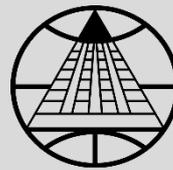


---

**Eurachem** 



**CITAC**  
Cooperation on International  
Traceability in Analytical Chemistry

---

## EURACHEM / CITAC Guide

# Bewertung der Leistung und Unsicherheit in der qualitativen chemischen Analyse



*Ich persönlich habe etwas gesehen, das wie ein Tier aussah, aber ich kann nicht mit absoluter Sicherheit sagen, dass es sich nicht um ein Mineral handelte. Ich denke, dass es sich eher um Energie als um Materie handelte. Relativ gesehen wäre es am einfachsten, das Ganze als ein Phänomen zu beschreiben, das irgendwo im Grenzbereich von Dimensionen und Bezeichnungen schwebt, im Grenzbereich von Farbe, Form, Geruch, Masse, Länge und Breite, Konturen, Schatten, Dunkelheit und so weiter und so fort.*

Slawomir Mrozek, "Streap-Tease"  
Übersetzung ins Englische von Edward Rotherth





## Bewertung der Leistung und Unsicherheit in der qualitativen chemischen Analyse

1. Auflage (2021)

### Herausgeber

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

### Übersetzung ins Deutsche

Die Übersetzung erfolgte durch Dr. Jan Sunderkötter, BAM Berlin unter Mitwirkung weiterer Kollegen von EuroLab-D.

Im Falle von Differenzen zwischen dieser Übersetzung und dem englischen Original ist die englische Fassung maßgeblich.

### Danksagung

Dieses Dokument wurde von einer gemeinsamen Eurachem/CITAC-Arbeitsgruppe erstellt, die sich wie folgt zusammensetzt (siehe rechts). Die Herausgeber sind all diesen Personen und Organisationen sowie anderen, die Kommentare, Ratschläge und Unterstützung beigetragen haben, zu Dank verpflichtet.

Die Erstellung dieses Handbuchs wurde teilweise von der Stiftung für Wissenschaft und Forschung, Portugal (Fundação para a Ciência e a Tecnologia) unterstützt.

### Zusammensetzung der Arbeitsgruppe \*

#### Eurachem

R. Bettencourt da Silva (chair)	<i>Univ. Lisboa, Portugal</i>
S. Ellison (secretary)	<i>LGC, Teddington, UK</i>
A. Togola	<i>BRGM, France</i>
D. Ivanova	<i>Eurachem Bulgaria</i>
E. Theodorsson	<i>LIU, Sweden</i>
E. Totu	<i>RUniversity Politehnica of Bucharest, Romania</i>
H. Emons	<i>European Commission, European Union</i>
I. Leito	<i>Univ. Tartu, Estonia</i>
M. Segal	<i>INRIM, Italy</i>
O. Levgar	<i>Ukrmetteststandart, Ukraine</i>
O. Pellegrino	<i>IPQ/DMET, Portugal</i>
P. Pereira	<i>IPST, Portugal</i>
R. Kaus	<i>Eurachem Germany</i>
S. Lardy-Fontan	<i>LNE, France</i>
W. Wegscheider	<i>Montanuniversität Leoben, Austria</i>

#### CITAC

A. Both	<i>NMISA, South Africa</i>
F. Lourenço	<i>Univ. São Paulo, Brazil</i>

*\*Zur Zeit der Genehmigung des Dokuments*

Diese Veröffentlichung sollte wie folgt zitiert werden\*:

"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. Available from <https://www.eurachem.org>"

*\*Je nach Anforderungen der entsprechenden Zeitschrift*

Bewertung der Leistung und Unsicherheit in der  
qualitativen chemischen Analyse

Deutsche Übersetzung (2023) der englischen Ausgabe

Erste Ausgabe (2021)

PDF: AQA\_2021\_EN\_v01a.pdf

Printed 2021-11-17

Copyright © 2021

Das Urheberrecht an diesem Dokument liegt bei den  
beteiligten Autoren. Alle Anfragen bezüglich der  
Vervielfältigung in jeglichem Medium, einschließlich  
Übersetzungen, sollten an das Eurachem-Sekretariat  
gerichtet werden.



## Inhaltsverzeichnis

<b>Vorwort</b> .....	1
<b>Anwendungsbereich</b> .....	2
<b>Abkürzungen und Symbole</b> .....	3
<b>1. Einführung</b> .....	4
<b>2. Arten der qualitativen Analysen</b> .....	5
<b>3. Leistungsbewertung qualitativer Analysen</b> .....	6
3.1. Allgemeine Betrachtungen .....	6
3.2. Quantifizierung der Leistung qualitativer Analysen .....	7
3.3. Bewertung der Falsch-positiv- und Falsch-negativ-Raten.....	10
3.4. Nachweisgrenze und Selektivität.....	15
<b>4. Mögliche Beschreibungen des Vertrauens in die qualitative Analyse</b> .....	17
4.1. Allgemeine Erwägungen.....	17
4.2. Likelihood Ratio .....	17
<b>4.3. A-posteriori-Wahrscheinlichkeit</b> .....	20
4.4. Zuverlässigkeit der Kennzahlen .....	21
4.5. Unsicherheit von Raten .....	21
<b>5. Berichterstattung qualitativer Analysenergebnisse</b> .....	23
<b>6. Schlussfolgerungen und Empfehlungen</b> .....	24
<b>7. Beispiele</b> .....	25
7.1. E1: Identifizierung von Verbindungen mit Hilfe der niedrig auflösenden Massenspektrometrie durch Suche in Datenbanken oder über das Vorhandensein charakteristischer Ionen.....	25
7.2. E2: Identifizierung von gereinigten Verbindungen durch Infrarotspektrometrie.....	30
7.3. E3: Identifizierung von Drogenmissbrauch im Urin mit Hilfe der Enzyme Multiplied Immunoassay Technique (EMIT) und einer alternativen Technik .....	33
7.4. E4: Identifizierung des menschlichen SRY-Gens in biologischem Material mittels qPCR ....	36
7.5. E5: Identifizierung von Pestizidrückständen in Lebensmitteln mittels GC-MS/MS auf der Grundlage von Retentionszeit und Ionenhäufigkeitsverhältnis .....	38
7.6. E6: Identifizierung von SARS-CoV-2-RNA durch Nukleinsäure-Amplifikationstest.....	43
<b>Anhang A - Satz von Bayes, Quoten und Likelihood Ratio</b> .....	47
A.1 Satz von Bayes.....	47
A.2 Wahrscheinlichkeit und Quoten.....	47
A.3 Die Quotenform des Satzes von Bayes und die Likelihood Ratio.....	48
<b>Anhang B - Qualitative Analyse im Zusammenhang mit der Bewertung der Konformität mit einem quantitativen Grenzwert</b> .....	50
<b>Literaturverzeichnis</b> .....	55





## **Vorwort**

Das Problem der Bewertung und Angabe der Unsicherheit bei qualitativen chemischen Analysen wird in der Literatur weit weniger behandelt als die Unsicherheit bei quantitativen Analysen (d. h. Messungen) [1]. Zwar haben sich einige Autoren mit dieser Thematik befasst [2] - [11], doch gibt es kaum allgemeine Anleitungen für die Bewertung der Leistung von qualitativen Analysen oder für die Bewertung und Angabe der Unsicherheit bei qualitativen Analysen.

Von akkreditierten Laboratorien wird derzeit nicht erwartet, dass sie Unsicherheiten im Zusammenhang mit qualitativen Analysenergebnissen bewerten oder berichten [12]. Gemäß ISO/IEC 17025 [13] und ISO 15189 [14] müssen Laboratorien jedoch sicherstellen, dass sie valide qualitative und quantitative Analysenergebnisse erzielen können. Für Laboratorien ist es außerdem von entscheidender Bedeutung, die Zuverlässigkeit qualitativer Analysenergebnisse zu kennen; so können sie gegebenenfalls vor Einschränkungen bei der Interpretation der Ergebnisse warnen und auf Kundenanfragen zur Zuverlässigkeit genau eingehen. Eine quantitative Bewertung der Zuverlässigkeit eines qualitativen Analysenergebnisses ist besonders dann nützlich, wenn falsche Ergebnisse wahrscheinlicher sind. Dieser Leitfaden sollte verwendet werden, wenn eine quantitative Bewertung der Zuverlässigkeit eines qualitativen Analysenergebnisses erstrebenswert ist.

Dieser Leitfaden stützt sich auf Erfahrungen aus verschiedenen analytischen Bereichen, in denen die qualitative Analyse häufig eingesetzt wird, z. B. in der Forensik [15] und im klinischen Bereich [16] - [18], sowie auf umfassenden allgemeinen Leitlinien [7].

## Anwendungsbereich

Dieser Leitfaden soll Laboratorien bei der Festlegung und Umsetzung geeigneter Methoden zur Bewertung der Leistungsfähigkeit qualitativer Analysenverfahren und zur Bewertung von Unsicherheiten in der qualitativen chemischen Analyse unterstützen. In diesem Leitfaden wird die qualitative Analyse als "Klassifizierung nach festgelegten Kriterien" definiert. Für die analytische Chemie und verwandte Disziplinen beziehen sich die "Kriterien" im Allgemeinen auf Informationen zur Bestimmung der chemischen Zusammensetzung, der Eigenschaften und/oder der Struktur der analysierten Gegenstände.

Die folgenden Arten von Kriterien werden in diesem Leitfaden berücksichtigt:

- Quantitative Kriterien, bei denen ein numerisches Ergebnis verwendet wird, um einen Prüfgegenstand einer vorher festgelegten Klasse zuzuordnen;
- Qualitative Kriterien wie das Vorhandensein oder Fehlen eines bestimmten Merkmals, eine Farbänderung bei einem Test usw.

Dieser Leitfaden erhebt keinen Anspruch auf Vollständigkeit bei der Beschreibung der verfügbaren Instrumente für die Leistungsbewertung qualitativer Analysenverfahren und der Unsicherheit qualitativer Analyseergebnisse. Die in diesem Leitfaden dargestellten Leistungsmerkmale beruhen auf gemessenen oder geschätzten Falschergebnisraten und berücksichtigen beispielsweise keine Maße für die Übereinstimmung zwischen qualitativen Verfahren oder die Behandlung von Klassifizierungen auf Ordinalskalen<sup>1</sup>, außer als richtige oder falsche Klassifizierung.

---

<sup>1</sup> Eine Ordinalskala ist eine Skala mit natürlich geordneten Kategorien, bei der der Abstand zwischen den Kategorien nicht bekannt ist. Die Mohs-Skala ist eine Ordinalskala für die Mineralhärte.

## Abkürzungen und Symbole

Die folgenden Abkürzungen und Symbole kommen in diesem Leitfaden vor. Die in diesem Dokument verwendeten Symbole sind nicht in allen Bereichen der Wissenschaft, in denen sie Anwendung finden, harmonisiert. Im medizinischen Laboratorium sind beispielsweise FN und FNR die Abkürzungen für "Anzahl der falsch-negativen Ergebnisse" bzw. "Falsch-negativ-Rate".

$A$	Ionenhäufigkeit eines Massenspektrums	$O(\cdot)$	Quoten für ein Ereignis, z. B. $O(A)$ bezeichnet Quoten für ein Ereignis A
$\bar{A}$	Mittlere Ionenhäufigkeit eines Massenspektrums	$p$	Anzahl positiver Ergebnisse
$AR$	Häufigkeitsverhältnis der Ionen im Massenspektrum	$P(\cdot)$	Wahrscheinlichkeit eines Ereignisses; z. B. $P(A)$ ist die Wahrscheinlichkeit des Ereignisses A
$c$	Gemessene Konzentration (oder eine andere Größe) des analysierten Gegenstands	$P(+)$	A-priori-Wahrscheinlichkeit eines positiven Falls
CI	Konfidenzintervall	$P(-)$	A-priori-Wahrscheinlichkeit eines negativen Falls
$c_{\max}$	Zulässige Höchstkonzentration	$pc$	Anzahl positiver Fälle
$c_{\min}$	Zulässige Mindestkonzentration	$PN$	A-Posteriori-Wahrscheinlichkeit eines negativen Falls (siehe Anhang A)
$DOR$	Diagnostische Odds Ratio	$PP$	A-Posteriori-Wahrscheinlichkeit eines positiven Falls (siehe Anhang A)
$E$	Effizienz	$PPV$	Positiver Vorhersagewert
$fn$	Anzahl falsch negativer Ergebnisse	qPCR	Quantitative Polymerase-Kettenreaktion
$FN$	Falsch-negativ-Rate bezogen auf positive Fälle	RT-PCR	Polymerase-Kettenreaktion mit reverser Transkription
$fp$	Anzahl falsch positiver Ergebnisse	$RA$	Relative Häufigkeit
$FP$	Falsch-positiv-Rate bezogen auf negative Fälle	Rn	Referenzierter Marker
GC-MS	Gaschromatographie-Massenspektrometrie	$RR$	Ergebnisrate
GC-MS/MS GUM	Gaschromatographie-Tandem-Massenspektrometrie	$SA$	Standardabweichung der Ionenhäufigkeit
GUM	Guide to the Expression of Uncertainty in Measurement (Leitfaden zur Angabe der Messunsicherheit)	$SP$	Spezifität
$HL_{RR,95}$	Obere Grenze des 95 %-Konfidenzintervalls für die Ergebnisrate $RR$ (z. B. $SS$ )	$SS$	Sensitivität
LC-MS	Flüssigchromatographie-Massenspektrometrie	$s_{Ri}$	Standardabweichung der Retentionszeit
$LL_{RR,95}$	Untere Grenze des 95 %-Konfidenzintervalls für die Ergebnisrate $RR$ (z. B. $SS$ )	$tn$	Anzahl wahrer negativer Ergebnisse
$LL_{RR,95}^{tg}$	Ziel- oder Mindestwert für $LL_{RR,95}$	$TN$	Wahr-Negativ-Rate bezogen auf negative Fälle
LOD	Nachweisgrenze	$tp$	Anzahl wahrer positiver Ergebnisse
LOQ	Bestimmungsgrenze	$TP$	Wahr-Positiv-Rate bezogen auf die positiven Fälle
$LR$	Likelihood Ratio	$t_R$	Retentionszeit
$LR(+)$	Likelihood Ratio positiver Ergebnisse	$\bar{t}_{Ri}$	Mittlere Retentionszeit
$LR(-)$	Likelihood Ratio negativer Ergebnisse	$u(c)$	Standardunsicherheit von $c$
$n$	Anzahl der negativen Ergebnisse	$w$	Masseanteil
$nc$	Anzahl negativer Fälle	$Y$	Youden-Index
$NPV$	Negativer Vorhersagewert	$\Delta Rn$	Wert des referenzierten Markers abzüglich des Wertes der Basislinie
		$\rho$	Spearman's Korrelationskoeffizient zwischen Paaren von Ionenhäufigkeiten

## 1. Einführung

Viele relevante sozioökonomische oder individuelle Interessen, wie industrielle Produktivität und Gesundheitszustand, hängen von chemischen Analysen ab. Einige dieser Analysen sind ausschließlich qualitativ oder beinhalten eine nachträgliche Quantifizierung der identifizierten chemischen Substanz. Die Interessen, die durch diese Analysen geschützt werden sollen, sind nur dann gewahrt, wenn die analytische Qualität für die beabsichtigte Verwendung geeignet ist.

In einigen Veröffentlichungen wird für die "qualitative Analyse" der Begriff "Untersuchung" [1], "Untersuchung einer nominellen Eigenschaft" [19] oder "Test" und "Prüfung" verwendet. In der internationalen Norm für die Akkreditierung medizinischer Laboratorien wird der Begriff "Untersuchung" sowohl für quantitative als auch für qualitative Analysen verwendet [14]. Da zwischen den verschiedenen relevanten internationalen Gemeinschaften kein Konsens über diese Begriffe erzielt wurde, wird in diesem Leitfaden der Begriff "qualitative Analyse" für die Bestimmung nominaler (qualitativer) Eigenschaften in der chemischen Analyse verwendet.

Vereinfacht gesagt ist ein qualitatives Analyseergebnis eine einfache Aussage oder Kategorisierung eines Prüfgegenstands oder Materials, d. h. eine Klassifizierung. Entscheidungen werden ausnahmslos auf Grundlage dieser Kategorisierung getroffen, z. B., ob eine Düngemittelcharge freigegeben wird, ob Wasser trinkbar ist, ob eine Person im Besitz einer geregelten Substanz ist oder nicht oder ob ein neu synthetisiertes Material die gemäß den Anforderungen richtige Struktur hat. Falsche Klassifizierungen - wie z. B. die "Annahme" eines Produkts, obwohl es für die Verwendung ungeeignet ist - bergen Risiken für alle Beteiligten. Um diese Risiken zu beherrschen, arbeiten die an der Analyse beteiligten Fachleute hart daran, dass ihre Verfahren zu einem akzeptabel niedrigen Risiko einer falschen Klassifizierung führen.

Daraus folgt, dass zu irgendeinem Zeitpunkt in der Entwicklung eines solchen Prüfverfahrens eine Bewertung des Risikos einer falschen Klassifizierung vorgenommen werden muss. Daher ist es vernünftig, von einem Laboratorium zu erwarten, dass es für die meisten derartigen Verfahren Informationen über die Risiken falscher Ergebnisse erstellt oder Zugang dazu hat. Eine wichtige Ausnahme ist die Verwendung standardisierter Prüfverfahren, die von Gruppen außerhalb des Laboratoriums als für den vorgesehenen Zweck geeignet festgelegt wurden [20]–[21][22] [23]. Es kann durchaus sein, dass das Laboratorium nur begrenzten oder gar keinen Zugang zu Leistungsdaten solcher Prüfverfahren hat. Allerdings spezifizieren diese Verfahren stets die Prüfung mit relevanten Details, und vom Labor wird im Allgemeinen erwartet, dass relevante Faktoren, die seiner Kontrolle unterliegen, tatsächlich den Anforderungen des Prüfverfahrens entsprechen. Dies wiederum kann bedeuten, dass nachgewiesen werden muss, dass die Unsicherheit der kontrollierten Parameter und die Prüfleistung im Verhältnis zum Zweck der Prüfung angemessen ist.

Die Bewertung von Unsicherheiten im Zusammenhang mit quantitativen Parametern oder Analyseergebnissen ist seit der Veröffentlichung des "Guide to the Expression of Uncertainty in Measurement" (GUM), der als ISO Guide 98 [24] sowie als JCGM-Dokument [25] vorliegt, Gegenstand erheblicher Anstrengungen. Andererseits wurde den Unsicherheiten bei der qualitativen Analyse weit weniger Aufmerksamkeit geschenkt. Seit Veröffentlichung der ersten Ausgabe der ISO/IEC 17025 [26] hat das Interesse an Unsicherheiten bei qualitativen Analysen zugenommen. Die Herausforderungen bei der Bestimmung der mit qualitativer Analyse verbundenen Unsicherheit, wie z. B. „bestanden/nicht bestanden“, Identitäts- oder vergleichende Identitätsanalysen, haben dementsprechend mehr Aufmerksamkeit erhalten, insbesondere in Bereichen, in denen die Auswirkungen falscher qualitativer Analyseergebnisse äußerst relevant sind, z. B. in der Forensik oder bei Dopinganalysen.

Es gibt eine Vielzahl von Kennzahlen zur Angabe der Unsicherheit bei qualitativen Ergebnissen [7]. Es besteht jedoch nur ein begrenzter Konsens darüber, welche Kennzahlen zu verwenden sind. Ausnahmen bilden die Bereiche Epidemiologie und klinisches Labor, wo die Begriffe "klinische Sensitivität" und "klinische Spezifität" durchweg als Parameter für klinische Genauigkeit verwendet werden [27].

Quantitative und qualitative Analysen unterscheiden sich wesentlich darin, wie die Ergebnisse und die damit verbundenen Unsicherheiten angegeben werden. Während quantitative Ergebnisse als ein Intervall angegeben werden, das den "wahren Wert" der Messgröße mit einem definierten Konfidenzniveau enthält, werden nominale Eigenschaften als eine Klassifizierung mit Kennzahlen angegeben, die die Möglichkeit einer korrekten oder falschen Klassifizierung ausdrücken. Diese "Möglichkeit" kann durch eine Wahrscheinlichkeit, Likelihood<sup>2</sup>, Quoten oder andere Kennzahlen beschrieben werden, die aus der Interpretation der Eingabedaten geschätzt werden. Die Qualität der berichteten Kennzahlen hängt von der Anzahl und Vielfalt der untersuchten Fälle ab. Die Bestimmung dieser Kennzahlen ermöglicht die Ermittlung von Fällen, in denen Verfahren verbessert werden sollten, um die Likelihood der Erzeugung falscher Ergebnisse zu verringern.

Dieser Leitfaden beschreibt die allgemeinen Grundsätze für die Leistungsbewertung qualitativer Analysen zur Angabe der Unsicherheit qualitativer Analyseergebnisse und stellt Anwendungsbeispiele der beschriebenen Theorie vor. Der Leitfaden befasst sich nicht mit der Fähigkeit des Prüfgegenstandes, eine Gruppe identischer Gegenstände oder ein größeres Objekt zu repräsentieren, d. h. er erörtert nicht die Auswirkungen der Probenahme bei diesen Bewertungen.

Ordinale Ergebnisse können auf binäre (Ja/Nein) Ergebnisse reduziert und unter Verwendung der Methoden in diesem Leitfaden behandelt werden, indem ordinale Klassifizierungsergebnisse als „korrekt“ oder „inkorrekt“ bezeichnet werden. Andere Methoden zur Behandlung ordinaler Skalen liegen außerhalb des Anwendungsbereichs dieses Leitfadens.

## 2. Arten der qualitativen Analysen

Wie im Anwendungsbereich dieses Leitfadens erwähnt, wird die Qualitative Analyse<sup>3</sup> als "Klassifizierung nach bestimmten Kriterien" definiert [28]. In Tabelle 1 sind einige Beispiele aufgeführt. Obwohl alle diese Fälle sehr unterschiedlich erscheinen, haben sie ein gemeinsames Merkmal: Sobald die Kriterien festgelegt sind, lässt sich die Leistung der Klassifizierungsmethode relativ einfach in Form ihrer Erfolgs- oder Misserfolgsquote beschreiben. Diese Erfolgs- und Misserfolgsquoten bilden die Grundlage der meisten Leistungskennzahlen für qualitative Analysen.

Die im Haupttext behandelten qualitativen Analysen werden in zwei Kategorien unterteilt, die auf verschiedenen Arten von Klassifizierungskriterien beruhen, qualitativ oder quantitativ. Tabelle 1 enthält Beispiele für jede Kategorie. In Abschnitt 3 werden Strategien zur Leistungsbewertung für verschiedene Arten von Klassifizierungskriterien beschrieben. Bei qualitativen Analysen, bei denen die Wahr- oder Falsch-Antwort-Rate von einer quantitativen Eigenschaft abhängt, wie z. B. dem Vorhandensein einer verbotenen Substanz, deren Nachweis von der vorhandenen Menge abhängt, wird auch eine Nachweisgrenze berücksichtigt (siehe Abschnitt 3.4).

