pro horní mez. Obrázek ukazuje relativní pozice oblastí přijetí a odmítnutí pro a) vysokou spolehlivost správného odmítnutí, b) vysokou spolehlivost správného přijetí. Úsek  $g$  se nazývá „ochranné pásmo“. Horní konec oblasti přijetí je mez přijetí**

V některých případech stanovuje specifikace horní a dolní mez, například ke kontrole složení produktu. Obrázek 4 znázorňuje oblasti přijetí a odmítnutí pro případ, kdy byla ochranná pásma zvolena tak, aby pro vzorek ve shodě existovala velká pravděpodobnost, že je měřená veličina uvnitř mezi daných specifikací, tedy vysoká spolehlivost správného přijetí.

#### 4.4 Rozhodovací pravidla s podmíněnými nebo nejednoznačnými výsledky

Některé rozhodovací postupy mohou zahrnovat možnost „podmíněného“ výstupu vyhovuje/nehovuje nebo „nejednoznačného“ výstup, typicky když je mez specifikace uvnitř intervalu rozšířené nejistoty (viz případy (ii), (iii) a (iv) na obr. 1). ILAC G8 [6] označuje tyto případy jako „nebinární“ rozhodovací pravidlo. Například na obr. 1 může být případ (ii) označen za „podmínečně nevyhovující“ a případy (iii) a (iv) lze považovat za „podmínečně vyhovující“. Rozhodovací pravidlo může alternativně tyto tři případy označit jako „nejednoznačné“.

V případě podmíněného nebo nejednoznačného výsledku mnohdy rozhodovací pravidlo zabezpečí další zkoušení.

ILAC G8 obsahuje další příklady podmíněného přijetí nebo odmítnutí.

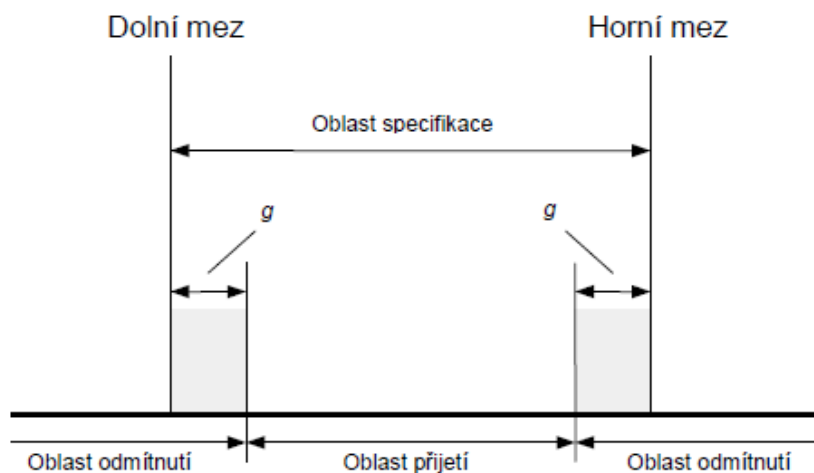
#### 4.5 Rozhodovací pravidlo určující dvoustupňový postup

Aby se snížilo riziko falešného přijetí anebo odmítnutí, přijímají některá rozhodovací pravidla dvoustupňový postup, ve kterém se v případě nejednoznačného výsledku provedou další měření. Obecně takovýto dvoustupňový postup popisuje dokument ISO [12].

Tyto postupy nevyžadují použití stejného měřicího postupu v každém stupni. Například z důvodu snížení nákladů na posuzování shody lze provést nejprve nízkonákladové měření s poměrně vysokou nejistotou (často se označuje jako „screeningová“ zkouška). Pokud je počáteční výsledek nejednoznačný nebo blízko mezi, použije se potvrzovací postup poskytující výsledky s menší nejistotou. Většina zkoušených produktů je pak hodnocena s dostatečnou spolehlivostí a nízkými náklady, zatímco hraniční případy jsou hodnoceny dražšími zkouškami s vyšší spolehlivostí.

## Použití informací o nejistotě k posuzování shody

Pravděpodobnosti falešného přijetí anebo falešného odmítnutí ve víceúrovňových postupech závisí na zvolených konkrétních stupních a jsou složitější než v případě jednoúrovňových postupů. Pravděpodobnosti víceúrovňových postupů jsou nad rozsah této příručky.



**Obrázek 4. Oblasti přijetí a odmítnutí pro oblast specifikace. Obrázek ukazuje relativní pozice specifikovaných mezí a oblastí přijetí a odmítnutí pro vysokou spolehlivost správného přijetí**



## 5 Výběr mezí pro oblast přijetí a odmítnutí

Rozsah ochranného pásma  $g$  závisí na hodnotě standardní nejistoty  $u$  získané z měření blízko mezi  $L$ . Volí se tak, aby vyhověl požadavkům rozhodovacího pravidla. Jestliže například rozhodovací pravidlo určuje, že pro neshodu má být pozorovaná hodnota větší než mez plus  $2u$ , pak je velikost ochranného pásma  $2u$ , což vede k mezi přijetí  $L + 2u$ .

Podobně když rozhodovací pravidlo stanoví nejméně 95% pravděpodobnost, že je hodnota měřené veličiny menší než  $L$ , pak se  $g$  volí tak, aby pro pozorovanou hodnotu  $L - g$  byla pravděpodobnost, při níž hodnota měřené veličiny leží pod mezí, 95 % viz obrázek 3b.

Obecně je hodnota ochranného pásma  $g$  jednoduchým násobkem  $u$ , kde  $u$  je standardní nejistota. V některých případech může rozhodovací pravidlo určit, jaká hodnota násobku se má používat. Mez přijetí bude obecně záviset na velikosti požadované pravděpodobnosti  $P$  a znalosti rozdělení hodnot měřené veličiny. Někdy může být  $g$  komplexní funkcí  $u$ . Některé typické případy popisuje příloha A.

## 6 Určení přijatelné hodnoty standardní nejistoty

Čím větší je hodnota standardní nejistoty  $u$  a čím blíže je pozorovaná hodnota mezi, tím větší je podíl vzorků, které budou posuzovány nesprávně. Například při posuzování shody s horní mezí, když není nastaveno žádné ochranné pásmo, leží-li pozorovaná hodnota více než  $3u$  pod mezí, pak riziko učinění špatného rozhodnutí je velmi malé (přibližně 0,1 %). S tím, jak se pozorovaná hodnota blíží mezi, zvyšuje se riziko: pokud je pozorovaná hodnota vzdálena od meze  $2u$ , je 2,3 %, pokud je s mezí totožná, je 50 %. Čím menší bude hodnota  $u$ , tím vyšší budou náklady na měření. Tedy v ideálním případě by se měla  $u$  zvolit tak, aby se minimalizovaly náklady na měření plus náklady špatných rozhodnutí.

V některých analytických oblastech je spolu s maximálními anebo povolenými mezemi pro měřenou veličinu definována cílová (tj. maximální) nejistota měření [13]. Příručka Eurachem/CITAC pro nastavení a používání cílové nejistoty měření [14] doporučuje způsoby definování cílové nejistoty v případech, kdy není regulována nebo dána zákazníkem. Cílovou nejistotu lze definovat z šířky povoleného intervalu pozorované koncentrace (viz oddíl 5.1.2 uvedené příručky) nebo z definované meze přijatelnosti pro naměřenou koncentraci na základě přijatelného rizika špatného rozhodnutí odběratele nebo výrobce (viz oddíl 5.1.4 uvedené příručky).

## 7 Doporučení

K rozhodnutí, zda přijmout či odmítnout produkt podle dodaného výsledku a jeho nejistoty, musí existovat:

- a) specifikace, která stanovuje horní anebo dolní meze kontrolovaných charakteristik (měřených veličin),
- b) nejistota měření a
- c) rozhodovací pravidlo, které popisuje, jak se bude používat nejistota měření při přijímání či odmítnutí produktu podle jeho specifikace a výsledku měření.

Rozhodovací pravidlo má mít jasně dokumentovanou metodu určení velikosti oblastí přijetí a odmítnutí, ideálně tak, aby zahrnula minimální přijatelnou mez pravděpodobnosti, že měřená veličina leží uvnitř mezi specifikace. Má také poskytnout postup pro vypořádání se např. s opakovanými měřeními a odlehlými hodnotami (viz oddíl 4.1).

Rozhodovací pravidlo může nastavit rozsah oblasti přijetí nebo odmítnutí pomocí vhodného ochranného pásma. Rozsah ochranného pásma vypočítáme s využitím znalostí o nejistotě měření a minimální přijatelné hladině pravděpodobnosti, že měřená veličina leží uvnitř mezi specifikace. Pro běžné případy, kdy je nejistota přibližně konstantní nebo předpokládané rozdělení chyb je symetrické a má směrodatnou odchylku úměrnou „pravé“ hodnotě, lze mez a nejistotu použít k výpočtu ochranného pásma. Toto popisují případy 1–3 v příloze A.

Nedílnou součástí zprávy o shodě musí být také odkazy na použitá rozhodovací pravidla.

## Příloha A: Určení velikosti ochranného pásma a meze přijetí

Rozsah ochranného pásma  $g$  a tudíž mez přijetí se volí tak, aby vyhovovaly požadavkům rozhodovacího pravidla. Závisí na:

- hodnotě nejistoty.
- minimální přijatelné hladině pravděpodobnosti  $P$ , že měřená veličina leží uvnitř mezi specifikace (nebo analogicky maximální přijatelná pravděpodobnost, že hodnota měřené veličiny neleží uvnitř mezi specifikace).
- dostupné znalosti rozdělení hodnot měřené veličiny.

Je-li relativní standardní nejistota menší než přibližně 15 % až 20 %, můžeme předpokládat normální rozdělení [15]. Rozsah  $g$  se bude rovnat  $ku$ , jako je tomu v případech 1a a 1b níže. Jsou-li známy efektivní stupně volnosti, hodnota  $k$  bude vzata z  $t$ -rozdělení, jako je tomu v případě 2. I v dalších případech, kde je známo, že hodnota měřené veličiny je větší než nula, ale relativní standardní nejistota je větší než 20 %, použití normálního rozdělení nemusí být na místě. Rozsah  $g$  se pak určí z tvaru rozdělení a očekávané hodnoty  $P$ , jako je tomu v případech 3 a 4. Existuje řada rozdělení poskytujících srovnatelné výsledky pro relativní nejistoty vyšší než 50 %, např. lze použít lognormální [15], beta [16] a gamma [9] rozdělení. Další návod pro přiřazení rozdělení na základě dostupných znalostí lze nalézt v JCGM 101:2008 [17]. Pokud se modelový vzorec pro výpočet hodnoty veličiny skládá z násobení/dělení kladných veličin, je to dobrým podkladem pro použití lognormálního rozdělení, jak popisuje případ 4 [15].

### Případ 1a: Je k dispozici standardní nejistota

V tomto případě, pro relativní standardní nejistotu menší než 20 %, bude velikost ochranného pásma  $ku$  a hodnota  $k$  bude buď specifikována v rozhodovacím pravidle, nebo se odvodí z rozdělení pravděpodobnosti hodnot přiřazených měřené veličině, o kterém se obvykle předpokládá, že je normální. Základy pro takový předpoklad a podmínky, za nichž je jeho použití vhodné, uvádí GUM [5] v příloze G. Předpoklad vychází z použití *centrální limitní věty*. V oddíle G 2.3 GUM se zdůrazňuje „...jestliže v kombinované standardní nejistotě  $u$  nepřevažuje složka standardní nejistoty získaná vyhodnocením typu A, které je založeno pouze na několika málo pozorováních, nebo složka standardní nejistoty získaná vyhodnocením typu B na základě rovnoměrného rozdělení, pak racionální první aproximace pro výpočet rozšířené nejistoty  $U$ , která poskytuje interval na konfidenční úrovni  $P$ , použije pro  $k$  hodnoty z normálního rozdělení.“

V mnoha případech je běžnou praxí používat  $k = 2$ . Za předpokladu, že je rozdělení přibližně normální, to znamená úroveň spolehlivosti přibližně 95 %, se pro pozorovanou hodnotu  $x$  nachází hodnota měřené veličiny v intervalu  $x \pm 2u$ . Na tomto základě je pak pravděpodobnost, že hodnota měřené veličiny je menší než  $x + 2u$ , přibližně 97,5 %. V běžně se vyskytujícím případě, kdy je požadován důkaz o shodě s horní mezí, jak ukazuje obrázek 3, jsou volba  $k = 2$  a požadavek jasné neshody rovnocenné stanovení ochranného pásma  $g = 2u$ . Jestliže přesahuje pozorovaná hodnota mez o více než  $g$ , pak je hodnota měřené veličiny nad mezí se spolehlivostí nejméně 97,5 %. To povede k menšímu počtu chybných rozhodnutí o neshodě než rozhodování založené na jednostranném testování spolehlivosti pro 95% spolehlivost (tj. s  $k = 1,64^*$ ). Když bude třeba rozhodovat na jiných úrovních spolehlivosti, pak lze hodnotu  $k$  získat z tabulek nebo statistického software pro příslušnou úroveň spolehlivosti.

V oddíle G 1.2 GUM [5] se však zdůrazňuje, že snaha odlišovat blízké úrovně spolehlivosti (řekněme 94 % a 96 %) nemá obvykle smysl, jelikož hodnota  $U$  je v nejlepším případě pouze přibližná. Kromě toho GUM uvádí, že je obzvláště obtížné získat intervaly pro úroveň spolehlivosti 99 % nebo vyšší.

---

\* Hodnota  $k$  je 1,64 se dvěma významnými číslicemi, zatímco hodnota 1,6449 s pěti významnými číslicemi.

Pro relativní standardní nejistotu  $u$  vyšší než 20 % by se mělo zvážit lognormální rozdělení [15]. To popisuje případ 4 níže a příklad 3 v příloze B.

#### **Případ 1b: Je k dispozici rozšířená nejistota s uvedeným koeficientem rozšíření $k$**

Dělte  $U$  danou hodnotou  $k$  (obvykle 2) a určete hodnotu ochranného pásma s použitím vhodné revidované hodnoty  $k$  jako v případě 1a.

#### **Případ 2: Je k dispozici standardní nejistota spolu s efektivními stupni volnosti**

V tomto případě je uznávanou praxí předpokládat, že hodnoty, které mohou být přisuzovány měřené veličině, mají  $t$ -rozdělení se známým počtem stupňů volnosti, a pro hodnotu  $k$  používat jednostranný 95% kvantil  $t$ -rozdělení pro koeficient rozšíření  $k$ . Velikost ochranného pásma bude  $ku$  a příklad shody s horní mezí ukazuje obrázek 3.

V oddílech G3 a G4 GUM je podrobně diskutováno  $t$ -rozdělení a stupně volnosti. Alternativní přístupy, které se vyhýbají problémům s používáním efektivního počtu stupňů volnosti, navrhl Williams [18] a Kacker a Jones [19].

#### **Případ 3: Jsou k dispozici jednotlivé složky a rozdělení**

O tomto případě pojednává GUM v oddíle 1.4 [5]. Uvádí, že pokud jsou známa rozdělení pravděpodobnosti vstupních veličin a hodnota měřené veličiny je lineárně závislá na těchto vstupních veličinách, pak se může rozdělení pravděpodobnosti hodnot přisuzovaných měřené veličině vypočítat konvolucí těchto rozdělení. To lze provést metodou Monte Carlo pro propagaci rozdělení pravděpodobnosti a rozdělení výsledků použitých k výpočtu požadovaného konfidenčního intervalu. Monte Carlo distribuce [20] by pro rutinní použití měla být porovnána s dalšími známými rozděleními a v mnoha případech je pravděpodobné, že se dobré proložení ukáže lognormální rozdělení. Stojí za to zmínit, že metoda Monte Carlo [17] lze použít pro každý model měření propojující měřenou veličinu se sadou vstupních veličin.

#### **Případ 4: asymetrická rozdělení**

Případem, kdy vstupní veličina má asymetrické rozdělení, se v obecné rovině zabývá GUM v oddíle G 5.3 [5]. Zdůrazňuje, že „to nemá vliv na výpočet  $u$ , ale může ovlivnit výpočet  $U$ “.

V obecnější rovině existují tři důležité situace, kdy jsou asymetrické konfidenční intervaly pro přijímání rozhodnutí nezbytné:

- a) (předpokládané) rozdělení měřené veličiny  $x$  je přirozeně asymetrické (jako je Poissonovo rozdělení s malým počtem stupňů volnosti),
- b) relativní standardní nejistota 1. je větší než 20 % a 2. je konstantní,
- c) měřená odezva  $x$  je blízko fyzikálnímu omezení (např. pozorované koncentrace blízké nule).

Pokud pro a) a b) použitá data ukazují na asymetrické rozdělení, lze spočítat meze konfidenčního intervalu stejně jako v případě 3 výše. První situace je známa například z měření radioaktivity při malém počtu detekovaných jevů.

Třetí situace c) se vyskytuje u měření blízko meze detekce nebo stanovitelnosti, nebo je-li definice proměnné omezena na specifický interval. V tomto případě může být nezbytné použití useknutého rozdělení (viz [20], příloha F).

Pro řadu analytických měření je známo, že hodnota měřené veličiny je kladná a modelová rovnice se skládá ze součinu nebo podílu kladných veličin. Pak je pro situace b) a c) vhodné použití lognormálního rozdělení.

Za předpokladu, že rozdělení hodnot přiřaditelných k měřené veličině je lognormální, lze vypočítat meze přijetí s použitím rozšířeného faktoru nejistoty  $^F U$  [3, 21]:

$$^F U = \exp(k s_G), \quad \text{Rovnice 1}$$

kde  $s_G$  je směrodatná odchylka v  $\log_e$  škále (přirozené logaritmy). Pro  $u_{\text{rel}}$  méně než 0,5 (50 %),  $s_G \approx u_{\text{rel}}$  [15] a faktor nejistoty lze spočítat jako:

$$^F U \approx \exp(k u_{\text{rel}}), \quad \text{Rovnice 2}$$

kde koeficient rozšíření  $k$  je horní kvantil standardního normálního rozdělení na zvolené konfidenční hladině.

Horní mez přijetí pro vysokou spolehlivost správného odmítnutí je pak:

$$L_u \times ^F U \quad \text{Rovnice 3}$$

a horní mez přijetí pro vysokou spolehlivost správného přijetí:

$$L_u / ^F U \quad \text{Rovnice 4}$$

Ochranné pásmo pro horní mez a vysokou spolehlivost správného odmítnutí se vypočítá:

$$g = L_u \times ^F U - L_u \quad \text{Rovnice 5}$$

Ve srovnání s normálním rozdělením se rozsah ochranného pásma pro horní mez a správné odmítnutí zvýší (viz obr. 3a) a pro správné přijetí se sníží (viz obr. 3b). Důvodem je asymetrie lognormálního rozdělení. Jako příklad jsou v tabulce 1 vypočtena ochranná pásma pro normální a lognormální rozdělení s horní mezí  $L = 100$ ,  $k = 1,64$  a  $u_{\text{rel}}$  rovno 0,3 a 0,5. Relativní standardní nejistoty 0,3 a 0,5 odpovídají rozšířeným relativním nejistotám 60 % a 100 %.

**Tabulka1: Mez přijetí pro horní mez za předpokladu normálního a lognormálního rozdělení a vysokou relativní nejistotu**

Předpokládané rozdělení	Relativní standardní nejistota $u_{\text{rel}}$	Horní mez	Meze přijetí	
			Správné přijetí	Správné odmítnutí
Normální	0,3	100	51	149
Lognormální	0,3	100	61	164
Normální	0,5	100	18	182
Lognormální	0,5	100	44	227

V hodnotách  $u_{\text{rel}}$  jsou meze přijetí pro normální rozdělení:

$$L(1 + k u_{\text{rel}}) \text{ a } L(1 - k u_{\text{rel}})$$

a pro lognormální rozdělení:

$$L(\exp(ku_{\text{rel}})) \text{ a } L(\exp(-ku_{\text{rel}})).$$

Pro lognormální rozdělení lze rovnice rozšířit (pomocí obvyklého rozšíření pro  $\exp(x)$ ) na:

$$L\left(1 + ku_{\text{rel}} + \frac{(ku_{\text{rel}})^2}{2} + \dots\right) \text{ a } L\left(1 - ku_{\text{rel}} + \frac{(-ku_{\text{rel}})^2}{2} + \dots\right),$$

kde “...” představuje vyšší členy rozšíření. Pokud jsou členy nad  $ku_{\text{rel}}$  významné, mělo by se zvážit použití lognormálního rozdělení [15]. Při  $u_{\text{rel}} = 20\%$  a  $k = 1,64$  vzroste faktor pro výpočet meze přijetí pro správné zamítnutí asi o 5 % ve srovnání s použitím  $(1 + ku_{\text{rel}})$ .

Příklad použití lognormálního rozdělení je v příloze B – příklad 3.

## Příloha B: Příklady

### Příklad 1 – uplatnění rozhodovacího pravidla uvedeného v případě 1b v příloze A

Případ 1b v příloze A se zabývá situací, kdy je dostupná rozšířená nejistota spolu s uvedenou hodnotou  $k$ . Hmotnostní zlomek niklu pro druh nerezové oceli musí ležet v interval od 16,0 % do 18,0 % Ni.

Měřená veličina	Hmotnostní zlomek niklu (Ni) v šarži oceli určené zákazníkovi.
Nejistota	Absolutní rozšířená nejistota $U$ je 0,2 % Ni, $k = 2$ (95 %). Standardní nejistota $u = 0,1$ % Ni. Tato nejistota zahrnuje jak nejistotu vzorkování šarže, tak i analytickou nejistotu.
Specifikace	Oblast specifikace je mezi dolní mezí 16,0 % Ni a horní mezí 18,0 % Ni.
Rozhodovací pravidlo <i>Vysoká spolehlivost správného přijetí</i>	<i>Oblast přijetí představuje interval hmotnostního zlomku, kde může být s hladinou jistoty ne menší než 95 % (<math>\alpha = 0,05</math>) rozhodnuto, že šarže má hmotnostní zlomek nad dolní mezí a pod horní mezí.</i>
Rozdělení	Předpokládá se normální rozdělení hodnot měřené veličiny.
Ochranné pásmo	Každé ochranné pásmo se vypočítá jako $1,64u \approx 0,17$ % s hodnotou $k$ 1,64 z jednostranného 95 % horního kvantilu normálního rozdělení.
Oblast přijetí	Od 16,2 % Ni do 17,8 % Ni, po zaokrouhlení na jedno desetinné místo.
Pozorovaná hodnota	16,1 % Ni

Obrázek 4 ukazuje umístění oblastí přijetí a odmítnutí pro specifikaci s horní a dolní mezí s vysokou spolehlivostí správného přijetí. Pozorovaná hodnota 16,1 % Ni je pod dolní mezí přijetí 16,2 %, leží tedy v oblasti odmítnutí. **Šarže je nevyhovující.**

#### POZNÁMKA

Pokud by rozhodovací pravidlo udávalo jednoduché přijetí, oblast přijetí by byla 16,0 % až 18,0 % a šarže by byla vyhovující.



## Příklad 2 – uplatnění rozhodovacího pravidla uvedeného v případě 2 v příloze A

Případ 2 v příloze A pojednává o situaci, kdy je k dispozici standardní nejistota  $u$  a efektivní počet stupňů volnosti. Ve vyráběné šarži má být koncentrace analytu méně než 200 ng/g.

Měřená veličina	Hmotnostní zlomek analytu ve vyráběné šarži určené zákazníkovi.
Nejistota	Absolutní standardní nejistota $u = 2,2$ ng/g. Nejistota zahrnuje složky z vzorkování. Hlavní příspěvek k nejistotě byl získán z 9 měření, tedy máme 8 stupňů volnosti. Lze předpokládat, že hodnoty přiřaditelné měřené veličině odpovídají $t$ -rozdělení.
Specifikace	Horní povolená mez $L_u$ je 200 ng/g.
Rozhodovací pravidlo <i>Vysoká spolehlivost správného odmítnutí</i>	<i>Šarže bude považována za nevyhovující, pokud pravděpodobnost hodnoty, že hmotnostní zlomek bude větší než 200 ng/g, převyšuje 95 %.</i>
Rozdělení	Předpokládá se normální rozdělení hodnot měřené veličiny.
Ochranné pásmo	Ochranné pásmo se vypočte s hodnotou $k$ rovnou 1,86 získanou jako jednostranný horní 95% kvantil $t$ -rozdělení s 8 stupni volnosti: $ku = 1,86 \times 2,2 = 4,1$ ng/g.
Mez přijetí	204,1 ng/g
Pozorovaná hodnota	203,7 ng/g

Obrázek 3a ukazuje umístění oblastí přijetí a odmítnutí pro specifikaci s horní mezí s vysokou spolehlivostí správného odmítnutí. Pozorovaná hodnota 203,7 ng/g je pod mezí přijetí 204,1 ng/g, leží tedy v oblasti přijetí. **Šarže je vyhovující.**

### POZNÁMKA

Pokud by rozhodovací pravidlo udávalo jednoduché přijetí, mez přijetí by se rovnala povolené mezi 200 ng/g a šarže by byla nevyhovující.

### Příklad 3 – uplatnění rozhodovacího pravidla uvedeného v případě 4 v příloze A

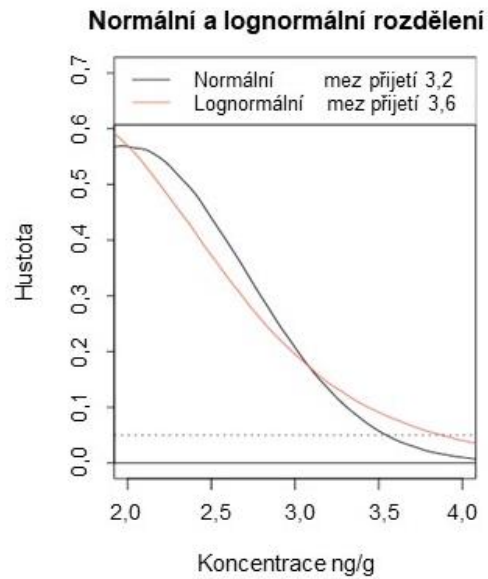
Případ 4 v příloze A pojednává o asymetrických rozděleních. V tomto případě je asymetrie způsobena velkou relativní nejistotou kontrolní analýzy zakázané látky a rozdělení nejistoty je přibližně lognormální.

Měřená veličina	Hmotnostní zlomek zakázané látky ve vzorku.
Nejistota	Relativní standardní nejistota je 35 %.
Specifikace	Horní povolená mez $L_u$ je 2 ng/g.
Rozhodovací pravidlo <i>Vysoká jistota správného odmítnutí</i>	<i>Koncentrace zakázané látky bude považována za nadlimitní, pokud pravděpodobnost hodnoty, že hodnota koncentrace bude větší než mez, bude <math>\geq 95</math> %.</i>
Rozdělení	Předpokládá se lognormální rozdělení hodnot měřené veličiny.
Ochranné pásmo	V případě lognormálního rozdělení lze ochranné pásmo vypočítat pomocí faktoru nejistoty $^F U$ ze vzorce $2 - ^F U \approx \exp(ku_{\text{rel}})$ ; $k$ je 1,64, z jednostranného horního 95% kvantilu normálního rozdělení a $u_{\text{rel}} = 0,35$ , což poskytne $^F U \approx \exp(1,64u_{\text{rel}}) = 1,78$ . Ochranné pásmo $g$ pro správné odmítnutí lze pak vypočítat jako: $g = L_u \times ^F U - L_u = 1,6$ ng/g.
Mez přijetí	3,6 ng/g
Pozorovaná hodnota	3,3 ng/g

Obázek 3a ukazuje umístění oblastí přijetí a odmítnutí pro specifikaci s horní mezí s vysokou spolehlivostí správného odmítnutí. Pozorovaná hodnota 3,3 ng/g je pod mezí přijetí rovné 3,6 ng/g, tedy v oblasti přijetí. **Vzorek je vyhovující.**

#### POZNÁMKA

Zásadní je předpoklad typu rozdělení. Pokud bychom v tomto případě předpokládali normální rozdělení, oblast přijetí by byla nižší – 3,2 ng/g a vzorek by byl nevyhovující,  $L + g = L(1 + ku) = 2(1 + 1,64 \times 0,35) = 3,2$ . Porovnání lognormálního a normálního rozdělení pravděpodobnosti je na obr. 5.



**Obrázek 5. Pravá strana pravděpodobnostních rozdělení pro mezní hodnotu 2 ng/g s relativní standardní nejistotou 35 % pro normální a lognormální rozdělení vykazuje rozdíl v horních koncích, což vede k odlišným mezím přijetí. Horizontální čára při hustotě 0,05 usnadňuje vizuální porovnání**

## Příloha C: Riziko výrobce a odběratele

### Úvod

Hlavní část příručky se věnuje rozhodovacím pravidlům a předpokládá existenci výsledku a přidružené rozšířené nejistoty. Některé pokyny, např. ILAC G8 [6], zahrnují doporučení pro nastavení rozhodovacích pravidel určených k řízení rizik výrobce anebo odběratele. Tato příloha poskytuje stručné vysvětlení rizika výrobce a odběratele a „specifického“ a „globálního“ rizika.

### Riziko výrobce a odběratele

Riziko výrobce a odběratele jsou pojmy vycházející z managementu výrobního procesu, lze je však dobře použít k mnoha situacím posuzování a využívají se např. při „výběrové přejímce“ výrobků. Ve výrobním prostředí je „riziko výrobce“ pravděpodobnost nesprávného odmítnutí přijatelných výrobků. Nazývá se tak, protože vede k nadbytečným nákladům výrobce. Obdobně „riziko odběratele“ je pravděpodobnost nesprávného přijetí nevyhovujícího produktu, možnosti, že odběratel může přijmout vadný výrobek, který prošel kontrolou.

Uvedený případ znázorňuje obrázek 6. Horní křivka „procesní rozdělení“ představuje rozdělení hodnot přidružených k produktům vyrobeným ve výrobním procesu. Hodnoty jsou pro nějakou důležitou měřenou charakteristiku, např. dávka účinné látky v léčivu, balená hmotnost potraviny, koncentrace alkoholu v nápoji.  $L_1$  a  $L_u$  představují povolené horní a dolní meze charakteristiky. Pro jednoduchost obrázek předpokládá, že jsou nastaveny jako meze přijetí pro kontrolu, tedy nepoužívá se žádné ochranné pásmo. Produkt mezi  $L_1$  a  $L_u$  je vyhovující produkt, mimo je nevyhovující. Hodnota A na obrázku je nevyhovující. Rozdělení přidružené k nevyhovující hodnotě A je rozdělení výsledků měření, které lze pozorovat na zkušebních výrobcích s touto hodnotou charakteristiky. Šedá část výsledků zapadá do mezí přijetí. Tato část představuje poměr falešného přijetí produktu A. Je příkladem rizika odběratele, pravděpodobnosti přijetí nevyhovujícího produktu.

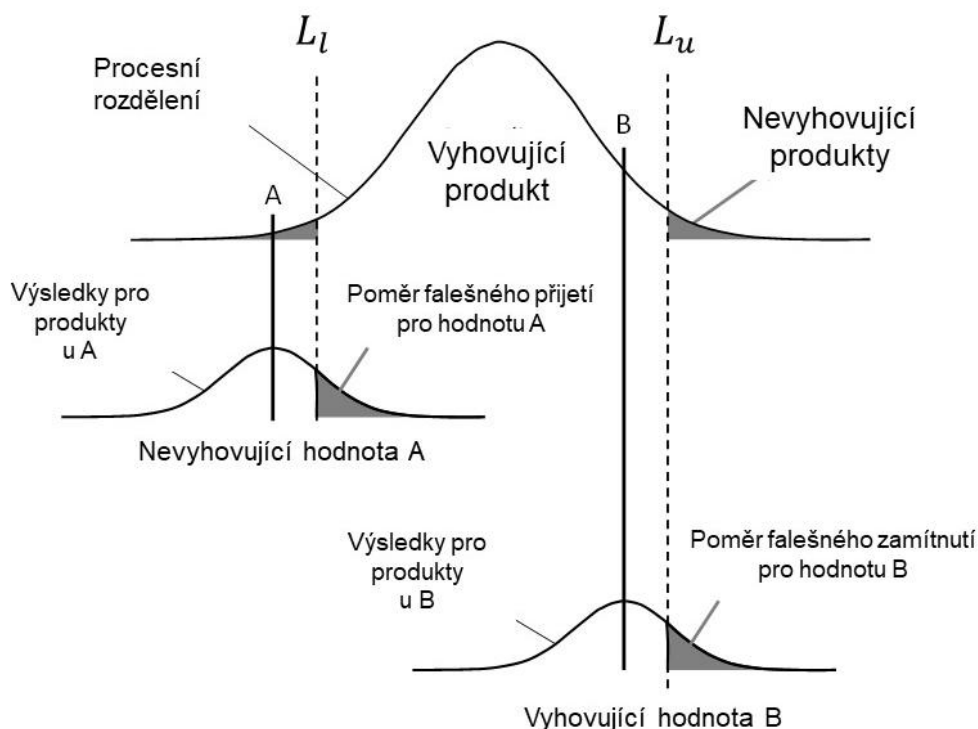
Hodnota B na obr. 6 ilustruje riziko výrobce. Hodnota B je uvnitř povolených mezí, ale existuje pravděpodobnost (šedá část rozdělení pro výsledky na hodnotě B), že výsledky padnou mimo meze přijetí. Tato část je riziko výrobce.

POZNÁMKA: výše uvedený popis vychází z tradičního modelu pro rizika výrobce a odběratele. Ta předpokládají proces, který generuje výsledky s pravými hodnotami jako A a B a následné rozdělení chyb měření vedoucí k rozdělení výsledků měření použitých pro rozhodování. Jedná se o teoretický model. Novější přístupy začínají u pozorovaných hodnot, nejistot a omezených informacích o procesu a zohledňují závěry a pravděpodobnosti, které z nich lze získat. Tento pohled je stručně popsán níže.

### Specifická a globální rizika

Část očividně vyhovujících výsledků na obr. 6 znázorněná šedou částí výsledků přidružených k hodnotě A je specifická pro produkty s touto hodnotou charakteristiky. Jedná se o příklad „specifického rizika“, tj. pravděpodobnosti nesprávného rozhodnutí o shodě spojené s produktem na určité hodnotě. V případě hodnoty A se jedná o specifické riziko výrobce pro hodnotu A. Obdobně riziko výrobce znázorněné šedou částí rozdělení pro hodnotu B je specifickým rizikem výrobce pro hodnotu B.

Důležitou vlastností specifického rizika je, že závisí téměř výhradně na rozdělení výsledků měření pro danou pravou hodnotu charakteristiky. Z pohledu zkušební laboratoře s výsledkem měření, laboratorní odhad specifického rizika závisí na pozorované hodnotě a nejistotě měření. Pro jakoukoliv hodnotu pro zkušební položku (tj. produkt) je specifické riziko menší, je-li menší nejistota.



**Obrázek 6. Riziko výrobce a odběratele. Obrázek znázorňuje rozdělení (nahore) hodnot zájmoví měřené veličiny generované výrobním nebo jiným procesem spolu s povoleným rozsahem mezi mezemi  $L_l$  a  $L_u$ , spolu s rozděleními pro pozorované hodnoty z produktů A a B. Viz text pro další podrobnosti**

Avšak specifické riziko nepopisuje celkovou pravděpodobnost nesprávných rozhodnutí každého druhu, protože zkušební položky (produkty) s různými hodnotami budou mít každá své vlastní specifické riziko. Proto existuje druhá důležitá pravděpodobnost nazývaná „globální riziko“. Globální riziko představuje pravděpodobnost nesprávných rozhodnutí získaných pro statistické rozdělení celé produkce. Pro případ rizika odběratele představuje globální riziko odběratele kombinovanou pravděpodobnost nesprávných rozhodnutí o přijetí, tedy kombinaci specifických rizik pro všechny možné hodnoty pro nevyhovující produkt vážené jejich četností výskytu. Podobně globální riziko výrobce vzniká z kombinace všech specifických rizik výrobce při všech různých hodnotách mezi  $L_l$  a  $L_u$ .

POZNÁMKA: Globální riziko vypočteme jako součet všech specifických rizik pro každou možnou hodnotu vynásobenou jejich četností výskytu. Pro spojitě rozdělení jako ta na obr. 6 je pravděpodobnost výskytu nahrazena výškou křivky popisující procesní rozdělení (hustotu) a součet pak integrací přes oba procesy a rozdělení měření. Matematické podrobnosti jsou uvedeny např. v JCGM 106 [9].

Důležitým rozdílem mezi specifickým a globálním rizikem je to, že globální riziko silně závisí na procesním rozdělení, zatímco specifické riziko ne. Například pro hypotetický výrobní proces produkující pouze vyhovující produkty, může být globální riziko odběratele pouze nula, protože není žádná možnost, že výsledkem prohlídky je propuštění nevyhovujícího produktu. Obdobně pro velmi špatný proces s vysokou pravděpodobností generování nevyhovujícího materiálu bude globální riziko odběratele srovnatelně vysoké.

Avšak pro zkušební laboratoř je často rozdělení hodnot produkovaných procesem neznámé. Z tohoto důvodu je pro zkušební laboratoř nejsnazší spolehnout se raději na specifické než globální riziko. Specifická rizika jsou udržována nízká, zejména udržováním malých nejistot, čímž i globální rizika budou malá. Proto ILAC G8 [6] doporučuje tam, kde není žádný jiný důvod pro nastavení rozhodovacích pravidel, nastavit rozhodovací pravidlo tak, aby zůstávalo specifické riziko nízké.

### Specifické riziko pro výsledek měření

Výše bylo uvedeno, že obecný popis obr. 6 je založen na teoretických rozděleních hodnot pocházejících z výrobního procesu a výsledků měření. Prvně uvedené jsou „pravé hodnoty“ pro výrobky opouštějící proces, posledně jmenované je rozdělení pozorovaných hodnot těch pravých hodnot pro zkušební položky. V praxi má laboratoř pouze výsledky měření s nejistotami a někdy může mít informaci z přírodního nebo komerčního procesu o očekávaných hodnotách. Analyzující laboratoř může tedy pouze odhadnout specifické a globální riziko z informací, které má.

Pokud je informace z výrobního procesu pouze málo přínosná nebo když je nejistota malá ve srovnání s šířkou procesního rozdělení, lze specifické riziko přiměřeně odhadnout z nejistoty měření a přidruženého rozdělení. Příslušné riziko je pouze částí rozdělení nejistoty nad povolenou hodnotou. To je stejné jako šedé regiony hodnot A a B na obr. 6.

Pokud je k dispozici dost informací o procesním rozdělení, lze vypočítat specifické riziko pro konkrétní zkoušenou položku na základě procesního rozdělení s využitím předchozí pravděpodobnosti, že zkoušená položka je vyhovující. To může někdy znatelně změnit odhadnutá specifická rizika. Kompletní vysvětlení je nad rozsah této příručky, podrobnosti lze však najít v JCGM 106 [9].

## Příloha D: Definice

Následující definice vycházejí z definic v ASME B89.7.3.1-2001 [7], VIM [4], GUM [5], ILAC G8 [6] nebo ISO/IEC 17025 [8].

*měřená veličina*: veličina, jež je předmětem měření

*rozšířená nejistota*: veličina definující interval hodnot zahrnující výsledek měření, o němž lze předpokládat, že obsahuje velký podíl distribuce hodnot, které mohou být důvodně přiřazeny měřené veličině.

*výsledek měření*: soubor hodnot veličiny přiřazený měřené veličině společně s jakoukoliv další dostupnou relevantní informací

POZNÁMKA: výsledek měření je obecně vyjádřen jako naměřená hodnota a nejistota měření.

*rozhodovací pravidlo*: pravidlo, které popisuje, jak se počítá s nejistotou měření při stanovování shody se specifickým požadavkem

*oblast přijetí*: množina hodnot charakteristiky pro specifikovaný měřicí proces a rozhodovací pravidlo, která má za následek přijetí produktu, pokud je výsledek měření uvnitř této oblasti.

*oblast odmítnutí*: množina hodnot charakteristiky pro specifikovaný měřicí proces a rozhodovací pravidlo, která má za následek odmítnutí produktu, pokud je výsledek měření uvnitř této oblasti.

*mez specifikace (toleranční mez)*: stanovená horní nebo dolní mez dovolených hodnot vlastnosti

*oblast specifikace (toleranční oblast)*: interval dovolených hodnot vlastnosti

*mez přijetí*: stanovená horní nebo dolní mez dovolených hodnot měřené veličiny

*jednoduché přijetí*: rozhodovací pravidlo, ve kterém je mez přijetí stejná jako mez specifikace

*ochranné pásmo*: interval mezi mezi specifikace a odpovídající mezi přijetí

## Literatura

Aktuální citované pokyny a další doporučenou literaturu najdete v Eurachem Reading List, který je dostupný na webové stránce Eurachem, [www.eurachem.org](http://www.eurachem.org).

1. I. Kuselman, F. Pennechi, R. Bettencourt da Silva, 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., (2020). Available from [doi.org/10.1515/pac-2019-0906](https://doi.org/10.1515/pac-2019-0906).
2. F. Pennechi, M. G. Cox, P. Harris, A. M. H. van der Veen, S. L. R. Ellison, Euramet project EMUE-D2-1-Multicomponent Materials (2020). Available from [www.euramet.org](http://www.euramet.org).
3. M. H. Ramsey and S. L. R. Ellison (eds.) Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches (2<sup>nd</sup> ed. 2019). ISBN 978-0-948926-35-8. Available from [www.eurachem.org](http://www.eurachem.org). [Český překlad je součástí: Milde D. (ed.): Kvalimetrie 25: Nejistota vzorkování. Eurachem-ČR, Ústí nad Labem 2020. (ISBN 978-80-86322-13-1), dostupné na [www.eurachem.cz](http://www.eurachem.cz).]
4. JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM). Available from [www.bipm.org](http://www.bipm.org). [Zavedeno v ČR jako: TNI 01 0115:2009 Mezinárodní metrologický slovník – Základní a všeobecné pojmy a přidružené termíny (VIM). ÚNMZ Praha 2009. Zdarma k dispozici na: <https://www.unmz.cz/metrologie/slovniky/>.]
5. JCGM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM). Available from [www.bipm.org](http://www.bipm.org). [Zavedeno v ČR jako: TNI 01 4109-3.1:2011 Nejistoty měření – Část 3: Pokyn pro vyjádření nejistoty měření (GUM:1995) (Pokyn ISO/IEC 98-3).]
6. ILAC G8:2019, Guidelines on Decision Rules and Statements of Conformity. Available from [www.ilac.org](http://www.ilac.org). [Český překlad: ILAC-G8:09/2019 Pokyny pro použití rozhodovacích pravidel a uvádění výroků o shodě. Dostupné na [www.cia.cz](http://www.cia.cz).]
7. ASME B89.7.3.1-2001, Guidelines for decision rules: considering measurement uncertainty in determining conformance with specifications ([asme.org](http://asme.org)).
8. ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories, ISO, Geneva, (2017). [Zavedeno v ČR jako: ČSN EN ISO/IEC 17025:2018. Všeobecné požadavky na kompetenci zkušebních a kalibračních laboratoří. ÚNMZ Praha, 2018.]
9. JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment. Available from [www.bipm.org](http://www.bipm.org). [Zavedeno v ČR jako: TNI POKYN ISO/IEC 98-4 Nejistota měření – Část 4: Úloha nejistoty měření při posuzování shody. ÚNMZ Praha, 2020.]
10. Á. Silva Ribeiro and M. Golze, EUROLAB Technical Report 1/2017: Decision rules applied to conformity assessment, Available from [www.eurolab.org](http://www.eurolab.org).
11. WADA Technical Document – TD2019DL, Decision limits for the confirmatory quantification of threshold substances (2019). Available from [www.wada-ama.org](http://www.wada-ama.org).
12. ISO 10576:2003 Statistical methods – Guidelines for the evaluation of conformity with specified requirements – Part 1: General principles, ISO, Geneva, (2003).