Die Konformitätsbewertung des Wertes einer quantitativen Eigenschaft eines Gegenstandes mit einem Grenzwert oder Intervall kann manchmal als eine Umwandlung eines Messergebnisses in ein qualitatives Ergebnis ("konform" oder "nicht konform") angesehen werden. Die Verwendung von Messwerten und deren Messunsicherheiten für die Konformitätsbewertung wird in einem anderen Eurachem/CITAC-Leitfaden [29] ausführlich behandelt und wird daher im vorliegenden Leitfaden nicht im Detail betrachtet. In Anhang B wird jedoch erörtert, wie einige der Kennzahlen, die zur Bewertung der Leistung oder Unsicherheit qualitativer Analysen verwendet werden, für die quantitative Konformitätsbewertung bestimmt werden können.

---

<sup>2</sup> Anmerkung d. Übers.: Der Begriff Likelihood wurde in diesem Dokument nicht übersetzt. Die Likelihood unterscheidet sich von der Wahrscheinlichkeit dadurch, dass die Wahrscheinlichkeit das mögliche Eintreten von bestimmten Beobachtungen beschreibt, die Likelihood dagegen beschreibt, wie gut ein Modell die beobachteten Daten erklärt.

<sup>3</sup> Qualitatives Prüfen" ist im Prinzip ein breiteres Feld als "qualitative Analyse", einfach weil die chemische Analyse, die häufig als analytische Arbeit bezeichnet wird, eine besondere Tätigkeit unter vielen Bereichen des Prüfens ist. In diesem Leitfaden für analytische Chemiker und Angehörige verwandter Disziplinen werden die Begriffe jedoch synonym verwendet.

**Tabelle 1.** Arten der qualitativen Analyse auf der Grundlage verschiedener Klassifizierungskriterien.

Kriterium für die Klassifizierung	Beispiel einer qualitativen Analyse
Qualitativ	<ol style="list-style-type: none"> <li>(1) Nachweis von aliphatischen Aldehyden in einer Lösung durch Farbumschlag nach Zugabe von Schiffschem Reagenz</li> <li>(2) Identifizierung der kristallinen Form eines Materials durch Beobachtung</li> <li>(3) Identifizierung der Marke und des Jahrgangs von Wein durch sensorische Analyse</li> <li>(4) Identifizierung einer biologischen Art durch Bestimmung oder Nachweis einer bestimmten DNA-Sequenz</li> <li>(5) Identifizierung der menschlichen Blutgruppe durch Beobachtung der Agglutination</li> </ol>
Quantitativ	<ol style="list-style-type: none"> <li>(1) Identifizierung eines Pestizidrückstands in Obst anhand gemessener Fragmentmassen und relativer Fragmenthäufigkeiten in GC-MS</li> <li>(2) Bestimmung der infrarotspektralen Äquivalenz zwischen einem neuen und einem bereits anerkannten industriellen Rohstoff anhand von Wellenlängen- und Intensitätskriterien</li> <li>(3) Identifizierung eines Diuretikums im Urin eines Sportlers anhand der Retentionszeit und der gemessenen Fragmentmassen im GC-MS</li> <li>(4) Identifizierung einer Droge im Blut anhand der Retentionszeit und der gemessenen Fragmentmassen im LC-MS</li> <li>(5) Nachweis eines Virus in einer klinischen Probe anhand der Fluoreszenzintensität in der quantitativen Echtzeit-Polymerase-Kettenreaktion (qPCR)</li> </ol>

### 3. Leistungsbewertung qualitativer Analysen

#### 3.1. Allgemeine Betrachtungen

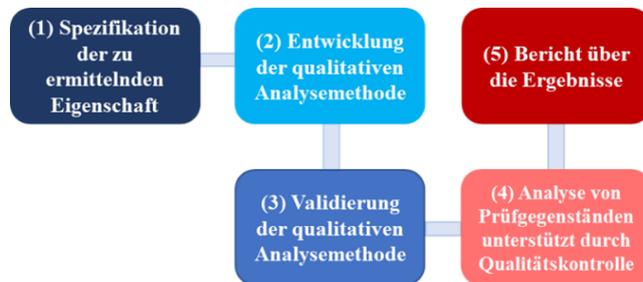
Dieser Abschnitt enthält Leitlinien für die Bewertung und Angabe der Leistung von Verfahren, die eine einfache Klassifizierung in zwei Klassen ("binäre Klassifizierung") ermöglichen sollen. Die "Klassen" werden hier als "positiv" oder "negativ" bezeichnet, um "der betreffenden Klasse zugehörig" oder "nicht zugehörig" zu kennzeichnen. Diese Klassifizierung deckt die meisten praktischen Situationen ab, einschließlich "oberhalb eines Grenzwertes", "akzeptabel", "inakzeptabel", "Identität als" oder "Vorhandensein einer bestimmten Art".

Es wird davon ausgegangen, dass die Klassen umfassend und exklusiv sind, um die Berechnung eindeutiger Falschantwortraten zu ermöglichen. Dies impliziert, dass kein Prüfgegenstand als einer dritten Klasse zugehörig klassifiziert werden darf. Dies kann im Allgemeinen durch eine sorgfältige Festlegung der Klassifizierungskriterien erreicht werden. Es ist jedoch möglich, dass ein Ergebnis keine ausreichende Sicherheit für die Klassifizierung bietet. Unter diesen Umständen ist es durchaus vertretbar, dass der Analytiker ein Prüfergebnis als "nicht eindeutig" im Sinne von unzureichend sicher bezeichnet. Nicht eindeutige Ergebnisse erfordern weitere Untersuchungen, damit die Ergebnisse als "eindeutig" bezeichnet werden können. Diese Ergebnisse werden in medizinischen Laboratorien als "Grauzone" oder "zweideutige Zone" bezeichnet.

Einige der beschriebenen Konzepte können im Prinzip auf mehrere Klassen ausgedehnt werden, z. B. bei einer Klassifizierung auf einer Ordinalskala, indem für alle Klassen richtige und falsche Klassifizierungsraten ermittelt werden. Eine sinnvolle Erweiterung besteht darin, die Identifizierung von Struktur oder Identität (formal ein Mehrklassenproblem) entweder als "richtig" oder "falsch" zu behandeln, und dieser Ansatz wird hier übernommen. Eine detaillierte Behandlung des Mehrklassenproblems, das mehrere gleichzeitige Zuordnungen oder Zuordnungen zu mehreren Klassen beinhalten kann, würde jedoch den Rahmen dieses Leitfadens sprengen.

Die qualitative Analyse umfasst verschiedene Phasen, nämlich (1) Problembeschreibung, (2) Verfahrensentwicklung und (3) Validierung, (4) Prüfung an unbekanntem Gegenständen, die durch

eine Qualitätskontrolle überprüft werden, und (5) Berichterstattung der Ergebnisse (Abbildung 1). Eine eindeutige Spezifikation der zu bestimmenden Eigenschaft und die Bewertung der Eignung der Analyse für den beabsichtigten Zweck sind von entscheidender Bedeutung. Der Bericht eines qualitativen Analysenergebnisses muss durch valide Verfahren und eine angemessene Qualitätskontrolle der Prüfung unterstützt werden. Wie die Ergebnisse berichtet werden, hängt vom Zweck der Analyse und dem Empfänger des Berichts ab. In diesem Leitfaden wird nicht detailliert beschrieben, wie das Verfahren zu entwickeln ist oder wie die Qualitätskontrolle gestaltet sein sollte.



**Abbildung 1.** Qualitativer Analyseprozess von der Problembeschreibung bis zur Berichterstattung der Ergebnisse.

## 3.2. Quantifizierung der Leistung qualitativer Analysen

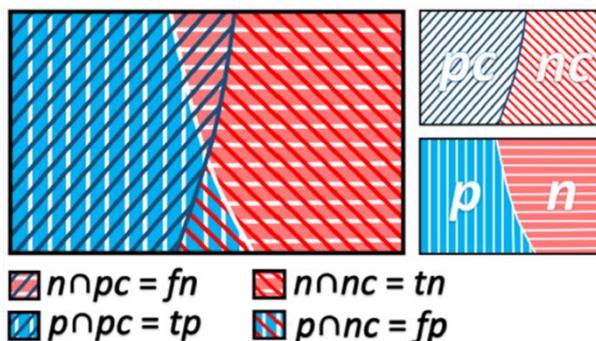
### 3.2.1. Festlegung der Grundlage für die Leistungsbewertung

Die grundlegendste Art, die Leistung eines qualitativen Analyseverfahrens zu quantifizieren, ist die Berechnung der Falsch-Ergebnis-Rate. Bei "positiven" oder "negativen" Ergebnissen ist es sinnvoll, "richtig positive" und "falsch positive" bzw. "richtig negative" und "falsch negative" Raten anzugeben. Diese Raten können jedoch entweder auf die Gesamtzahl einer bestimmten Art von Fällen oder Ergebnissen oder auf die Gesamtzahl der möglichen Fälle oder Ergebnisse bezogen werden.

Die Falsch-Positiv-Rate kann zum Beispiel definiert werden als:

- i. Der Anteil der negativen Fälle, die fälschlicherweise als positiv gemeldet werden ( $fp/nc$ ), wobei  $fp$  und  $nc$  die Anzahl der falsch positiven Ergebnisse bzw. der negativen Fälle sind. In Abbildung 2 wird die Überschneidung der verschiedenen Arten von Fällen und Ergebnissen grafisch dargestellt. Die  $(fp/nc)$  wird durch das Verhältnis der Flächen der Schnittmenge ( $\cap$ ) der positiven Ergebnisse " $p$ " mit " $nc$ " ( $p \cap nc = fp$ ) und der Fläche von " $nc$ " dargestellt. Die  $FP$  der Tabelle 2 zeigt diese Berechnung.
- ii. Der Anteil der positiven Ergebnisse, die fälschlicherweise als positiv gemeldet werden ( $fp/p$ ), wobei  $p$  die Anzahl der positiven Ergebnisse ist. In Abbildung 2 wird diese Rate durch das Verhältnis zwischen den Flächen ( $p \cap nc = fp$ ) und " $p$ " dargestellt.
- iii. Der Anteil an der Gesamtzahl der Fälle oder Ergebnisse, die fälschlicherweise als positiv gemeldet werden ( $fp/(pc + nc) = fp/(p + n)$ ), wobei  $pc$  und  $n$  die Anzahl der positiven Fälle bzw. negativen Ergebnisse darstellen. In Abbildung 2 wird diese Rate durch das Verhältnis zwischen der als ( $p \cap nc = fp$ ) gekennzeichneten Fläche und der Gesamtfläche der Abbildung dargestellt.

Abbildung 2. Grafische Darstellung eines Beispiels für die Überschneidung der Anzahl der positiven,  $pc$ , oder negativen,  $nc$ , Fälle mit der Anzahl der positiven,  $p$ , oder negativen,  $n$ , Ergebnisse. Das Symbol " $\cap$ " steht für die Schnittmenge von Gruppen; zum Beispiel bezeichnet  $n \cap pc$  hier die Menge der negativen Ergebnisse von positiven Fällen. Die Symbole " $n \cap pc$ ", " $p \cap pc$ ", " $n \cap nc$ " und " $p \cap nc$ " definieren  $fn$ ,  $tp$ ,  $tn$  bzw.  $fp$ .



Der Unterschied zwischen diesen Definitionen ist entscheidend. Im Fall i) variiert die Rate nicht mit dem Anteil der "negativen Fälle" in der Population, d. h.  $nc/(nc+pc)$ , da  $FP = fp/nc$ . Für die Fälle ii) und iii) hängt die Falsch-Positiv-Rate jedoch von  $nc/(nc+pc)$  ab, da mehr  $fp$  in Populationen mit mehr  $nc$  beobachtet werden. Daher charakterisieren diese Definitionen die Leistung der qualitativen Analyse auf unterschiedliche Weise und führen zu unterschiedlichen Interpretationen ihrer Werte.

Die Wahr-Positiv-,  $TP$ , ( $tp/pc$ ) und die Wahr-Negativ-,  $TN$ , ( $tn/nc$ ) Raten, bezogen auf eine relevante Anzahl von Fällen, sind in der klinischen Chemie als "Sensitivität" bzw. "Spezifität" bekannt [7] (Tabelle 2). Die Bestimmung der klinischen Sensitivität und Spezifität erfordert die ordnungsgemäße Bestimmung der untersuchten Fälle durch eine eindeutige klinische Diagnose. In der quantitativen Analyse hat der Begriff "Sensitivität" [1] oder "analytische Sensitivität" [30] eine andere Bedeutung.<sup>4</sup>

**Tabelle 2.** Alternative Leistungsmerkmale zur Angabe der Qualität qualitativer Analyseergebnisse.

Leistungscharakteristika	Terminus
Wahr-positiv-Rate, $TP$ (Sensitivität, $SS$ )	$tp/pc = tp/(tp+fn) = 1 - FN$
Falsch-positiv-Rate, $FP$	$fp/nc = fp/(tn+fp) = 1 - TN$
Wahr-negativ-Rate, $TN$ (Spezifität, $SP$ )	$tn/nc = tn/(tn+fp) = 1 - FP$
Falsch-negativ-Rate, $FN$	$fn/pc = fn/(tp+fn) = 1 - TP$
Präzision" oder "positiver Vorhersagewert", $PPV$	$TP/p = tp/(tp+fp)$
"Negativer Vorhersagewert", $NPV$	$tn/n = tn/(tn+fn)$
Effizienz, $E$	$(tp+tn)/(p+n)$
Youden-Index, $Y$	$SS(\%)+SP(\%)-100$
Likelihood Ratio positiver Ergebnisse, $LR (+)$	$TP/FP$
Likelihood Ratio negativer Ergebnisse, $LR (-)$	$TN/FN$
A-posteriori-Wahrscheinlichkeit	Siehe Anhang A

$tp$  - Anzahl der wahr positiven Ergebnisse;  $fp$  - Anzahl der falsch positiven Ergebnisse;  $tn$  - Anzahl der wahr negativen Ergebnisse;  $fn$  - Anzahl der falsch negativen Ergebnisse;  $p$  - Anzahl der positiven Ergebnisse ( $tp + fp$ );  $n$  - Anzahl der negativen Ergebnisse ( $tn + fn$ );  $pc$  - Anzahl der positiven Fälle und  $nc$  - Anzahl der negativen Fälle.

Die Wahr-Positiv-Rate bezogen auf die positiven Fälle ( $tp/p$ ) wird auch als "Präzision" oder "positiver Vorhersagewert" ( $PPV$ ) der qualitativen Analyse bezeichnet [30]. Der Begriff "negativer Vorhersagewert" ( $NPV$ ) wird für die Wahr-negativ-Rate bezogen auf die Gesamtzahl der negativen Ergebnisse (d. h.  $tn/n$ ) verwendet. Die Effizienz der qualitativen Analyse ist als der Anteil der richtigen Ergebnisse jeder Art an allen Ergebnissen (d. h.  $(tp + tn)/(p + n)$ ) definiert. Der Youden-

<sup>4</sup> Nach dem Internationalen Vokabular für Metrologie [1] ist die "Sensitivität eines Messsystems" der "Quotient aus der Änderung einer Anzeige eines Messsystems und der entsprechenden Änderung eines Wertes einer Messgröße".

Index ist eine alternative Methode zur Quantifizierung des Erfolgs qualitativer Analysen (Tabelle 2) [31].

Obwohl die Kennzahlen, die sich auf die Anzahl der positiven oder negativen Fälle beziehen, nicht von der Häufigkeit der Falltypen abhängen, können diese Zahlen allein nicht die Wahrscheinlichkeit angeben, dass ein bestimmtes Ergebnis korrekt ist. Zur Abschätzung der Wahrscheinlichkeit, dass ein Ergebnis korrekt ist, müssen auch eine relevante Ergebnisrate und die Häufigkeit der Fälle berücksichtigt werden. Dies und andere Kennzahlen für das Vertrauen in qualitative Ergebnisse werden in Abschnitt 4 erörtert.

### 3.2.2. Definition der Referenz für die Leistungsbewertung

Die zur Quantifizierung der Leistung qualitativer Analysen verwendeten Kennzahlen können zusätzliche Besonderheiten aufweisen. Die positiven und negativen Fälle können auf unterschiedliche Weise ermittelt werden. Einige Fälle oder Proben, die als Referenz verwendet werden, können aufgrund ihrer Herkunft oder ihrer Rezeptur als "positiv" für ein Merkmal bekannt sein. Bei anderen Fällen kann es sich, wie von AOAC International definiert, um Fälle handeln, in denen die Ergebnisse eines Bestätigungsverfahrens und eines anderen Analysenverfahrens beide positiv sind" [32]. Beispiele für eine geeignete Herkunft positiver Fälle können Patienten sein, bei denen eine bestimmte Krankheit diagnostiziert wurde, oder Böden, die bekanntermaßen kontaminiert sind. Positive Prüfgegenstände können hergestellt werden, indem die zu identifizierende Spezies in einer Matrix hinzugefügt wird, die den analysierten Gegenständen entspricht, z. B. ein Pestizid in einem Lebensmittel, in dem ggf. bereits ein Gehalt an Pestiziden gefunden wurde. Wenn die Identifizierungsleistung erheblich von einer quantitativen Eigenschaft abhängt (z. B. von der Konzentration des zu identifizierenden oder nachzuweisenden Stoffes), sollte die Zugabe die Bestimmung dieses Wertes ermöglichen. Ein negativer Fall kann auch als ein Fall interpretiert werden, von dem bekannt ist, dass er von seiner Herkunft, seiner Rezeptur her negativ ist, oder der als negativ definiert wird, weil ein "Bestätigungsverfahren und ein anderes Analysenverfahren beide negativ sind". Die Definitionen der AOAC International für Positiv- und Negativfälle haben eine umfassendere Anwendung, da es der einzige Ansatz ist, der für die Analyse komplexer Gegenstände anwendbar ist, die schwer reproduzierbar sind. Allerdings hängt er von der Qualität der Ergebnisse der verwendeten Analysentechniken ab. In einigen Bereichen ist es schwierig, Gegenstände mit dem untersuchten Analyten und möglichen Störsubstanzen für die Prüfung der Analysenleistung künstlich herzustellen, da die Matrizes der Gegenstände unbekannt und unvorhersehbar sind.

Die positiven und negativen Fälle können auch als Referenzdaten zur Verfügung gestellt werden, z. B. Spektren, die von einer bestimmten Verbindung stammen. Nach der Festlegung von Identifizierungskriterien kann die Wahrscheinlichkeit ermittelt werden, ob die Zusammensetzung im Vergleich zu diesen Kriterien richtig oder falsch ist. Bei der Massenspektrometrie kann die Identifizierung beispielsweise auf der Bewertung des Vorhandenseins oder des Vorhandenseins und der Häufigkeit charakteristischer Ionen beruhen. Die Wahrscheinlichkeit einer spektroskopischen Übereinstimmung kann durch binomische oder hypergeometrische Statistiken vorhergesagt werden, wie in den Beispielen E1 und E2 beschrieben.

### 3.2.3. Darstellung der Leistung eines Verfahrens

#### 3.2.3.1 Kontingenz-Tabellen

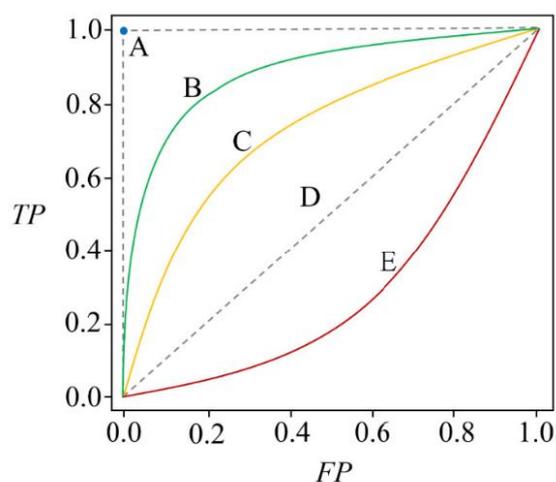
Eine sehr bequeme Art, die Leistung eines qualitativen Analysenverfahrens, das innerhalb des analytischen Anwendungsbereichs nicht signifikant variiert, darzustellen, ist eine Kontingenztafel. Tabelle 3 enthält ein Beispiel für eine solche Tabelle. In diesem Beispiel liegen die *TP*, *FP*, *TN* und *FN* bei 97,8 % (228/233), 0,33 % (1/301), 99,7 % (300/301) bzw. 2,1 % (5/233).

Typischerweise kann der analytische Anwendungsbereich verschiedene Niveaus der untersuchten Spezies oder Eigenschaft und verschiedene Matrizes des analysierten Gegenstands umfassen. Dies kann separate Kontingenztabellen für verschiedene Teile des analytischen Anwendungsbereichs erfordern.

**Tabelle 3.** Ein spezifisches Beispiel für eine Kontingenztabelle, die die Leistung eines qualitativen Analysenverfahrens beschreibt, die innerhalb des analytischen Anwendungsbereichs ungefähr konstant sein sollte.

		Fall		
		positiv ( <i>pc</i> )	negativ ( <i>nc</i> )	Gesamtergebnis
Ergebnis	positive ( <i>p</i> )	<i>tp</i> = 228	<i>fp</i> = 1	<i>p</i> = 229
	negative ( <i>n</i> )	<i>fn</i> = 5	<i>tn</i> = 300	<i>n</i> = 305
Gesamtzahl der Fälle		<i>pc</i> = 233	<i>nc</i> = 301	

**Abbildung 3.** Fünf Beispiele für ROC-Kurven, in denen die Variation von *TP* und *FP* mit der Variation der quantitativen Identifikationskriterien dargestellt wird. Kurve A (blauer Punkt) zeigt eine perfekte Prüfung, bei der die Identifizierungskriterien keinen Einfluss auf die Ergebnisse haben und *TP* und *FP* 100 % bzw. 0 % betragen. Kurve B und C stellen geeignete Methoden dar, bei denen  $TP \geq FP$  ist. Von diesen drei Methoden ist Methode B der Methode C vorzuziehen. Kurve D stellt die Zufallsdiagonale dar, bei der  $TP = FP$  für alle Entscheidungsschwellenwerte gilt; dies wäre kein brauchbarer Klassifikator. Kurve E scheint auf den ersten Blick ein sehr schlechter Klassifikator zu sein, da die Falsch-Positiv-Rate durchweg höher als die Wahr-Positiv-Rate ist. Ein einfacher Wechsel der gemeldeten Ergebnisse würde jedoch eine ROC-Kurve erzeugen, die der von C nahekommt; der Klassifikator könnte sich dann als brauchbar erweisen.



### 3.2.3.2 Receiver-Operating-Characteristic-Kurve (ROC)

Bei qualitativen Analysen, die auf der Bewertung eines quantitativen Merkmals beruhen, kann die Auswahl der Klassifizierungskriterien, die die Wahr- und Falsch-Ergebnis-Raten ausgleicht, typischerweise *TP* und *FP*, mit Hilfe von Receiver Operating Characteristic (ROC)-Kurven erfolgen. Diese stellen das Paar (*TP*, *FP*) als Klassifizierungskriterium dar (d. h. ein Unterscheidungsschwellenwert variiert). Diese Kurven können auch zum Vergleich verschiedener qualitativer Analysenverfahren verwendet werden [31]. Obwohl die detaillierte Beschreibung dieser Kurven den Rahmen dieses Leitfadens sprengen würde, zeigt Abbildung 3 fünf schematische Beispiele für ROC-Kurven. Jede Kurve zeigt, wie sich die Wahr-positiv-Rate und die Falsch-positiv-Rate verändern, wenn das Identifizierungskriterium von einer strengeren zu einer weniger strengen Identifizierung positiver Fälle, die mit einer niedrigen bzw. hohen *TP* verbunden sind, variiert. Neben der visuellen Darstellung der Leistung kann die Fläche unter der Kurve (oft mit "AUC" für „area under the curve“ abgekürzt) als Zusammenfassung der Klassifikatorleistung verwendet werden [31].