- [Zavedeno v ČR jako: ČSN ISO 10576-1 Statistické metody – Směrnice pro hodnocení shody se specifikovanými požadavky – Část 1: Obecné principy. ČNI Praha, 2004.]
13. Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2002/657/EC).
  14. R. Bettencourt da Silva and A. Williams (eds), Eurachem/CITAC Guide: Setting and using target uncertainty in chemical measurement, (1<sup>st</sup> ed. 2015). Available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad je 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. (ISBN 978-80-86322-09-4)]
  15. A. Williams, Calculations of the expanded uncertainty for large uncertainties using the lognormal distribution, *Accred. Qual. Assur.*, **25**, 335–338 (2020).
  16. A. M. H. van der Veen and G. Nieuwenkamp, Revision of ISO 19229 to support the certification of calibration gases for purity, *Accred. Qual. Assur.*, **24**, 375–380 (2019).
  17. JCGM 101:2008, Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method. Available from [www.bipm.org](http://www.bipm.org).  
[Zavedeno v ČR jako: TNI Pokyn ISO/IEC 98-3/Doplněk 1: Šíření rozdělení užitím metod Monte Carlo. ÚNMZ Praha, 2011.]
  18. A. Williams, An alternative to the effective number of degrees of freedom, *Comparability and Reliability in Chemical Measurement*, *Accred. Qual. Assur.*, **4**, 14–17 (1999).
  19. R. Kacker and A. Jones, On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent, *Metrologia*, **40**, 235–248 (2003).
  20. S. L. R. Ellison and A. Williams (eds), Eurachem/CITAC Guide: Quantifying uncertainty in analytical measurement, (3<sup>rd</sup> ed. 2012), ISBN 0-948926-15-5. Available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad: Suchánek M., Milde D. (eds.): Kvalimetrie 19: Stanovení nejistoty analytického měření. Eurachem-ČR, Praha 2014. (ISBN 978-80-86322-07-0)]
  21. M. H. Ramsey S. L. R. Ellison S.L.R Uncertainty Factor: an alternative way to express measurement uncertainty in chemical measurement. *Accred. Qual. Assur.*, **20**, 153–155 (2015).



---

# KVALIMETRIE

27

2. část

**Výběr, použití a interpretace programů  
zkoušení způsobilosti (PT)**

Pokyn Eurachem

---



# Výběr, použití a interpretace programů zkoušení způsobilosti (PT)

Třetí vydání 2021

## Editoři

Brian Brookman (LGC, Spojené království)  
Ian Mann (SAS, Švýcarsko)

## Složení pracovní skupiny\*

Brian Brookman (předseda) LGC, Spojené království  
Natalie Mason (sekretariát), LGC, Spojené království  
Stéphanie Albarède, CTCB, Francie  
Frank Baumeister, ISWA, Německo  
Ute Braun, MUVA, Německo  
Laura Ciaralli, ISS, Itálie  
Dana Dominová, ČIA, Česká republika  
Magnus Holmgren, RISE, Švédsko  
Hans D. Jensen, DANAK, Dánsko  
Christian Lehmann, DAkkS, Německo  
Ulrich Leist, DRRR, Německo  
Mirja Leivuori, SYKE, Finsko  
Caroline Leonard, Merck, Francie  
Ian Mann, SAS, Švýcarsko  
Richard McFarlane, UKAS, Spojené království  
Raquel Murtula Corbi, LABAQUA, Španělsko  
Sabrina Pepa, Accredia, Itálie  
Erika Sarkany, QualiCont, Maďarsko  
Anne Vegard Stavelin, Noklus, Norsko  
Johannes van de Kreeke, BAM, Německo

*\*v době schválení dokumentu*

## Poděkování

Toto vydání bylo vytvořeno pracovní skupinou EEE-PT (EA-Eurolab-Eurachem). Do této revize byla rovněž zapojena EQALM (Evropská organizace pro externí programy zajišťování kvality v laboratorní medicíně) jako přidružený člen.

## Citování

Tato publikace by měla být citována\* jako: „B. Brookman and I. Mann (eds.) Eurachem Guide: Selection, Use and Interpretation of Proficiency Testing (PT) Schemes (3<sup>rd</sup> ed. 2021). Dostupné z [www.eurachem.org](http://www.eurachem.org).“

*\*Dle požadavků časopisu*



## Obsah

<b>Zkratky a symboly</b> .....	<b>38</b>
<b>1 Úvod</b> .....	<b>40</b>
<b>2 Účel</b> .....	<b>41</b>
<b>3 Definice</b> .....	<b>42</b>
<b>4 Úvod do zkoušení způsobilosti</b> .....	<b>45</b>
4.1 Úloha PT v systému managementu .....	45
4.2 Druhy programů PT.....	45
<b>5 Výběr vhodných programů PT</b> .....	<b>47</b>
5.1 Úvod .....	47
5.2 Strategie účasti v PT .....	47
5.3 Dostupnost programů PT.....	48
5.4 Jak rozhodnout zda zvolený program PT je vhodný pro daný účel? .....	49
<b>6 Použití PT v laboratořích</b> .....	<b>50</b>
6.1 Úvod .....	50
6.2 Identifikace problémů měření.....	50
6.3 Porovnání postupů měření .....	50
6.4 Porovnání schopností obsluhy .....	50
6.5 Porovnání analytických systémů .....	50
6.6 Zlepšování výkonnosti.....	51
6.7 Vzdělávání personálu .....	51
6.8 Výměna informací s poskytovatelem PT.....	51
6.9 Získání důvěry zaměstnanců, vedení a externích uživatelů laboratorních služeb .....	51
6.10 Nejistota měření.....	51
6.11 Použití položek zkoušení způsobilosti jako materiálů pro interní řízení kvality .....	51
6.12 Stanovení preciznosti a/nebo pravdivosti měření .....	52
6.13 Splnění požadavků regulačních a akreditačních orgánů.....	52
<b>7 Jak poskytovatel PT hodnotí výkonnost laboratoře</b> .....	<b>53</b>
7.1 Úvod .....	53
7.2 Základní prvky hodnocení výsledků PT .....	53
<b>8 Interpretace výsledků PT v laboratořích</b> .....	<b>58</b>
8.1 Úvod .....	58
8.2 Hodnocení výkonnosti laboratoří .....	58
8.3 Prověřování nevyhovujících nebo problematických výsledků PT .....	60
<b>Příloha A - Výběr nejvhodnějšího programu PT</b> .....	<b>63</b>
<b>Příloha B - Prověřování nevyhovujících a problematických výsledků PT</b> .....	<b>64</b>
<b>Příloha C - Interpretace údajů PT koncovými uživateli</b> .....	<b>65</b>
<b>Příloha D - Statistické aspekty PT</b> .....	<b>67</b>
<b>Literatura</b> .....	<b>70</b>

## Zkratky a symboly

V této příručce se vyskytují následující zkratky a symboly.

### Zkratky

CIPM	International Committee for Weights and Measures (Mezinárodní výbor pro váhy a míry)
CRM	certifikovaný referenční materiál
DIN	Deutsches Institut für Normung (Německý institut normalizace)
EA	Evropská spolupráce pro akreditaci
EPTIS	mezinárodní databáze programů zkoušení způsobilosti
EQA	externí hodnocení kvality
EU	Evropská unie
ICP-MS	hmotnostní spektrometrie s indukčně vázaným plazmatem
IEC	International Electrotechnical Commission (Mezinárodní elektrotechnická komise)
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine (Mezinárodní federace klinické chemie a laboratorní medicíny)
ILAC	International Laboratory Accreditation Cooperation (Mezinárodní spolupráce v oblasti akreditace laboratoří)
ILC	mezilaboratorní porovnání
IQC	vnitřní kontrola kvality
ISO	International Organization for Standardization (Mezinárodní organizace pro normalizaci)
IUPAC	International Union of Pure and Applied Chemistry (Mezinárodní unie pro čistou a užitou chemii)
MRA	Dohoda o vzájemném uznávání
PCR	polymerázová řetězová reakce
PT	zkoušení způsobilosti
QA	zajišťování kvality
QC	řízení kvality
RM	referenční materiál
SDPA	směrodatná odchylka pro posuzování způsobilosti (standard deviation for proficiency assessment)

### Symboly

$E_n$	$E_n$ skóre výkonnosti
$s^*$	robustní směrodatná odchylka výsledků účastníků
$\sigma_{pt}$	směrodatná odchylka pro posuzování způsobilosti (SDPA)
$p$	počet účastníků
$x_i$	výsledek uváděný účastníkem $i$
$x_{pt}$	vztažná hodnota
$u(x_i)$	standardní nejistota výsledku účastníka $i$
$u(x_{pt})$	standardní nejistota vztažné hodnoty
$U(x_i)$	rozšířená nejistota výsledku uvedená účastníkem $i$



## Výběr, použití a interpretace programů zkoušení způsobilosti (PT)

---

$U(x_{pt})$	rozšířená nejistota vztažné hodnoty
$z$	výkonnostní skóre $z$
$z'$	modifikované $z$ (primární) výkonnostní skóre
$\zeta$	výkonnostní skóre zeta

### 1 Úvod

Pravidelné nezávislé hodnocení technické výkonnosti laboratoře je nezbytné k zajištění platnosti výsledků a mělo by být součástí celkové strategie kvality. Běžným přístupem k tomuto nezávislému hodnocení je využívání nezávislých programů zkoušení způsobilosti (PT) nebo programů externího hodnocení kvality (EQA), jak se tyto často označují ve zdravotnictví. Program PT je systém pro objektivní vyhodnocení výkonnosti laboratoře za použití externích prostředků a zahrnuje pravidelné srovnávání výsledků laboratoře s výsledky jiných laboratoří. Dosahuje se toho tím, že poskytovatel PT distribuuje účastníkům k analýze homogenní a stabilní PT položky a sděluje výsledky měření. Každá distribuce položek PT se uvádí jako „cyklus“. Hlavním cílem programu PT je pomoci účastníkovi posoudit platnost jeho měření. Programy PT se mohou zabývat také preanalytickými a postanalytickými aspekty fáze měření, tj. specifickými postupy prováděnými před analytickým postupem nebo po něm. Může jít například o odběr vzorků nebo interpretaci výsledků měření. Kromě toho se od laboratoří, které usilují o uznání své kompetence prostřednictvím akreditace podle norem ISO/IEC 17025 [2] nebo ISO 15189 [3], vyžaduje účast v příslušném programu PT, včetně účasti na malých mezilaboratorních porovnáních (ILC) [1], pokud příslušný program PT není k dispozici. V některých odvětvích může být účast ve specifických programech povinná. Programy PT se uskutečňují proto, aby přinášely prospěch účastníkům. Avšak i další strany mají legitimní zájem o programy PT. Mezi ně patří zákazníci služeb analytických laboratoří, akreditační orgány, regulační orgány a ostatní koncoví uživatelé laboratorních výsledků. Je důležité, aby poskytovatelé PT měli na paměti potřeby těchto organizací, aby mohli být schopni využít výsledky programů PT k tomu, aby jim pomohly pochopit schopnosti a kompetence laboratoří (viz příloha C).

Je důležité, aby laboratoře měly souhrnnou informaci o dostupnosti a předmětu PT programů v oblastech, ve kterých působí. To jim umožní činit příslušná rozhodnutí o tom, kterého programu se mají účastnit. Je důležité, aby tento druh informací byl široce dostupný a bylo tak možné vybrat nejvhodnější program PT.

Laboratoře musí také dobře rozumět PT, znát cíle PT programů, případná omezení, způsob vyhodnocování údajů poskytovatelem PT a jak jsou údaje poskytovatelem PT vyhodnocovány a jak by se měly údaje z PT programů interně vyhodnocovat a používat.

V této příručce se pojem „měření“ používá obecně pro měření prováděná při kalibraci i zkouškách. Termínem „vyšetření“ se rozumí zkouška kvantitativních i kvalitativních charakteristik (viz ISO 15189). Pro stručnost, pokud není uvedeno jinak, je třeba ustanovení o měření položky PT, postupech měření atd. chápat tak, že se vztahují na jakoukoli laboratorní činnost vedoucí k vykázání výsledku zkoušky způsobilosti, ať už kvantitativního, kvalitativního nebo interpretačního.

Existuje řada klíčových zásad, o kterých pojednává tento dokument, které mohou napomoci zajistit vhodnost účasti v PT programech a u kterých je třeba, aby je zainteresované strany braly v úvahu a rozuměly jim:

- a) program PT, kterého se laboratoř účastní, by se měl co nejvíce podobat rutinní práci laboratoře, například pokud jde o matici, charakteristiky a úroveň vzorků; případné rozdíly by měly být zaznamenány a zohledněny;
- b) laboratoře by měly s PT položkami zacházet jako s rutinními vzorky;
- c) hodnocení a interpretace výkonnosti v programu PT by měly zohledňovat riziko spojené s měřením;
- d) všechny nevyhovující nebo opakovaně problematické výsledky se musí důkladně prozkoumat, aby laboratoř mohla pochopit příčiny špatného výkonu a podle potřeby je napravit;
- e) výkonnost laboratoře v průběhu několika cyklů programu PT a analýza trendů jsou pro určení úspěšnosti účasti nejdůležitější;
- f) dokumentace programu PT, jako jsou protokoly programu, musí poskytovat jasné informace, aby všechny strany pochopily, jak program PT funguje;
- g) poskytovatel PT by měl být otevřený diskusi mezi zúčastněnými stranami, aby se získalo přesnější pochopení programu PT a jeho fungování;
- h) laboratoře by měly považovat účast na PT za vzdělávací nástroj a využívat výsledky programu PT v procesu zlepšování a poskytovat zpětnou vazbu zaměstnancům.

## 2 Účel

Tento dokument má za cíl poskytnout laboratořím směrnici pro:

- a) cíle a přínosy účasti v programech PT;
- b) výběr nejvhodnějšího programu PT;
- c) pochopení základních statistik a skóre výkonnosti, které používají poskytovatelé zkoušení způsobilosti;
- d) používání a interpretaci výsledků PT pro zlepšování celkové výkonnosti laboratoře.

Tento dokument je určen všem organizacím, které provádějí odběr vzorků, zkoušení, kalibrace a vyšetření, například zkušebními laboratořím, kalibračním laboratořím, inspekčním orgánům, biobankám atd. Zahrnuje měření, laboratorní vyšetření a interpretace.

Tento dokument se nezabývá těmi mezilaboratorními porovnáváními, které jsou zaměřeny na hodnocení výkonnostních charakteristik měřicího postupu (např. „kruhové testy“ nebo „mezilaboratorní studie“) nebo přiřazení hodnot referenčním materiálům (např. „certifikace“ RM), ani porovnáváními prováděnými v rámci CIPM MRA [4], jako jsou „klíčová porovnání“, která jsou určena národním metrologickým institutům.

## 3 Definice

### 3.1 Mezilaboratorní porovnání (ILC)

Organizování, provádění a vyhodnocení měření nebo zkoušek stejné nebo podobné položky dvěma nebo více laboratořemi za předem stanovených podmínek

[ISO/IEC 17043, definice 3.4] [5]

### 3.2 Zkoušení způsobilosti (proficiency testing – PT)

Vyhodnocení výkonnosti účastníka vůči předem stanoveným kritériím pomocí mezilaboratorního porovnání

**POZNÁMKA** Někteří poskytovatelé zkoušení způsobilosti v oblasti laboratorní medicíny používají pro své programy zkoušení způsobilosti nebo rozsáhlejší programy termín Externí hodnocení kvality – EHK (External Quality Assessment – EQA).

[ISO/IEC 17043, definice 3.7] [5]

### 3.3 Program zkoušení způsobilosti (program PT)

Zkoušení způsobilosti navržené a prováděné pro specifickou oblast zkoušení, měření, kalibrace nebo inspekce v jednom či více cyklech

[ISO/IEC 17043, definice 3.11] [5]

### 3.4 Položka zkoušky způsobilosti (položka PT)

Vzorek, výrobek, artefakt, referenční materiál, část zařízení, měřicí etalon (standard), soubor dat nebo jiná informace použitá pro zkoušení způsobilosti

[ISO/IEC 17043, definice 3.8] [5]

### 3.5 Poskytovatel zkoušení způsobilosti (poskytovatel PT)

Organizace, která odpovídá za všechny úkoly při přípravě a provádění programu zkoušení způsobilosti

[ISO/IEC 17043, definice 3.9] [5]

### 3.6 Účastník

Laboratoř, organizace či fyzická osoba, která obdrží položky zkoušky způsobilosti a předá poskytovateli zkoušení způsobilosti výsledky k posouzení

**POZNÁMKA** V některých případech může být účastníkem inspekční orgán.

[ISO/IEC 17043, definice 3.6] [5]

### 3.7 Vztažná hodnota

Hodnota přiřazená konkrétní vlastnosti položky zkoušky způsobilosti

[ISO/IEC 17043, definice 3.1] [5]

**POZNÁMKA** U pre- a postanalytických programů zkoušení způsobilosti to může mít například formu odborného posudku - viz bod 3.13.

### 3.8 Směrodatná odchylka pro posuzování způsobilosti (standard deviation for proficiency assessment – SDPA)

Míra rozptýlení používaná při vyhodnocení výsledků zkoušení způsobilosti vycházející z dostupných informací

**POZNÁMKA 1** Směrodatná odchylka se týká pouze poměrné a diferenciální míry výsledků.

POZNÁMKA 2 Ne všechny programy zkoušení způsobilosti vyhodnocují způsobilost na základě rozptýlení výsledků.

[ISO/IEC 17043, definice 3.13] [5]

### 3.9 Měření

Proces experimentálního získávání jedné nebo více hodnot veličiny, které mohou být důvodně přiřazeny veličině

POZNÁMKA 1 Měření se nepoužívá pro jmenovité vlastnosti.

POZNÁMKA 2 Měření v sobě obsahuje porovnání veličin nebo zjišťování počtu entit.

POZNÁMKA 3 Měření předem předpokládá popis veličiny přiměřený určenému použití výsledku měření, popis postupu měření a kalibrovaného měřicího systému pracujícího v souladu se specifikovaným postupem měření, včetně podmínek měření.

[JCGM 200:2012, definice 2.1] [6]

### 3.10 Nejistota měření / nejistota

Nezáporný parametr charakterizující rozptýlení hodnot veličiny přiřazených k měřené veličině na základě použité informace

POZNÁMKA 1 Nejistota měření zahrnuje složky pocházející ze systematických vlivů, jako například složky související s korekcemi a přidělenými hodnotami veličiny etalonů, stejně jako definiční nejistotu. Někdy nejsou odhadnuté systematické vlivy korigovány, ale místo toho jsou začleněny jako složky přidružené nejistoty měření.

POZNÁMKA 2 Parametrem může být například směrodatná odchylka nazývaná standardní nejistota měření (nebo její stanovený násobek) nebo poloviční šířka intervalu s uvedenou pravděpodobností pokrytí.

POZNÁMKA 3 Nejistota měření se obecně skládá z mnoha složek. Některé z těchto složek smějí být vyhodnoceny vyhodnocením nejistoty měření způsobem A ze statistického rozdělení hodnot veličiny z řady měření a mohou být charakterizovány směrodatnými odchylkami. Jiné složky, které mohou být vyhodnoceny vyhodnocením nejistoty měření způsobem B, mohou být také charakterizovány směrodatnými odchylkami vypočtenými z funkcí hustoty pravděpodobností založených na zkušenosti nebo jiné informaci.

POZNÁMKA 4 Obecně se pro daný soubor informací předpokládá, že nejistota měření je přidružena ke stanovené hodnotě veličiny přiřazené k měřené veličině. Modifikace této hodnoty má za následek modifikaci přidružené nejistoty.

[JCGM 200:2012, definice 2.26] [6]

### 3.11 Postup měření

Podrobný popis měření podle jednoho nebo více měřících principů a dané metody měření založený na modelu měření a zahrnující jakýkoliv výpočet k získání výsledku měření.

POZNÁMKA 1 Postup měření je obvykle dostatečně podrobně dokumentován, aby umožnil obslužnému personálu provést měření.

POZNÁMKA 2 Postup měření může obsahovat vyjádření týkající se cílové nejistoty měření.

POZNÁMKA 3 Postup měření se někdy nazývá standardní operační postup, zkráceně SOP.

[JCGM 200:2012, definice 2.6] [6]

POZNÁMKA 4 EA-4/18 [7] používá ekvivalentní termín „proces měření“, který označuje: „Proces měření charakteristiky včetně jakékoli požadované předúpravy vzorku dodaného laboratoři pro provedení měření na příslušném měřicím zařízení.“

### 3.12 Laboratorní vyšetření/vyšetření

Soubor úkonů, jejichž cílem je stanovení hodnoty nebo charakteristik vlastnosti

[ISO 15189] [3]

### 3.13 Charakteristika

Parametr, který je měřen

[EA-4/18] [7]

POZNÁMKA 1 Příklady: arsen, tuk, kreatinin, délka, tvrdost, síla.

POZNÁMKA 2 V kontextu tohoto dokumentu lze tento pojem dále rozšířit o další charakteristiky, jako jsou stanoviska, barva, chuť, přítomnost/nepřítomnost.

### 3.14 Měřená veličina

Veličina, která má být měřena

[JCGM 200:2012, definice 2.3, kromě poznámek] [6]

### 3.15 Předmět

Položka, na kterou se aplikuje měřicí proces

[EA-4/18] [7]

POZNÁMKA Příklady: půda, zelenina, sérum, polystyren, beton.

### 3.16 Oblast odborné kompetence

Oblast odborné způsobilosti vymezená minimálně jedním měřicím procesem, charakteristikou a předmětem, které spolu souvisejí. Příklad: množství arsenu v půdě pomocí ICP-MS

[EA-4/18] [7]

### 3.17 Úroveň účasti

Počet konkrétních činností, které organizace identifikuje v rámci svého rozsahu akreditace, a tedy počet konkrétních zkoušení způsobilosti, pro které by měla být zvážena účast.

[EA-4/18] [7]

### 3.18 Četnost účasti

Počet zkoušení způsobilosti za časovou jednotku, kterých se laboratoř účastní v rámci činnosti uvedené v jejím rozsahu akreditace.

[EA-4/18] [7]

## 4 Úvod do zkoušení způsobilosti

### 4.1 Úloha PT v systému managementu

Pro sledování platnosti svých měření je důležité, aby laboratoř zavedla systém zajišťování kvality (QA), který zahrnuje sledování její výkonnosti porovnáváním s výsledky jiných laboratoří, tam, kde je to dostupné a vhodné. Pro ty laboratoře, které jsou akreditovány nebo usilují o akreditaci, jsou tato opatření důležitou součástí požadavků. Účast laboratoří v PT nebo jiném mezilaboratorním porovnávání je normami ISO/IEC 17025 [2] a ISO 15189 [3] požadována. Pouze v normě ISO 15189 je specifikováno použití alternativních přístupů, pokud není k dispozici mezilaboratorní porovnávání. Poskytovatelé PT, kteří splňují požadavky normy ČSN EN ISO/IEC 17043 [5], jsou považováni za kompetentní (odborně způsobilé).

PT hraje velmi cennou úlohu, protože poskytuje objektivní důkaz kompetence účastníka. Tyto důkazy lze využít ke zlepšení výkonnosti účastníka anebo poskytnutí důvěry v jeho schopnost provádět konkrétní měření.

Navíc účast v programech PT neposkytuje pouze informace o výkonu postupu měření, ale také o dalších aspektech systému managementu, jako je odběr vzorků, příjem/zpracování vzorku, zpracování údajů, interpretace výsledků, hlášení výsledků atd. Nejdůležitější je, aby si laboratoř vytvořila odpovídající strategii pro účast v programech PT (viz 5.2).

PT poskytuje možnost porovnávat data účastníků se vztažnými hodnotami (nebo dalšími výkonnostními kritérii) nebo s výkonností srovnatelných laboratoří. Výsledky z takové účasti pak buď potvrdí účastníkům, že jejich výkonnost je vyhovující, nebo je upozorní, že je třeba zabývat se potenciálními problémy.

Je důležité zdůraznit, že cílem účasti v PT není jen skóre výkonnosti, ale také umožnit účastníkovi, aby se poučil ze své účasti v programech PT a získané informace využil pro zlepšení kvality svých měření.

Ačkoliv hlavním cílem programu PT je vyhodnotit výkonnost účastníků, existuje mnoho dalších přínosů, o kterých pojednává podrobně kapitola 6.

### 4.2 Druhy programů PT

K dispozici jsou různé druhy programů PT, každý z nich je založen na nejméně jednom prvku z následujících čtyř vlastností:

#### 4.2.1 Typ očekávaného výsledku

- a) kvalitativní: výsledky kvalitativních zkoušek se uvádějí v kategorickém (nominálním) nebo ordinálním měřítku. To zahrnuje:
  - PT programy, které vyžadují vykazování na kategoriální stupnici (někdy nazývané „nominální“), kde charakteristická hodnota nemá žádnou velikost (např. typ látky nebo organismu);
  - PT programy na přítomnost nebo nepřítomnost charakteristiky, ať už je určena subjektivními kritérii nebo velikostí signálu z měřicího postupu. To lze považovat za zvláštní případ kategoriální nebo ordinální stupnice, která má pouze dvě hodnoty (nazývané také „dichotomická“ nebo binární);
  - Programy PT vyžadující výsledky uváděné na ordinální stupnici, které lze seřadit podle velikosti, ale u nichž neexistují aritmetické vztahy mezi různými výsledky. Například „vysoký, střední a nízký“ tvoří ordinální stupnici.
- b) kvantitativní: výsledky kvantitativních měření jsou číselné a zpravidla se udávají v intervalové nebo poměrné stupnici.

**POZNÁMKA 1** Intervalová stupnice je měřicí stupnice, kde určitá vzdálenost na stupnici znamená totéž bez ohledu na to, kde se na stupnici nacházíte, ale kde „0“ na stupnici neznamená nepřítomnost měřeného parametru a kde poměry nemají jednotný výklad. Příkladem intervalových stupnic jsou stupnice podle Fahrenheita a Celsia.

**POZNÁMKA 2** Poměrová stupnice je měřicí stupnice, kde určitá vzdálenost na stupnici znamená totéž bez ohledu na to, kde se na stupnici nacházíte, a kde „0“ na stupnici představuje nepřítomnost měřeného parametru a kde poměry mají jednotný význam, např. „4“ na takové stupnici znamená dvakrát více

měřeného parametru než „2“ a „8“ znamená dvakrát více než „4“. Kelvinova stupnice teploty je příkladem poměrové stupnice.

- c) interpretační: neobsahují žádná měření. Položka PT je předmět, výsledek měření, soubor údajů nebo jiný soubor informací vyžadující posouzení v rámci kompetence účastníka.

### 4.2.2 Četnost

- a) příležitostný program: položky PT jsou poskytovány jednorázově;  
b) kontinuální: položky PT jsou poskytovány pravidelně.

### 4.2.3 Distribuční formát

- 1) sekvenční: měřená položka PT se postupně předává od jednoho účastníka k druhému. V tomto případě může být položka PT před předáním dalšímu účastníku vrácena poskytovateli PT aby se zjistilo, zda nedošlo k nějakým změnám položky PT. Je také možné, aby se účastníci sešli na společném místě a měřili stejnou položku PT.  
2) simultánní: při nejběžnějších PT se náhodně vybrané (dílčí) vzorky (podvzorky) z homogenního zdrojového materiálu současně rozesílají účastníkům pro souběžná měření. Po obdržení výsledků poskytovatel PT provede vyhodnocení výkonnosti každého jednotlivého účastníka a skupiny jako celku pomocí statistických technik.

### 4.2.4 Procesy

- a) preanalytické [8]: u tohoto druhu programu PT položkou PT může být předmět (např. hračka), u kterého účastník musí rozhodnout, jaká měření se mají provést nebo soubor dat nebo jiná informace (např. případová studie). Může to být také postup odběru vzorků, který se používá k odběru vzorků pro měření v laboratoři;  
b) analytické: zaměřuje se zejména na analytický proces;  
c) postanalytické: u tohoto druhu programu PT položkou PT může být soubor dat, k nimž má účastník vyjádřit svůj názor nebo interpretaci.

Tyto tři procesy představují tři fáze kompletního cyklu měření; program PT může být navržen tak, aby se zabýval jednou nebo více fázemi nebo částmi určité fáze.

Jedno ze speciálních použití PT, často nazývané jako „slepé“ PT (blind PT), je situace, kdy položka PT je nerozlišitelná od běžných položek zákazníků nebo vzorků, který obdrží účastník. Všechny výše zmíněné druhy programů PT mohou být organizovány jako slepé PT.



## 5 Výběr vhodných PT programů

### 5.1 Úvod

Výběr programu PT od kompetentního poskytovatele PT má zásadní význam pro zajištění co největšího prospěchu účastníka v programu PT. Je proto nezbytné, aby účastník vyhodnotil kompetenci poskytovatele PT.

Účast v programu PT doplňuje vlastní postupy interního řízení kvality (IQC) laboratoře tím, že poskytuje další externí míru její měřicí schopnosti. Všechny laboratoře si proto musí vytvořit odpovídající strategii účasti v PT s cílem účastnit se příslušných programů PT s četností odpovídající jejich podmínkám [7].

Při výběru vhodného programu PT v rámci své odborné kompetence by si laboratoř měla odpovědět alespoň na následující otázky, aby se ujistila, že daný program PT je pro daný účel vhodný (viz 5.4):

- 1) Jakou úroveň zkoušení způsobilosti a četnosti účasti potřebujeme?
- 2) Existují nějaké programy PT pro různé oblasti odborné kompetence?
- 3) Je konkrétní program PT relevantní?
- 4) Je poskytovatel PT kompetentní, např. pracuje podle normy ISO/IEC 17043 [5]?
- 5) Je vybraný program PT nezávislý na jakýchkoli výrobních nebo obchodních zájmech v oblasti zařízení, činidel nebo kalibrátorů v oblasti své působnosti?

V případech, kdy je účast v PT povinným požadavkem stanoveným regulačními orgány, laboratoř nemusí mít možnost volby, pro výběr programu PT a četnost účasti, ačkoli možnost výběru poskytovatele PT je možná.

### 5.2 Strategie účasti ve zkoušení způsobilosti

Před vlastním výběrem programu PT by laboratoře měly zhodnotit svou úroveň a četnost účasti a stanovit si strategii účasti v PT. Toto zhodnocení by mělo být provedeno se zohledněním odborné kompetence laboratoře v různých oblastech. Laboratoř si pak může vybrat nejvhodnější programy PT od kompetentního poskytovatele PT. Poskytovatel PT, který pracuje podle ISO/IEC 17043, může být považován za kompetentního, ale pokud poskytovatel PT není akreditován, laboratoř by si měla ověřit jeho kompetenci vhodnými prostředky (např. posouzením na místě, auditem na dálku, přezkoumáním dokumentů atd.). Akreditace poskytovatelů PT podle ISO/IEC 17043 poskytuje laboratoři potřebný důkaz.

Laboratoř by měla určit svou úroveň a četnost účasti po pečlivé analýze svých ostatních opatření řízení kvality (zejména těch, které jsou schopny zjistit, kvantifikovat a monitorovat vývoj vychýlení (bias) definované velikosti). Úroveň a četnost účasti by měla záviset na rozsahu, v jakém již byla přijata ostatní opatření. Mezi další opatření zajištění kvality mimo jiné patří:

- pravidelné používání (certifikovaných) referenčních materiálů ((C)RM);
- porovnání analýzy nezávislými postupy měření;
- účast na vývoji metod/validačních studiích a/nebo studiích charakterizace referenčních materiálů;
- použití opatření k vnitřnímu řízení kvality;
- další mezilaboratorní nebo intralaboratorní porovnávání, např. analýza slepých (blind) vzorků v laboratoři.

Je třeba si uvědomit, že existují odvětví, kde může být účast v PT obtížná vzhledem k technickým charakteristikám měření, nedostupnosti programů PT, nízkému počtu existujících laboratoří v daném odvětví atd. Někdy může být zkoušení způsobilosti možné nebo ekonomicky proveditelné pouze pro části prováděného postupu měření. V těchto případech je nejdůležitější vhodnost jiných opatření zajišťování/řízení kvality (QA/QC). Laboratoř by měla zvážit případné veškeré legislativní požadavky na úroveň a četnost účasti.

Laboratoře by měly být schopny zdůvodnit a tam, kde se to vyžaduje, i zdokumentovat odborné důvody, které vedly k jejich rozhodnutí o úrovni a četnosti účasti ve zkoušení způsobilosti.

### 5.2.1 Úroveň účasti

Laboratoř by měla nejprve stanovit oblasti své odborné kompetence. Oblasti odborné kompetence laboratoře lze vymezit jedním postupem měření, jednou charakteristikou a jediným předmětem. Některé oblasti mohou obsahovat více než jeden postup měření, charakteristiku nebo předmět, pokud lze prokázat rovnocennost a srovnatelnost.

Při určování oblasti odborné kompetence může být vhodné zvážit postup po jednotlivých krocích počínaje měřicí technikou přes charakteristiky k předmětům. A to proto, že je pravděpodobnější, že v dané oblasti odborné kompetence bude s jedním postupem měření spojeno několik předmětů anebo charakteristik než naopak. S odkazem na:

- a) postup měření: je možné, nikoliv běžné, zahrnout do stejné oblasti odborné kompetence různé postupy měření, pokud je lze považovat za rovnocenné;
- b) charakteristika, tj. parametr, který má být měřen, stanoven nebo identifikován: v téže oblasti odborné kompetence je možné uvést více než jednu charakteristiku;
- c) analyzované předměty: do stejné oblasti odborné kompetence je možné zahrnout různé předměty za předpokladu, že zahrnuté matrice, objekty nebo materiály jsou srovnatelné povahy.

Jakmile laboratoř vymezí své oblasti odborné kompetence, lze „úroveň účasti“ považovat za definovanou.

### 5.2.2 Četnost účasti

Laboratoř by měla zvážit míru rizika, která se jí týká, odvětví, v němž působí, nebo používaných postupů měření. To lze určit například zvážením:

- počtu prováděných měření;
- fluktuace odborného personálu;
- zkušenosti a znalostí odborného personálu;
- zdrojů metrologické návaznosti (např. CRM, národní etalony/standards);
- známé nestability postupu měření;
- významu a konečného využití naměřených údajů (např. forenzní věda představuje oblast vyžadující vysokou úroveň jistoty).

POZNÁMKA EA 4/18 [7] obsahuje pokyny a některé případové studie ilustrující volbu úrovně a četnosti účasti.

## 5.3 Dostupnost programů PT

Informace o poskytovatelích a/nebo dostupnosti programů PT lze zjistit z různých zdrojů:

- a) Z různých mezinárodních databází, které uvádějí dostupné programy PT, např.:
  - databáze EPTIS [9], uvádějící stovky programů PT provozovaných po celém světě;
  - databáze IFCC [10], uvádějící řadu programů PT pro zdravotnictví;
- b) Národní akreditační orgány mohou poskytovat podrobnosti o akreditovaných poskytovatelích PT a jejich zaměření;
- c) Od obdobně zaměřených laboratořích, které se již zúčastňují zkoušení způsobilosti, nebo mají informace o odpovídajících programech zkoušení způsobilosti.
- d) Poskytovatelé PT ve vlastní zemi laboratoře mohou mít rovněž informace o programech jiných poskytovatelů PT;
- e) Vyhledáváním na internetu s použitím vhodných klíčových slov může též poskytnout užitečné informace.

## 5.4 Jak rozhodnout, zda zvolený program PT je vhodný pro daný účel

Pokud jsou dostupné podobné programy PT a je třeba provést výběr, laboratoř má vzít v úvahu, že různé programy PT budou poskytovat různou míru vhodnosti pro daný účel a že nebude časté, aby existoval program PT, který bude dokonale laboratoři vyhovovat. V praxi se tedy musí vybrat ten program, který se nejvíce hodí pro daný účel.

Kritéria používaná k určení vhodnosti programu PT pro daný účel by měla zahrnovat mimo jiné body vyjádřené následujícími otázkami:

- a) Jsou počet a velikost PT položek dodávaných poskytovatelem PT vhodné pro prováděné zkoušky?
- b) Jsou typy PT položek a/nebo úrovně/koncentrace nabízených programem PT podobné těm, se kterými se laboratoř setkává?
- c) Je popsán statistický model a bere v úvahu různé měřicí postupy?

POZNÁMKA 1 Statistický model zahrnuje proces plánování, sběru, hodnocení výkonnosti a uvádění údajů z programu PT.

POZNÁMKA 2 Mnozí poskytovatelé PT mají na svých webových stránkách příklady svých zpráv z PT nebo kopie protokolů, umožňující posoudit používané hodnocení výkonnosti (určení skóre).

POZNÁMKA 3 Jsou specifikace výkonnosti přizpůsobeny činnosti zákazníka? Většina programů PT používá specifikace výkonnosti odrážející současný stav (statistický rozptýlení), ale ve zdravotnictví je důležité zohlednit i dopad výsledku na péči o pacienta. V tomto případě jsou vhodnější specifikace výkonnosti založené na klinickém hodnocení nebo biologické variabilitě (pevně stanovené přijatelné meze) [11].

POZNÁMKA 4 Je směrodatná odchylka pro posuzování způsobilosti (SDPA) používaná poskytovatelem PT vhodná pro daný účel laboratoře?

- d) Je počet a složení účastníků vhodné pro daný program PT?

POZNÁMKA V některých případech jsou výsledky závislé na postupu měření a je třeba vzít v úvahu počet účastníků používajících konkrétní postup měření (srovnatelná skupina).

- e) Je četnost cyklů dostatečná?

- f) Má poskytovatel PT odpovídající zkušenosti a je kompetentní?

POZNÁMKA Akreditace poskytovatele PT akreditačním orgánem podle normy ISO/IEC 17043 je důkazem jeho kompetence.

Programy PT mají důležitou vzdělávací roli. Pokud poskytovatel PT poskytne informace o výkonnosti různých postupů měření a zohlední také nejistotu výsledků udávanou účastníky, může to poskytnout cenné informace pro další hodnocení postupů měření.

Je důležité poznamenat, že je na odpovědnosti laboratoře rozhodnout, která kritéria použije k porovnání a posouzení významu daného programu zkoušení způsobilosti. Diagram v příloze A poskytuje užitečný návod pro provedení tohoto procesu výběru. Při posuzování programu PT na základě otázek obsažených v diagramu by laboratoř měla například zvážit, které aspekty jsou povinné, užitečné nebo nejsou nutné. To by mělo pomoci určit, zda jsou program PT a potřeby laboratoře jsou dostatečně srovnatelné. Pokud ano, měla by laboratoř seriózně zvážit účast v programu PT. Řada poskytovatelů PT umožňuje účast pouze v jednom cyklu. Pokud není laboratoř plně přesvědčena o vhodnosti daného programu PT je toto dobrá volba. Někdy je možné zakoupit položky PT z předchozích cyklů programu PT spolu s příslušnou zprávou o daném PT. To je také dobrá možnost, jak posoudit vhodnost konkrétního programu PT předtím, než padne rozhodnutí zúčastnit se všech cyklů programu PT.

## 6 Použití zkoušení způsobilosti v laboratořích

### 6.1 Úvod

V kontextu této příručky se pod slovem laboratoř chápou všechny organizace provádějící zkušební nebo kalibrační činnosti, například zkušební, kalibrační a zdravotnické laboratoře, inspekční orgány, biobanky a výrobci RM. Základním využitím PT pro laboratoř je posouzení její výkonnosti při provádění konkrétních zkoušek, měření nebo kalibrací. Účast v programu PT poskytuje laboratoři možnost porovnat své výsledky s ostatními laboratořemi prostřednictvím nezávislého externího hodnocení.

Výsledky a informace získané na základě účasti v programech PT poskytují laboratořím buď potvrzení, že jejich výkonnost je vyhovující nebo signál, že existují potenciální problémy a že by se měla provést jejich náprava. Pro maximalizaci výhod plynoucích z účasti v programech PT je nezbytné, aby účastníci věnovali pozornost dokumentům poskytnutým poskytovatelem PT.

Avšak užití PT by mělo být mnohem širší než pouhé konstatování, že laboratoř je kompetentní či nikoliv. Jak bylo uvedeno v kapitole 4.1, laboratoře mohou těžit z účasti v programech PT různými způsoby. Některé z nich jsou uvedeny níže [12].

### 6.2 Identifikace problémů při měření

Pokud výsledek laboratoře v programu PT ukazuje na neuspokojivou výkonnost, měl by být zahájen proces zkoumání možných zdrojů chyb (viz 8.3). Bez účasti v programu PT by mohly takové zdroje chyb zůstat neodhaleny a laboratoř by nebyla schopna podniknout odpovídající nápravná opatření. To by mohlo mít za následek, že laboratoř by nadále poskytovala špatné výsledky svým zákazníkům nebo dalším partnerům. Kromě toho by mohly takové chyby vést ke ztrátě dobrého jména laboratoře nebo k právním nebo jiným úkonům ze strany zákazníků nebo jiných zúčastněných stran, jako jsou regulační orgány. Z tohoto pohledu je možno považovat používání PT za nástroj managementu rizik a zlepšování kvality.

### 6.3 Porovnání postupů měření

Pro některé laboratoře může jejich účast v programech PT sloužit k vyzkoušení jejich výkonnosti při použití nových, upravených nebo zřídka prováděných měření. V jiných případech může účast poskytovat příležitost porovnat výsledky dosahované laboratoři za použití různých postupů měření (nebo při stanovení na různých koncentračních úrovních atd.) s těmi, které se normálně v laboratoři běžně používají.