## 3.3. Bewertung der Falsch-positiv- und Falsch-negativ-Raten

### 3.3.1. Anwendungsbereich des Verfahrens und Validierungsdetails

Die Validierung eines qualitativen Analysenverfahrens beinhaltet die Festlegung von Leistungsanforderungen und die Überprüfung, ob diese erfüllt werden [30].

Vor dieser Leistungsbewertung sollte der Analysenumfang in Bezug auf die Art der Klassifizierung (z. B. Vorhandensein von Pentachlorphenol über  $1 \text{ mg kg}^{-1}$ ) und die analysierten Gegenstände (z. B. Lederprodukte) klar definiert werden. Auch die Klassifizierungsmethode sollte angegeben werden,

nämlich die Analysetechnik (z. B. GC-MS/MS), die Art der Anwendung dieser Technik (z. B. Probenvorbereitung und instrumentelle Bedingungen) und die Klassifizierungskriterien. Die Klassifizierungskriterien müssen eindeutig beschrieben werden, um zu gewährleisten, dass die gesammelten Leistungsdaten auch für nachfolgende Analysen gelten.

Bei einigen qualitativen Analysen wird das Analysenverfahren aus Gründen der Effizienz in zwei Stufen unterteilt: ein vorläufiges, schnelleres und billigeres Screening-Verfahren, dem bei Bedarf ein zeitaufwändigeres und teureres Bestätigungsverfahren folgt. Die Bestätigung wird durchgeführt, wenn die erste Bewertung zu Ergebnissen führt, die den erwarteten widersprechen oder relevante Auswirkungen auf ein individuelles oder kollektives Interesse haben können. Es ist jedoch wichtig, die Falsch-negativ- und Falsch-positiv-Raten für das gesamte Verfahren zu bewerten, das die Screening- und Bestätigungstests umfasst. Zum Beispiel ist es wichtig zu überprüfen, ob die Falsch-negativ-Rate in der Screening-Phase ausreichend niedrig ist, wenn nur positive Ergebnisse einer Bestätigung unterliegen.

Was die Details der Verfahrensvalidierung anbelangt, so sollte bei Verfahren, die für eine Vielzahl von Gegenständen (z. B. verschiedene Lebensmittel) anwendbar sind, die Leistung für einen repräsentativen Satz von Gegenständen geprüft werden. Art und Anzahl der geprüften Gegenstände hängen von den Auswirkungen der analysierten Matrix auf die Leistung ab. In einigen Fällen ermöglicht das Verständnis der Klassifizierungsprinzipien die Vorhersage von Gruppen von Gegenständen mit gleichwertiger qualitativer Analysenleistung, aus denen ein repräsentativer Gegenstand ausgewählt und untersucht werden kann. Die Analysenleistung des repräsentativen Gegenstands kann dann auf die Gruppe von Gegenständen mit gleichwertiger qualitativer Analysenleistung extrapoliert werden. Wenn die Leistung der Klassifizierungstechnik es zulässt, kann die Entscheidung getroffen werden, die Analysenleistung von Gegenständen und/oder Merkmalswerten zu untersuchen, bei denen die Falsch-Ergebnis-Rate die höchsten Werte erreicht. Das Labor sollte die Gründlichkeit der Leistungsbewertung unter Berücksichtigung der für diese Bewertung zur Verfügung stehenden Zeit und Ressourcen im Auge behalten. In einigen Fällen kann es akzeptabel sein, eine kontinuierliche Validierungsstrategie durchzuführen, bei der jedes Mal, wenn ein für das Laboratorium neuer Gegenstand geprüft wird, zusätzliche und spezifische Kontrollen der Qualität der Analyse durchgeführt werden.

### 3.3.2. Verwendung von Informationen aus der Literatur

Bei häufig verwendeten qualitativen Analysenverfahren kann davon ausgegangen werden, dass Leistungsdaten öffentlich zugänglich sind. Bevor die Leistung eines etablierten Analysenverfahrens untersucht wird, sollte ein angemessenes Studium der einschlägigen Fachliteratur durchgeführt werden, um unabhängige Informationen über die Eignung des Verfahrens für den beabsichtigten Zweck zu erhalten. Veröffentlichte Falsch-Antwort-Raten sollten jedoch mit Vorsicht verwendet werden; sie könnten mit bestimmten Geräten, Reagenzien und Personal erzielt worden und auf bestimmte Probenmatrices und charakteristische Werte bezogen sein, so dass der Analytiker abwägen muss, ob seine Situation vergleichbar ist. Wenn beispielsweise die in der Literatur untersuchten Gegenstände charakteristische Werte aufweisen, die weit von den Schwellenwerten entfernt sind, die zur Unterscheidung zwischen Klassen verwendet werden, und wenn ihre Matrices relativ frei von Störungen sind, kann die ermittelte Identifizierungsleistung im Vergleich zu den "realen" analytischen Problemen des Labors zu optimistisch sein. Daher hängen die Raten der wahren und falschen Ergebnisse stark von den verfügbaren Daten ab.

In einigen Fällen lässt sich vorhersagen, ob die der Literatur entnommene Leistung besser oder schlechter sein wird als die bei der qualitativen Analyse im Labor beobachtete Leistung. Wenn man zu dem Schluss kommt, dass das qualitative Analysenverfahren für Worst-Case-Szenarien gültig ist, d. h. Ergebnisse liefert, die für den beabsichtigten Zweck geeignet sind, kann das Verfahren ohne Einschränkungen für die Analyse unbekannter Gegenstände verwendet werden.

In Abschnitt 4 wird erörtert, wie Kriterien für die Entscheidung, ob eine Analyse für den beabsichtigten Zweck geeignet ist, festgelegt werden können.

### 3.3.3. Bewertung ausschließlich durch Experimente

Unabhängig von der in Abschnitt 2 erwähnten Art der qualitativen Analyse können die *FP* und *FN* direkt aus der Anzahl der falschen Ergebnisse einer Reihe von Analysen geschätzt werden. Bei qualitativen Analysen, die ausschließlich auf qualitativen Eingaben basieren (Tabelle 1), ist dies die einzige Möglichkeit, die Unsicherheit der qualitativen Analyse zu schätzen. Wenn jedoch falsche Antworten unwahrscheinlich sind, erfordert dieser Ansatz eine große Anzahl von Prüfungen.

Angesichts der Tatsache, dass die Zahl falscher Antworten im Idealfall gering sein sollte, stellt sich die Frage, wie viele Proben zu prüfen sind, um mit ziemlicher Sicherheit eine von Null abweichende Zahl falscher Antworten zu finden.

Aus veröffentlichten Informationen (siehe z. B. Ferrara *et al.* [33]) geht hervor, dass Falsch-positiv- oder Falsch-negativ-Raten 0,5 % betragen können und in einigen Fällen sogar noch niedriger sein können [6, 8, 9]. Tabelle 4 zeigt für eine Reihe von Falsch- Ergebnis-Wahrscheinlichkeiten die Anzahl der Proben, die analysiert werden müssten, um sicherzustellen, dass innerhalb des angegebenen Konfidenzniveaus mindestens ein oder mehrere falsche Ergebnisse gefunden werden. Tabelle 4 verwendet die Binomialverteilung und zeigt, dass für eine 95-prozentige Wahrscheinlichkeit, ein oder mehrere falsche Ergebnisse zu entdecken, die Anzahl der durchzuführenden Prüfungen dreimal so hoch sein muss bezogen auf die Anzahl der Prüfungen, die durchschnittlich ein falsches Ergebnis liefern. Bei einem Verfahren mit einer Falsch-positiv-Rate von 1 % wird zum Beispiel festgestellt, dass (im Durchschnitt) bei 100 Analysen negativer Fälle ein positives Ergebnis festgestellt wird. Jedoch müssen zur Sicherstellung einer "95-prozentigen Gewissheit", dass ein falsch positives Ergebnis beobachtet wird, 299 (etwa  $3 \times 100$ ) Prüfungen an negativen Fällen durchgeführt werden.

**Tabelle 4.** Die Mindestanzahl von Analysen, um ein oder mehrere falsche (positive oder negative) Ergebnisse zu finden.

Falsch-Ergebnis-Rate	Konfidenzniveau	
	95 %	99 %
0,5 %	598	919
1 %	299	459
5 %	59	90

Die Werte in Tabelle 4 reichen nicht aus, um eine gute Schätzung der Falsch-Ergebnis-Raten vorzunehmen oder verschiedene Verfahren zu vergleichen. Selbst für eine ungefähre Schätzung wäre in der Regel das Fünf- bis Zehnfache der in Tabelle 4 angegebenen Mindestanzahl von Beobachtungen erforderlich. Diese Tabelle kann auch als Mindestanzahl von Analysen interpretiert werden, die erforderlich ist, um die Einhaltung verschiedener akzeptabler Falsch-Ergebnis-Raten zu überprüfen. Dies wird im Folgenden erläutert.

Bei dem Versuch, die Falsch-Ergebnis-Raten für ein neues Verfahren direkt aus Experimenten zu ermitteln, steht der Analytiker häufig vor einem Dilemma. Einerseits ist für ein gegebenes Verfahren die interessierende Falsch-Antwort-Rate unbekannt, und daher können alle durchgeführten Klassifizierungen unzuverlässig sein. Andererseits würde sich, analysiert man lediglich bis zum Auftreten der ersten Falsch-Antwort, nicht unbedingt ein wahrheitsgetreues Bild der Falsch-Antwort-Rate ergeben. Um dieses Problem zu lösen, wird vorgeschlagen, dass der Analytiker im Voraus akzeptable Werte für die beiden Falsch-Antwort-Raten festlegt. Für ein gewähltes Konfidenzniveau kann die Binomialverteilung verwendet werden, um die Anzahl der Versuche zu schätzen, die erforderlich ist, um eine oder mehrere falsche Antwort(en) mit ausreichender Sicherheit zu finden. Dieser Ansatz garantiert nicht, dass eine exakte Zahl für die Falsch-Antwort-Rate ermittelt wird, legt jedoch eine Grenze dafür fest. Angenommen, der Analytiker entscheidet, dass eine *FP* von 5 % akzeptabel ist, und nach Durchführung von 59 Experimenten (Tabelle 4), die den wahrscheinlichen Bereich der Matrices abdecken, werden keine falsch-positiven Ergebnisse gefunden. In diesem Fall kann der Schluss gezogen werden, dass die *FP* nicht größer als 5 % ist. Als Qualitätskontrollmaßnahme für das validierte Verfahren wird außerdem empfohlen, zu den Proben hier und da Leerproben und Referenzmaterialien, die das Zielmerkmal (z. B. den Analyten) in relevanten Merkmalskonzentrationen enthalten, einzufügen. Es sollte immer bedacht werden, dass die

Falsch-Ergebnis-Raten sehr stark von den Unwägbarkeiten und/oder Besonderheiten der zu beprobenden Population und von der Probenahmestrategie an dieser Population abhängt.

Tabelle 4 zeigt, dass es bei niedrigen Falsch-Antwort-Raten möglicherweise nicht praktikabel ist, eine ausreichende Anzahl von Proben zu analysieren, um eine falsche Antwort zu beobachten. Wenn eine Prüfung kostengünstig ist und/oder in großem Umfang eingesetzt werden soll, z. B. als Drogen-Screening-Test, kann es daher akzeptabel sein, zunächst experimentell festzustellen, dass die Falsch-Antwort-Rate eine Obergrenze von z. B. 5 % nicht überschreitet, und diese Zahl dann aufgrund der Erfahrungen mit weiteren Proben zu verfeinern.

Wenn die Anzahl der Proben voraussichtlich relativ gering und/oder die Prüfung teuer ist, sollten alle Prüfungen parallel zu einer Bestätigungsprüfung durchgeführt und die Falsch-Antwort-Raten von Zeit zu Zeit neu berechnet werden.

Die mathematische Verarbeitung verfügbarer Informationen kann dazu verwendet werden, einige Einschränkungen der experimentellen Bestimmung der Falsch-Antwort-Raten zu überwinden (siehe Abschnitte 3.3.4 und 3.3.5).

### 3.3.4. Bewertung anhand einer Datenbank

Eine Alternative zur Bestimmung von Falsch-Ergebnis-Raten aus Experimenten sind Zufalls-Mismatch-Studien in Referenzdatenbanken wie Massenspektren- oder Infrarotspektren-Datenbanken. In einigen Fällen ermöglicht dies das Äquivalent von vielen Tausenden von Experimenten. Diese Datenbanken sind zwar informativ und leistungsfähig, haben aber den Nachteil, dass sie oft nicht repräsentativ sind für die Grundgesamtheit der Tests. Zum Beispiel variieren die Materialien in der allgemeinen Verwendung stark, während eine typische Referenzdatenbank nur eines enthält. Dies kann zu erheblich abweichenden Wahrscheinlichkeitsschätzungen führen; auch hier sind die erhaltenen Werte wahrscheinlich nicht mehr als Schätzungen der Größenordnung. Die Beispiele E1 und E2 veranschaulichen die Anwendung dieser Methodik zur Bewertung der qualitativen Analysenleistung.

### 3.3.5. Bewertung anhand quantitativer Datenmodelle

Die Bewertung der Leistung hochselektiver und zeitaufwändiger und/oder teurer qualitativer Analysen ausschließlich auf der Grundlage von Experimenten in einem einzigen Labor ist nicht machbar.

Bei qualitativen Analysen, die sich auf ein quantitatives Klassifizierungskriterium für quantitative Ergebnisse stützen (z. B. eine instrumentelles Analysenverfahren), können Modelle zur Streuung der Ergebnisse verwendet werden, um die Wahr- und Falsch-Ergebnis-Raten zu schätzen. Anhang B enthält weitere Einzelheiten dazu. Wenn beispielsweise das relevante instrumentelle Signal, wie die Retentionszeit des Analyten in einem chromatographischen Verfahren, normalverteilt ist, kann die Wahrscheinlichkeit, dass eine störende Komponente eine Retentionszeit innerhalb des akzeptierten Retentionszeitintervalls für den Analyten hat, vorhergesagt werden (siehe die folgende Kurzanleitung 1).

#### Kurzanleitung 1 – Beispiel einer Signalmodellierung

Wenn für die Identifizierung von Deltamethrin in Olivenöl mittels GC-MS die geschätzte Standardabweichung der Wiederholbarkeit der Retentionszeit  $s_{tRi}$  0,022 min mit  $\nu = 32$  Freiheitsgraden beträgt, kann die Retentionszeittoleranz zur Identifizierung dieser Verbindung in einer Probe wie folgt berechnet werden:  $(t_R \pm t \cdot s_{tRi}) = (t_R \pm 2,04 \times 0,022) = (t_R \pm 0,045)$  min; wobei  $t_R$  die aus einer einzelnen täglichen Injektion einer Standardlösung beobachtete Retentionszeit und  $t$  der zweiseitige kritische Wert für die  $t$ -Verteilung mit 95 % Konfidenz und 32 Freiheitsgraden ist. Daher würde für  $t_R$  von 36,055 min das Akzeptanzintervall für einen Probenpeak  $(36,055 \pm 0,045)$  min betragen. Unter der Annahme, dass ein Störfaktor eine um 0,05 min kürzere Retentionszeit als Deltamethrin aufweist und die Präzision beider Retentionszeiten gleich ist, würde die Wahrscheinlichkeit, dass der Störfaktor eine Retentionszeit innerhalb des Akzeptanzintervalls hat, 1,5 % betragen.

Dieser Wert wird anhand der kumulativen  $t$ -Verteilung für einen  $t$ -Wert von  $(-0,05/0,022)$  und  $v$  geschätzt (MS-Excel-Formel: T.DIST(-0,05/0,022;32;TRUE)).

Die Modellierung hängt jedoch von der Gültigkeit der Modellannahme und der Werte der Eingangsvariablen ab. Da beispielsweise relative Retentionszeiten möglicherweise nicht normalverteilt sein können, kann die Annahme der Normalverteilung die Falsch-Ergebnis-Raten unterschätzen. Die Simulation instrumenteller Signale mit der Monte-Carlo-Methode ist eine bequeme Methode zur Schätzung von  $FP$  und  $FN$  aus nicht normalverteilten Parametern [8, 9]. Beispiel E5 veranschaulicht die Modellierung der instrumentellen Signalstreuung zur Schätzung der Falsch-Ergebnis-Rate von hochselektiven GC-MS/MS-Identifizierungen.

### 3.3.6. Bewertung der qualitativen Prüfleistung in Abhängigkeit von einer kontinuierlichen Variable

Viele Bestätigungs- oder Nachweisprüfungen weisen eine starke Abhängigkeit der Nachweiswahrscheinlichkeit oder der Falsch-Antwort-Rate von einer kontinuierlichen Variablen auf. Beispielsweise hängen die Nachweisraten oft von der Konzentration oder der Anzahl der Partikel des gesuchten Materials ab. In diesem Fall kann es sinnvoll sein, die Abhängigkeit der Falsch-Antwort-Rate von den kontinuierlichen (oder anderen) Variablen zu modellieren.

Logistische Regression und Probit-Regression [34, 35] werden häufig bei solchen Probleme angewendet und wurden (mit Beispielen) für die Leistungsbewertung von qualitativen Analysenverfahren vorgeschlagen [36]. Die logistische Regression wurde beim Nachweis von DNA mit geringer Kopienzahl angewandt [37]. Das Verfahren ist in Lehrbüchern gut dokumentiert und in praktisch allen statistischen Softwarepaketen verfügbar, weshalb es hier nicht im Detail vorgestellt wird. Die einfache logistische Regression modelliert die Wahrscheinlichkeit einer binären Antwort als Funktion einer kontinuierlichen Variablen. Das Modell lautet:

$$p = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)} \quad (1a)$$

Oder entsprechend:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x \quad (1b)$$

wobei  $p$  die interessierende Wahrscheinlichkeit ist (z. B. die Wahrscheinlichkeit eines positiven Ergebnisses),  $x$  die kontinuierliche Variable (normalerweise die Konzentration des Analyten) und  $b_0$  und  $b_1$  die Regressionskoeffizienten. Die meisten Statistikpakete bieten eine Anpassungsmethode entweder anhand von Rohdaten (Konzentration/qualitative Ergebnisaare) oder anhand von Proportionen, die aus der Anzahl der Ergebnisse berechnet werden. Beachten Sie, dass die erste Methode nur eine Folge von Ja/Nein- (oder 1/0-) Werten erfordert und keine Proportionen benötigt. Dadurch ist es möglich, die Methode auf eine Reihe von Prüfgegenständen mit unterschiedlichen (bekannten oder unabhängig gemessenen) Konzentrationen anzuwenden, die jeweils nur einmal dem qualitativen Prüfverfahren unterzogen werden.

Sobald eine Beziehung hergestellt ist, ist es möglich, die Nachweisgrenzen (siehe unten) aus der angepassten Beziehung zwischen Konzentration und Nachweiswahrscheinlichkeit zu schätzen, indem einfach eine geeignete Grenze für die Nachweiswahrscheinlichkeit gewählt wird, die der Definition des verwendeten Nachweisvermögens entspricht.

Beispiel E4 liefert ein praktisches Beispiel für eine logistische Regression.

### 3.3.7. Expertenurteilung

Wenn keine Daten über die Leistung des Analysenverfahrens von Dritten vorliegen und es nicht möglich ist, die Leistung ausschließlich anhand von Experimenten (Abschnitt 3.3.3) oder

Modellierungen (Abschnitte 3.3.4 und 3.3.5) zu bewerten, kann der Analytiker seine praktische Erfahrung in der Klassifizierungstechnik für die untersuchten oder ähnliche Gegenstände nutzen, um zu entscheiden, ob das Verfahren für die beabsichtigte Verwendung geeignet ist.

Wenn immer möglich, sollte die Entscheidung über die Eignung eines Verfahrens für den beabsichtigten Verwendungszweck durch objektive Beweise gestützt werden.

Der Prozess der Formulierung von Expertenurteilen ist Gegenstand mehrerer Studien. Die Urteile werden von vielen Faktoren beeinflusst, die zu entsprechenden Schätzungen der Unsicherheit des Ergebnisses führen [38].

### 3.4. Nachweisgrenze und Selektivität

#### 3.4.1. Nachweisgrenze

Die Nachweisgrenze<sup>5</sup> (LOD) beschreibt in der Regel die niedrigste Konzentration einer Substanz, die zu einem zuverlässigen Nachweis führt. Bei Prüfungen, bei denen die Klassifizierung die Bewertung eines quantitativen Merkmals beinhaltet und der Wert dieses Merkmals die qualitativen Ergebnisse beeinflusst, sollte die "Nachweisgrenze" (LOD) und/oder die "Bestimmungsgrenze" (LOQ), die in der qualitativen und/oder quantitativen Analyse berücksichtigt werden, in Bezug auf die qualitative Analysenleistung überprüft werden [30]. Das Ergebnis der qualitativen Analyse sollte für die beabsichtigte Verwendung auf diesem Niveau geeignet sein.

HINWEIS: In den Verordnungen (EU) Nr. 589/2014 [39] und Nr. 152/2009 [40] der Kommission wird ein LOQ definiert als "der niedrigste Gehalt des Analyten, der mit hinreichender statistischer Sicherheit gemessen werden kann und der die in international anerkannten Normen beschriebenen Identifizierungskriterien erfüllt" [41].

Für ausschließlich qualitative Analysen kann die LOD ermittelt werden, indem das Verfahren auf Prüfgegenstände angewandt wird, die schrittweise geringere Konzentrationen des Merkmals enthalten, bis die Likelihood falscher Ergebnisse einen vorab festgelegten Schwellenwert erreicht. Bei dieser Art der Bewertung können auch logistische und Probit-Regression verwendet werden (Abschnitt 3.3.4).

#### 3.4.2. Selektivität

Selektivität, in dem Sinne, wie dieser Begriff in der analytischen Chemie üblicherweise verwendet wird, bezieht sich auf "das Ausmaß, in dem ein bestimmtes Verfahren zur Bestimmung von Analyten unter gegebenen Bedingungen in Gegenwart anderer Komponenten mit ähnlichem Verhalten verwendet werden kann" [42]. Das International Vocabulary of Metrology (VIM) definiert diesen Begriff entsprechend als eine Eigenschaft des Messsystems [1].

HINWEIS: Der Begriff "Spezifität" wird im Zusammenhang mit der quantitativen Analyse für eine perfekt selektive Analyse verwendet [42, 43], was in der Chemie nur sehr selten vorkommt. Es gibt jedoch eine klare alternative Verwendung des Begriffs "Spezifität" im Zusammenhang mit der qualitativen Analyse (siehe Tabelle 2). In diesem Leitfaden wird der Begriff "Selektivität" in einem allgemeinen Sinne verwendet und der Begriff "Spezifität" für den in Tabelle 2 genannten Zweck vorbehalten.