Zpráva o PT může v některých případech obsahovat shrnutí a srovnání všech měřicích postupů nebo komerčních souprav používaných účastníky, jako je tomu v případě systému vigilance diagnostických zdravotnických prostředků in vitro platného v EU [13]. U nových nebo neobvyklých činností tyto údaje mohou být velmi cenné a mohou laboratoři pomoci při budoucím výběru vhodných měřicích postupů nebo by mohly naznačit potřebu dalšího šetření před zavedením nových měřicích postupů.

### 6.4 Porovnání schopností obsluhy

Pokud je k dispozici dostatečné množství PT položek, aby analýzu mohl provádět více než jeden pracovník v rámci laboratoře, má laboratoř další výhodu v tom, že může porovnávat výsledky svých pracovníků. Kromě toho by to mohlo poskytnout určité vstupy pro vyhodnocení nejistoty měření laboratoři u příslušných měření.

Mohlo by to také umožnit laboratoři porovnat preciznost v rámci jejich pracovníků s publikovanými (nebo jinak dostupnými) údaji pro daná měření.

Samotný program PT může v některých případech umožňovat uvádění výsledků jednoho nebo více pracovníků.

### 6.5 Porovnání analytických systémů

Výsledky PT mohou poskytnout objektivní externí posouzení relativní výkonnosti analytických systémů používaných v laboratoři (na stejných nebo různých místech).

### 6.6 Zlepšování výkonnosti

Když laboratoř není spokojena se svými výsledky v programu PT, poskytuje to managementu laboratoře vhodnou příležitost zabývat se oblastmi, kde by se jeho zkoušky mohly v budoucnosti zlepšit. To může zahrnovat například další výcvik obsluhy, zavedení nových nebo upravených postupů měření, zlepšení vnitřního řízení kvality údajů nebo úpravy, kalibraci či výměnu zařízení (viz oddíl 8.3).

### 6.7 Vzdělávání personálu

Mnohé z programů PT mají jako jeden ze svých cílů poskytnout informace o měřicích postupech, interpretaci dat, stanovení nejistoty atd., vycházející z celkových výsledků programu PT nebo poskytnuté experty, kteří jsou zapojeni do vyhodnocování těchto výsledků. Některé programy PT zahrnují komplexní vzdělávací funkci pro účastníky a jednotlivé pracovníky obsluhy.

### 6.8 Výměna informací s poskytovatelem PT

Po vydání zprávy o zkoušení způsobilosti mají obvykle laboratoře možnost spojit se s poskytovatelem zkoušení způsobilosti a získat dodatečné informace o výsledcích nebo doporučení týkající se potenciální příčiny nevyhovujících výsledků. Někteří poskytovatelé též pořádají „setkání účastníků“, která mohou poskytovat laboratořím velmi užitečné informace.

### 6.9 Získání důvěry zaměstnanců, vedení a externích uživatelů laboratorních služeb

Úspěšný výkon v programu PT může jednotlivým zaměstnancům a jejich přímým nadřízeným poskytnout dodatečnou sebedůvěru. Další členové vedení, včetně těch bez příslušných odborných znalostí, se mohou též ujistit o úspěšné výkonnosti personálu laboratoře, často v oblastech se zásadním významem pro činnost a odpovědnost jejich organizace.

Externí uživatelé služeb laboratoře, včetně jejich zákazníků a stran, jichž se výstupy měření dotýkají, mohou též získat větší důvěru, když si uvědomí, že laboratoř je ochotna nechat pravidelně hodnotit svoji výkonnost účastí v PT.

### 6.10 Nejistota měření

Výsledky laboratoře z účasti v PT mohou být s patřičnou opatrností použity ke kontrole vyhodnocené nejistoty měření, protože tato nejistota by měla být srovnatelná s rozptýlením výsledků získaných laboratořmi v průběhu několika cyklů PT.

„Přístup PT“ se může ve specifických případech použít i k vyhodnocení nejistoty. Pokud například všichni účastníci programu PT používají stejný postup měření, je daná směrodatná odchylka ekvivalentní odhadu reprodukovatelnosti a v zásadě ji lze použít stejným způsobem jako směrodatnou odchylku reprodukovatelnosti získanou z mezilaboratorního porovnání provedeného za účelem charakterizace výkonnosti postupu měření [14–17]. Výsledky PT pro vzorkování lze použít k vyhodnocení nejistoty měření vyplývající z vychýlení/bias mezi jednotlivými vzorkaři [18].

### 6.11 Použití položek PT jako materiálů pro interní řízení kvality

V některých programech PT, kde je účastníkům poskytnut dostatek stabilního materiálu, by mohl být nevyužitý materiál použit jako materiál pro řízení kvality k monitorování výkonnosti měření v rámci postupů interního řízení kvality laboratoře.

V případě potřeby lze vztažné hodnoty pro položku PT považovat za platné jako interní referenční hodnoty pro řízení kvality měření, výcvik obsluhy atd.

Informační leták Eurachem [19] poskytuje další informace o používání přebytečných materiálů z PT.

### 6.12 Stanovení preciznosti a/nebo pravdivosti měření

V závislosti na návrhu programu PT mohou sloužit některé programy pro stanovení preciznosti (opakovatelnosti a reprodukovatelnosti) nebo pro porovnání pravdivosti metod měření použitých v daném

programu PT. Ve většině případů není stanovení preciznosti a pravdivosti metod primárním cílem programu PT. Pro tento účel je často třeba mít další informace, které mohou být získány od poskytovatele PT.

### **6.13 Splnění požadavků regulačních a akreditačních orgánů**

Úspěšná výkonnost laboratoře v programu PT (nebo efektivní náprava problémů měření po nevyhovující výkonnosti) může regulačním a akreditačním orgánům poskytovat důvěru v laboratoře, jejichž údaje schvalují nebo jinak uznávají. Zřejmým přínosem pro laboratoře je zachování jejich postavení jako kompetentních organizací.

Největší hodnotu by však měly mít interní přínosy pro laboratoře, jejich zaměstnance a vedení, pokud budou PT považovat za zásadní nástroj pro trvalé udržování a zlepšování důvěry bez ohledu na to, zda se laboratoř musí účastnit programu PT pro účely její akreditace.

## 7 Jak poskytovatel PT hodnotí výkonnost laboratoře

### 7.1 Úvod

Výsledky z programů PT mohou nabývat různých forem, zahrnujících široký rozsah druhů dat a odpovídajících statistických rozdělení. Účelem tohoto oddílu je tedy ukázat na hlavní aspekty statistických modelů používaných poskytovateli PT, aby laboratoře mohly lépe porozumět prováděným hodnocením. To by mělo pomoci laboratoři při výběru vhodného programu PT a interpretaci výsledků. Avšak vzhledem k rozsahu používaných různých technik není v tomto dokumentu možné zabývat se všemi statistickými hledisky. Je důležité, aby statistický model použitý poskytovatelem PT odpovídal druhu a účelu organizovaného programu PT. Dále plný popis statistického modelu použitého poskytovatelem PT zpracování by měl být účastníkům plně k dispozici. Statistické techniky, kterým se dává přednost, byly popsány v normě ISO 13528 [20], ale je možné používat i další platné postupy rozhodnutí.

Výchozí předpoklady statistických přístupů používaných pro programy PT jsou většinou založeny na normálním rozdělení dat. Je však běžné, že soubor výsledků účastníků s přibližně normálním rozdělením vykazuje výrazné chvostování a malý podíl odlehlých výsledků. Původní přístup používaný poskytovateli PT (a stále ještě používaný pro některé programy PT) zahrnoval použití statistického testování pro odhalení přítomnosti odlehlých hodnot v daném datovém souboru. Běžnějším přístupem nyní používaným poskytovateli PT a doporučený normou ISO 13528, je však robustní statistika [21, 22]. Robustní statistika má tu výhodu, že snižuje příspěvek odlehlých hodnot u vypočtených statistických parametrů, jako je průměr a směrodatná odchylka. Existuje řada robustních statistických přístupů, z nichž některé jsou popsány v normě ISO 13528.

### 7.2 Základní prvky hodnocení výsledků PT

#### 7.2.1 Obecně

Jedním ze základních prvků každého programu PT je zhodnocení výkonnosti každého účastníka. To vyžaduje kritéria pro hodnocení vykazovaných výsledků. Pro hodnocení kvantitativních výsledků musí poskytovatel PT stanovit dvě hodnoty, které se používají pro takové hodnocení:

- 1) vztažnou hodnotu;
- 2) směrodatnou odchylku pro posuzování způsobilosti (SDPA).

Kromě toho se očekává, že poskytovatel PT uvede nejistotu měření a prohlášení o metrologické návaznosti vztažné hodnoty, jak je uvedeno v normě ISO/IEC 17043 [5]. Význam, potřeba a proveditelnost hodnocení nejistoty se musí určit na základě návrhu programu PT.

Pro stanovení vztažné hodnoty a SDPA lze použít různé metody popsané v normě ISO 13528. Neexistuje žádný striktní standardizovaný protokol podrobně předepisující statistický model zpracování, které se má použít, nicméně toto zpracování by mělo být v zásadní shodě s modely popsány v normě ISO 13528. Statistický model by měl být popsán poskytovatelem PT, běžně buď ve formě protokolu programu PT nebo ve zprávě z PT a laboratoře by jej měly vzít v úvahu při výběru vhodného programu PT.

POZNÁMKA Hodnocení kvalitativních výsledků je uvedeno v 7.2.6.

#### 7.2.2 Vztažná hodnota

Norma ISO 13528 popisuje v podstatě pět metod, jak získat vztažnou hodnotu, pracovní odhad skutečné hodnoty:

- 1) z přípravy položky PT,
- 2) použitím CRM,
- 3) z výsledků z jedné laboratoře,
- 4) z konsenzuální hodnoty z expertních laboratoří,
- 5) z konsenzuální hodnoty z výsledků účastníků.

Další informace o těchto přístupech jsou uvedeny v příloze D.

### 7.2.3 Směrodatná odchylka pro posuzování způsobilosti (SDPA)

Norma ISO 13528 popisuje v podstatě pět metod, jak stanovit směrodatnou odchylku pro posuzování způsobilosti, tedy přijatelného rozpětí výsledků účastníků.

- 1) na základě expertního úsudku,
- 2) na základě zkušeností z předchozích cyklů programu PT,
- 3) pomocí obecného modelu,
- 4) použitím směrodatných odchylek opakovatelnosti a reprodukovatelnosti z předchozího mezilaboratorního porovnání preciznosti metody měření,
- 5) z údajů získaných v témže cyklu programu PT.

Další informace o těchto přístupech jsou uvedeny v příloze D.

V současné době obvyklým přístupem, jak stanovit vztažnou hodnotu a SDPA je použití výsledků účastníků PT k výpočtu obou hodnot (běžně nazývaných "konsenzuální" hodnoty). V Harmonizovaném protokolu pro PT v analytických laboratořích (Harmonised Protocol for the Proficiency Testing of Analytical Chemistry Laboratories) vydaném IUPAC [23] se však důrazně doporučuje, aby metody tvorby skóre byly založeny na kritériu vhodnosti pro daný účel a odpovídající konkrétním okolnostem stanovení. Tam, kde je to možné, by poskytovatel PT měl zakládat SDPA na hodnotě vhodné pro daný účel spíše než na hodnotě měnící se od jednoho cyklu ke druhému v závislosti na rozptylu výsledků poskytnutých účastníky. Používání hodnoty vhodné pro daný účel usnadní sledování výkonnostních skóre během následných cyklů programu PT.

### 7.2.4 Hodnocení výkonnosti

Vyhodnocením výkonnosti (nebo přidělením skóre) dává poskytovatel PT přidanou hodnotu k analytickým výsledkům uváděným účastníkem. Vyhodnocení výkonnosti se poskytuje normalizovaným způsobem proto, aby se staly všechny výsledky porovnatelné a účastník mohl okamžitě pochopit smysl hodnocení.

Používání nejistoty měření při hodnocení výkonnosti je čím dál častější a zlepšuje se porozumění tomuto hledisku. Musí se brát v úvahu dva zdroje nejistoty měření:

- 1) nejistota měření vztažné hodnoty,
- 2) nejistota měření výsledku účastníka.

Vzhledem k rozmanitým účelům programů PT není možné určit jedinou univerzální metodu hodnocení. Proto je k dispozici řada různých metod používaných pro hodnocení výkonnosti. Nejčastější z nich jsou uvedeny níže. V ISO 13528 jsou uvedena i další statistická zpracování, která nejsou zahrnuta do tohoto dokumentu.

- a) „z skóre“ je nejčastěji používaný způsob, který udává míru odchylky výsledku od vztažné hodnoty. Nejistoty měření vztažné hodnoty a uváděného výsledku se neberou v úvahu. Vypočítá se jako:

$$z = (x_i - x_{pt}) / \sigma_{pt}$$

$x_i$  je výsledek uváděný účastníkem  $i$

$x_{pt}$  je vztažná hodnota

$\sigma_{pt}$  je směrodatná odchylka pro posuzování způsobilosti (SDPA)

- b) „z' skóre“ se používá v případě zohlednění nejistoty vztažné hodnoty (viz 7.2.6), takže se zahrnuje standardní nejistota vztažné hodnoty:

$$z' = (x_i - x_{pt}) / \sqrt{\sigma_{pt}^2 + u^2(x_{pt})}$$



kde:

- $x_i$  je výsledek uváděný účastníkem  $i$   
 $x_{pt}$  je vztažná hodnota  
 $\sigma_{pt}$  je směrodatná odchylka pro posuzování způsobilosti (SDPA)  
 $u(x_{pt})$  je standardní nejistota vztažné hodnoty

- c) „ $\zeta$  skóre“ je užitečné pro hodnocení schopnosti účastníků poskytovat výsledky blízké vztažné hodnotě v rámci jimi deklarované nejistoty měření. Zohledňuje se tedy standardní nejistota vztažné hodnoty  $i$  výsledku účastníka:

$$\zeta = (x_i - x_{pt}) / \sqrt{u^2(x_i) + u^2(x_{pt})}$$

kde:

- $x_i$  je výsledek uváděný účastníkem  $i$   
 $x_{pt}$  je vztažná hodnota  
 $u(x_i)$  je standardní nejistota výsledku účastníka  $i$   
 $u(x_{pt})$  standardní nejistota vztažné hodnoty

- d) " $E_n$  skóre" je další systémové skóre, které zohledňuje rozšířenou nejistotu vztažné hodnoty a výsledku účastníka:

$$E_n = (x_i - x_{pt}) / \sqrt{U^2(x_i) + U^2(x_{pt})}$$

kde:

- $x_i$  je výsledek uváděný účastníkem  $i$   
 $x_{pt}$  je vztažná hodnota  
 $U(x_i)$  je rozšířená nejistota výsledku uváděného účastníkem  $i$   
 $U(x_{pt})$  je rozšířená nejistota vztažné hodnoty

**POZNÁMKA** Norma ISO 13528 upozorňuje, že kombinování rozšířených nejistot, jako je tomu u skóre  $E_n$ , neumožňuje konzistentní interpretaci, pokud obě nemají stejný faktor pokrytí a stupně volnosti.

Pro skóre  $z$ ,  $z'$  a  $\zeta$  se běžně používá následující ocenění:

- i)  $|skóre| \leq 2,0$  takové skóre ukazuje na „vyhovující“ výkonnost a nedává vzniknout žádnému hlášení;
- ii)  $2.0 < |skóre| < 3,0$  skóre poukazuje na „problematickou“ výkonnost a signalizuje varování.
- iii)  $|skóre| \geq 3,0$  takové skóre ukazuje na „nevyhovující“ výkonnost a vyžaduje opatření.

Při interpretaci skóre  $z'$  a  $\zeta$  je třeba postupovat opatrně [24]:

- Skóre  $z'$  správně slouží ke standardizaci odchylky od vztažné hodnoty, ale nedokáže rozlišit mezi špatným výsledkem a špatnou vztažnou hodnotou.
- Skóre  $\zeta$  se zvyšuje buď s rostoucí odchylkou od vztažné hodnoty, nebo s klesající vykázanou nejistotou, takže větší skóre  $\zeta$  může znamenat velkou chybu, podhodnocenou nejistotu nebo obojí.

Pro  $E_n$  skóre se běžně používá následující ocenění:

- i)  $|E_n| \leq 1,0$  skóre by mohlo znamenat „vyhovující“ výkon a nevytváří žádný signál, pokud jsou nejistoty platné a odchylka ( $x_i - x_{pt}$ ) je menší, než účastník potřebuje;
- ii)  $|E_n| > 1,0$  skóre by mohlo znamenat „nevyhovující“ výkon a potřebu přezkoumat vyhodnocenou nejistotu nebo napravit problém s postupem měření.

Hodnocení musí probíhat konzistentně v jednotlivých cyklech programů PT, aby byly výsledky v po sobě jdoucích cyklech srovnatelné. Jedině tímto způsobem může účastník sledovat dlouhodobé trendy své výkonnosti. Proto je vhodnější, aby směrodatná odchylka pro posuzování způsobilosti vycházela z kritérií vhodnosti pro daný účel (viz 7.2.3), a nikoli z rozptýlení výsledků účastníků.

### 7.2.5 Alternativní přístupy k hodnocení výkonnosti

Některé programy PT používají jako ukazatel výkonnosti prostý rozdíl mezi vztažnou hodnotou a výsledkem účastníka, často označovaný  $D$ . Ten lze také vyjádřit jako procento vztažné hodnoty,  $D\%$ .

$$D_i = x_i - x_{pt} \qquad D_i\% = 100(x_i - x_{pt})/x_{pt}$$

Rozdíl  $D$  nebo  $D\%$  se obvykle porovnává s kritériem založeným na vhodnosti pro daný účel nebo na očekávané výkonnosti. Jeho výhodou je jednoduchost pro analytika obeznámeného s danou oblastí, ale nemá konzistentní výklad pro různé charakteristiky.

### 7.2.6 Vliv nejistoty vztažné hodnoty

Standardní nejistota vztažné hodnoty závisí na řadě faktorů. Patří mezi ně metoda použitá k odvození vztažné hodnoty, a pokud je odvozena z měření provedených v několika laboratořích, též na počtu laboratoří. Metody výpočtu standardní nejistoty vztažné hodnoty jsou obsaženy v ISO 13528.

Pokud je standardní nejistota ( $u(x_{pt})$ ) příliš velká ve srovnání se směrodatnou odchylkou pro posuzování způsobilosti SDPA pak existuje riziko, že některé laboratoře obdrží problematické nebo nevyhovující hodnocení výkonnosti v důsledku nepřesnosti stanovení vztažné hodnoty, nikoliv vlivem nějakých příčin v laboratořích. Proto je třeba stanovit standardní nejistotu vztažné hodnoty a seznámit s ní laboratoře účastníci se v programu PT.

Pokud je  $u(x_{pt}) < 0,3\sigma_{pt}$  [20], pak je standardní nejistota vztažné hodnoty zanedbatelná a nemusí se s ní počítat při interpretaci výsledků zkoušky způsobilosti.

Pokud výše uvedené kritérium není splněno, poskytovatel PT již obvykle učinil jeden z následujících kroků:

- a) použil odlišnou metodu pro stanovení vztažné hodnoty tak, aby její nejistota odpovídala výše uvedenému kritériu;
- b) použil nejistotu vztažné hodnoty při interpretaci výsledků dané zkoušky způsobilosti (viz výše pro  $z'$  skóre,  $\zeta$  skóre nebo  $E_n$  skóre);
- c) uvedl samostatné hodnoty a nejistoty pro každý dílčí soubor (například pro účastníky používající různé postupy měření), pokud byla vztažná hodnota odvozena z výsledků účastníků a velká nejistota vznikla z rozdílů mezi identifikovatelnými dílčími populacemi účastníků;
- d) informoval účastníky zkoušky způsobilosti, že nejistota vztažné hodnoty není zanedbatelná.

### 7.2.7 Kvalitativní a interpretační programy PT

Kvalitativní programy PT (jak je uvedeno v ISO 13528) a interpretační programy PT vyžadují zvláštní pozornost ve fázích návrhu, přiřazení hodnot a hodnocení výkonnosti (přiřazování skóre), protože:

- vztažné hodnoty jsou velmi často založeny na odborném posudku a
- statistické zpracování určené pro spojitě a počítatelné údaje není použitelné pro kvalitativní údaje. Například nemá smysl brát průměry a směrodatné odchylky výsledků ordinální škály, i když je lze seřadit.

V normě ISO 13528 jsou uvedeny následující mechanismy pro odvození vztažných hodnot v kvalitativních a interpretačních programech PT:

- a) na základě posouzení odborníky;
- b) použitím referenčních materiálů jako položek zkoušky způsobilosti;
- c) ze znalosti původu nebo přípravy položky (položek) PT;
- d) použitím modu nebo mediánu výsledků účastníků (medián není vhodný pro nominální hodnoty).

Hodnocení výkonnosti obvykle vychází z následujícího:

- účastníci jsou posuzováni pouze podle toho, zda se jejich výsledek přesně shoduje s vztažnou hodnotou pro příslušnou položku PT;
- účastníci jsou hodnoceni prostřednictvím odborného hodnocení, které může zahrnovat určitou formu systému známkování, jehož výsledkem je určitý typ výkonnostního skóre.

### 7.2.8 Odlehlé výsledky

Odlehlá hodnota je pozorování, které je numericky vzdálené od zbytku dat. Pravděpodobnost náhodného výskytu odlehlých hodnot je velmi nízká. Odlehlé hodnoty obvykle nepatří do stejné populace dat, tj. do stejného rozdělení. Často tedy svědčí o chybě v měření nebo v jiné fázi analytického procesu. Poskytovatel PT by měl tyto odlehlé hodnoty buď vyřadit, nebo použít statistiku, která je vůči odlehlým hodnotám odolná-robustní. Pokud existuje podezření, že rozdělení není normální, měly by se použít robustní statistické techniky, které jsou odolné vůči asymetrii. Směs dvou nebo více rozdělení, která mohou být dvěma nebo více odlišnými subpopulacemi, např. v důsledku použití dvou různých postupů měření, může někdy také vypadat, jako by existovalo více odlehlých hodnot. Norma ISO 13528 poskytuje návod pro možná následná řešení.

Norma ISO 13528 se výslovně zabývá problematikou odstraňování hrubých chyb. Zjevné hrubé chyby, jako je vykazování výsledků v nesprávných jednotkách nebo záměna výsledků z různých položek PT, se vyskytují ve většině cyklů PT a zhoršují výkonnost následných statistických metod. Doporučuje se, aby poskytovatelé PT odstranili zjevné hrubé chyby ze souboru dat v rané fázi analýzy dat, před použitím jakýchkoli robustních postupů nebo aplikací jakýchkoli testů k identifikaci statistických odlehlých hodnot.

Poskytovatel PT by měl uvést, jakým způsobem zohledňuje odlehlé hodnoty a jak se vypořádává s hrubými chybami při zpracování údajů z jednotlivých cyklů PT.

## 8 Interpretace výsledků zkoušení způsobilosti v laboratořích

### 8.1 Úvod

Účast v programu PT má pro laboratoř omezenou hodnotu, pokud se v laboratoři nevyužije hodnocení její výkonnosti a obecné informace obsažené ve zprávě programu PT.

Je důležité, aby laboratoř nejen vzala na vědomí získané hodnocení výkonnosti, ale vyhodnotila a interpretovala je, a vyvarovala se jakýchkoli chybných či nadměrných interpretací. Vyhodnocení výkonnosti laboratoří by se mělo provádět po každém cyklu a u kontinuálních programů by se měla hodnotit též výkonnost v čase.

### 8.2 Hodnocení výkonnosti laboratoří

#### 8.2.1 Význam hodnocení výkonnosti

Interpretace výkonnosti v PT se týká všech úrovní managementu laboratoře, od techniků až po vrcholové vedení. Pracovníci odpovědní za měření by měli být obeznámeni s fungováním programu PT a měli by obvykle provést počáteční hodnocení. Pokud se provádějí jakákoliv zkoumání, měla by probíhat v rámci postupu pro neshodné práce systému managementu kvality laboratoře. Vedení laboratoře nemusí vždy být obeznámeno s výkonností v PT, ale je velice záhodno, aby získalo odpovídající úroveň porozumění programům PT.

Vzhledem k tomu, že laboratoř by měla používat validované postupy měření spolu s postupy vnitřního řízení kvality, je třeba brát vážně každý špatný výsledek, protože naznačuje, že existuje problém s platností postupů měření a/nebo postupů vnitřního řízení kvality.

Existují určité základní body interpretace výsledků PT, o kterých je vhodné se zmínit před tím, než se přistoupí k podrobnému rozboru tohoto tématu. Jak již bylo uvedeno, PT není o tom, zda měření „projde“ či „neuspěje“, ale o poučení z výsledků. Vyhovující výkonnost laboratoře v jednom cyklu, ve kterém všichni účastníci dosáhli vyhovující výkonnosti, neznamená nutně vysokou úroveň kompetence. Je tak možné, že směrodatná odchylka pro posuzování způsobilosti mohla být příliš velká. Na druhou stranu ani nevyhovující výkonnost v jednom cyklu neznamená, že laboratoř je nekompetentní; tento výsledek se musí prověřit a získat z něj poučení, aby se již neopakoval. Avšak trvale špatná výkonnost poukazuje na zásadní problémy v měřicích postupech laboratoře a v takovém případě by měla laboratoř pečlivě zvážit, zda může pokračovat v nabídce takového měření do doby, než sporné otázky vyřeší.

#### 8.2.2 Přezkoumání výsledků z jednoho cyklu PT

Výsledky každého cyklu PT se mají hodnotit bez ohledu na dosažený výkon, protože vyhovující výsledek nemusí nutně znamenat dobrý výkon.

Měly by se zhodnotit všechny informace uvedené ve zprávě ze zkoušení způsobilosti, nikoliv pouze skóre výkonnosti. Například nevyhovující výkonnost v rámci cyklu, kde většina účastníků dosáhla vyhovující úrovně, by se měla odlišit od nevyhovující výkonnosti v případě, kdy významný podíl účastníků dosáhl jen nevyhovující výkonnosti. Na obě situace by se mělo nahlížet s vážností, protože obě signalizují problémy v měřicích postupech.

Součástí tohoto přezkoumání by vždy mělo být, aby personál laboratoře zkontroloval, zda výsledky uvedené ve zprávě se shodují s těmi, které laboratoř dodala a zvláště pak zda systém hodnocení výkonnosti v programu daným skóre je dobře chápán a je vhodný pro daný účel. Pokud je to potřeba, je nutné se spojit s poskytovatelem PT, aby se vyloučily chybné interpretace výkonnosti.

V odůvodněných případech se laboratoř může rozhodnout přepočítat své výkonnostní skóre pomocí kritéria odpovídajícího jejím podmínkám. Například zákazníci laboratoře nemusí vyžadovat takovou úroveň výkonnosti, jakou předpokládá obvyklá směrodatná odchylka pro posuzování způsobilosti programu, nebo směrodatná odchylka pro posuzování způsobilosti programu může být založena na směrodatné odchylce neobvykle podobného souboru výsledků účastníků (viz 8.2.3).

Pokud po důkladném prozkoumání laboratoř dojde k závěru, že výsledek je vskutku nevyhovující, měla by přistoupit k nápravným opatřením (viz oddíl 8.3).

Výsledky laboratoře z její účasti v PT lze rovněž použít ke kontrole platnosti nejistoty měření laboratoře (viz oddíl 6 a 7.2.4 c).

### 8.2.3 Vyhodnocení vhodnosti výpočtu skóre hodnocení z

Pro hodnocení výkonnosti se nejčastěji používá skóre  $z$  (a související skóre  $z'$  a  $\zeta$ ). Skóre  $z$  se vypočítá vydělením rozdílu mezi výsledkem zkoušky laboratoře a vztažnou hodnotou hodnotou směrodatné odchylky pro posuzování způsobilosti. Posouzení, zda je výsledek laboratoře považován za vyhovující, sporný nebo nevyhovující, proto silně závisí na hodnotě směrodatné odchylky pro posuzování způsobilosti použité ve jmenovateli rovnice pro skóre  $z$ . Čím větší je směrodatná odchylka pro posuzování způsobilosti, tím menší je skóre  $z$ , a tedy tím lepší je zjevný výkon laboratoře v cyklu PT. Laboratoř musí odpovědět na otázku, zda byla směrodatná odchylka pro posuzování způsobilosti vybrána na základě správných kritérií a zda je vhodná pro účely laboratoře. Mezi otázky, které je třeba zvážit, patří:

- a) Je směrodatná odchylka pro posuzování způsobilosti odvozena na čistě statistickém základě?  
V tomto případě se odhad směrodatné odchylky pro posuzování způsobilosti stává nespolehlivým při malém počtu účastníků, velkém rozptylu výsledků nebo při obojím. Malý počet účastníků proto může vyžadovat stanovení směrodatné odchylky pro posuzování způsobilosti pomocí alternativního přístupu (viz oddíl 7.2.3 a příloha D). Totéž platí, pokud je směrodatná odchylka pro posuzování způsobilosti odvozena ze statistických údajů z předchozích cyklů programu PT, pokud v těchto cyklech PT byl malý počet účastníků, velké rozptýlení výsledků nebo obojí.
- b) Je směrodatná odchylka pro posuzování způsobilosti odvozena z legislativních nebo normativních dokumentů nebo z nich vychází?  
Právní předpisy nebo normy často obsahují informace o opakovatelnosti a/nebo reprodukovatelnosti postupů měření. Na základě těchto informací lze posoudit, zda je směrodatná odchylka pro posuzování způsobilosti vhodná pro daný účel.
- c) Je směrodatná odchylka pro posuzování způsobilosti v souladu s vlastními údaji o zajišťování kvality v laboratoři?  
Odhady opakovatelnosti a reprodukovatelnosti odvozené z validačních nebo verifikačních studií, výsledky analýzy materiálů pro řízení kvality a informace o směrodatné odchylce pro posuzování způsobilosti používané jinými poskytovateli PT mohou napovědět, zda je směrodatná odchylka pro posuzování způsobilosti realistická.
- d) Je směrodatná odchylka pro posuzování způsobilosti realistická v porovnání s požadavky trhu?  
Jako ukazatele lze použít požadavky zákazníka na opakovatelnost a nejistotu měření týkající se výsledků nebo podmínek pro analýzy, jakož i požadavky na přesnost postupů měření odvozené z rozhodnutí o hodnocení výrobků jako nevyhovující či vyhovující.

Pokud laboratoř dojde k závěru, že směrodatná odchylka pro posuzování způsobilosti je pro účel laboratoře příliš malá nebo velká, měla by zvážit volbu alternativní hodnoty a přepočítat své výkonnostní skóre. Pokud laboratoř používá menší směrodatnou odchylku pro posuzování způsobilosti, je třeba ji používat s opatrností v tom smyslu, že by to bylo možné pouze tehdy, pokud by úroveň homogenity a stability položek PT byla stále přiměřená [20]. Ve většině případů to bude vyžadovat diskusi s poskytovatelem PT, protože informace potřebné k ověření této skutečnosti nebudou ve zprávě PT k dispozici. Pokud úroveň homogenity a/nebo stability není dostatečná nebo ji nelze zkontrolovat, PT není vhodný pro daný účel a laboratoř by měla zvolit jiný PT, který odpovídá jejím požadavkům.

### 8.2.4 Sledování výkonnosti ve zkoušení způsobilosti v průběhu času

Kromě pečlivého hodnocení výsledků jednotlivých cyklů je třeba sledovat výkonnost v PT v průběhu času, aby se rozpoznaly potenciální problémy ve vztahu k špatné preciznosti, systematickým chybám nebo chybám obsluhy. Jednotlivé laboratoře mohou sledovat svou výkonnost v průběhu času různými způsoby, například pomocí grafických přístupů nebo výpočtem dlouhodobých složek vychýlení (bias) [25, 26].

Velmi užitečné pro sledování výkonnosti laboratoře v PT je grafické vynesení skóre výkonnosti v jednotlivých cyklech. Obvykle toto uvádí ve své zprávě z PT poskytovatel PT, nebo si to může vyneset sám účastník. Tento postup umožňuje zvýraznit neobvyklé nebo neočekávané výsledky, stejně jako pomoci při určení trendu. Dá se očekávat, že opatření vnitřního řízení kvality laboratoře by měly běžně objevit trendy spojené například

s nesprávnou kalibrací instrumentace nebo údržbou, nebo s nesprávnou manipulací/nevhodným používáním reagensů. Sledování výkonnosti PT v průběhu času doplňuje běžné postupy interního řízení kvality. Příklady typických diagramů lze nalézt v normě ISO 13528.

Aby bylo možné určit, zda se výkonnost laboratoře v průběhu času zlepšuje nebo zhoršuje, musí být údaje z následných cyklů PT porovnatelné. Pokud se však směrodatná odchylka pro posuzování způsobilosti získává ze souboru dat pro každý cyklus PT, může mít stejné měření ze dvou různých cyklů PT různou hodnotu směrodatné odchylky pro posuzování způsobilosti, což vede k rozdílnému výpočtu skóre výkonnosti. Pokud se směrodatná odchylka pro posuzování způsobilosti použitá v po sobě jdoucích cyklech PT výrazně liší, mohli by si účastníci vypočítat vlastní skóre  $z$  (nebo jiné skóre výkonnosti) pomocí vhodné směrodatné odchylky pro posuzování způsobilosti. Lze použít směrodatnou odchylku pro posuzování způsobilosti získanou z literatury (např. ze standardního postupu měření zveřejněného národním nebo mezinárodním normalizačním orgánem, jako je ISO nebo DIN). Pokud takové hodnoty nejsou k dispozici, může laboratoř stanovit vlastní kritéria na základě účelu účasti v programu PT nebo významu měření. Laboratoř si může zvolit jakoukoli vhodnou hodnotu, pokud dokáže svou volbu zdůvodnit. Poznamenejme, že zvolená směrodatná odchylka pro posuzování způsobilosti nemusí být konstantní, tedy může být koncentračně závislá. Při výběru jiné směrodatné odchylky pro posuzování způsobilosti (viz 8.2.3), než který používá poskytovatel PT, je třeba postupovat s určitou opatrností. Jestliže se laboratoř rozhodne přepočítat své skóre výkonnosti, měla by svoji volbu zdůvodnit a zdokumentovat.

Použití kombinovaných výsledků (například zprůměrovaných nebo sečtených u stejných nebo různých položek PT) by se mělo používat pouze s opatrností [20]. Pokud je poskytl poskytovatel PT, měla by být omezení takového přístupu jasně uvedena ve zprávě o PT.

### 8.3 Prověřování nevyhovujících a problematických výsledků zkoušení způsobilosti

#### 8.3.1 Potřeba prověřování

Všechny laboratoře vykáží příležitostně nevyhovující nebo problematické výsledky ve zkoušení způsobilosti. Když k tomu dojde, laboratoř by je měla zřetelně označit a zdokumentovat.

Hloubka přezkumu, který je třeba provést, závisí na řadě faktorů. Patří mezi ně kritičnost postupu měření, četnost neuspokojivých výsledků a důkaz o vychýlení (bias). V každém případě by laboratoř měla zdokumentovat vyhodnocení výsledků, a to i tehdy, když se rozhodne nepřijímat žádná opatření.

Základním principem je, že každé nevyhovující skóre výkonnosti by mělo být přešetřeno a šetření zdokumentováno, protože to zřetelně ukazuje na problém. Laboratoř by měla mít obecné zásady (politiku) pro prověřování:

- Pochybných skóre výkonnosti u stejných měření;
- Následných skóre výkonnosti stejného měření, která vykazují stejné znaménko vychýlení vůči vztažné hodnotě.

Je však důležité poznamenat, že je na laboratoři, aby si stanovila vlastní kritéria [2, 3] pro zahájení prověřování, a to s ohledem na četnost účasti, vhodnost účelu programu PT, kritičnost měření atd. Klíčovou otázkou je nutnost prošetřit nevyhovující výkonnost a prozkoumat trendy.

#### 8.3.2 Prověřování hlavní příčiny

Pokud je nezbytné provést celkovou prověrku, upřednostňuje se postupný přístup, aby se maximalizovala šance na určení hlavní příčiny problému. Schéma podporující tento přístup je uvedeno v příloze B.

Vhodný postup postupné prověrky by měl zahrnovat následující kroky a měl by zahrnovat pracovníky, kteří měření prováděli:

- Analyzujte problém na základě výchozích údajů, celkové výkonnosti cyklu PT, výsledků z po sobě jdoucích cyklů PT a údajů interního řízení kvality;
- Vypracujte plán nápravného opatření (nápravných opatření);
- Proveďte a zaznamenejte nápravné opatření (nápravná opatření);

d) Zkontrolujte, zda nápravné opatření (nápravná opatření) byla účinná.

### 8.3.3 Příčiny nízké výkonnosti

#### 8.3.3.1 Typické příčiny

Důvody špatné výkonnosti mohou být bohužel četné, což má za následek časovou náročnost a komplexnost prověrky. Protože by však tato prověřování měla vést ke zlepšení výkonnosti laboratoře, je vhodné vynaložit potřebné úsilí. Pro usnadnění prověrky je vhodné si uvědomit, jaké bývají hlavní důvody špatné výkonnosti, aby bylo možné šetření lépe zaměřit. Mezi typické příčiny špatné výkonnosti patří [27]:

- a) příprava vzorku (např. navažování, sušení, extrakce, rozklad, čištění, ředění atd.);
- b) postupy měření;
- c) lidské chyby (např. nevhodný výcvik, chyby při přepisu);
- d) kalibrace;
- e) výběr postupu měření;
- f) chyba ve výpočtu;
- g) problémy uvádění údajů (formát, jednotka, interpretace);
- h) problém vyplývající z položky PT;
- i) přeprava a skladování vzorku;
- j) primární vzorkování;
- k) sledování vzorku (označení, dodavatelský řetězec-chronologická dokumentace);
- l) problém, který vzniká u poskytovatele PT.

Pro identifikaci hlavní příčiny špatné výkonnosti je důležité se zaměřit na potenciální příčiny, které lze rozdělit do následujících skupin:

- a) administrativní chyba;
- b) technický problém (např. postup měření, vybavení, výcvik, interní řízení kvality);
- c) problém mající vztah k programu PT (např. neodpovídající program, nevhodné hodnocení).

Je možné, že ani důkladná prověrka nepovede ke zjištění původu špatné výkonnosti. Pokud se jedná o opakovanou špatnou výkonnost, je třeba zpochybnit laboratorní postupy (technické nebo manažerské).

#### 8.3.3.2 Administrativní chyba

I když administrativní chyby nemají přímý vztah ke kompetenci laboratoře, je třeba zdůraznit, že laboratoře mohou mít potenciální problémy při uvádění výsledků zákazníkům.

Administrativní chyby mohou spočívat v následujícím:

- a) chybách při přepisování;
- b) chybném označení;
- c) chybách desetinné čárky;
- d) výsledcích vykazovaných v nesprávných jednotkách.

Zjistit, zda nedošlo k administrativní chybě, je důležitým prvním krokem šetření. Pokud jsou administrativní chyby pravidelnou příčinou nevyhovujících výsledků, mělo by se šetření zaměřit na hlediska kvality systému managementu.

### 8.3.3.3 Technický problém

Vzhledem ke složitosti laboratorní činnosti se problémy mohou vyskytnout na všech úrovních laboratorních postupů a během prověrky by měl být přezkoumán každý z následujících prvků:

- a) skladování a předběžná úprava položky PT;
- b) postup měření/údaje interního řízení kvality;
- c) vybavení/reagencie/kalibrace;
- d) podmínky prostředí;
- e) zpracování dat.

Pokud prošetření laboratorních postupů neumožní laboratoři zjistit hlavní příčinu, může být potřebné přezkoumat validaci postupu měření.

### 8.3.3.4 Problémy související s programem zkoušení způsobilosti

Špatná výkonnost může též být zapříčiněna skutečností, že vybraný program PT není vhodný, nebo že problém spočívá v položkách PT. Je třeba prošetřit následující body:

- a) rozdíl mezi položkou PT a rutinními vzorky;
- b) potenciální degradace položky PT;
- c) úrovně koncentrací mimo oblast použití postupu měření;
- d) nedostatečná stabilita nebo homogenita položek PT;
- e) nevhodné pokyny pro účastníky;
- f) problémy skladování položky PT;
- g) nevhodná skupina laboratoří;
- h) neodpovídající vztažná hodnota;
- i) nevhodná směrodatná odchylka pro posuzování způsobilosti;
- j) nesprávná vstupní data poskytovatele PT.

Laboratoři se doporučuje, aby svá zjištění projednala s poskytovatelem PT nebo aby posoudila, zda je vybraný program PT vskutku vhodný.



## Příloha A: Výběr nejvhodnějšího programu PT

### Položka PT

- Jaká je matrice?
- Je položka PT reálná nebo simulovaná (modelová)?
- Jsou všechny charakteristiky, které jsou rutinně zkoušeny, dostupné?
- Jsou hodnoty charakteristik (např. koncentrace) odpovídající?
- Jsou použity při uvádění standardní jednotky?

### Účastníci

- Je základna účastníků národní nebo mezinárodní?
- Je počet účastníků nebo velikost podskupiny rovnocenných účastníků přiměřená?
- Jaké měřicí postupy používají účastníci?
- Jaký druh laboratoří se účastní?

### Distribuce položek PT

- Jsou data o distribuci dostupná a vhodná?
- Odpovídá četnost distribuce potřebám laboratoře?
- Umožňuje poskytovatel PT flexibilní účast?

### Výsledky

- Jsou k dispozici termíny pro zasílání výsledků a jsou odpovídající?
- Jak vyžaduje poskytovatel PT, aby byly hlášeny výsledky?
- Mohou účastníci uvádět výsledky získané pomocí jimi zvoleného postupu měření?
- Mohou se uvádět nejistoty měření a budou zohledněny při vyhodnocování výkonnosti?
- Je použitý statistický přístup dostupný a vhodný?