Die Selektivität kann bewertet werden, indem ein oder mehrere Prüfgegenstände mit bekannten oder wahrscheinlichen Störmerkmalen analysiert werden, d. h. mit Merkmalen, die nicht das Ziel der Analyse sind, aber wahrscheinlich zu einer positiven Prüfantwort führen.

Manchmal ist es möglich, störende Spezies oder Szenarien zu identifizieren, die besonders wahrscheinlich zu falsch positiven Ergebnissen führen. So könnte zum Beispiel erwartet werden, dass Prüfungen auf Ammoniak auf primäre Amine reagieren, und bei Prüfungen auf bestimmte Bakterienstämme ist zu erwarten, dass sie auf alle Bakterien derselben allgemeinen Spezies reagieren.

---

<sup>5</sup> Anm. des Übers.: Die in Deutschland häufig gebräuchliche DIN 32645 verwendet hierfür den Begriff Erfassungsgrenze.

Wenn die qualitative Analyse in Worst-Case-Situationen relativ gut abschneidet, kann man davon ausgehen, dass das Verfahren für alle Arten von Prüfgegenständen valide ist.

Obwohl die Falsch-Antwort-Rate für jedes unterschiedliche Material oder jede vorhandene störende Spezies gemessen werden kann, ist es unwahrscheinlich, dass Selektivitätsstudien einen einzigen definitiven Wert für die Selektivität ergeben. Dies liegt daran, dass die Reaktion von den in die Studie einbezogenen potenziellen kreuzreagierenden Spezies und von der Menge dieser Spezies abhängt. Daher sollten Selektivitätsstudien am besten als grobe Anhaltspunkte für die Angemessenheit des qualitativen Analysenverfahrens bei unterschiedlichen Herausforderungen angesehen werden.

## 4. Mögliche Beschreibungen des Vertrauens in die qualitative Analyse

### 4.1. Allgemeine Erwägungen

Während Aussagen zur Messunsicherheit in der quantitativen Analyse in der Regel zu einem Wertebereich führen, wie z. B. einem erweiterten Unsicherheitsintervall oder einer Mindestreinheit, können Klassifizierungsaussagen in der Regel nicht mit einem Bereich verknüpft werden. Im Allgemeinen kann man nicht sagen, dass das Material zu 90 % "bestanden" hat, dass ein Analyt zu 99 % vorhanden ist oder dass eine chemische Spezies innerhalb einer zusammenhängenden Sequenz liegt. Stattdessen ist die typische Form der Unsicherheitsinformation probabilistischer Natur. Das heißt, man gibt einen Hinweis auf die Wahrscheinlichkeit, dass eine bestimmte Klassifizierung richtig ist, oder auf typische Wahrscheinlichkeiten einer Fehlklassifizierung von Gegenständen, deren korrekte Klasse bekannt ist.

Die Leistungswerte, die aus Validierungsstudien gewonnen werden können, können zusammen mit dem qualitativen Prüfergebnis angegeben werden. Im Allgemeinen liefern sie jedoch nur selten direkte Informationen über beispielsweise die Wahrscheinlichkeit, dass ein qualitatives Ergebnis richtig ist. In diesem Abschnitt werden zwei Kennzahlen beschrieben, die für diesen Zweck vorgeschlagen wurden, um das Verständnis zu erleichtern und den Stand der Technik bei der Angabe der Unsicherheit für qualitative Analyseergebnisse zu verbessern. Die hier vorgestellten Kennzahlen verwenden Varianten des Satzes von Bayes [4] (siehe Anhang A). Diese können verwendet werden, um a) einen Hinweis auf die Beweiskraft zu geben, die durch ein oder mehrere qualitative Ergebnisse zugunsten einer möglichen Klassifizierung gegenüber einer anderen geliefert wird; b) in Verbindung mit fundierten Informationen über die Wahrscheinlichkeiten, auf abweichende (wahre) Werte qualitativer Merkmale in einer Population zu stoßen, eine Angabe über die Wahrscheinlichkeit zu machen, dass eine bestimmte Klassifizierung, basierend auf einem bestimmten qualitativen Analyseergebnis, wahr ist.

### 4.2. Likelihood Ratio

Die bekannteste und am weitesten verbreitete Form der Berichterstattung über die Leistung qualitativer Analysen sind die Falsch-Ergebnis-Raten, insbesondere *FP* und *FN* bzw. ihre Komplementärraten, *TN* und *TP* (z. B. = 1 - *FP*). Zwei dieser Raten lassen sich bequem zu demselben Leistungsmerkmal zusammenfassen: der Likelihood Ratio, *LR*. Wenn ein positives Ergebnis gemeldet wird, wird *LR*(+) mit Gleichung (2) geschätzt:

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

*LR*(+) ist ein Verhältnis zweier Wahrscheinlichkeiten: der Wahrscheinlichkeit, ein positives Ergebnis anzugeben, wenn der Fall positiv ist, dividiert durch die Wahrscheinlichkeit, ein positives Ergebnis anzugeben, wenn der Fall negativ ist. Im Großen und Ganzen ist die Likelihood Ratio ein Maß für die Veränderung der Wahrscheinlichkeit, dass die Probe tatsächlich positiv ist, nachdem ein positives Prüfergebnis angezeigt wurde. Mathematisch gesehen ist die Likelihood Ratio die Veränderung der Wahrscheinlichkeit, ausgedrückt als "Quote" (siehe Anhang A). Eine hohe Likelihood Ratio bei einer Prüfung deutet darauf hin, dass der Prüfgegenstand mit größerer Wahrscheinlichkeit positiv ist, als man vor der Durchführung der Prüfung sagen konnte. Gelegentlich wird dies als "Beweiskraft" eines positiven Prüfergebnisses interpretiert, dass das Prüfobjekt tatsächlich positiv ist.

#### Kurzanleitung 2 – Interpretation der Likelihood Ratio

Wenn ein positives Ergebnis gemeldet wird, wird die Wahrscheinlichkeit, dass der Fall tatsächlich positiv ist, mit Gleichung (Q2.1) berechnet (siehe unten). Bei dieser Gleichung handelt es sich um den bekannten Satz von Bayes (Anhang A), bei dem die bedingten Wahrscheinlichkeiten durch die Wahr-positiv- und Falsch-positiv-Raten ersetzt werden.

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

wobei  $P(+)$  die Wahrscheinlichkeit ist, dass der Fall vor der Prüfung positiv ist. Dies kann auch in Form von "Quoten" ausgedrückt werden (siehe Anhang A):

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

In Q2.2 stellt das Verhältnis  $P(+)/P(-)$  die Wahrscheinlichkeit für einen positiven Fall vor Anwendung des qualitativen Tests dar, d. h. die "A-priori-Quote". Das Verhältnis  $TP/FP$  ist die berechnete Likelihood Ratio  $LR(+)$ .

Die Likelihood Ratio beschreibt also, wie sich die Wahrscheinlichkeit (dargestellt durch die Quote) nach einem positiven Prüfergebnis verändert; es kann als Maß für die zusätzliche Information angesehen werden, die die Prüfung liefert.

In dem speziellen Fall, in dem  $P(+)=P(-)=0,5$  ist, so dass die „A-priori-Quote“ 1,0 beträgt, stellt  $LR(+)$  das Verhältnis der A-posteriori-Wahrscheinlichkeiten dar, dass der Fall positiv oder negativ ist. Bei gleicher Häufigkeit (oder angenommener Häufigkeit) positiver und negativer Fälle würde beispielsweise ein positives Ergebnis mit  $LR(+)$  von 1000 bedeuten, dass die A-posteriori-Wahrscheinlichkeit, dass der Fall tatsächlich positiv ist, 1000-mal größer ist als die Wahrscheinlichkeit, dass der Fall negativ ist.

In dem speziellen Fall, in dem sowohl positive als auch negative Fälle gleich wahrscheinlich sind ( $P(+)=P(-)=0,5$ ; wobei  $P(+)$  und  $P(-)$  die Häufigkeit positiver bzw. negativer Fälle sind), kann die  $LR(+)$  so verstanden werden, dass sie angibt, wie viel wahrscheinlicher das gemeldete positive Ergebnis wahr als falsch ist (siehe Kurzanleitung 2). Wenn beispielsweise positive und negative Fälle vor der Prüfung als gleich wahrscheinlich angesehen werden, bedeutet ein positives Ergebnis, das mit einem  $LR(+)$  von 7300 verbunden ist, dass das positive Ergebnis 7300-mal wahrscheinlicher wahr als falsch ist.

Wenn ein negatives Ergebnis gemeldet wird, ist das  $LR(-)$ :

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

Für den Sonderfall, dass positive und negative Fälle vor der Prüfung gleich wahrscheinlich sind, gibt das  $LR(-)$  an, wie viel wahrscheinlicher ein negatives Ergebnis wahr als falsch ist.

Einige Autoren kombinieren beide Likelihood Ratios in dem Parameter "Diagnostische Chancenverteilung (Diagnostische Odds Ratio)",  $DOR$  ( $DOR = LR(+)/LR(-)$ ) [30].

Eine der nützlichsten Eigenschaften der Likelihood Ratio ( $LR(+)$  oder  $LR(-)$ ) besteht darin, dass, wenn die Klassifizierung von zwei unabhängigen Nachweisen abhängt (d. h. das Ergebnis wird nur gemeldet, wenn zwei unabhängige Analysen mit unabhängigen Verfahren dies bestätigen), das  $LR_{(1\&2)}$  des Ergebnisses beider Analysen durch Multiplikation der LR geschätzt wird, die die Unsicherheit jedes Nachweises ( $LR_{(1)}$  und  $LR_{(2)}$ ) quantifiziert:

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

Wird beispielsweise das Vorhandensein einer Verunreinigung in einem Lebensmittel mittels GC-MS auf der Grundlage einer Retentionszeit mit einem  $LR(+)$  von 99,9 und Daten eines Massenspektrums mit einem  $LR(+)$  von 490 bestimmt, so beträgt  $LR(+)$  der Identifizierungen auf der Grundlage dieser beiden Verfahren  $4,9 \times 10^4$  (d. h.  $99,9 \times 490$ ). Die Gleichung (4) ergibt sich aus der Tatsache, dass die Wahrscheinlichkeit der Konvergenz zweier unabhängiger Ergebnisse durch Multiplikation der jeweiligen Einzelwahrscheinlichkeiten geschätzt wird.

Wenn  $m$  unabhängige Nachweise ( $i = 1$  bis  $m$ ) berücksichtigt werden, um ein positives oder negatives Ergebnis anzugeben, d. h. ein Ergebnis wird nur angegeben, wenn es von den  $m$  Nachweisen angezeigt wird, wird die  $LR$  aus den kombinierten Nachweisen durch Gleichung (5) geschätzt.

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

wobei  $\Pi$  das Produkt einer Folge von Variablen bezeichnet und  $LR_{(i)}$  die Likelihood Ratio aus der  $i$ -ten qualitativen Analyse ( $LR_{(i)(+)}$  oder  $LR_{(i)(-)}$ ) ist.

Wenn die Nachweise nicht unabhängig sind, unterschätzt oder überschätzt Gl. (5) die gemeinsame Wahrscheinlichkeit. Die Kurzanleitung 3 zeigt, wie nicht unabhängige Wahrscheinlichkeiten kombiniert werden können.

Likelihood Ratios können schwierig zu interpretieren sein, insbesondere für Nichtfachleute. Für forensische Anwendungen wurde die Skala in Tabelle 5 für verbale Angaben zur Beweiskraft empfohlen [44]. Nach dieser Tabelle werden gesammelte Nachweise nur dann als "extrem stark" eingestuft, wenn die LR größer als 106 ist. Grundsätzlich kann dieser Ansatz für andere Situationen angepasst werden, wenn eine allgemeine Angabe zur Beweiskraft erforderlich ist. Für die Identifizierung des Polymertyps von Mikroplastik, das im Rahmen der Umweltüberwachung aus Sedimenten gesammelt wurde, sind die in Tabelle 5 aufgeführten Kriterien beispielsweise zu streng.

### Kurzanleitung 3 - Wahrscheinlichkeit für nicht-unabhängige Nachweise

Die Wahrscheinlichkeit des Eintretens zweier unabhängiger Ereignisse A und B,  $P(A \cap B)$ , wird mit Gleichung (Q3.1) geschätzt.

$$P(A \cap B) = P(A)P(B) \quad (\text{Q3.1})$$

wobei  $P(A)$  und  $P(B)$  die Wahrscheinlichkeiten des Eintretens der Ereignisse A und B sind, z. B. ein positives Ergebnis bei der Analyse eines positiven Falls (d. h. einer TP). Wenn jedoch  $P(A)$  und  $P(B)$  miteinander verknüpft sind, wird die Wahrscheinlichkeit des Zusammentreffens beider Ereignisse mit Gleichung (Q3.2) bestimmt, die die bedingte Wahrscheinlichkeit des Eintretens von Ereignis B unter der Voraussetzung beinhaltet, dass Ereignis A eingetreten ist:

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

Oder äquivalent dazu

$$P(A \cap B) = P(A)P(A|B) \quad (\text{Q3.2b})$$

Bei direkten oder inversen Korrelationen, die mit  $r_{AB} > 0$  bzw.  $r_{AB} < 0$  verbunden sind, ist  $P(A \cap B)$  größer bzw. kleiner als in Fällen, in denen A und B unabhängig sind.

**Tabelle 5.** Auslegung der vom Europäischen Netzwerk forensischer Institute vorgeschlagenen Likelihood Ratios für die Kriminaltechnik [44].

Wert der Likelihood Ratio	Verbale Entsprechung
1	Die Ergebnisse sprechen nicht für die eine oder die andere These
2 – 10	Schwache Unterstützung für den ersten Vorschlag statt für die Alternative
10 – 100	Mäßige Unterstützung für den ersten Vorschlag statt für die Alternative
100 – 1000	Mäßig starke Unterstützung für den ersten Vorschlag statt für die Alternative
1000 – 10 000	Starke Unterstützung für den ersten Vorschlag statt für die Alternative
10 000 – 1 000 000	Sehr starke Unterstützung für den ersten Vorschlag statt für die Alternative
> 1 000 000	Extrem starke Unterstützung für den ersten Vorschlag statt für die Alternative

Für diese Analysen sollten Ergebnisse mit einem  $LR(+)$  größer als 19 angemessen sein (d. h. mit einer  $TP \geq 95\%$  und  $FP \leq 5\%$ ), da die Kontamination nach der Identifizierung vieler Partikel aus mehreren Proben bestimmt wird [45].

Obwohl die Bestimmung einer binären Eigenschaft nur eines von zwei Ergebnissen liefern kann, kann für den Fall, dass das wahrscheinlichste Ergebnis (z. B. ja oder nein) mit einer niedrigen  $LR$  verbunden ist, die Entscheidung getroffen werden, ein Ergebnis als nicht eindeutig zu bezeichnen, anstatt das verbale Äquivalent von Tabelle 5 anzugeben. So kann z. B. entschieden werden, ein positives oder negatives Ergebnis anzugeben, wenn die entsprechende  $LR$  größer als beispielsweise  $10^5$  ist, während bei einer niedrigeren  $LR$  nicht eindeutig gemeldet wird. Diese "Grauzone" für den  $LR$ -Wert kann je nach Zweck der Analyse auf einen Wert unter  $10^5$ , 19 oder einen beliebigen anderen Wert festgelegt werden. Die Angabe eines Ergebnisses als nicht eindeutig ist sinnvoll, wenn sowohl falsch-positive als auch falsch-negative Ergebnisse eine relevante Auswirkung haben. Bei Untersuchungen auf Dopingsubstanzen im Urin eines Athleten sind falsch-positive Ergebnisse viel schwerwiegender als falsch-negative Ergebnisse, was bedeutet, dass das Ergebnis bei fehlendem Nachweis von Doping als negativ (d. h. kein Nachweis von Doping) gemeldet werden kann [9, 46]. Sowohl falsch-positive als auch falsch-negative Ergebnisse können jedoch ein Problem für die Identifizierung von Mutterschaft oder Vaterschaft darstellen, was darauf hindeutet, dass eine positive Übereinstimmung mit einer niedrigen  $LR$  nicht als eindeutige "Nicht-Übereinstimmung" gemeldet werden sollte [47].

### 4.3. A-posteriori-Wahrscheinlichkeit

Wenn es zuverlässige Informationen über die Häufigkeit eines bestimmten Merkmals gibt (z. B. eine Population mit einer gut dokumentierten Häufigkeit einer bestimmten Krankheit), kann die mit einem Testergebnis verbundene  $LR(+)$  in die Wahrscheinlichkeit  $PP$  umgewandelt werden, dass der getestete Gegenstand positiv ist, wenn das Testergebnis positiv ist. Dies wird als A-posteriori-Wahrscheinlichkeit bezeichnet und mit Hilfe des Satzes von Bayes geschätzt (Anhang A). Eine Form davon, die die Likelihood Ratio verwendet, lautet:

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

Dabei sind  $P(+)$  und  $P(-)$  A-priori-Wahrscheinlichkeiten, d. h. Informationen, die vor dem Test verfügbar sind, und  $PP$  und  $PN$  sind A-posteriori-Wahrscheinlichkeiten.

Nimmt man das vorangegangene Beispiel der Analyse einer Verunreinigung in einem Lebensmittel mittels GC-MS, bei dem ein positives Ergebnis mit einem  $LR(+)$  von  $4,9 \times 10^4$  verbunden ist, und geht von  $P(+)=P(-)=0,5$  aus, so ergibt sich ein  $PP$  von 99,998 % ( $PP = 4,9 \times 10^4 / (4,9 \times 10^4 + 1)$ ).

Wird ein negatives Ergebnis gemeldet, wird die A-posteriori-Wahrscheinlichkeit, dass die Probe tatsächlich negativ ist,  $PN$ , wie folgt geschätzt:

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

Diese Gleichung ist ähnlich wie Gleichung (6).  $P(+)$  und  $P(-)$  drücken die Häufigkeit der positiven bzw. negativen Fälle aus.

Im Allgemeinen kann die A-posteriori-Wahrscheinlichkeit, da sie sich auf eine gemeldete Klassifizierung bezieht, als Maß für die Wahrscheinlichkeit betrachtet werden, dass der gemeldete Wert korrekt ist.

A-posteriori-Wahrscheinlichkeiten können in der Praxis schwierig anzuwenden sein. Manchmal sind keine ausreichend relevanten und zuverlässigen A-priori-Wahrscheinlichkeiten verfügbar. Obwohl einige Autoren vorgeschlagen haben, dass dieses Problem dadurch gelöst werden kann, dass positive und negative Ergebnisse als gleich wahrscheinlich angenommen werden, so dass  $P(+)/P(-) = 1$  ist, ist dies nicht immer sinnvoll. Manchmal, vor allem in der Forensik, kann es unangemessen sein, aus der Kenntnis nicht verwandter Fälle auf A-priori-Wahrscheinlichkeiten für einen bestimmten Fall zu schließen. In solchen Fällen kann eine Likelihood Ratio (Abschnitt 4.2) eine nützliche

Zusammenfassung des Vertrauens in ein Testergebnis liefern, ohne dass A-priori-Wahrscheinlichkeiten bestimmt werden müssen.

In einigen Bereichen, z. B. in der Medizin, kann die Berücksichtigung der Häufigkeit einer Erkrankung oder eines Merkmals bei Entscheidungen über qualitative Analyseergebnisse für die Diagnose entscheidend sein. Die Diagnose einer Krankheit oder einer klinischen Situation auf der Grundlage von klinischen Analyseergebnissen stützt sich auch auf zusätzliche Informationen wie Schleimhautfarbe, Ort und Intensität von Schmerzen, Alter und Geschlecht, Risikoverhalten usw. Die Art und Weise, wie diese Informationen zur endgültigen Entscheidung über das beobachtete Ergebnis beitragen, kann durch die Berechnung einer  $PP$  oder  $PN$  veranschaulicht werden, obwohl Klinikärzte diese Berechnungen nicht routinemäßig durchführen; vielmehr wird von ihnen erwartet, dass sie sich der allgemeinen Bedeutung der Prävalenz bewusst sind, wenn sie eine Diagnose auf der Grundlage eines Testergebnisses stellen.

Weitere Einzelheiten zu diesen Kennzahlen sind im Literaturverzeichnis [4, 8, 9] zu finden.

#### 4.4. Zuverlässigkeit der Kennzahlen

Die Zuverlässigkeit der berechneten  $LR$ ,  $PP$  oder  $PN$  hängt von der Zuverlässigkeit der betrachteten Ergebnisrate und bei A-posteriori-Wahrscheinlichkeiten von der Zuverlässigkeit der verwendeten A-priori-Wahrscheinlichkeit ab. Tabelle 4 zeigt die Anzahl der erforderlichen Tests für eine zuverlässige Erkennung einer oder mehrerer falscher Antworten bei unterschiedlichen Wahrscheinlichkeiten für falsche Antworten. Eine zuverlässige Quantifizierung einer solchen Rate erfordert in der Regel wesentlich mehr Tests (siehe Abschnitt 3.3.3). Die Zahl der untersuchten Fälle muss unter Umständen weiter erhöht werden, um den gesamten Anwendungsbereich des Prüfverfahrens abzudecken; so kann es beispielsweise bei der Prüfung von Lebensmitteln erforderlich sein, mehrere verschiedene Lebensmittelmatrices zu untersuchen. Die Modellierung eines instrumentellen Signals, das bei der qualitativen Analyse berücksichtigt wird, kann die Quantifizierung niedriger Falsch-Ergebnisraten ermöglichen, hängt jedoch von der Angemessenheit der Eingangsdaten und des Modellierungsalgorithmus ab.

Für die Bestimmung von  $PP$  oder  $PN$  anhand sehr unterschiedlicher  $P(+)$  und  $P(-)$  kann Tabelle 4 herangezogen werden, um die Anzahl der Fälle (aus der Zielpopulation) festzulegen, die untersucht werden sollten.

Die Qualität der Eingangsdaten für die Schätzung dieser Kennzahlen ist sogar noch kritischer, wenn verschiedene Nachweise kombiniert und Kennzahlen zur Quantifizierung der Stärke der kombinierten Informationen berechnet werden. Daher sollten die vorgestellten Kennzahlen mit Vorsicht verwendet werden, wobei relevante Details der Eingangsdaten, die Verwendung des gemeldeten Ergebnisses und die jeweiligen Konsequenzen zu berücksichtigen sind. Eine Überinterpretation der Leistungsdaten qualitativer Analysen kann ebenso schädlich sein wie das Ignorieren der Grenzen einer bestimmten qualitativen Analyse.

#### 4.5. Unsicherheit von Raten

Die statistische Qualität der geschätzten Ergebnisrate, die von der Anzahl der für ihre Bestimmung verwendeten Prüfungen abhängt, kann als Konfidenzintervall (CI) für die berechnete Rate ausgedrückt werden. Dieses Konfidenzintervall wird auch als "bedingte Unsicherheit" bezeichnet (4.4.6 von [18]) und wird in der Regel für das 95 %-Konfidenzniveau (95 %-CI) berechnet.