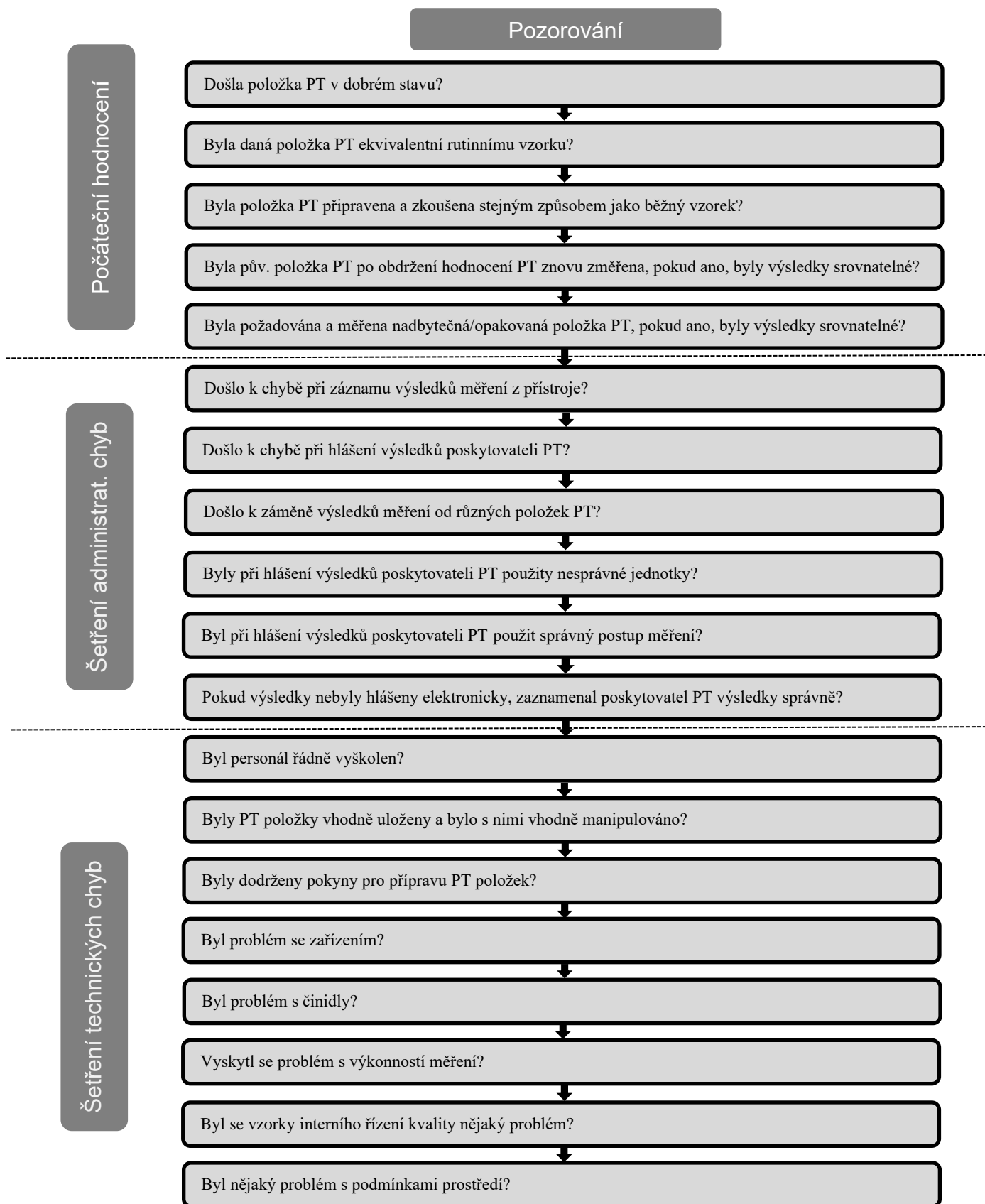
### Zprávy PT

- Jak rychle po uzavěře výsledků jsou poskytovány zprávy o PT?
- Jaké informace jsou poskytovány ve zprávách z PT?
- Odpovídají kritéria hodnocení (např. směrodatná odchylka pro posuzování výkonnosti) účelům laboratoře?
- Je zpráva z PT dostupná v papírové formě, elektronicky nebo online?
- Obsahuje zpráva snadno interpretovatelné grafické shrnutí?
- Je jazyk používaný ve zprávách z PT srozumitelný příslušným pracovníkům?

### Poskytovatelé PT

- Jaký je rámec nabízených programů PT?
- Je poskytována odpovídající zpětná vazba a pomoc?
- Jsou laboratořím poskytovány „nadbytečné PT položky/položky k opakování“ pro provádění šetření špatných výsledků a hodnocení účinnosti nápravných opatření?
- Splňují požadavky normy ISO/IEC 17043?
- Jsou akreditováni akreditačním orgánem podle ISO/IEC 17043?

## Příloha B: Prověřování nevyhovujících a problematických výsledků zkoušení způsobilosti



## Příloha C: Interpretace údajů PT koncovými uživateli

### C.1 Úvod

Laboratoře potřebují prokázat svou kompetenci zainteresovaným stranám, jako jsou akreditační a regulační orgány a zákazníci. Výsledky PT spolu s dalšími činnostmi řízení kvality jsou jedním z prostředků, jak kompetenci prokázat. Protože zkoušení způsobilosti je hodnocení třetí stranou, zainteresované strany stále častěji doporučují či požadují, aby se laboratoře účastnily programů PT, aby měly k dispozici nezávislé hodnocení výkonnosti laboratoře. Účast v PT je mechanismus pro externí řízení kvality výsledků a pro porovnání s jinými laboratořemi, jak vyžadují normy ISO/IEC 17025 [2] a ISO 15189 [3].

Laboratoř je odpovědná za to, že při předávání svých výsledků v PT poskytne zainteresovaným stranám též všechny další potřebné informace (např. přepočítané skóre výkonnosti, prověřování).

### C.2 Akreditační orgány

Akreditační orgány a odborní posuzovatelé angažovaní akreditačními orgány zpravidla dobře chápou úlohu PT a jsou odborně kvalifikovaní v interpretaci výsledků programů zkoušení způsobilosti, získaných laboratořemi, které jsou buď akreditovány, nebo o akreditaci usilují. Odborní posuzovatelé jsou též obecně seznámeni s programy PT, kterých se laboratoře zúčastňují. V případě potřeby akreditační orgán po prostudování protokolů a další dokumentace programu PT kontaktuje poskytovatele PT, aby s ním projednal nebo vyjasnil případné nevyřešené otázky. Úroveň výkonnosti laboratoře v programu PT se stanoví pomocí kritérií stanovených poskytovatelem zkoušení způsobilosti. V některých případech to, co je pro laboratoř nevyhovující výkonnost v programu PT, může ještě být přijatelné nebo vhodné pro daný účel pro potřeby akreditace laboratoře a naopak.

### C.3 Regulační orgány

Regulační orgány musí zajistit, aby měření prováděná v laboratořích, na které se vztahují předpisy nebo směrnice, měla uspokojivou kvalitu. Regulační orgány mohou tudíž používat výkonnost v programu PT jako jeden ze způsobů posuzování kvality, kromě dalších způsobů jako jsou referenční analýzy nebo zaslání kontrolních vzorků k analýze.

Pokud se na vývoji programu PT podílel regulační orgán, bude obsahovat prvky, které jsou pro něj přímo relevantní a budou snadno pochopitelné. V takových případech, kdy regulační orgán používá nezávislý program PT pro své vlastní účely, se doporučuje, aby v plné míře projednal předmět a operační parametry programu s poskytovatelem PT. To umožní dávat výsledky dotyčných laboratoří do souvislosti. Je třeba rozumět statistickým postupům, které používá poskytovatel PT k výpočtu výkonnosti laboratoře, aby bylo možné posuzovat výkonnost laboratoře ve vztahu k mezím, které jsou povoleny v regulacích. V takových situacích může být od poskytovatele PT vyžadováno poradenství, aby nedošlo k nesprávné interpretaci údajů o výkonnosti z programu PT.

### C.4 Zákazníci účastníků se laboratoří

Zákazník laboratoře, která se zúčastňuje programu PT, může používat výkonnost v programu PT jako jeden z nástrojů pro sledování kvality této laboratoře. Je potřeba, aby zákazník dobře chápal, jak program PT funguje a jak poskytovatel PT vypočítává výkonnost v daném programu PT. Ačkoli některé způsoby, jak stanovit výkonnost v programu PT jsou hojně rozšířené, jako je použití z skóre, mohou se používat mnohé další způsoby. Zákazníci by si měli kromě toho uvědomit, že způsob, jakým se vypočítává z skóre a další charakteristiky výkonnosti se může mezi jednotlivými programy PT lišit.

Zákazníci stále častěji zahrnují kritéria výkonnosti v programu PT do zadávací dokumentace a využívají informace o výkonnosti v programu PT poskytnuté potenciálními dodavateli, aby pomohli při rozhodování o tom, které laboratoři bude zakázka zadána. Když použijí výkonnost v programu PT jako kritéria ve výběrovém řízení, měli by zákazníci zajistit, že tam, kde vyhláší „standard výkonnosti“, je tento standard realistický a dosažitelný. Například požadovat, aby laboratoře dosahovaly vyhovující výsledky pro všechny charakteristiky ve všech cyklech programu PT, je nereálné. Poskytovatelé PT poskytují běžně příslušné informace o celkové výkonnosti v programu PT ve zprávě z PT, což umožňuje stanovit dobré měřítko pro srovnání. Zákazníci by

těž měli věnovat pozornost tomu, aby jasně formulovali, o které matrice (charakteristiky) mají zájem. Program totiž může mít širší zaměření a výkonnost laboratoře pro matrice (charakteristiky), které nejsou předmětem přímého zájmu, by mohla být irrelevantní.

Zákazníci musí veškeré údaje týkající se výkonnosti v programu PT od smluvní laboratoře zasadit do správného kontextu; laboratoře by mohly zákazníkovi předložit údaje způsobem, který by vytvářel nerealisticky pozitivní obraz.

Zákazníkům se doporučuje, aby v případě potřeby provedli následující kroky, a získali tak přesný obraz o skutečné výkonnosti laboratoře:

- a) získejte informace o zaměření a fungování dotyčného programu PT (např. protokol z programu PT) od laboratoře nebo poskytovatele PT.
- b) prozkoumejte výkonnost laboratoře v čase, protože pouze jeden cyklus programu PT poskytuje jen úsečný snímek výkonnosti laboratoře.
- c) přezkoumejte celkovou výkonnost všech účastníků při posouzení, jak daná laboratoř pracuje;
- d) vyžádejte si kopie zpráv programu PT (tam, kde nejsou důvěrné) pro potvrzení údajů shrnujících provádění programu PT. Poskytovatel PT může poskytnout tyto údaje, i když obecně to vyžaduje souhlas účastníka.

Jeden nevyhovující výsledek v některém cyklu programu PT neznámá, že laboratoř je špatná, stejně tak jako dosažení na 100 % vyhovujících výsledků v některém cyklu nutně neznámá, že laboratoř je výtečná.

Způsob, jakým laboratoř reaguje na nevyhovující výsledek, poskytne obvykle více informací o laboratoři než samotný výskyt nevyhovujícího výsledku.

## Příloha D: Statistická hlediska zkoušení způsobilosti

### D.1. Hlavní parametry

Jedním ze základních prvků každého zkoušení způsobilosti je vyhodnocení výkonnosti každého účastníka. Za tímto účelem stanovuje poskytovatel PT dvě hodnoty, které se používají pro takové hodnocení:

- 1) Vztažnou hodnotu;
- 2) Směrodatnou odchylku pro posuzování způsobilosti (SDPA).

Jak se uvádí v oddílu 7.2, musí poskytovatel PT též vyhodnotit nejistotu měření vztažné hodnoty.

### D.2 Vztažná hodnota a standardní nejistota vztažné hodnoty

Jak popisuje ISO 13528 [20] existuje v podstatě pět způsobů, jak získat vztažnou hodnotu a k ní přidruženou standardní nejistotu. Za výběr metody je zodpovědný poskytovatel PT:

- 1) Příprava: smíchání materiálů s různými známými koncentracemi ve stanovených poměrech nebo přidání známého množství nebo koncentrace analytu do základního materiálu, který žádný analyt neobsahuje. Tato metoda vyhovuje v mnoha případech, zvláště pak tam, kde je předmětem měření spíše celkové množství analytu spíše než koncentrace. Avšak tato metoda nemůže modelovat obtíže postupů přípravy normálního vzorku (které mohou zahrnovat kroky, jakými jsou extrakce a speciace), kde mohou nastat problémy s výtěžností.

Standardní nejistota se vyhodnocuje kombinací nejistot spojených s přípravou položky PT pomocí vhodného modelu.

- 2) CRM: V případě kdy položkou PT je certifikovaný referenční materiál (CRM) použije se jeho certifikovaná hodnota jako vztažná hodnota. Výhodou tohoto způsobu je, že poskytuje návaznou vztažnou hodnotu, je ale nákladný a vhodné certifikované referenční materiály nejsou často dostupné. Kromě toho jsou CRM často intenzivně zpracovávány, aby byla zajištěna jejich dlouhodobá stabilita, což může ohrozit komutabilitu [28] položek PT.

Standardní nejistota je odvozena z údaje o nejistotě obsažené v certifikátu certifikovaného referenčního materiálu.

- 3) Výsledky z jedné laboratoře: vybrané připravené položky PT jsou změřeny ve vybrané laboratoři buď za použití primární metody, nebo souběžně s certifikovaným referenčním materiálem. Vztažná hodnota se získá přímo použitou primární metodou nebo z kalibrace za použití certifikované hodnoty certifikovaného referenčního materiálu (CRM). Získá se tak návazná hodnota prostřednictvím primární metody, nebo k použitému CRM, ale opírá se o výsledky jediné laboratoře a vhodné primární metody nebo CRM nemusí být dostupné. Referenční metoda nebo CRM používaná jako reference by měly být komutabilní vůči všem metodám měření používaným účastníky [28].

Standardní nejistota se odvodí z výsledků měření vybrané laboratoře a nejistot certifikovaných hodnot použitých CRM.

- 4) Konsenzuální hodnota z expertních laboratoří: stanovení konsenzuální hodnoty získané na základě výsledků skupiny expertních laboratoří, které jsou zběhlé v použitých postupech měření. Jiným problémem je, že je často těžké nebo i nemožné najít skupinu expertních či referenčních laboratoří, jejichž odborné schopnosti jsou mimo jakoukoliv pochybnost a jsou uznávány všemi účastníky PT. Ještě více to platí pro velké mezinárodní programy PT s účastníky z mnoha zemí. U řady měřených veličin (například extrahovatelné olovo v půdě) je skutečná hodnota měřeného parametru v zásadě definována použitým postupem měření. V takových případech se často používá definice „měřená veličina definovaná postupem“ a takové postupy (metody) měření se označují jako „empirické metody“. V těchto případech by měly všechny expertní laboratoře používat stejný postup měření a měly by jej dodržovat do všech detailů. Výsledky takové skupiny expertních laboratoří mohou vykazovat neznámé vychýlení (bias). Expertní laboratoře a používané postupy měření by se měly vyhlásit před zavedením programu PT.

Pokud expertní laboratoře uvádějí nejistoty u výsledků, je odhad hodnoty na základě konsensu výsledků složitým problémem a byla navržena celá řada přístupů, včetně například vážených průměrů, nevážených

průměrů, postupů, které zohledňují nadměrný rozptyl, a postupů, které zohledňují možné odlehlé nebo chybné výsledky a vyhodnocené nejistoty. Poskytovatel PT by proto měl stanovit postup pro odhad, který:

- a) zahrnuje kontroly platnosti uváděných nejistot, například kontrolu, zda uváděné nejistoty plně zohledňují pozorované rozptýlení výsledků;
  - b) používá postup vážení odpovídající rozsahu a spolehlivosti uváděných nejistot, který může zahrnovat rovnoměrné vážení, pokud jsou uváděné nejistoty buď podobné, nebo málo spolehlivé či neznámé;
  - c) umožňuje zohlednit možnost, že uváděné nejistoty nemusí plně zohledňovat pozorované rozptýlení („nadměrné rozptýlení“), například zahrnutím dodatečného členu, který zohledňuje nadměrné rozptýlení;
  - d) připouští možnost neočekávaných odlehlých hodnot u vykazovaného výsledku nebo nejistoty;
  - e) má solidní teoretický základ;
  - f) byl ověřen, například na zkušebních datech nebo v simulacích, aby se ukázalo, že je pro účely programu PT dostatečně vhodný.
- 5) Konsenzuální hodnota z výsledků účastníků: použití konsenzuální hodnoty vytvořené v každém cyklu programu PT a založené na výsledcích získaných účastníky. Konsenzuální hodnota se obvykle odhaduje za použití technik robustní statistiky. Konsenzuální přístup je nepochybně v některých případech nejpřímochařejším přístupem. Když se používají položky PT s přírodní matricí, může být jediným způsobem jak stanovit odhad pravé hodnoty.

Běžný odhad nejistoty pro konsenzuálně přiřazenou vztahnou hodnotu získanou robustním statistickým postupem je:

$$u(x_{pt}) = 1,253 \times \left( \frac{s^*}{\sqrt{p}} \right)$$

kde:

$s^*$  je robustní odhad směrodatné odchylny účastníků

$p$  je počet účastníků

Faktor 1,25 vychází z rozptylu mediánu pro normálně rozdělená data.

Další platné přístupy jsou popsány například v normě ISO 13528.

Mezi omezení tohoto přístupu patří, že:

- a) nemusí existovat reálný konsenzus mezi účastníky;
- b) konsenzus může být vychýlený na základě obecného použití chybné metodologie a toto vychýlení (bias) nebude zohledněno ve standardní nejistotě vztahné hodnoty vypočítané výše popsaným způsobem.

### D.3 Směrodatná odchylna pro posuzování způsobilosti (SDPA)

Jak popisuje norma ISO 13528, existuje v podstatě pět metod, jak stanovit směrodatnou odchylnu pro posuzování způsobilosti, tedy přijatelného rozpětí výsledků účastníků.

- 1) Dle názoru expertů: směrodatnou odchylnu pro posuzování způsobilosti lze nastavit na hodnotu, která odpovídá úrovni výkonnosti, kterou regulační orgán, akreditační orgán nebo odborní experti poskytovatele PT považují pro účastníky za přiměřenou.
- 2) Na základě zkušeností z předchozích cyklů programu PT: směrodatná odchylna pro posuzování způsobilosti může být stanovena hodnotou, která vychází ze zkušeností z předchozích cyklů programu PT. To odpovídá úrovni výkonnosti, kterou by si poskytovatel PT přál, aby ji laboratoře byly schopny dosáhnout.
- 3) Použití obecného modelu: hodnotu směrodatné odchylny pro posuzování způsobilosti lze odvodit z obecného modelu reprodukovatelnosti postupu měření, například z modelu závislého na koncentraci.

Výhodou této metody je objektivita a konzistence napříč měřeními charakteristikami a také to, že je založena na empirických poznatcích.

Nevýhodou tohoto přístupu je, že skutečná reprodukovatelnost konkrétního měřicího postupu se může podstatně lišit od hodnoty vyplývající z modelu, protože použití obecného modelu předpokládá, že reprodukovatelnost závisí jen na koncentrační úrovni analytu a nikoliv na vlastním analytu, postupu měření anebo velikosti vzorku.

- 4) Použití směrodatných odchylek opakovatelnosti a reprodukovatelnosti získaných z předchozího mezilaboratorního porovnání provedeného k posouzení výkonnostních charakteristik postupu měření: pokud je postup měření, který má být použit v programu PT, standardizován a jsou k dispozici informace o opakovatelnosti a reprodukovatelnosti postupu měření, lze směrodatnou odchylku pro posuzování způsobilosti vypočítat na základě těchto informací.
- 5) Z údajů získaných ve stejném cyklu programu PT: při tomto přístupu se směrodatná odchylka pro posuzování způsobilosti použitá v cyklu programu PT odvozuje z výsledků vykázaných účastníky ve stejném cyklu PT. Jedná se o robustní směrodatnou odchylku výsledků vykázaných všemi účastníky.

Nevýhodou tohoto přístupu je to, že taková hodnota se může významně lišit od jednoho cyklu k druhému, to činí obtíž, pokud laboratoř chce použitím hodnot z skóre hledat trendy, které se projevují v průběhu několika cyklů.

## Literatura

Aktualizaci aktuálních nejdůležitějších odkazů naleznete v seznamu literatury Eurachem umístěném v sekci Publikace na webových stránkách Eurachem, [www.eurachem.org](http://www.eurachem.org). Informace o dostupnosti českého znění je uvedena za odkazem v hranatých závorkách.

1. EA-4/21 INF:2018, Guidelines for the assessment of the appropriateness of small interlaboratory comparisons within the process of laboratory accreditation, available from [www.european-accreditation.org](http://www.european-accreditation.org)  
[Český překlad Pokyn pro posuzování vhodnosti malých mezilaboratorních porovnání v rámci procesu akreditace laboratoře, ČIA, Praha 2019 je dostupný na [www.cai.cz](http://www.cai.cz)].
2. ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories, ISO, Geneva.  
[Zavedeno v ČR jako: ČSN EN ISO/IEC 17025:2018. Všeobecné požadavky na kompetenci zkušebních a kalibračních laboratoří. ČAS Praha 2018].
3. ISO 15189:2012, Medical laboratories – Requirements for quality and competence, ISO, Geneva.  
[Zavedeno v ČR jako: ČSN EN ISO 15189 ed. 2:2013. Zdravotnické laboratoře - Požadavky na kvalitu a způsobilost. ÚNMZ Praha 2013].
4. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, CIPM, October 1999, available from [www.bipm.org/en/cipm-mra](http://www.bipm.org/en/cipm-mra).
5. ISO/IEC 17043:2010, Conformity assessment – General requirements for proficiency testing, ISO, Geneva  
[Zavedeno v ČR jako: ČSN EN ISO/IEC 17043:2010 Posuzování shody – Všeobecné požadavky na zkoušení způsobilosti. ÚNMZ Praha 2010].
6. JCGM 200:2012, International vocabulary of metrology — Basic and general concepts and associated terms (VIM, 3<sup>rd</sup> edition), available from [www.bipm.org/en/publications](http://www.bipm.org/en/publications).  
[Zavedeno v ČR jako: TNI 01 0115:2009 Mezinárodní metrologický slovník – Základní a všeobecné pojmy a přidružené termíny (VIM). ÚNMZ Praha 2009. Zdarma dostupné na: <https://www.unmz.cz/metrologie/slovniky/>].
7. EA-4/18 G:2021, Guidance on the level and frequency of proficiency testing participation, available from [www.european-accreditation.org](http://www.european-accreditation.org).  
[Český překlad Návod k určení úrovně a četnosti účasti ve zkoušení způsobilosti, ČIA, Praha 2022 je dostupný na [www.cai.cz](http://www.cai.cz)].
8. Eurachem - Information leaflet on pre- and post-analytical proficiency testing, First English edition, 2009-05-14, available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad Zkoušení způsobilosti – preanalytické a postanalytické užití publikován jako Metodický list 6, 2010 je dostupný na <http://www.eurachem.cz/metodicke-listy.php>].
9. EPTIS PT scheme database, [www.eptis.org](http://www.eptis.org).
10. IFCC laboratory medicine PT database, available from [www.ifcc.org/ifcc-scientific-division/sd-committees/cmd/externalqualityassessment-proficiencytestinginmoleculardiagnosics](http://www.ifcc.org/ifcc-scientific-division/sd-committees/cmd/externalqualityassessment-proficiencytestinginmoleculardiagnosics).
11. Sandberg, S. et al., Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53(6):833-835.
12. ILAC Brochure: 2019, Benefits for laboratories participating in proficiency testing programs, available from [www.ilac.org/publications-and-resources/ilac-promotional-brochures](http://www.ilac.org/publications-and-resources/ilac-promotional-brochures).
13. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU.  
[Český překlad Nařízení Evropského parlamentu a Rady (EU) 2017/746 ze dne 5. dubna 2017



- o diagnostických zdravotnických prostředcích in vitro a o zrušení směrnice 98/79/ES a rozhodnutí Komise 2010/227/EU je dostupný na <https://eur-lex.europa.eu/homepage.html?locale=cs>].
14. Eurolab Technical Report 1/2007 – Measurement uncertainty revisited, Alternative approaches to uncertainty evaluation, March 2007, available from [www.eurolab.org/pubs-techreports](http://www.eurolab.org/pubs-techreports).  
[Český překlad Technická zpráva č. 1/2007 Revize nejistot měření: alternativní přístupy k vyhodnocení nejistot EUROLAB-CZ, Praha 2008 je dostupný na <http://www.4e.cz/dokumenty.htm>].
  15. ISO 21748:2017, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, ISO, Geneva.  
[Zavedeno v ČR jako: ČSN ISO 21748:2012. Návod pro použití odhadů opakovatelnosti, reprodukovatelnosti a pravdivosti při odhadování nejistoty měření. Neplatná, zrušena 2021. ÚNMZ Praha 2012].
  16. Ellison, S. L. R and Williams, A. (eds), Eurachem/CITAC Guide: Quantifying uncertainty in Analytical Measurement, 3<sup>rd</sup> Edition, 2012, available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad KVALIMETRIE 19 Stanovení nejistoty analytického měření. Editoři: Miloslav Suchánek a David Milde, Čtvrté české přepracované vydání EURACHEM-ČR Praha 2014. ISBN 978-80-86322-07-0].
  17. NORDTEST Technical Report 537: Handbook for calculation of measurement uncertainty in environmental laboratories, NORDTEST 2012, available from [www.nordtest.info/index.php/technical-reports/category/environment](http://www.nordtest.info/index.php/technical-reports/category/environment).  
[Český překlad je součástí: KVALIMETRIE 26 První část: Příručka pro výpočet nejistoty měření v environmentálních laboratořích. Milde D. (ed.): Eurachem-ČR, Ústí nad Labem 2020. (ISBN 978-80-86322-15-5), dostupné na [www.eurachem.cz](http://www.eurachem.cz)].
  18. Ramsey, M. H., Ellison, S. L. R. and Rostron, P. (eds.) Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches. Second Edition, Eurachem (2019). ISBN (978-0-948926-35-8), available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad je součástí: KVALIMETRIE 25 První část: Nejistota měření vyplývající z odběru vzorků Příručka metod a přístupů. Milde D. (ed.): Eurachem-ČR, Ústí nad Labem 2020. (ISBN 978-80-86322-13-1), dostupné na [www.eurachem.cz](http://www.eurachem.cz)].
  19. Eurachem – Information leaflet on use of surplus proficiency test items, First English edition, November 2019, available from [www.eurachem.org](http://www.eurachem.org).  
[Český překlad Použití zbylých položek ze zkoušení způsobilosti publikován jako Metodický list 21, 2020 je dostupný na <http://www.eurachem.cz/metodicke-listy.php>].
  20. ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparison, ISO, Geneva.  
[Zavedeno v ČR jako: ČSN ISO 13528:2017. Statistické metody používané při zkoušení způsobilosti mezilaboratorním porovnáváním. ÚNMZ Praha 2017].
  21. Analytical Methods Committee – Robust Statistics Part I & II. Analyst 1989:114:1693-1702.
  22. Thompson, M. and Ellison, S. L. R., Fitness for purpose – the integrating theme of the revised Harmonised Protocol for Proficiency Testing in Analytical Chemistry Laboratories. Accred. Qual. Assur. 2006:11:373-378.
  23. Thompson, M., Ellison, S. L. R. and Wood, R., The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories. Pure Appl. Chem. 2006:78:1:145-196.
  24. AMC technical brief No. 74: Z-Scores and other scores in chemical proficiency testing their meanings, and some common misconceptions, Anal. Methods 2016:8:5553.
  25. Meijer, P., DE Maat, M. P. M, Klufft, C., Haverkate, F., and van Houwelingen, H. C., Long-Term Analytical Performance of Hemostasis Field Methods as Assessed by Evaluation of the Results of an External Quality Assessment Program for Antithrombin, Clin. Chem. 2002: 48:7:1011-1015.
  26. Meijer, P., Haverkate, F., Klufft, C., Performance goals for the laboratory testing of antithrombin, protein C and protein S. Thromb. Haemost. 2006:96:5:584-589.
-

27. Ellison, S. L. R. and Hardcastle, W. A., Causes of error in analytical chemistry: results of a web-based survey of proficiency testing participants, *Accred. Qual Assur.* 2012: 17, 453–464, available from <https://doi.org/10.1007/s00769-012-0894-2>.
28. ISO Guide 35:2017, Reference materials – Guidance for characterization and assessment of homogeneity and stability, ISO, Geneva.  
[Zavedeno v ČR jako: TNI POKYN ISO 35:2019. Referenční materiály – Návod pro charakterizaci a posouzení homogeneity a stability. ÚNMZ Praha 2019].

---

# KVALIMETRIE

27

3. část

## Hodnocení výkonnosti a nejistota v kvalitativní chemické analýze

Pokyn Eurachem/CITAC

---



## 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, Sweden</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>"

\* Závisí na požadavcích časopisů



## Obsah

<b>Obsah</b>	<b>77</b>
<b>Předmluva</b>	<b>79</b>
<b>Účel</b>	<b>80</b>
<b>1 Úvod</b>	<b>83</b>
<b>2 Typy kvalitativní analýzy</b>	<b>85</b>
<b>3 Hodnocení výkonnosti kvalitativní analýzy</b>	<b>86</b>
3.1 Obecné úvahy	86
3.2 Kvantifikace výkonu kvalitativní analýzy	86
3.3 Hodnocení falešně pozitivních a falešně negativních poměrů	89
3.4 Mez detekce a selektivita	92
<b>4 Vyjádření spolehlivosti v kvalitativní analýze</b>	<b>94</b>
4.1 Obecné úvahy	94
4.2 Věrohodnostní poměr	94
4.3 Posteriorní pravděpodobnost	96
4.4 Spolehlivost metrik	97
4.5 Nejistota poměrů	97
<b>5 Uvádění kvalitativního analytického výsledku</b>	<b>99</b>
<b>6 Závěry a doporučení</b>	<b>100</b>
<b>7 Příklady</b>	<b>101</b>
7.1 E1: Identifikace sloučenin hmotnostní spektrometrií s nízkým rozlišením pomocí vyhledávání v databázi nebo přítomnosti charakteristických iontů	101
7.2 E2: Identifikace čištěných sloučenin infračervenou spektrometrií	105
7.3 E3: Identifikace zneužívaných drog v moči technikou enzymové multiplikované imunoanalýzy (EMIT) a alternativní technikou	107
7.4 E4: Identifikace lidského genu SRY v biologickém materiálu pomocí qPCR	109
7.5 E5: Identifikace reziduí pesticidů v potravinách pomocí GC-MS/MS na základě retenčního času a zastoupení iontů	111
7.6 E6: Identifikace SARS-CoV-2 RNA testováním amplifikace nukleové kyseliny	114
<b>Příloha A – Bayesova věta, šance a věrohodnostní poměr</b>	<b>117</b>
A.1 Bayesova věta	117
A.2 Pravděpodobnost a šance	117
A.3 Forma Bayesovy věty pro šance a věrohodnostní poměr	118
<b>Příloha B – Kvalitativní analýza spojená s posouzením shody a mezí stanovitelnosti</b>	<b>119</b>
B.1 Posouzení shody jako kvalitativní analýza	119
B.2 Pozitivní a negativní výsledky posouzení shody	120
B.3 Příklad – Posouzení shody pro rezidua pesticidů v ovoci	120
B.4 Vzorce v tabulkovém procesoru pro posouzení shody pravděpodobností	121
<b>Literatura</b>	<b>122</b>

Úmyslně 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 forenzních [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 položek<sup>1</sup>. 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 stupnicích<sup>2</sup> jinak než jako správnou nebo nesprávnou klasifikaci.

---

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

<sup>2</sup> 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
$\Delta Rn$	Normalizovaná uváděná hodnota minus odezva základní linie
$\rho$	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/nehověl“, 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“<sup>3</sup> 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 nejistot 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).

<sup>3</sup> „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

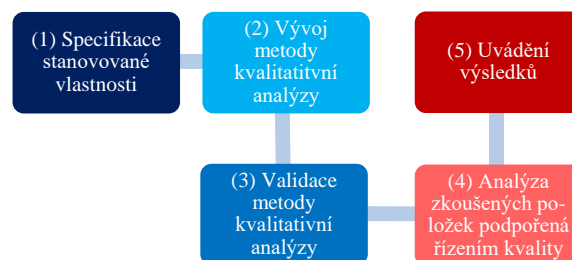
Tento oddíl poskytuje návod k posouzení a vyjádření výkonnosti pro postupy určené k jednoduché klasifikaci do dvou tříd („binární klasifikace“). „Třídy“ jsou zde označeny jako „pozitivní“ nebo „negativní“ k označení „člena zájmové třídy“ nebo „nečlena“. Tato klasifikace pokrývá většinu praktických situací,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

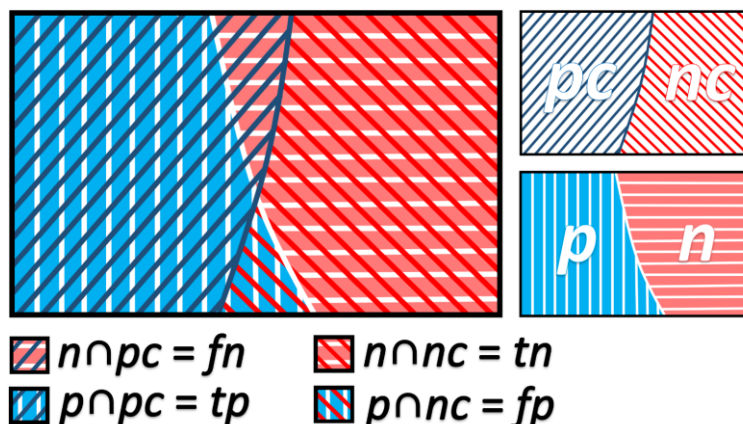
Základním způsobem kvantifikace výkonnosti metody kvalitativní analýzy je výpočet četnosti chybných výsledků. U „pozitivních“ nebo „negativních“ výsledků je užitečné hlásit poměr „skutečně pozitivní“ a „falešně pozitivní“ nebo „skutečně negativní“ a „falešně negativní“. Tyto poměry však mohou být vztaženy 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,  $n \cap nc$ ,  $p \cap pc$ ,  $n \cap nc$  a  $p \cap nc$  definují  $fn$ ,  $tn$  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.<sup>4</sup>

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.

<sup>4</sup> 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

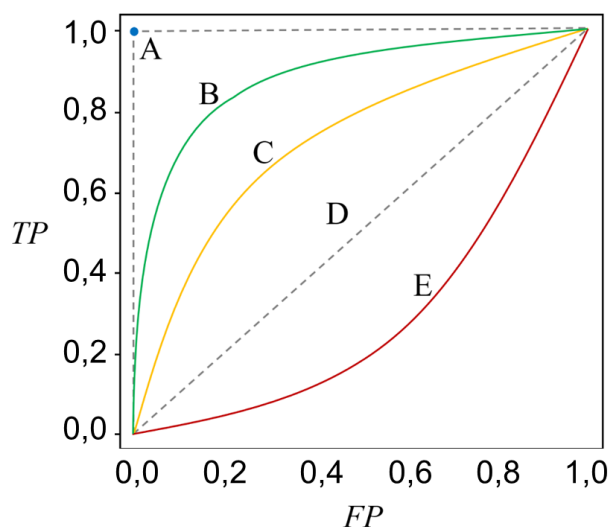
Velmi jednoduchý způsob uvádění výkonnosti kvalitativní metody analýzy, která se významně neliší v rámci analytického rozsahu, je pomocí kontingenční tabulky. Tabulka 3 představuje příklad takové tabulky. V tomto příkladu  $TP$ ,  $FP$ ,  $TN$  a  $FN$  jsou 97,8 % (228/233), 0,33 % (1/301), 99,7 % (300/301) 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<sup>-1</sup>) 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  $3 \times 100$ ) 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 matic, 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 spojitě proměnné

Mnoho konfirmačních nebo detekčních zkoušek ukazuje silnou závislost na pravděpodobnosti detekce nebo poměru falešné odpovědi na nějaké spojitě 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 spojitě 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),  $x$  spojitá 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  $\Pi$  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 matic 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 diethylstilboestrolu<sup>5</sup> (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

---

<sup>5</sup> 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) <sup>a</sup>
	239	50 – 70	
	145	45 – 65	

*RA* Relativní zastoupení (procento nejzastoupenějšího iontu, základní pík)

<sup>a</sup> 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)  
<sup>a</sup> 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  $10^6$ ). Předpokládáme, 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

Předpokládáme, 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. Předpokládáme, 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 <sup>a</sup>	$TP$ (%) <sup>b</sup>	$FP$ (%) <sup>c</sup>	LR TP / FP
DES	Masné výrobky	270 000	1 v $N$	~100	$3.7 \times 10^{-4}$	$2.7 \times 10^5$
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).

<sup>a</sup> Počet shod s ohledem na definované tolerance pro  $RA$  specifických iontů.

<sup>b</sup> Optimistický odhad u  $TP$  (cca 100 %).

<sup>c</sup> 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]  $\text{cm}^{-1}$

**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]  $\text{cm}^{-1}$ . Bylo zjištěno, že průměrný počet pásů na spektrum v intervalu [500, 1800]  $\text{cm}^{-1}$ ,  $M$ , byl 16. Dostupné spektrální rozlišení bylo 4  $\text{cm}^{-1}$  a to implikovalo existenci  $1300/4 = 325$ ,  $N$ , diskretních poloh píku v intervalu [500, 1800]  $\text{cm}^{-1}$ . 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 <sup>a</sup> možné shody	Předpovězené shody <sup>b</sup>	Pozorované shody	LR
1-Chlor-3-(1-naftyloxy)-2-propanol	23	$3,19 \times 10^{-4}$	19	192	311
Methylester $\alpha$ -kyanoskořicové kyseliny	17	$5,03 \times 10^{-5}$	3	29	2056
Fenyl v-triazolo-[1,5- $\alpha$ ]-pyridin-3-yl keton	24	$4,19 \times 10^{-4}$	25	190	314
Benzo- $\beta$ -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
$p$ -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

<sup>a</sup> 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]).

<sup>b</sup> 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.

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

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)

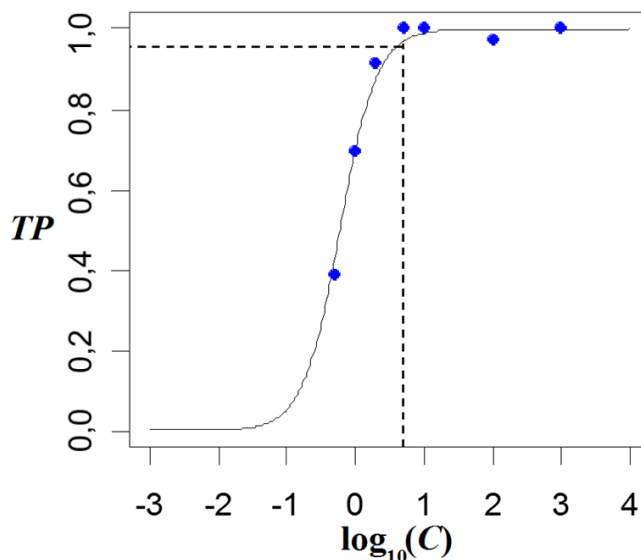
**Typ klasifikačního kritéria:** Neuvedeno

**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 <sup>-1</sup> )	Retenční čas		Zastoupení				$\rho$
		$\bar{t}_{Ri}$ (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 <sup>-1</sup> )	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<sup>-1</sup>);  $\bar{t}_{Ri}$  – 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;  $\rho$  – Spearmanův korelační koeficient.

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

Analyt	Extrakt	Interval hmotnostních zlomků, $w$ (mg kg <sup>-1</sup> )	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  $s_{Ai}$  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<sup>-1</sup> nebo 0,38 mg kg<sup>-1</sup> pro methyl chlorpyrifos, respektive methyl malasparyifos 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 <sup>-1</sup> )	$t_R$	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.

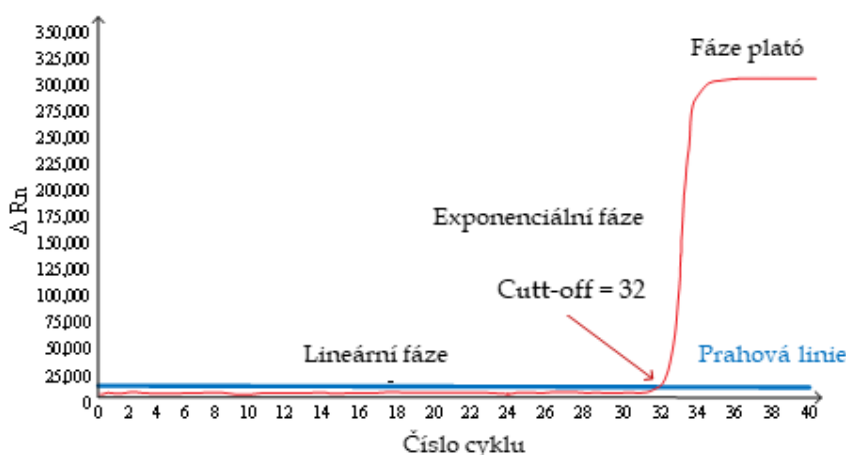
**Technika/přístroj:** Polymerázová řetězová reakce s reverzní transkripcí (RT-PCR)

**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.  $\Delta Rn$  (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“<sup>6</sup> (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:

<sup>6</sup> 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 specifikacním 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.
- 

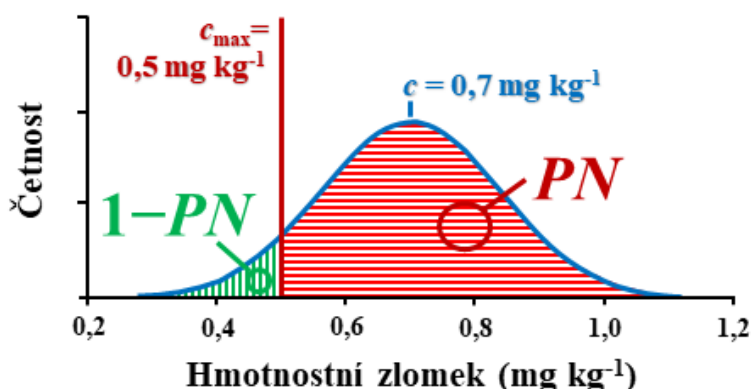
Použití rozhodovacích pravidel a nejistoty měření při posuzování shody je podrobně diskutováno v příručce Eurachem/CITAC „Použití informací o nejistotě při posuzování shody“ [29] („Příručka shody“). Pro úplnost však tato příloha pojednává o tom, jak lze informace o nejistotě nebo výkonnosti pro kvantitativní 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 specifikacní 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 volnosti<sup>7</sup>. 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

<sup>7</sup> 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.	<b>Vyhovující</b> (pozitivní) $c \leq c_{\max}$		$PP$	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE})$
2	Max.	<b>Nevyhovující</b> (negativní) (%) $c$		$PN$	$\text{NORM.DIST}(c_{\max}, c, u(c), \text{TRUE})$
3	Min.	<b>Vyhovující</b> (pozitivní) $c \geq c_{\min}$		$PP$	$\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$
4	Min.	<b>Nevyhovující</b> (negativní) (%) $c$		$PN$	$\text{NORM.DIST}(c_{\min}, c, u(c), \text{TRUE})$
5	Inter val	<b>Vyhovující</b> (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	<b>Nevyhovující</b> (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.
- [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. Encarnação, "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.



Záměrně ponecháno prázdné.

ISBN 978-80-86322-16-2