Ein breites 95 %-CI für die Sensitivität  $SS$  deutet beispielsweise darauf hin, dass der "wahre" Wert der  $SS$  stark von der Schätzung abweichen könnte. Die gleiche Logik lässt sich auch auf andere Ergebnisraten, wie z. B.  $SP$ , anwenden. Da die Ergebnisraten nicht auf Grundlage von Vorkenntnissen über die Grundgesamtheit der Fälle geschätzt werden, charakterisieren diese Intervalle lediglich die geschätzte analytische Leistungsqualität.

Die Interpretation des 95 %-CI ähnelt in gewisser Weise der der erweiterten Messunsicherheit [1]. Für ein 95 %-CI besteht eine 5-prozentige Wahrscheinlichkeit, dass der "wahre" Wert der Ergebnisrate außerhalb der CI-Grenzen liegt. Ebenso liefert das 95 %-CI für eine experimentell bestimmte Rate die statistische Unsicherheit für die berechnete Rate.

Wenn zum Beispiel die Fähigkeit eines qualitativen Analysenverfahrens, positive Fälle korrekt zu identifizieren, anhand der Analyse von 400 solcher Fälle getestet wird und alle 400 Ergebnisse positiv sind, ist die geschätzte  $SS$  von 100 % mit einem 95 %-CI verbunden, der zwischen 99 % und 100 % liegt; d. h. der wahre Wert der  $SS$  schwankt mit 95-prozentiger Sicherheit zwischen 99 % und 100 %. Wird das Verfahren mit nur 5 positiven Fällen getestet, wird das 95 %-CI der  $SS$  durch 57 % und 100 % begrenzt. Mit dem 95 %-CI kann die Qualität der Leistungsparameter des Analysenverfahrens ausgedrückt werden, was für seine korrekte Interpretation erforderlich ist. In den obigen Beispielen werden beide  $SS$  von 100 % angegeben, aber die Schätzung der  $SS$  ist im ersten Fall viel zuverlässiger. Die Berechnung des 95 %-CI für  $SS$  und  $SP$  ist ein Standardverfahren im klinischen Labor (10.1.3 von [27]).

Es wurden mehrere Modelle zur Berechnung des CI [48]–[49, 50, 51, 52, 53, 54] veröffentlicht. Das Wilson-Score-Intervall [54] wurde der Einfachheit halber und wegen seiner Anwendbarkeit auf kleine Zahlen verwendet. Die Gleichungen 8 und 9 können zur Berechnung der unteren ( $LL_{SS,95}$ ) und oberen ( $HL_{SS,95}$ ) Grenze des 95 %-CI für die  $SS$  oder  $TP$  verwendet werden.

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

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

wobei:

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

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

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

Die Gleichungen 10 und 11 werden zur Berechnung der unteren,  $LL_{SP,95}$ , und oberen,  $HL_{SP,95}$ , Grenzen des 95 %-CI für  $SP$  oder  $TN$  verwendet.

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

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

wobei:

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

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

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

Ein Ziel- oder Mindestwert von  $LL_{SS,95}$  oder  $LL_{SP,95}$  (d. h.  $LL_{SS,95}^{tg}$  or  $LL_{SP,95}^{tg}$ ) sollte entsprechend dem Zweck der Analyse festgelegt werden. Der Zielwert ist besonders wichtig, wenn die Auswirkungen falscher Ergebnisse kritisch sind. Bei Blutkomponenten, die für Transfusionen verwendet werden, sollte das Screening auf Infektionskrankheiten beispielsweise mit Tests mit einem  $LL_{SS,95}$  von nahezu 100 % durchgeführt werden, was nur bestätigt werden kann, wenn bei der Validierung viele positive Fälle getestet werden.

Wenn die Ergebnisrate mit einem angestrebten Mindestwert verglichen wird oder wenn entweder eine Zunahme oder eine Abnahme des Parameters untersucht wird, sollte eine einseitige Bewertung durchgeführt werden. Für einen 95 %-Konfidenztest sollte der Faktor 1,96 durch 1,64 ersetzt werden.

## 5. Berichterstattung qualitativer Analyseergebnisse

Derzeit sind akkreditierte Laboratorien nicht verpflichtet, qualitative Analyseergebnisse mit einer Unsicherheit anzugeben. Die Beispiele in diesem Abschnitt sollen dementsprechend mögliche Ansätze für die Berichterstattung vorschlagen, wenn sich ein Laboratorium entscheidet, dies zur Unterstützung eines Kunden zu tun.

Ein positives Ergebnis kann mit *TP* und *FP*, *LR(+)* oder *PP* und ein negatives Ergebnis mit *TN* und *FN*, *LR(-)* oder *PN* angegeben werden. Die anderen in Tabelle 2 dargestellten Kennzahlen können ebenfalls verwendet werden, um das Vertrauen in das Ergebnis anzugeben.

Diese Kennzahlen liefern in der Regel Informationen über ein einzelnes Prüfergebnis. In Fällen, in denen der Wert einer Kennzahl für den Analysenbereich konstant ist, können solche Parameter als Charakterisierung des Analyseverfahrens interpretiert werden. Die folgenden vier Beispiele veranschaulichen, wie qualitative Ergebnisse mit der jeweiligen Leistung oder Unsicherheit angegeben werden können.

### **Beispiel 1** *(der kursive Text verweist auf die Unsicherheit der qualitativen Analyse)*

Frau A. B. ist mit dem SARS-CoV-2-Virus infiziert.

*(Test mit einer Sensitivität von 90 % und einer Spezifität von 99 %)*

### **Beispiel 2** *(der kursive Text verweist auf die Unsicherheit der qualitativen Analyse)*

Der Urin von Herrn C. D. enthält Rückstände von Canrenon.

*(Identifizierung mit einer Likelihood Ratio von  $4,9 \times 10^4$ )*

### **Beispiel 3** *(der kursive Text verweist auf die Unsicherheit der qualitativen Analyse)*

Kokain ist in Probe 123 enthalten.

*(Identifizierung mit einer Likelihood Ratio von  $4,9 \times 10^4$  und als "sehr starker" Nachweis für das Vorhandensein des Analyten erachtet)*

### **Beispiel 4** *(der kursive Text verweist auf die Unsicherheit der qualitativen Analyse)*

Im Brandschutt wurden Benzinrückstände mit dem Probencode 456 identifiziert.

*(Identifizierung mit einer A-posteriori--Wahrscheinlichkeit von 99,998 %, geschätzt aus der Signalmodellsimulation und unter der Annahme, dass das Vorhandensein oder Nichtvorhandensein des Analyten gleich wahrscheinlich ist)*

## 6. Schlussfolgerungen und Empfehlungen

Für die Laboratorien ist es wichtig, zumindest die kritischste Falsch-Antwort-Rate zu überprüfen. Für einige Kennzahlen müssen sowohl die falsch-positiven als auch die falsch-negativen Raten ermittelt werden.

Es kann realistischerweise davon ausgegangen werden, dass die meisten Laboratorien die relevanten Parameter ihrer qualitativen Analysenverfahren (d. h. die Analysenbedingungen) angemessen unter Kontrolle haben. Ein Nachweis dafür beinhaltet in der Regel:

- Einen eindeutigen Nachweis der angemessenen metrologischen Rückführung der Werte der zu kontrollierenden Parameter aufgrund ihrer Bedeutung für die Prüfung;
- Einen Nachweis, dass die Unsicherheiten dieser Parameter für diesen Zweck ausreichend gering sind.

Man kann davon ausgehen, dass die Laboratorien, soweit verfügbar, veröffentlichte Leitlinien für bewährte Verfahren in der qualitativen Analyse befolgen, einschließlich der Verwendung geeigneter Referenzdaten und -materialien.

Quantitative (d. h. numerische) Berichte über Unsicherheiten bei qualitativen Prüfergebnissen sollten so dargestellt werden, dass Fehlinterpretationen vermieden werden.

Wann immer die Schlussfolgerung gezogen wird, dass das erhaltene Analysenergebnis mit zu niedrigen bzw. zu hohen Wahr- oder Falsch-Ergebnis-Raten verbunden ist, ist es durchaus angemessen, das Prüfergebnis als "nicht eindeutig" im Sinne von nicht ausreichend sicher anzugeben.

## 7. Beispiele

Die Beispiele werden im Anschluss an ihren Geltungsbereichs beschrieben.

### 7.1. E1: Identifizierung von Verbindungen mit Hilfe der niedrig auflösenden Massenspektrometrie durch Suche in Datenbanken oder über das Vorhandensein charakteristischer Ionen

#### 7.1.1. Einleitung

Dieses Beispiel ist unterteilt in Fall A und B, bei denen unterschiedliche Verfahren zur Identifizierung von Verbindungen in komplexen Matrices durch niedrig auflösende Massenspektrometrie eingesetzt werden. Die parallele Darstellung der beiden Fälle verdeutlicht den alternativen Charakter der Identifizierungsmöglichkeiten. Beachten Sie, dass die Identifizierung in der Praxis in der Regel anhand mehrerer Kriterien erfolgt, z. B. einer Kombination aus Massenspektren und chromatographischer Retentionszeit. Dieses Beispiel konzentriert sich nur auf die Massenspektrometrie-Komponente. Beispiel E5 zeigt ein Beispiel für mehrere Kriterien.

##### **Geltungsbereich:**

---

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien

**Gegenstand/Matrix:** A) Fleischerzeugnisse und B) forensische oder Umweltproben

**Parameter/Analyt:** A) Diethylstilbestrol, DES (verbotenes Wachstumshormon für Rind- und Geflügelfleisch) oder

B) Heroin, DES und Dichlordiphenyltrichlorethan (DDT)

**Art des Klassifizierungskriteriums:** 1) Identifizierung auf der Grundlage von Toleranzen für die relativen Häufigkeiten (*RA – relative abundances*) bestimmter Ionen des Massenspektrums; 2) Identifizierung auf der Grundlage des Vorhandenseins bestimmter Ionen des Massenspektrums unabhängig von den *RA*-Werten.

**Technik/Geräteausstattung:** Gaschromatographie gekoppelt mit niedrig auflösender Massenspektrometrie mit Elektronenstoßionisation (GC-MS)

**Art der Berichterstattung:** Likelihood Ratio

---

Dieses Beispiel beschreibt die Bewertung der Unsicherheit bei der Identifizierung von Verbindungen mittels GC-MS unter Verwendung verschiedener Identifizierungskriterien (Abschnitte 7.1.1 und 7.1.2). Die Beispiele zeigen Ergebnisse für die Identifizierung von drei Verbindungen (d. h. DES, Heroin und DDT) in zwei Arten von Proben (Fleischprodukte und forensische oder Umweltproben).

Massenspektrometrie, insbesondere in Kombination mit einer chromatographischen Trennstufe, ist ein leistungsstarkes Werkzeug, das zur Identifizierung unbekannter Verbindungen beitragen kann. Für die meisten Zwecke ist die niedrig auflösende Massenspektrometrie mit Elektronenstoßionisation (EI) die Methode der Wahl, wenn es um Identifizierung und nicht um Quantifizierung geht. Ein Massenspektrum kann viele Ionen enthalten, von denen nicht alle für diagnostische Zwecke nützlich sind. Dies wirft die Frage auf, ob es eine Mindestanzahl von Ionen gibt, die für eine eindeutige Identifizierung ausreichen würde.

ANMERKUNG: In einigen analytischen Bereichen ist die Mindestzahl der für die Identifizierung von Verbindungen erforderlichen Ionen festgelegt [20]–[21][22][23].

#### 7.1.2. Identifizierung anhand der relativen Häufigkeit charakteristischer Ionen

Sphon [55] untersuchte die Mindestzahl der Ionen, die betrachtet werden müssen, um eine eindeutige Identifizierung von Diethylstilbestrol<sup>6</sup> (DES) in Fleischprodukten zu ermöglichen. Die Daten einer

---

<sup>6</sup> DES wurde als Wachstumshormon für Rinder und Geflügel verwendet und später verboten, nachdem seine krebserregenden Eigenschaften nachgewiesen wurden.

nachfolgenden Studie [56], die auf einer kommerziellen Massenspektralbibliothek mit etwa 270 000 Einträgen basiert, sind in Tabelle E1.1 aufgeführt.

Tabelle E1.1 zeigt die Anzahl der in der verwendeten Bibliothek vorhandenen Spektren, die bestimmte Kriterien für die relative Häufigkeit (RA) eines oder mehrerer Ionen erfüllen. Die RA wird geschätzt, indem die Häufigkeit des untersuchten Ions durch die Häufigkeit des am häufigsten vorkommenden Ions (d. h. des Basispeaks) geteilt wird. Diese Normalisierung zielt darauf ab, einen Identifizierungsparameter zu erhalten, der weniger von der Analytmenge (z. B. der Analytkonzentration) abhängig ist. In Tabelle E1.1 ist zu sehen, dass die Anzahl der Übereinstimmungen in der Datenbank drastisch abnimmt, wenn mehr Ionen und engere Häufigkeitsbereiche berücksichtigt werden. Das Identifizierungskriterium Nr. 6 isoliert das Massenspektrum von DES, was zu einer einzigen Übereinstimmung führt.

**Tabelle E1.1.** Anzahl der Spektren einer Wiley-Bibliothek mit 270 000 Einträgen, die bestimmten Kriterien für die relativen Häufigkeiten einiger Ionen entsprechen.

#	Kriterium zur Identifizierung		Übereinstimmungen
	Ion ( $m/z$ )	RA (%) Akzeptanzintervall	
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: Relative Häufigkeit (prozentualer Anteil des am häufigsten vorkommenden Ions, des Basispeaks)

<sup>a</sup> Die einzige Übereinstimmung entspricht dem Massenspektrum von DES

Zum Vergleich mit einer alternativen Bibliothek zeigt Tabelle E1.2 die Anzahl der Übereinstimmungen aus einer öffentlich zugänglichen Referenzbibliothek, die damals 62 235 Spektren enthielt unter Berücksichtigung von Toleranzen für die relative Häufigkeit (RA) eines oder mehrerer Ionen [57]. Je enger die Toleranzen für die relative Häufigkeit mehrerer Ionen sind, desto weniger Massenspektren von Verbindungen werden isoliert. Tabelle E1.2 zeigt die Anzahl der Spektren, die zu drei verschiedenen Zielverbindungen passen, nämlich DES, Heroin und DDT. Heroin und DDT sind für die Analyse einiger forensischer Proben bzw. Umweltproben relevant.

Der in den Tabellen E1.1 und E1.2 beschriebene Vergleich der Isolierung des Massenspektrums von DES in beiden Bibliotheken lässt den Schluss zu, dass die Anzahl der Übereinstimmungen erwartungsgemäß von der Anzahl der Spektren in der Bibliothek abhängt (siehe Identifizierungskriterien #1 und #2 in den Tabellen E1.1 und E1.2). Wird die Anzahl der Übereinstimmungen durch die Gesamtzahl der Einträge geteilt, verringern sich die in den Tabellen E1.1 und E1.2 beobachteten Unterschiede.

Tabelle E1.3 fasst die in den Tabellen E1.1 und E1.2 gesammelten Informationen zu den selektivsten Identifizierungen zusammen. In Tabelle E1.3 werden die gesammelten Informationen in *TP* und *FP* umgerechnet und zu einer *LR(+)* kombiniert, die die Unsicherheit eines positiven Ergebnisses (d. h. die Angabe des Vorhandenseins des Analyten) schätzt.

Die geschätzte *TP* (d. h. etwa 100 %) setzt voraus, dass die festgelegten Toleranzen für die *RA* der Ionen deren Variabilität berücksichtigen. Idealerweise sollten die Toleranzen anhand von Modellen der Signalvariabilität festgelegt werden, die aus Wiederholungsspektren von Probenlösungen mit relevanten Analytkonzentrationen erstellt wurden [8, 9] (Beispiel E5).

Die in Tabelle E1.3 dargestellte *FP* geht davon aus, dass es sich bei den möglichen Störfaktoren um alle Verbindungen handelt, deren Massenspektrum in der verwendeten Bibliothek verfügbar ist. Viele Verbindungen, die in den Probenlösungen vorhanden sind, werden durch GC-MS nicht nachweisbar sein oder bei der Probenvorbereitung eliminiert. Andererseits werden viele Verbindungen, deren Spektrum in der Bibliothek enthalten ist, aufgrund chemischer Unverträglichkeit oder Unabhängigkeit von Quellen oder Herkunft wahrscheinlich nicht in den analysierten Proben vorkommen. Da bekannt ist, dass die *FP* nicht gleich Null sein wird, wird eine Worst-Case-*FP* geschätzt, die eins durch Gesamtzahl der Spektren, *N*, minus 1 ( $FP = 1/(N - 1)$ ) entspricht. Bei der Verwendung dieses Wertes sollten die beschriebenen Einschränkungen bei der Bewertung der *FP* berücksichtigt werden. *FP* kann alternativ aus Modellen für Signalrauschen geschätzt werden, wie in Beispiel E5 [8, 9] erörtert.

Aus den Daten in Tabelle E1.3 ist ersichtlich, dass bei Verwendung derselben Bibliothek alle Analyten mit einer einzigen Übereinstimmung dieselben *TP*, *FP* und *LR(+)* aufweisen, die nur von der Anzahl der Bibliothekseinträge abhängen. Nach den Kriterien des Europäischen Netzwerks forensischer Institute (Tabelle 5) liefert die Identifizierung eines Analyten durch Massenspektrometrie unter Verwendung eines beschriebenen Identifizierungsverfahrens einen "sehr starken" Nachweis des Vorhandenseins des Analyten (*LR(+)* zwischen  $10^4$  und  $10^6$ ). Angenommen, die Identifizierung wird auch durch die Retentionszeit des Analyten im chromatographischen System (d. h. in der GC) unterstützt und das Retentionszeitfenster ist angemessen. In diesem Fall kann sich die *LR(+)* der Identifizierung erhöhen (Beispiel E5).

**Tabelle E1.2.** Die Anzahl der Spektren in einer öffentlich zugänglichen Bibliothek mit 62 235 Einträgen, die bestimmten Kriterien für die relativen Häufigkeiten einiger Ionen entsprechen.

#	Kriterium zur Identifizierung		Übereinstimmungen
	Ion ( $m/z$ )	RA (%) Akzeptanzintervall	
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	
	268	55 – 95	
5	239	30 – 70	1 (DES)
	145	60 – 100	
	268	55 – 95	
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: Relative Häufigkeit (prozentualer Anteil des am häufigsten vorkommenden Ions, des Basispeaks)  
<sup>a</sup> Die einzige Übereinstimmung entspricht dem Massenspektrum von DES, Heroin oder DDT

### 7.1.3. Identifizierung anhand des Vorhandenseins charakteristischer Ionen - geschätzte zufällige Übereinstimmungswahrscheinlichkeiten

Angenommen, die Identifizierung des Analyten erfolgt nicht durch die Verwendung von Toleranzen für die RA spezifischer Ionen, sondern durch das einfache Vorhandensein von drei selektiven Ionen. In diesem Fall kann die folgende Mathematik für eine grobe Bewertung der Identifizierungsunsicherheit verwendet werden. Angenommen, es wird ein niedrig auflösendes Massenspektrometer verwendet, das nur  $m/z$ -Einheiten (d. h. Einheiten des Masse-zu-Ladungs-Verhältnisses) unterscheiden kann, und Ionen müssen  $m/z$ -Werte zwischen 180  $m/z$  und 480  $m/z$  haben. In diesem Fall können etwa 300 mögliche  $m/z$ -Werte in einem Spektrum beobachtet werden. Da die Anzahl der Kombinationen von 300 Objekten, von denen jeweils drei gleichzeitig betrachtet werden,  $300!/[3! \cdot (300 - 3)!] = 4\,455\,100$  beträgt, würde die Wahrscheinlichkeit, dass drei zufällige Peaks mit den drei ausgewählten Ionen übereinstimmen, unter der Annahme, dass alle  $m/z$ -Verhältnisse gleich wahrscheinlich sind,  $1/4,6 \times 10^6$  oder etwa  $2,2 \times 10^{-7}$  betragen. Dies berücksichtigt jedoch nicht die Tatsache, dass die meisten Massenspektren in der Regel viel mehr als drei Ionen im interessierenden Bereich enthalten; dies

erhöht die Wahrscheinlichkeit, dass ein Massenspektrum mit  $n$  Ionen mit den drei ausgewählten Ionen übereinstimmt, um den Faktor  $n!/[3! \cdot (n - 3)!]$ . Nimmt man zehn als typische Anzahl von Peaks an, erhöht sich die Wahrscheinlichkeit einer zufälligen Übereinstimmung um  $10!/3! \cdot 7! = 120$ . Die geschätzte Wahrscheinlichkeit einer zufälligen Übereinstimmung ist demnach  $120 \times 2.2 \times 10^{-7}$  oder ungefähr  $2.6 \times 10^{-5}$ . Dies ergibt eine ungefähre Falsch-positiv-Rate,  $FP$ .

Unter der Annahme, dass die Verunreinigungen hoch genug sind, um zuverlässige Massenspektren zu erhalten, kann man davon ausgehen, dass der  $TP$ -Wert etwa 1, oder 100 %, beträgt.

Die oben geschätzten  $TP$  und  $FP$  können kombiniert werden, um eine  $LR(+)$  von  $1/2,6 \times 10^{-5} = 3,8 \times 10^5$  zu erhalten, die einen oberen Grenzwert für die Likelihood Ratio für Identifizierungen auf der Grundlage des beschriebenen Verfahrens darstellt.

**Tabelle E1.3.** Unsicherheit der Identifizierung mehrerer Analyten in verschiedenen Matrices mittels GC-MS, geschätzt aus der Anzahl der Übereinstimmungen von Spektren aus der verwendeten Bibliothek, unter Berücksichtigung der Toleranzen für die relativen Häufigkeiten von zwei oder drei Ionen.

Analyt	Analysenobjekt	Anzahl der Spektren der Bibliothek, $N$	Anzahl der Übereinstimmungen <sup>a</sup>	$TP$ (%) <sup>b</sup>	$FP$ (%) <sup>c</sup>	$LR(+)$ ( $TP/FP$ )
DES	Fleischprodukte	270 000	1 in $N$	~100	$3,7 \times 10^{-4}$	$2,7 \times 10^5$
DES	F&E	62 235	1 in $N$	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$
Heroin	F&E	62 235	1 in $N$	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$
DDT	F&E	62 235	1 in $N$	~100	$1,6 \times 10^{-3}$	$6,2 \times 10^4$

DES: Diethylstilbestrol; F&E: Forensische und Umweltproben;  $TP$ : Wahr-Positive-Rate;  $FP$ : Falsch-Positiv-Rate;  $LR(+)$ : Likelihood Ratio ( $TP/FP$ ) (siehe Tabelle 2).

<sup>a</sup> Anzahl der Übereinstimmungen unter Berücksichtigung festgelegter Toleranzen für die  $RA$  spezifischer Ionen.

<sup>b</sup> optimistische Schätzung von  $TP$  (ungefähr 100 %)

<sup>c</sup> geschätzt als  $1/(N - 1)$  (Worst-Case-Szenario aus gesammelten Informationen)

ANMERKUNG: Bei dieser Schätzung wird davon ausgegangen, dass alle Ionenkombinationen gleich wahrscheinlich und möglich sind und dass ihr Auftreten unabhängig voneinander ist; dies ist bekanntermaßen eine grobe Annäherung (siehe Beispiel 7.2). Die tatsächliche Wahrscheinlichkeit einer zufälligen Übereinstimmung ist daher wahrscheinlich sehr viel höher als die berechnete Zahl, was zu einer niedrigeren Likelihood Ratio führt.

#### 7.1.4. Schlussbemerkungen

Die in diesem Beispiel vorgestellten Methoden zur Schätzung der  $LR(+)$  von GC-MS-Identifizierungen neigen dazu, die Gültigkeit qualitativer Analyseergebnisse zu optimistisch einzuschätzen (siehe Beispiel 7.2 für einen Vergleich mit den tatsächlichen Übereinstimmungswahrscheinlichkeiten). Daher sollten diese Berechnungen nur als erste Einschätzung der Gültigkeit der Identifizierung verwendet werden. In Beispiel E5 werden alternative und realistischere Bestimmungen der Unsicherheit von Identifizierungen mittels GC-MS/MS erörtert.

Auch wenn die hier vorgestellten Methoden zur Schätzung der Wahrscheinlichkeit einer zufälligen Übereinstimmung und der Likelihood Ratio optimistisch sein mögen, kann man dennoch schlussfolgern, dass z. B. die Identifizierung von Heroin in Proben von einem Tatort, die allein auf dem Vorhandensein von Ionen mit  $m/z$  von 369 und 327 beruht, unzureichend wäre, da sie mit einer  $LR(+)$  von 118 ( $118 = TP/\{FP\} = 100/\{[526/(62235-1)] \cdot 100\}$ ) einhergeht (Methodik von Abschnitt 7.1.1: Fall 7 der Tabelle E1.2) oder  $4,5 \times 10^4$  ( $4,5 \times 10^4 = 100/\{[1/(44850-1)] \cdot 100\}$ ); wobei  $44850 = 300!/[2! \cdot (300 - 2)!]$ ) (Methodik von Abschnitt 7.1.2), je nachdem, welcher Ansatz zur Bewertung des Vertrauens in die Ergebnisse verwendet wird. In der Praxis würde dies bedeuten, dass zusätzliche

Kriterien oder weitere Bestätigungstests erforderlich wären, um ein ausreichendes Vertrauen zu gewährleisten.

Auch wenn Sphon [55] und andere vorgeschlagen haben, dass Tierarzneimittelrückstände in Rindern durch drei Massenspektrum-Ionen identifiziert werden können, verlangt die Europäische Union (EU) für die amtliche Überwachung nicht zugelassener Substanzen zusätzliche Nachweise für das Vorhandensein dieser Verbindungen [20]. Wenn es beispielsweise nur möglich ist, zwei charakteristische Ionen mittels GC-MS bei entsprechenden Gehalten des Analyten zu überwachen, müssen zwei unabhängige chromatographische Läufe basierend auf Elektronenstoß- oder chemischer Ionisierung in Betracht gezogen werden, um das Vorhandensein des Analyten zu bestätigen [20].

## 7.2. E2: Identifizierung von gereinigten Verbindungen durch Infrarotspektrometrie

### Geltungsbereich:

---

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien

---

**Gegenstand/Matrix:** gereinigte chemische Verbindung

---

**Parameter/Analyt:** Eine Verbindung aus der verfügbaren Infrarotspektren-Datenbank

---

**Art des Klassifizierungskriteriums:** Übereinstimmung der Wellenzahl von drei oder sechs Banden der Infrarotspektren im Intervall  $[500, 1800] \text{ cm}^{-1}$

---

**Technik/Geräteausstattung:** Infrarotspektrometrie

---

**Art der Berichterstattung:** Likelihood Ratio

---

Mehrere Autoren haben die Verwendung von Datenbankstatistiken zur Bewertung von Kriterien für die qualitative Analyse untersucht. De Ruig *et al.* [58] schlugen Kriterien vor, die erfüllt sein müssen, um Tierarzneimittelrückstände in Fleischprodukten zu identifizieren (siehe Abschnitt 7.1.2). Die Autoren geben Richtwerte für zufällige Übereinstimmungswahrscheinlichkeiten auf der Grundlage eines einfachen Binomialmodells an. Ellison *et al.* haben gezeigt, dass eine hypergeometrische Verteilung ein geeigneteres Modell für zufällige Übereinstimmungen in Spektren ist, da sie eine geringe Anzahl übereinstimmender Peaks zwischen zwei Spektren zulässt, die beide eine größere Anzahl von Peaks enthalten [5]. Die letztgenannten Autoren konzentrierten sich auf zufällige Übereinstimmungen beim Vergleich eines Infrarotspektrums mit einer Spektralbibliothek.

Ellison *et al.* [5] untersuchten die Zuverlässigkeit der Identifizierung von gereinigten Verbindungen durch Vergleich des erhaltenen Infrarotspektrums mit Spektren aus einer Bibliothek. Bei der von Ellison *et al.* verwendeten Bibliothek handelt es sich um die Sadtler-Bibliothek, die Spektren von 59 626 verschiedenen Materialien enthält. Aus dieser Bibliothek wurde eine zufällige Teilmenge von dreißig Verbindungen ausgewählt und die Anzahl der Banden,  $m$ , im Intervall  $[500, 1800] \text{ cm}^{-1}$  für jede Verbindung notiert. Es wurde festgestellt, dass die durchschnittliche Anzahl der Banden pro Spektrum im Intervall  $[500, 1800] \text{ cm}^{-1}$ ,  $M$ , 16 beträgt. Die verfügbare spektrale Auflösung betrug  $4 \text{ cm}^{-1}$ , was auf das Vorhandensein von  $1300/4 = 325$ ,  $N$ , diskreten Peakpositionen im Intervall  $[500, 1800] \text{ cm}^{-1}$  hinweist. Für jedes unterschiedliche Spektrum in der ausgewählten Teilmenge wurde die gesamte Datenbank zweimal durchsucht – zuerst nach einem Minimum von  $n = 3$  übereinstimmenden Peaks und beim zweiten Mal nach einem Minimum von  $n = 6$  übereinstimmenden Peaks.

**Tabelle E2.1.** Zufallsabgleich mit sechs Banden in einer Infrarot-Datenbank.

Verbindung	Anzahl der Banden $m$ im Bereich	Wahrscheinlichkeit der Übereinstimmung <sup>a</sup>	vorhergesagte Übereinstimmungen <sup>b</sup>	Beobachtete Übereinstimmungen	$LR(+)$
1-Chlor-3-(1-naphthoxy)-2-propanol	23	$3,19 \times 10^{-4}$	19	192	311
$\alpha$ -Cyanozimtsäure-Methylester	17	$5,03 \times 10^{-5}$	3	29	2056
Phenyl- <i>v</i> -triazolo-[1,5- $\alpha$ ]-pyridin-3-yl-ke-ton	24	$4,19 \times 10^{-4}$	25	190	314
Benzo- $\beta$ -thiophen-6-Acrylsäure	20	$1,34 \times 10^{-4}$	8	52	1147
3-((Dipropylamino)methyl)1-5-nitroindol	17	$5,03 \times 10^{-5}$	3	29	2056
2-Mesityl-5-Phenyloxazol	22	$2,52 \times 10^{-4}$	15	99	602
<i>p</i> -Hydroxybenzoesäure	18	$6,71 \times 10^{-5}$	4	44	1355
Capronsäure-Isobutylester	8	$1,36 \times 10^{-7}$	0	1	59626
1-Bromadamantan	10	$9,64 \times 10^{-7}$	0	1	59626
Phenylpropylether	17	$5,03 \times 10^{-5}$	3	47	1269

<sup>a</sup> Die Wahrscheinlichkeit der zufälligen Übereinstimmung ist die Wahrscheinlichkeit, dass  $n = 6$  Peaks in zwei Spektren mit  $m$  Banden zufällig übereinstimmen, wobei ein zufälliges Auftreten von Banden im Spektrum angenommen wird. Die Wahrscheinlichkeit wurde anhand der hypergeometrischen Verteilung berechnet (siehe [5]).

<sup>b</sup> Die vorhergesagte Anzahl der Übereinstimmungen ist die Wahrscheinlichkeit der zufälligen Übereinstimmung multipliziert mit der Anzahl der Spektren in der Datenbank

Für  $n \geq 3$  war die Anzahl der beobachteten Übereinstimmungen etwa doppelt so hoch wie die durch die hypergeometrische Verteilung vorhergesagte Anzahl. Für  $n \geq 6$  war die Anzahl der Übereinstimmungen zwar erwartungsgemäß deutlich geringer, aber die beobachteten Übereinstimmungen übertrafen die vorhergesagten um das Zehnfache. Ein Teil der Daten für sechs übereinstimmende Peaks ist in Tabelle E2.1 dargestellt.

Die berechneten Wahrscheinlichkeiten für zufällige Übereinstimmungen für Sechs-Peak-Treffer lagen im Intervall  $[10^{-7}, 10^{-5}]$ . Die Wahrscheinlichkeit der zufälligen Übereinstimmung für eine Verbindung, multipliziert mit der Anzahl der Einträge in der Datenbank, schätzt die Anzahl der Verbindungen, die den Suchkriterien entsprechen. Bei zwei der in Tabelle E2.1 aufgeführten Verbindungen, nämlich Capronsäure-Isobutylester und 1-Bromadamantan, führen die Suchkriterien zu einer einzigen Übereinstimmung und scheinen daher für einen Verdacht auf diese Verbindungen ausreichend zu sein. Für die übrigen Verbindungen ergeben sich sehr viel mehr Übereinstimmungen, was auf die Notwendigkeit strengerer Kriterien hinweist.

Unter der Annahme, dass  $TP$  ungefähr 100 % beträgt (da IR-Spektren reiner Verbindungen zuverlässig mit ihren eigenen Referenzspektren übereinstimmen), und  $FP$  als das Verhältnis zwischen der Anzahl der beobachteten Übereinstimmungen und der Gesamtzahl der Spektren der Bibliothek (d. h. 59 626) definiert wird, kann man  $LR(+)$  der Identifizierung schätzen. Die letzte Spalte der Tabelle E2.1 zeigt die geschätzte  $LR(+)$  (berechnet als  $TP/FP$ , mit  $TP = 1,0$ ). Die angegebenen  $LR(+)$  liegen weit unter dem Mindestwert von  $10^6$ , der für die Einstufung eines Nachweises als "extrem stark" gilt (Tabelle 5), was darauf hindeutet, dass eine einfache Übereinstimmung von sechs Peaks allein anhand der Wellenlänge ohne zusätzliche Kriterien möglicherweise kein ausreichendes Vertrauen bietet. Weitere Kriterien (wie z. B. die zusätzliche Übereinstimmung mit der Peakintensität, das Fehlen von Peaks, die in der Zielverbindung nicht vorhanden sind, die Übereinstimmung mit der chromatographischen

Retentionszeit oder die Anforderung einer engen visuellen Übereinstimmung mit vollständigen Spektren) könnten erforderlich sein, um eine ausreichend eindeutige Identifizierung zu gewährleisten.

Dieses Beispiel unterstreicht, dass Referenzdatenbanken, zu denen auch Spektralbibliotheken gehören, nur unverbindliche Informationen über Falsch-Antwort-Raten liefern können. Die Antwort-Raten sind nur für Populationen zuverlässig, die der erwarteten Probenpopulation ähnlich sind. Es liegt auch in der Verantwortung des Analytikers, zu entscheiden, welche ggf. vorhandenen Übereinstimmungen einer unbekannt Probe entsprechen.

### 7.3. E3: Identifizierung von Drogenmissbrauch im Urin mit Hilfe der Enzyme Multiplied Immunoassay Technique (EMIT) und einer alternativen Technik

#### Geltungsbereich:

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien (untersucht anhand qualitativer Informationen)

**Gegenstand/Matrix:** Urin

**Parameter/Analyt:** Kokain, Methadon oder Opiate

**Art des Klassifizierungskriteriums:** keine Angaben

**Technik/Geräteausstattung:** Enzyme Multiplied Immunoassay Technique, EMIT, und eine alternative, proprietäre Technik.

**Art der Berichterstattung:** Likelihood Ratio und die Wahrscheinlichkeit, dass das positive Ergebnis richtig ist

Obwohl EMIT-Bestimmungen die Verarbeitung eines instrumentellen Signals beinhalten, wird in diesem Beispiel die Leistung dieser qualitativen Analyse anhand experimentell ermittelter Raten falsch positiver und falsch negativer Ergebnisse bewertet. Daher veranschaulicht dieses Beispiel die Bestimmung der Qualität einer qualitativen Analyse auf der Grundlage qualitativer Informationen.

In der Literatur wurde über die Verwendung von Datenbanken mit Probenergebnissen zur Ermittlung der relevanten Wahrscheinlichkeiten für eine Bayessche Bewertung qualitativer Analysenleistung berichtet. Ferrara *et al.* [33] haben für die Untersuchung von Urin auf Drogenmissbrauch eine Datenbank zusammengestellt, die Informationen über Drogentypen, Analysentechniken, Fehlreaktionsraten für die Techniken und die Prävalenz der Drogen enthält. Für das Labor der zitierten Autoren fasst Tabelle E3.1 einen Teil dieser Daten für EMIT zusammen. Die Tabelle zeigt auch die geschätzte A-posteriori-Wahrscheinlichkeit, dass die Testprobe wirklich positiv ist,  $PP$ , wie in Gleichung (6) beschrieben. Für die Berechnung wird die Häufigkeit negativer Ergebnisse als  $(1 - P(+))$ , die  $LR(+) = TP/FP$  und die  $TP = 1 - FN$  angenommen.

**Tabelle E3.1.** Wahrscheinlichkeiten für den EMIT-Nachweis von Drogenmissbrauch im Urin

Wahrscheinlichkeit	Werte der Leistungsmerkmale für verschiedene Drogen oder Drogenklassen		
	Opiate	Methadon	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
$P(+)$ – Häufigkeit positiver Ergebnisse			

Für die Identifizierung von Methadon beispielsweise wird  $PP$  nach Gleichung (E3.1) mit 0,988 bewertet ( $PP$  ist die aus dem Satz von Bayes ermittelte A-posteriori-Wahrscheinlichkeit). Mit anderen Worten: Der Analytiker könnte zu über 98 % sicher sein, dass eine positive Reaktion auf Methadon tatsächlich auf das Vorhandensein dieser Droge hinweist. Es ist jedoch zu beachten, dass dies zum Teil von der hohen beobachteten Prävalenz in der jeweiligen Stichprobenpopulation abhängt. In der

Allgemeinbevölkerung, d. h. in einer Bevölkerung, in der Drogenkonsum nicht so häufig vorkommt, könnte die Prävalenz sehr viel niedriger sein.

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

Tabelle E3.2 zeigt ähnliche Daten für ein anderes, nicht-immunochemisches Verfahren. Man beachte, dass *FP* für Kokain bei dieser Technik mit Null angegeben wird. Es ist jedoch strittig, ob die Falsch-Antwort-Rate bei solchen Screening-Tests tatsächlich gleich Null sein kann. In diesem Fall wurden keine falsch-positiven Ergebnisse gefunden, aber wenn mehr Proben analysiert worden wären, hätten ein oder mehrere falsch-positive Ergebnisse auftreten können. Daher wurde bei der Berechnung der A-posteriori-Wahrscheinlichkeit eine geschätzte Falsch-Antwort-Rate verwendet.

**Tabelle E3.2.** Wahrscheinlichkeiten für den Nachweis von Drogenmissbrauch im Urin durch die proprietäre Technik

Werte der Leistungsmerkmale für verschiedene Drogen oder Drogenklassen			
Wahrscheinlichkeit	Opiate	Methadon	Kokain
<i>P</i> (+)	0,44	0,26	0,20
<i>FP</i>	0,038	0,012	0,000
<i>FN</i>	0,276	0,179	0,247
<i>PP</i>	0,937	0,960	0,995 §

§ *PP* berechnet unter Verwendung eines geschätzten Worst-Case-*FP*-Werts von 0,001 (einer von 1000 Tests)

Bei der Betrachtung von Methadon liegt die *PP* bei 0,960. Dies ist eine recht hohe Wahrscheinlichkeit, wenn auch etwas weniger überzeugend als die des EMIT-Tests.

Zur Veranschaulichung der Auswirkungen einer Kombination von Daten nehmen wir an, dass beide Screening-Tests durchgeführt wurden. Wenn in beiden Fällen eine positive Reaktion erhalten wird, beträgt die kombinierte *PP* 0,9998 (siehe Gleichung (E3.2) auf der Grundlage der Kombination von Gleichung (5) und (6).

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

In diesem Beispiel sind zuverlässige Prävalenzwerte (d. h. A-priori-Wahrscheinlichkeiten) verfügbar. Hätten diese nicht zur Verfügung gestanden oder hätte der Analytiker es vorgezogen, sie nicht zu verwenden, hätten stattdessen Likelihood Ratios verwendet werden können; die entsprechenden Werte wären 246 (EMIT) und ~68 (proprietär), was eine kombinierte Likelihood Ratio von etwa 17 000 (Gleichung (E3.3)) ergibt. Nach Tabelle 5 entspricht das einem "sehr starken" Nachweis für das Vorhandensein von Methadon.

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

In allen Fällen wurde GC-MS als Referenztechnik verwendet, um die Falsch-Antwort-Rate zu ermitteln. Die hier genannte Datenbank ist für die untersuchten Analyten recht umfangreich und wurde so konzipiert, dass sie eine repräsentative Sammlung darstellt, die eine Bayessche Analyse der Daten ermöglicht. Es gibt zwangsläufig einige fehlende Werte, aber mit zunehmenden Daten sollten diese reduziert und die Genauigkeit der Vorhersagen weiter verbessert werden.

Ein weiterer Vorteil einer Datenbank, in der repräsentative Daten für mehrere verschiedene Techniken erfasst werden, sind Informationen, die sie zur Optimierung der analytischen Leistung liefert. Die Auswahl eines Screening-Verfahrens mit einer niedrigen Falsch-Positiv-Rate sollte beispielsweise die Kosten für teure Bestätigungsanalysen minimieren. Es müssen jedoch auch andere Faktoren berücksichtigt werden, wie die Nachweisgrenze der Technik, die Falsch-Negativ-Rate sowie die Geschwindigkeit und die Kosten der Analyse.

## 7.4. E4: Identifizierung des menschlichen SRY-Gens in biologischem Material mittels qPCR

### Geltungsbereich:

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien (Schwellenwert überschreitende Fluoreszenz)

**Gegenstand/Matrix:** Biologisches Material

**Parameter/Analyt:** SRY-Gen (geschlechtsbestimmender Bereich Y)

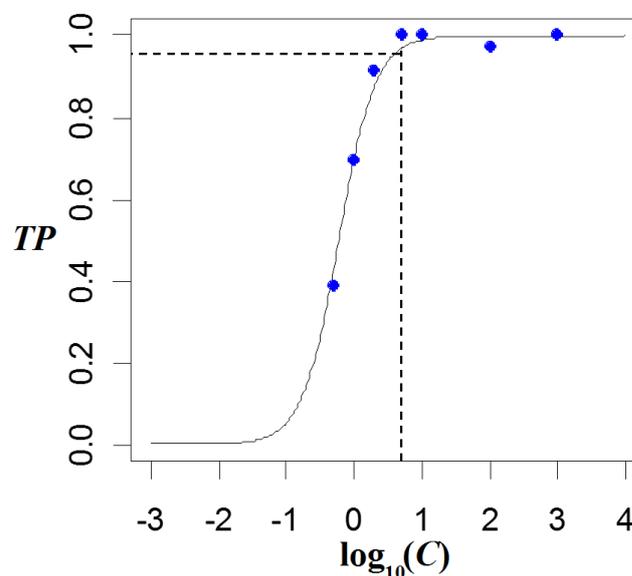
**Art des Klassifizierungskriteriums:** keine Angaben

**Technik/Geräteausstattung:** Quantitative Polymerase-Kettenreaktion (qPCR)

**Art der Berichterstattung:** Wahr-Positiv-Rate

Obwohl die Auswertung der qPCR die Verarbeitung eines instrumentellen Signals beinhaltet, wird in diesem Beispiel die Leistung anhand der Bestimmung der *TP* bei verschiedenen DNA-Konzentrationen bewertet.

Abbildung E4.1 zeigt einige experimentelle Daten aus einer Studie zum Nachweis des menschlichen SRY-Gens mittels quantitativer PCR (qPCR) in biologischem Material [37]. Die Daten stammen aus einem 3-Platten-Assay unter Verwendung eines 5'-Nuklease-Assays mit einer zweifach markierten fluorogenen "TaqMan"-Sonde, die auf die männerspezifische SRY-Region des menschlichen Genoms gerichtet ist. Ein Ergebnis wurde als "positiv" eingestuft, wenn die beobachtete Fluoreszenz innerhalb von 55 Amplifikationszyklen einen vorgegebenen Schwellenwert überschritt. Obwohl im Rahmen der Validierungsstudie unterschiedliche Präparationsverfahren verwendet wurden, zeigten statistische Tests keine signifikanten Unterschiede, so dass die Daten als ein einziger Datensatz behandelt wurden.



**Abbildung E4.1.** Durch logistische Regression verarbeitete DNA-Nachweisdaten. Die *TP* wird als Funktion von  $\log_{10}(C)$  aufgetragen, wobei  $C$  die Kopienzahl des Gens ist, in einer Studie über die DNA-Nachweis- und Klassifizierungsfähigkeit. Die Punkte zeigen den Anteil der positiven Ergebnisse von insgesamt 36 Wiederholungen bei jeder Kopienzahl. Die durchgezogene Linie zeigt die logistische Regressionsanpassung mit  $b_0 = 0,85$  und  $b_1 = 3,75$  (siehe Abschnitt 3.3.6). Die gestrichelte Linie zeigt  $TP = 0,95$  und die entsprechende logarithmische Konzentration von  $\log_{10}(C) = 0,56$  für  $C = 3,6$  Kopien.

Beachten Sie, dass der Punkt bei  $\log_{10}(C) = 2$  keine 100-prozentige *TP* aufweist, während die Punkte um ihn herum dies tun, was eine genaue Bewertung der Klassifizierungsfähigkeit erschwert. Die Regressionskurve ermöglicht jedoch eine recht genaue Lokalisierung der effektiven Nachweisgrenze, wie in der Abbildung dargestellt; durch die Verwendung eines kontinuierlichen Modells wurden die zufälligen Zählraten effektiv geglättet. Dies ist ein beträchtlicher Vorteil, denn so kann der Analytiker eine große Anzahl von Konzentrationen mit relativ wenigen Wiederholungen pro Stufe untersuchen, anstatt viele Wiederholungen bei wenigen Stufen zu benötigen, und dennoch relativ zuverlässige Wahrscheinlichkeitsschätzungen erhalten.

Dieser Datensatz veranschaulicht eine wichtige Einschränkung bei der Modellierung. Die Konzentrationsdaten werden logarithmisch aufgetragen und modelliert; eine gängige Praxis bei der Arbeit mit experimentellen DNA-Konzentrationen oder Keimzahlbestimmungen. Bei einer abhängigen Variable wird die Transformation häufig durch die Fehlerverteilung vorgegeben. In diesem Fall gibt es jedoch keinen zwingenden Grund, die logarithmische Transformation für die unabhängige Variable zu wählen; die Wahl ist im Wesentlichen willkürlich. Das Gleiche gilt für die Wahl des logistischen Modells; andere Modelle können ebenfalls recht gut die Daten anpassen. Wenn verschiedene, aber glatte Modelle die Daten ähnlich gut anpassen, ist die Interpolation nicht sehr empfindlich gegenüber der Wahl des Modells. Allerdings können extreme Wahrscheinlichkeiten sehr empfindlich auf die Wahl des Modells reagieren. Daraus folgt, dass es selbst dann, wenn ein Modell eine gute Beschreibung der Daten und relativ verlässliche Wahrscheinlichkeiten und Nachweisgrenzen innerhalb dieses Bereichs liefert, sehr unsicher ist, Wahrscheinlichkeitsschätzungen weit über den untersuchten Bereich hinaus zu extrapolieren, ohne wesentliche Beweise für die Gültigkeit des Modells zu haben.

## 7.5. E5: Identifizierung von Pestizidrückständen in Lebensmitteln mittels GC-MS/MS auf der Grundlage von Retentionszeit und Ionenhäufigkeitsverhältnis

### Geltungsbereich:

---

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien

---

**Gegenstand/Matrix:** Lebensmittel pflanzlichen Ursprungs

---

**Parameter/Analyt:** Chlorpyrifos-Methyl und Malathion

---

**Art des Klassifizierungskriteriums:** Intervalle für Retentionszeiten und Häufigkeitsverhältnisse charakteristischer Ionen

---

**Technik/Geräteausstattung:** GC-MS/MS

---

**Art der Berichterstattung:** Likelihood Ratio und die Wahrscheinlichkeit, dass das positive Ergebnis richtig ist

---

In diesem Beispiel geht es um die Schätzung der FP von hochselektiven Bestimmungen von Pestizidrückständen in Lebensmitteln mittels GC-MS/MS durch Modellierung des instrumentellen Signals unter Verwendung der Monte-Carlo-Methode. Die Monte-Carlo-Simulationen wurden in einer MS-Excel-Tabelle durchgeführt.

Die Identifizierung der Analyten basiert auf ihrer Retentionszeit  $t_R$  im chromatographischen System und auf dem Häufigkeitsverhältnis  $AR = A_1/A_2$  von zwei charakteristischen Ionen des Massenspektrums des Analyten. Während die  $t_R$  annähernd normalverteilt ist, kann  $AR$  erheblich von der Normalverteilung abweichen. Das Verhältnis von korrelierten Variablen ist bekanntermaßen nicht normalverteilt, insbesondere wenn die Variable mit geringerer Präzision (d. h. größerer Streuung der Werte) im Nenner steht [8].

Die Entwicklung und Validierung eines Verfahrens zur Identifizierung von Spuren der Analyten in Lebensmitteln mittels GC-MS/MS beginnt mit der Festlegung des qualitativen Analyseverfahrens, einschließlich der Probenvorbereitung und der GC-MS/MS-Bedingungen. Die Spezifikation der GC-MS/MS-Bedingungen umfasst die Auswahl charakteristischer Ionen des Massenspektrums des Analyten (Chlorpyrifos-Methyl-Ionen: 208  $m/z$  und 271  $m/z$ ; Malathion-Ionen: 99  $m/z$  und 127  $m/z$ ). Anschließend werden Wiederholungsinjektionen von Analyt-Stammlösungen und Lebensmittelextrakten durchgeführt. Die Injektionen von Analyt-Stammlösungen werden zur Untersuchung der Streuung von  $t_R$  und der Ionenhäufigkeiten  $A_1$  und  $A_2$  sowie der Korrelation der Ionenhäufigkeiten der einzelnen Analyten verwendet. Die Leistungsdaten wurden bei verschiedenen Analytkonzentrationen erfasst, da Wert und Streuung von  $A_1$  und  $A_2$  mit der Konzentration variieren. Tabelle E5.1 enthält eine Zusammenfassung der Leistungsparameter  $t_R$ ,  $A_1$  und  $A_2$ . Die Wiederholungsanalyse von Extrakten ohne nachweisbare Analytkonzentrationen wurde verwendet, um Modelle für die Streuung des Signalrauschens im Retentionszeitfenster zu definieren (Tabelle E5.1). Es wurden Signale von Lebensmittelextrakten untersucht, die für Lebensmittel pflanzlichen Ursprungs mit hohem Wassergehalt repräsentativ sind.

Aus den Daten in Tabelle E5.1 wurden Modelle für die  $t_R$ - und  $AR$ -Variabilität entwickelt.  $t_R$ -Modelle wurden aus Konfidenzintervallen auf der Grundlage der Studentschen  $t$ -Verteilung ( $\bar{t}_{Ri} \pm t_{S_{Ri}}$ , wobei  $\bar{t}_{Ri}$  und  $s_{Ri}$  der Mittelwert und die Standardabweichung von  $t_R$  und  $t$  der  $t$ -Wert der  $t$ -Verteilung für das festgelegte Konfidenzniveau und die Freiheitsgrade von  $\bar{t}_{Ri}$  und  $s_{Ri}$  sind) erstellt. Die  $AR$ -Modelle wurden anhand von Monte-Carlo-Simulationen erstellt. Aus der beobachteten Streuung der verschiedenen  $A_1$  und  $A_2$  des Analyten wurden Verhältnisse korrelierter  $A_1$  und  $A_2$  (d. h.  $AR$ ) simuliert. Aus Blindextrakten wurden das Signalrauschen und anschließend die  $AR$  in Blindproben simuliert. Das Signalrauschen wurde als eine bei null abgeschnittene Normalverteilung modelliert, da chromatographische Peaks keine negativen Flächen haben. Tabelle E5.2 zeigt die geschätzte Streuung von  $t_R$  und  $AR$ . Diese Tabelle enthält auch die MS-Excel-Formeln, die bei der Simulation von  $A_1$

und  $A_2$  verwendet wurden. Die Konfidenzgrenzen für  $t_R$  und  $AR$  wurden auf ein Konfidenzniveau von 99,9 % bzw. 98 % festgelegt, was der  $TP$  entspricht.

Die  $FP$  von Identifizierungen auf der Basis von  $t_R$  wurde auf der Grundlage der Erfahrung des Analytikers auf 10 % festgelegt. Diese  $FP$  stellt die Wahrscheinlichkeit dar, dass ein Peak, der nicht als vom Analyten stammend bestätigt wurde, innerhalb des festgelegten Retentionszeitfensters für den Analyten beobachtet wird. Die  $FP$  von  $AR$  wurde anhand von Monte-Carlo-Simulationen des Signalrauschens und der Bestimmung, wie oft das simulierte Rauschen  $AR$  innerhalb des Akzeptanzintervalls für diesen Parameter erzeugt, geschätzt. Da die  $FP$  bei niedrigen Analytkonzentrationen extrem groß sein kann, wurde sie bei verschiedenen Analyt-Massenanteilen bestimmt, indem eine Mindesthäufigkeit für jedes Ion festgelegt wurde. Tabelle E5.3 zeigt die geschätzten  $TP$ ,  $FP$  und ihre Kombination in  $LR(+)$ . In der letzten Spalte der Tabelle wird die Leistung der Identifizierungen auf der Grundlage von  $TP$  und  $AR$  angegeben. Tabelle E5.3 zeigt auch die A-posteriori-Wahrscheinlichkeit  $PP$ , dass ein Prüfgegenstand positiv ist, unter der Annahme, dass positive und negative Ergebnisse gleich wahrscheinlich sind (d. h.  $P(+)=P(-)=0,5$ ).

**Tabelle E5.1.** Leistungsparameter, die für die Identifizierung von Chlorpyrifos-Methyl und Malathion in Extrakten pflanzlichen Ursprungs mittels GC-MS/MS relevant sind. Alle Parameter wurden mit 11 Freiheitsgraden geschätzt.

Analyt: Chlorpyrifos-Methyl								
Extrakt	$w$ ( $\text{mg kg}^{-1}$ )	Retentionszeit, $t_R$		Häufigkeit				$p$
		$\bar{t}_{Ri}$ (min)	$s_{tRi}$ (min)	Ion: 208 $m/z$		Ion: 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	105668	13,3	138678	6,14	0,9956
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	$w$ ( $\text{mg kg}^{-1}$ )	Retentionszeit, $t_R$		Häufigkeit				$p$
		$\bar{t}_{Ri}$ (min)	$s_{tRi}$ (min)	Ion: 208 $m/z$		Ion: 271 $m/z$		
				$\bar{A}$ (a. u.)	$s_A$ (a. u.)	$\bar{A}$ (a. u.)	$s_A$ (a. u.)	
E§	3,33	19,45	0,070	226592	7,85	226765	10,3	0,9988
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	-

§ - Die Streuung der Ionenhäufigkeiten wurde durch Kombination von Signalen des Analyten in reinem Lösungsmittel mit Signalen von Pflanzenextrakten geschätzt. Extrakt-Matrix: G – Ingwer, O – Frühlingszwiebel, I – Irische Moosalgen; E – unspezifische Matrix  
 $w$  – Massenanteil des Analyten ( $\text{mg kg}^{-1}$ ),  $\bar{t}_{Ri}$  – mittlere Retentionszeit (min) (dieser Parameter kann je nach Tag der Injektion variieren);  $s_{tRi}$  – Standardabweichung der unter Wiederholbarkeitsbedingungen geschätzten Retentionszeit (min);  $\bar{A}$  – mittlere Ionenhäufigkeiten (willkürliche Einheiten, a.u.);  $s_A$  – Standardabweichung der Ionenhäufigkeit;  $p$  Spearmans Korrelationskoeffizient

**Tabelle E5.2.** Akzeptanzintervalle für Retentionszeit und Ionenhäufigkeitsverhältnis

Analyt	Extrakt	Intervall der Massenanteile, $w$ ( $\text{mg kg}^{-1}$ )	Maximale $\bar{t}_{Ri}$ - Differenz (min)(c. l. 99,9 %)§	AR-Intervall (c. l. 98 %)†
Chlorpyrifos-Methyl	E	0,04 – 3,33	0,18	0,439 – 1,18
Malathion	E	0,11 – 3,33	0,54	0,467 – 1,54

E – unspezifische Matrix; c. l. – Konfidenzniveau;

§ – Maximale Differenz zwischen der Retentionszeit des Analyten in einer Standardlösung und in der analysierten Probe ( $\sqrt{2ts_{tRi}}$ );

† – in der Simulation verwendete MS-Excel-Formel: Erstes Ion:  $A_1 = \bar{A}_1 + s_{A1} * TINV(R1, v_1)$  und Zweites Ion:  $A_2 = \bar{A}_2 + s_{A2} * (TINV(R1, v_2) * \rho + TINV(R2, v_2) * (1 - \rho^2)^{0,5})$ , wobei  $v_i$  die zu  $A_i$  und  $s_{Ai}$  gehörenden Freiheitsgrade und R1 und R2 zwei unabhängige Zufallswertgeneratoren U(0,1) sind (Excel-Formel ZUFALLSZAHL())

Nach Tabelle E5.3 werden Identifizierungen nur dann durch "sehr starke" Nachweise unterstützt (d. h.  $10^5 < LR(+) < 10^6$ ), wenn die Identifizierungen an oder über der Bestimmungsgrenze ( $0,14 \text{ mg kg}^{-1}$  bzw.  $0,38 \text{ mg kg}^{-1}$  für Chlorpyrifos-Methyl und Malathion) durchgeführt werden. Alle in Tabelle E5.3 dargestellten Leistungsmerkmale (*TP* und *FP*, *LR(+)* und *PP*) sind gültige Alternativen für die Angabe der Leistung oder Unsicherheit der qualitativen Analyse bei verschiedenen Analyt-Masseanteilen. Die Angabe von *LR(+)* hat den Vorteil, *TP* und *FP* in einer einzigen Kennzahl zusammenzufassen und keine Annahme einer Prävalenz des Pestizids in den analysierten Proben zu erfordern.

Tabelle E5.3. Leistungsmerkmale der Identifizierung von Chlorpyrifos-Methyl und Malathion durch GC-MS/MS.

Analyt	<i>w</i> (mg kg <sup>-1</sup> )	Leistungsmerkmale bei verschiedenen Analytkonzentrationen, <i>w</i>				
		<i>t<sub>R</sub></i>	<i>AR</i>	<i>t<sub>R</sub></i> & <i>AR</i>		
Chlorpyrifos-Methyl	<i>TP</i> (%)	≥ 0,04	99,9	98	97,9	
	<i>FP</i> (%)	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	
	<i>LR</i> (+)	0,04 §	9,99	3,24	32,4	
		0,08	9,99	490	4895	
		0,14 †	9,99	98000	979020	
		0,27	9,99	9800	929020	
	<i>PP</i> (%)	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	<i>TP</i> (%)	≥ 0,11	99,9	98	97,9
		<i>FP</i> (%)	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	
<i>LR</i> (+)		0,11	9,99	29,8	2,98	
		0,23	9,99	0,001	0,0001	
		0,38 †	9,99	0,001	0,0001	
		0,77	9,99	0,001	0,0001	
<i>PP</i> (%)		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	

§ – Nachweisgrenze; † – Bestimmungsgrenze; *t<sub>R</sub>* - Retentionszeit des Analyten; *AR* – Häufigkeitsverhältnis zweier charakteristischer Ionen des Massenspektrums des Analyten

Dieses Beispiel veranschaulicht, wie die Monte-Carlo-Simulation von Signalen die Schwierigkeit der experimentellen Bestimmung der Falsch-Positiv-Raten bei hochselektiven Identifizierungen überwinden kann.

## 7.6. E6: Identifizierung von SARS-CoV-2-RNA durch Nukleinsäure-Amplifikationstest

### Geltungsbereich:

---

**Art der qualitativen Analyse:** Analyse auf der Grundlage quantitativer Kriterien

---

**Gegenstand/Matrix:** Nasenabstriche, nasopharyngeale und oropharyngeale Abstriche

---

**Parameter/Analyt:** SARS-CoV-2 RNA

---

**Art des Klassifizierungskriteriums:** Zyklusschwelle, Ct, Werte, die gleich oder niedriger als der Ct-Grenzwert sind, werden als positiv eingestuft; höhere Werte werden als negativ eingestuft

---

**Technik/Geräteausstattung:** Reverse-Transkriptase-Polymerase-Kettenreaktion (RT-PCR)

---

**Art der Berichterstattung:** Wahr-Positiv-Rate und Wahr-Negativ-Rate.

---

Das schwere akute respiratorische Syndrom-Coronavirus 2 (SARS-CoV-2) ist das Virus, das COVID-19 verursacht, die für die COVID-19-Pandemie verantwortliche Atemwegserkrankung. Einer der Screening-Tests für das Vorhandensein dieses Virus in Nasenabstrichen umfasst eine Reverse-Transkriptase-Polymerase-Kettenreaktion (RT-PCR), eine Art von Nukleinsäure-Amplifikationstest (NAAT). Die "hausinterne" Validierung dieses Tests umfasst die Bestimmung des Zyklusschwellenwerts, des LOD, der (klinischen) Sensitivität (*SS*) und der Spezifität (*SP*) (Tabelle 2) sowie weiterer Leistungsparameter (in diesem Beispiel nicht dargestellt). In der klinischen Analyse werden *SS* und *SP* als Schätzwert für die klinische Genauigkeit verwendet. Die klinische Genauigkeit wird durch *FN* und *FP* sowie durch die epidemiologische Prävalenz, die Arten und Unterarten von Erregern, Mutationen und andere biologische Faktoren begrenzt.

### 7.6.1. Zyklusschwellenwert

Die "Anzahl der Zyklen, die benötigt werden, damit ein Amplikon über dem Untergrund nachweisbar ist", wird als Zyklusschwellenwert (Ct) definiert [59] - die Anzahl der Zyklen, die erforderlich sind, um virale RNA zu amplifizieren, um ein nachweisbares Niveau zu erreichen. Um die Anwendung des Ct-Zyklusschwellenwerts zu verstehen, sollten einige Variablen berücksichtigt werden.  $R_n$  (normalisiertes Reportersignal) ist die Größe des unter den gegebenen PCR-Bedingungen erzeugten Signals.  $\Delta R_n$  (Abbildung E6.1) ist das normalisierte Reportersignal abzüglich des Grundlinienniveaus. Der Schwellenwert ist der Signalpegel, der einen statistisch signifikanten Anstieg gegenüber dem berechneten Basisliniensignal darstellt (siehe Abbildung E6.1). Dieser Entscheidungsweg wird festgelegt, um relevante Verstärkungssignale vom Untergrund zu unterscheiden. In diesem Beispiel setzt die Software den Schwellenwert auf das 10-fache der Standardabweichung des Basislinien-Fluoreszenzwertes. Der Grenzwert wird im Bereich eines exponentiellen Wachstums des PCR-Produkts festgelegt. Abbildung E6.1 veranschaulicht die positive Klassifizierung (kleiner oder gleich dem Ct-Schwellenwert von 32) einer menschlichen Probe. Fluoreszenzergebnisse, die höher als 32 sind, werden als negativ eingestuft. Beachten Sie, dass nur sigmoidale Amplifikationskurven auf eine echte Amplifikation hinweisen (siehe Abbildung E6.1).

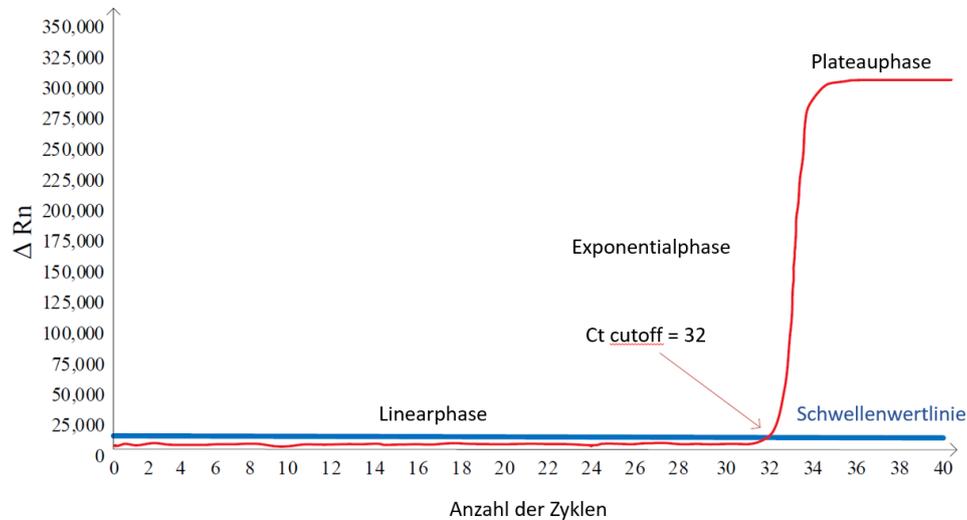


Abbildung E6.1 Nachweis von SARS-CoV-2-RNA aus einer wahr-positiven Probe.

### 7.6.2. LOD-Schätzung

Die Nachweisfähigkeit in der Nähe des Ct-Schwellenwerts wird anhand der *LOD* bewertet. Die *LOD* ist definiert als das Konzentrationsvielfache (z. B. ausgedrückt in Anzahl der Kopien/mL), das mit einer *TP* von 95 % verbunden ist ( $LOD_{95\%}$ ). Die  $LOD_{95\%}$  wird geschätzt, indem die Variation der *TP* mit der Konzentration modelliert und die Konzentration, bei der die *TP* 95 % beträgt, mittels Probit-Regression geschätzt wird [60] - [63]. In diesem Beispiel wird die  $LOD_{95\%}$  in einer Serie von sieben Verdünnungen einer Probe mit einer Konzentration von 500 Kopien/mL auf 114 Kopien/mL geschätzt. Weitere Einzelheiten zum vorgestellten Verfahren zur Bestimmung der  $LOD_{95\%}$  sind im Literaturverzeichnis (5.5 in [63]) zu finden.

### 7.6.3. Klinische Genauigkeit

Die Bewertung der klinischen Genauigkeit beinhaltet die Festlegung von Ziel- oder Mindestwerten für die unteren Grenzwerte des 95 %-CI von *SS* und *SP*,  $LL_{SS,95}^{tg}$  und  $LL_{SP,95}^{tg}$ , und zu prüfen, ob die geschätzten unteren Grenzwerte gleich oder höher als der jeweilige Zielwert sind (d. h. ob  $LL_{SS,95} \geq LL_{SS,95}^{tg}$  und  $LL_{SP,95} \geq LL_{SP,95}^{tg}$ ). Für diesen Test gilt  $LL_{SS,95}^{tg} = 95\%$  und  $LL_{SP,95}^{tg} = 90\%$ , d. h. der untere Grenzwert der Sensitivität sollte größer als 95 % und der untere Grenzwert der Spezifität sollte größer als 90 % sein.

Für die Verfahrensvalidierung wurden 200 Nasen-, Nasopharyngeal- und Oropharyngeal-Abstriche analysiert: 100 von Personen, die bekanntermaßen mit SARS-CoV-2 infiziert waren, und 100 von Personen, bei denen eine Infektion mit diesem Virus ausgeschlossen werden konnte. Tabelle E6.1 zeigt die Kontingenztabelle, die sich aus den 200 Tests ergibt. Aus dieser Tabelle geht hervor, dass keine falsch-negativen und drei falsch-positive Ergebnisse gemeldet wurden, so dass die klinische *SS* und *SP* 100 % bzw. 97 % betragen. Die Grenzwerte des 95 %-CI von *SS* und *SP*, berechnet aus Gl. (8) bis (11), betragen [96,3, 100] bzw. [91,6, 99,0] (Tabelle E6.2). Da  $LL_{SS,95}$  und  $LL_{SP,95}$  über 95 % bzw. 90 % liegen, wird das Analysenverfahren als gültig angesehen. *SS* wird durch die Serokonversionssensitivität ergänzt [64].

**Tabelle E6.1** Kontingenztafel, die die Leistung des Verfahrens zum Nachweis von SARS-Cov-2-RNA in Nasenabstrichen, nasopharyngealen und oropharyngealen Abstrichproben zusammenfasst.

		Fall		Ergebnis gesamt
		Positiv (pc)	Negativ (nc)	
Ergebnis	Positiv (p)	$tp = 100$	$fp = 3$	$p = 103$
	Negativ (n)	$fn = 0$	$tn = 97$	$n = 97$
Fall gesamt		$pc = 100$	$nc = 100$	

**Tabelle E6.2** Klinische Genauigkeit des Verfahrens zum Nachweis von SARS-Cov-2-RNA in Humanserum oder -plasma

	Klinische Sensitivität	
$SS = 100 \%$	$LL_{SS,95} = 96,3 \%$	$HL_{SS,95} = 100 \%$
	Klinische Spezifität	
$SP = 97 \%$	$LL_{SP,95} = 91,6 \%$	$HL_{SP,95} = 99,0 \%$

Absichtlich leer

## Anhang A - Satz von Bayes, Quoten und Likelihood Ratio

### A.1 Satz von Bayes

Der Satz von Bayes beschreibt, wie sich die Wahrscheinlichkeit eines Ereignisses  $A$  (z. B., dass ein Prüfgegenstand tatsächlich positiv ist) mit neuen Informationen  $E$  (z. B. einem positiven Prüfergebnis) ändert. Der Satz von Bayes wird meist für zwei Ereignisse,  $A$  und  $E$ , wie folgt geschrieben:

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

Dabei sind  $P(A)$  und  $P(E)$  die Wahrscheinlichkeiten des alleinigen Eintretens der Ereignisse  $A$  und  $E$ ,  $P(A|E)$  ist die Wahrscheinlichkeit des Ereignisses  $A$ , wenn  $E$  eingetreten ist, und  $P(E|A)$  ist die Wahrscheinlichkeit des Ereignisses  $E$ , wenn  $A$  eingetreten ist. In der Statistik werden  $P(A|E)$  und  $P(E|A)$  üblicherweise als "bedingte Wahrscheinlichkeiten" bezeichnet; beispielsweise kann  $P(E|A)$  als die bedingte Wahrscheinlichkeit des Ereignisses  $E$  unter der Voraussetzung bezeichnet werden, dass  $A$  eingetreten ist.

Im Zusammenhang mit der qualitativen Analyse, am Beispiel des positiven Falls, kann  $P(A)$  als die Wahrscheinlichkeit verstanden werden, dass ein zufällig ausgewählter Prüfgegenstand tatsächlich positiv ist, bevor irgendwelche Prüfungen durchgeführt werden.  $P(E|A)$  ist die Wahrscheinlichkeit, dass ein wahr positiver Prüfgegenstand ein positives Prüfergebnis erzeugt - die Wahr-Positiv-Rate  $TP$  in Tabelle 2.  $P(E)$  ist die Wahrscheinlichkeit eines positiven Prüfergebnisses unabhängig vom Zustand des Prüfgegenstands.  $P(A|E)$  schließlich ist die Wahrscheinlichkeit, dass der Prüfgegenstand wirklich positiv ist, *nachdem* die durch das positive Prüfergebnis hinzugefügte Information berücksichtigt wurde. Da sie berechnet wird, nachdem der Nachweis  $E$  verfügbar ist, wird  $P(A|E)$  gewöhnlich als "A-posteriori-Wahrscheinlichkeit" für  $A$  bezeichnet. Eine geschätzte A-posteriori-Wahrscheinlichkeit gibt einen direkten Hinweis auf das Vertrauen, das in eine Klassifizierung gesetzt werden kann.

Es ist wichtig, sich daran zu erinnern, dass (um beim positiven Fall zu bleiben)  $P(E)$  sowohl wahr positive Ergebnisse als auch falsch positive Ergebnisse einschließt und dass  $P(E)$  für die gesamte Population der Prüfgegenstände gilt. Das bedeutet, dass  $P(E)$  sowohl auf die wahren als auch auf die falschen Positivraten sowie auf die Anteile der tatsächlich positiven und negativen Prüfgegenstände sensitiv reagiert. Quantitativ lässt sich  $P(E)$  für zwei Fälle  $A$  und  $\neg A$  (für "Nicht- $A$ ", einen wahr negativen Prüfgegenstand) als gewichtete Summe schreiben:

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

Betrachtet man  $E$  als ein positives Testergebnis, so besagt Gleichung A.2, dass die kombinierte Wahrscheinlichkeit von  $E$  der Wahr-Positiv-Rate mal dem Anteil der tatsächlich positiven Proben plus der Falsch-Positiv-Rate mal dem Anteil der tatsächlich negativen Proben entspricht. Aus diesem Grund verringert eine hohe Falsch-Positiv-Rate das Vertrauen in positive Prüfergebnisse. Unter Bezugnahme auf Gleichung A.1 bedeutet dies, dass bei einer hohen Wahrscheinlichkeit positiver Ergebnisse aus wahr negativen Prüfgegenständen  $P(A|E)$  abnimmt, weil  $P(E)$  zunimmt. Dies entspricht der Intuition: Wie hoch die Rate der wahr positiven Ergebnisse auch sein mag, die Wahrscheinlichkeit einer großen Anzahl von falsch-positiven Ergebnissen sollte uns weniger sicher machen, dass ein positives Ergebnis auf einen wirklich positiven Prüfgegenstand hinweist.

### A.2 Wahrscheinlichkeit und Quoten.

Eine Wahrscheinlichkeit wird in der Regel als Zahl zwischen 0 und 1 ausgedrückt. Sie kann aber auch in Form von "Quoten" ausgedrückt werden, ein Begriff, der vielleicht am bekanntesten aus dem Bereich Sportwetten ist. Wenn die Wahrscheinlichkeit eines Ereignisses  $A$  gleich  $P(A)$  und die alternative Möglichkeit einfach "Nicht  $A$ " ist, kann die Quote  $O(A)$  zugunsten von  $A$  berechnet werden mit Hilfe:

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

Im Gegensatz zu Wahrscheinlichkeiten können Quoten jeden nicht-negativen Wert annehmen; Quoten von  $10^6$  oder "eine Million zu eins" sind möglich. Quoten können in Wahrscheinlichkeiten umgewandelt werden, indem man A.3 umformt zu:

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

### A.3 Die Quotenform des Satzes von Bayes und die Likelihood Ratio

Wenn es nur zwei alternative und komplementäre Hypothesen gibt,  $A$  und  $\neg A$  (d. h. "Nicht- $A$ "), und ein Hinweis  $E$  (z. B. ein positives Prüfergebnis für  $A$ ) verwendet wird, um die Wahrscheinlichkeiten der einzelnen Hypothesen zu aktualisieren, ergeben sich nach dem Satz von Bayes die A-posteriori-Wahrscheinlichkeiten wie folgt:

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

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

Das Verhältnis ihrer Wahrscheinlichkeiten ist dann:

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

oder um die Begriffe aus Gründen der Klarheit zu trennen:

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

Da es nur zwei Hypothesen gibt,  $A$  und  $\neg A$ , müssen sich die A-priori- und die A-posteriori-Wahrscheinlichkeiten zu 1 summieren, d.h.  $P(\neg A) = 1 - P(A)$  und  $P(\neg A|E) = 1 - P(A|E)$ . Dies bedeutet, dass die linke Seite von A.6 gleich  $P(A|E)/[1 - P(A|E)]$  ist. Im Vergleich zu A.3 ist dies nur die Quote zugunsten von  $A$ , wenn  $E$  gegeben ist, oder  $O(A|E)$ , die "A-posteriori-Quote" zugunsten der Hypothese  $A$ . In ähnlicher Weise erscheint die A-priori-Quote  $O(A)$  auf der rechten Seite von A.6 als  $P(A)/P(\neg A) = P(A)/[1 - P(A)] = O(A)$ . Das verbleibende Verhältnis,  $P(E|A)/P(E|\neg A)$ , ist bekanntlich die „Likelihood Ratio“. Für den Fall der qualitativen Analyse, in dem  $\neg A$  einem tatsächlich negativen Prüfgegenstand entspricht, gibt Tabelle 5 die (geschätzte) Likelihood Ratio als  $TP/FP$  an. Die Quotenform des Satzes von Bayes kann daher wie folgt geschrieben werden

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

oder schematisch

$$\text{A-posteriori-Quote} = \text{A-priori-Quote} \times \text{Likelihood Ratio}$$

Die Likelihood Ratio kann daher quantitativ als die Veränderung der Quote zugunsten einer bestimmten Hypothese interpretiert werden.

## Anhang B – Qualitative Analyse im Zusammenhang mit der Bewertung der Konformität mit einem quantitativen Grenzwert

### B 1 Konformitätsbewertung als qualitative Analyse

Die Bewertung der Übereinstimmung des Wertes eines quantitativen Parameters des analysierten Gegenstands mit einem Grenzwert oder Intervall kann als qualitative Analyse unter Verwendung eines einzigen quantitativen Kriteriums (Abschnitt 2) mit den Ergebnissen "konform" oder "nicht konform" betrachtet werden. Tabelle B.1 enthält einige Beispiele für diese Art von Analysen.

**Tabelle B.1.** Beispiele für qualitative Analysen, die auf der Bewertung der Übereinstimmung des Wertes eines quantitativen Parameters des analysierten Gegenstandes mit einem Grenzwert oder Intervall beruhen.

- 
- (1) Bewertung der Farbe von Rohstoff durch Vergleich der Absorptionsmessungen mit einem Schwellenwert.
  - (2) Bewertung der Konformität einer Legierung mit einem Mindestgehalt für ihren Hauptbestandteil.
  - (3) Bewertung der Konformität eines Arzneimittels mit dem Spezifikationsintervall für die Konzentration des Wirkstoffs.
  - (4) Bewertung der Konformität eines Pestizidrückstandes in Obst angesichts eines Rückstandshöchstgehalts
  - (5) Bewertung des Gesundheitszustands einer Person durch Vergleich eines gemessenen Blutbestandteils mit einem Intervall von Werten gesunder Personen.
- 

Die Verwendung von Entscheidungsregeln und Messunsicherheit bei der Konformitätsbewertung wird im Eurachem/CITAC-Leitfaden "Use of uncertainty information in compliance assessment" [29] ("The compliance guide") ausführlich behandelt. Der Vollständigkeit halber wird in diesem Anhang jedoch erörtert, wie Unsicherheits- oder Leistungsdaten für quantitative Analysen verwendet werden können, um einige der Kennzahlen in Tabelle 2 abzuleiten. Diese können dann verwendet werden, um die Leistung qualitativer Analyseverfahren zu charakterisieren, die ganz oder teilweise auf dem Vergleich von Messergebnissen mit einem Grenzwert oder einer Spezifikation beruhen.

Wenn die Analyse die Bewertung beinhaltet, ob eine gemessene Eigenschaft oberhalb, unterhalb oder innerhalb einer Spezifikationsgrenze oder eines Intervalls liegt, kann die Messunsicherheit zur Quantifizierung der Zuverlässigkeit der Konformitätsbewertung verwendet werden.

**ANMERKUNG:** In diesem Leitfaden wird nicht erörtert, wie die Messunsicherheit zu bewerten ist. Die Bewertung der Messunsicherheit wird ausführlich im Eurachem/CITAC-Leitfaden "Quantifying uncertainty in analytical measurement" [65] beschrieben.

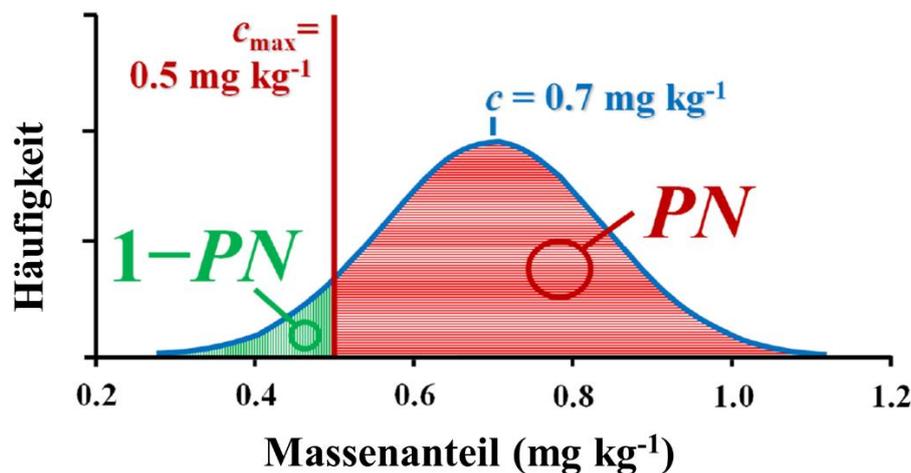
Die Verwendung der Messunsicherheit für Konformitätsentscheidungen, wie sie im Eurachem/CITAC-Konformitätsleitfaden [29] beschrieben wird, beinhaltet die Festlegung eines Kriteriums für die Entscheidung, ob ein Gegenstand konform oder nicht konform ist, bei einer maximalen Wahrscheinlichkeit falscher Konformitätsentscheidungen von  $x$  %. Der Konformitätsleitfaden unterscheidet zwischen "spezifischen" und "globalen" Risiken. Das "spezifische Risiko" quantifiziert die Wahrscheinlichkeit einer falschen Entscheidung über die Konformität eines bestimmten Gegenstands; es basiert ausschließlich auf der Verteilung, die mit dem Messergebnis für diesen Gegenstand verbunden ist. Im Gegensatz dazu quantifiziert das "globale Risiko" die Wahrscheinlichkeit von Fehlentscheidungen über die Konformität eines zufällig ausgewählten zukünftigen Gegenstands [66]. Das globale Risiko berücksichtigt die Verteilung der möglichen Werte für die gemessenen Gegenstände, wie z. B. die Verteilung der Werte von Gegenständen aus einer Produktionslinie oder einem Umweltbereich. Die Berechnung des globalen Herstellerrisikos erfordert beispielsweise die Wahrscheinlichkeit, dass eine Produktionslinie Produkte mit einem Wert nahe dem Grenzwert produziert, so dass sie fälschlicherweise als nicht konform angesehen werden können.

Daher muss für die Bestimmung des globalen Risikos die Verteilung der Werte für die Grundgesamtheit der Gegenstände gut charakterisiert sein.

In den meisten Fällen sind Analytiker daran interessiert, die Konformität eines bestimmten analysierten Gegenstands zu bewerten. Wie können in solchen Fällen Kennzahlen zur Quantifizierung der Zuverlässigkeit anderer Arten von qualitativen Analysen bestimmt werden? Im Folgenden werden eine Fallstudie und die zur Bestimmung dieser Parameter verwendeten Formeln vorgestellt.

## B 2 Positive und negative Ergebnisse der Konformitätsbewertung

Wird der Gegenstand aufgrund eines "positiven Ergebnisses" als konform angesehen, kann die Verteilung, die die Messunsicherheit beschreibt, verwendet werden, um entweder eine Likelihood Ratio  $LR(+)$  zugunsten der Konformität anzugeben oder, unter bestimmten Umständen, die A-posteriori-Wahrscheinlichkeit  $PP$ , dass der Gegenstand tatsächlich konform ist ("positiver Fall"), und die Wahrscheinlichkeit, dass der Gegenstand tatsächlich nicht konform ist ("negativer Fall")  $(1 - PP)$ . Entsprechend kann, wenn ein Ergebnis im Vergleich zu einem Grenzwert oder Intervall als "nicht konform" oder "negativ" gemeldet wird, die Messunsicherheit verwendet werden, um die entsprechende Likelihood Ratio  $LR(-)$  oder die A-posteriori-Wahrscheinlichkeiten  $PN$  und  $(1 - PN)$  zu erhalten. Im folgenden Beispiel werden die A-posteriori-Wahrscheinlichkeiten betrachtet.



**Abbildung B.1** Grafische Darstellung der Wahrscheinlichkeit  $PN$  dafür, dass ein analysierter Gegenstand in Bezug auf einen oberen Grenzwert,  $c_{\max}$ , nicht konform ist, ausgehend von einem Messwert,  $c$ , mit zugehöriger Standardunsicherheit,  $u(c) = 0,14 \text{ mg kg}^{-1}$ , und der entsprechenden Wahrscheinlichkeit  $1 - PN$ , dass er konform ist<sup>7</sup>.

Anmerkung: Die Definition von "positiven Ergebnissen" und "negativen Ergebnissen" als "konform" bzw. "nicht konform" ist willkürlich; die umgekehrte Vereinbarung kann befolgt werden.

## B.3 Beispiel - Konformitätsbewertung für Pestizidrückstände in Obst

Angenommen, die Konformität einer Weintraubenprobe wird anhand eines Rückstandshöchstgehalts,  $c_{\max}$ , von  $0,5 \text{ mg kg}^{-1}$  für Acetamiprid [67] bewertet, und der gemessene Massenanteil in der Probe beträgt  $0,70 \text{ mg kg}^{-1}$ ,  $c$ .<sup>8</sup> Das Messergebnis hat eine Normalverteilung mit einer Standardunsicherheit,  $u(c)$ , von  $0,14 \text{ mg kg}^{-1}$ , die mit einer sehr hohen Anzahl von Freiheitsgraden geschätzt wurde. Da  $c >$

<sup>7</sup> Anm. d. Übers.: Fehler in der englischen Originalfassung korrigiert.

<sup>8</sup> Das Symbol  $c$  für Konzentration wird in diesem Leitfaden für Fälle verwendet, die sich auf verschiedene Arten von Messgrößen wie Massenkonzentration, Massenanteil und pH-Wert beziehen.

$c_{\max}$ , ist die wahrscheinlichste Schlussfolgerung bezüglich der Konformität die "Nicht-Konformität" der Trauben (ein "negatives Ergebnis").

Formal setzt eine A-posteriori-Wahrscheinlichkeit wie  $PP$  oder  $PN$  eine A-priori-Wahrscheinlichkeit voraus. In diesem Fall gibt es keine Information über die allgemeine Verteilung von Acetamid in Weintrauben. Bei einem Verdacht auf Kontamination ist es jedoch oft sinnvoll anzunehmen, dass die Verteilung so breit ist, dass sie im Bereich des Messergebnisses im Wesentlichen nicht aussagekräftig ist. Wenn dies der Fall ist, kann die Messunsicherheit als Näherung an die A-posteriori-Verteilung betrachtet werden. Nach diesem Ansatz ist die A-posteriori-Wahrscheinlichkeit einer negativen (nicht konformen) Stichprobe,  $PN$ , die Fläche unter der Wahrscheinlichkeitsdichtefunktion des Messergebnisses rechts von  $c_{\max}$ , dargestellt in Abbildung B.1 in Rot. Die Fläche ist der obere Wahrscheinlichkeitsbereich einer Normalverteilung mit Mittelwert  $c$  und Standardabweichung gleich  $u(c)$ . Dies kann mit einer Tabellenkalkulation oder einem Statistikpaket berechnet werden; in Microsoft Excel lautet die erforderliche Formel beispielsweise  $1 - \text{NORM.VERT}(c_{\max} = 0,5, c = 0,7, u(c) = 0,14, \text{WAHR})$  (siehe Tabelle B.2). Für dieses Beispiel beträgt die Fläche 0,923, also 92,3 %. Die entsprechende Wahrscheinlichkeit  $PP$ , dass die Stichprobe positiv (konform) ist, beträgt  $(1 - PN)$  oder 7,6 %.

Wenn aussagekräftige A-priori-Informationen verfügbar sind und es angemessen ist, diese zu nutzen, beinhalten die Berechnungen eine Integration über die A-priori-Verteilung. Integrale für normalverteilte A-priori-Verteilungen und Messunsicherheiten sind z. B. in JCGM 106 [66] enthalten, zusammen mit Hinweisen zu anderen Verteilungen.

**ANMERKUNG** Es kann überraschend schwierig sein, eine wirklich wenig informative A-priori-Verteilung zu erstellen. Zum Beispiel würde eine einfache gleichmäßige Verteilung über den Bereich von Abbildung B.1 ( $0,2 - 1,2 \text{ mg kg}^{-1}$ ) einer 30-prozentigen A-priori-Wahrscheinlichkeit entsprechen, dass ein Prüfgegenstand den Grenzwert einhält, einfach weil nur 30 % dieses Bereichs unter dem Grenzwert von  $0,5 \text{ mg kg}^{-1}$  liegen. Eine vollständige Bayes'sche Analyse umfasst daher eine Überprüfung, die in der Regel alternative Auswahlmöglichkeiten einer A-priori-Verteilung beinhaltet, um sicherzustellen, dass die Schlussfolgerung nicht übermäßig anfällig für die angenommene A-priori-Verteilung ist.

## B.4 Tabellenkalkulationsformeln für Wahrscheinlichkeiten der Konformitätsbewertung

Tabelle B.2 enthält die MS-Excel-Formeln, die verwendet werden sollten, wenn verschiedene Konformitätsgrenzen oder -intervalle berücksichtigt werden und der gemessene Wert unter, über, innerhalb oder außerhalb der Grenze(n) liegt.

Wenn die Standardunsicherheit  $u(c)$  mit einer kleinen Anzahl von Freiheitsgraden  $\nu$  geschätzt wird, sollte, anstatt die Streuung der Schätzung der Messgröße durch eine Normalverteilung zu beschreiben, die Student- $t$ -Verteilung in Betracht gezogen werden. In diesem Fall sollte die in Tabelle B.2 verwendete allgemeine Formel  $\text{NORM.VERT}(C, c, u(c), \text{WAHR})$  durch  $\text{TVERT}(\text{ABS}(C - c)/u(c), \nu, \text{WAHR})$  ersetzt werden.

**Tabelle B.2** MS Excel-Formeln zur Abschätzung der Wahrscheinlichkeit einer Konformitäts- ( $PP$ ) oder Nicht-Konformitäts-Entscheidung ( $PN$ ) für einen bestimmten analysierten Gegenstand. Die Formeln können verwendet werden, um  $LR(+)=PP/(1-PP)$  und  $LR(-)=PN/(1-PN)$  zu berechnen.

S	Grenzwert	Konformität des Gegenstands ( <i>Ergebnisart</i> ) Szenario	Konformität Zuverlässigkeit	MS-Excel-Formel (basierend auf der kumulativen Normalverteilung)
1	Max.	konform ( <i>positiv</i> ) $c \leq c_{\max}$	$PP$	$NORM.VERT(c_{\max}, c, u(c), WAHR)$
2	Max.	nicht konform ( <i>negativ</i> ) $c > c_{\max}$	$PN$	$1 - NORM.VERT(c_{\max}, c, u(c), WAHR)$
3	Min.	konform ( <i>positiv</i> ) $c \geq c_{\min}$	$PP$	$1 - NORM.VERT(c_{\min}, c, u(c), WAHR)$
4	Min.	nicht konform ( <i>negativ</i> ) $c < c_{\min}$	$PN$	$NORM.VERT(c_{\min}, c, u(c), WAHR)$
5	Inter.	konform ( <i>positiv</i> ) $c_{\min} \leq c \leq c_{\max}$	$PP$	$NORM.VERT(c_{\max}, c, u(c), WAHR) - NORM.VERT(c_{\min}, c, u(c), WAHR)$
6	Inter.	nicht konform ( <i>negativ</i> ) $c > c_{\max}$ or $c < c_{\min}$	$PN$	$1 - NORM.VERT(c_{\max}, c, u(c), WAHR) + NORM.VERT(c_{\min}, c, u(c), WAHR)$

S – Szenario; Inter., Max. or Min. – Intervall, maximaler oder minimaler Grenzwert; positives oder negatives Ergebnis – ein konformes oder nicht konformes Ergebnis;  $c$  und  $u(c)$  – gemessene Konzentration und zugehörige Standardunsicherheit;  $c_{\max}$  oder  $c_{\min}$  – maximal oder minimal zulässige Konzentration

Absichtlich leer

## Literaturverzeichnis

- [1] JCGM, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (3rd edn.) (JCGM 200:2012), Sevres: BIPM, 2012.
- [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.
- [13] ISO, IEC, General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025), Geneva: ISO, 2017.
- [14] ISO, Medical laboratories – Requirements for quality and competence (ISO 15189), Geneva: ISO, 2012
- [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.
- [26] ISO, IEC, General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025)(superseded), Geneva: ISO, 1999.
- [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.
- [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.
- [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.
- [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.
- [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.

Absichtlich leer

