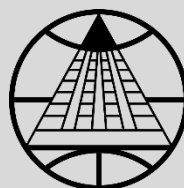

Eurachem 



CITAC
Cooperation on International
Traceability in Analytical Chemistry

Руководство Еврахим / СИТАК

**Оценка
эффективности и
неопределенности в
качественном химическом
анализе**

INTENTIONALLY BLANK

*Я лично видел что-то похожее на
животное, но я не могу быть
абсолютно уверен, что это не минерал.
Я думаю, что речь шла скорее об
энергии, чем о материи. Условно говоря,
проще всего было бы описать все это
как явление, парящее где-то на границе
размеров и обозначений, на стыке
цвета, формы, запаха, массы, длины и
ширины, контуров, теней, темноты и
т. д. и т. п.*

Славомир Мрожек, «Стриптиз»

INTENTIONALLY BLANK

Оценка эффективности и неопределенности в качественном химическом анализе

Первое издание (2021)

Редакторы

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

Состав рабочей группы*

Eurachem members

R. Bettencourt da Silva (Chair)	<i>Univ. Lisboa, Portugal</i>
S. Ellison (Secretary)	<i>LGC, United Kingdom</i>
A. Togola	<i>BRGM, France</i>
D. Ivanova	<i>Eurachem Bulgaria</i>
E. Theodorsson	<i>LIU, Sweden</i>
E. Totu	<i>University 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. Levbar	<i>Ukrmetrteststandart, 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>Montanuniversitaet Leoben, Austria</i>

CITAC members

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

*At time of document approval

Благодарности

Этот документ был подготовлен объединенной рабочей группой Еврахим/СИТАК в указанном составе (справа).

Редакторы выражают благодарность всем специалистам и организациям за комментарии, советы и содействие.

Выпуск этого Руководства частично поддержан Fundação para a Ciência e a Tecnologia, Portugal.

Данную публикацию следует цитировать как:

“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>”

* Subject to journal requirement.

Assessment of performance and uncertainty in qualitative chemical analysis

English edition

First Edition (2021)

Copyright © 2021

Copyright in this document is held by the contributing authors. All enquiries regarding reproduction in any medium, including translation, should be directed to the Eurachem secretariat.

Оценка эффективности и неопределенности в качественном химическом анализе

Перевод с английского Е.В. Мурашовой и И.В. Болдырева
Компьютерная верстка Е.В. Мурашовой

Ассоциация аналитических центров «Аналитика»

Под редакцией Р.Л. Кадиса, Ph.D., и М.С. Вонского, канд. биол. наук
Всероссийский научно-исследовательский институт метрологии
им. Д.И. Менделеева, Санкт-Петербург

Содержание

Предисловие	1
Область применения	2
1 Введение	5
2 Виды качественного анализа	7
3 Оценка эффективности качественного анализа	8
3.1 Общие соображения	8
3.2 Количественная оценка эффективности качественного анализа	8
3.3 Оценивание доли ложноположительных и ложноотрицательных результатов	12
3.4 Предел обнаружения и селективность	17
4 Выражение достоверности в качественном анализе	18
4.1 Общие соображения	18
4.2 Отношение правдоподобия	18
4.3 Апостериорная вероятность	21
4.4 Надежность показателей	22
4.5 Неопределенность долей	22
5 Представление результата качественного анализа	24
6 Выводы и рекомендации	25
7 Примеры	26
7.1 E1: Идентификация соединений с помощью масс-спектрометрии низкого разрешения с использованием поиска в базе данных или наличия характеристических ионов	26
7.2 E2: Идентификация очищенных соединений с помощью инфракрасной спектрометрии	31
7.3 E3: Идентификация наркотических веществ в моче с помощью метода иммуноферментного анализа (EMIT) и альтернативного метода	33
7.4 E4: Идентификация гена SRY человека в биологическом материале с помощью количественной ПЦР (qPCR)	35
7.5 E5: Идентификация остатков пестицидов в пищевых продуктах с помощью ГХ-МС/МС на основе времени удерживания и отношения интенсивностей ионов	37
7.6 E6: Идентификация РНК SARS-CoV-2 методом амплификации нуклеиновых кислот	41
Приложение А – Теорема Байеса, шансы и отношение правдоподобия	44
A.1 Теорема Байеса	44
A.2 Вероятность и шансы	44
A.3 Форма шансов теоремы Байеса и отношение правдоподобия	45
Приложение В - Качественный анализ, связанный с оценкой соответствия количественному пределу	46
Библиография	51

INTENTIONALLY BLANK

Предисловие

Проблема оценки и выражения неопределенности в качественном химическом анализе получила гораздо меньшее освещение в литературе, чем проблема неопределенности в количественном анализе (т. е. измерениях) [1]. Хотя некоторые авторы обращались к этой области [2] – [11], общие рекомендации по оценке эффективности качественного анализа или оценке и представлению неопределенности в качественном анализе недостаточны.*

В настоящее время нет ожидания, что аккредитованные лаборатории будут оценивать или сообщать о неопределенностях, связанных с результатами качественного анализа [12]. Тем не менее, ISO/IEC 17025 [13] и ISO 15189 [14] требуют, чтобы лаборатории могли получать достоверные качественные и количественные результаты анализа. Для лабораторий также крайне важно знать о надежности результатов качественного анализа; это позволяет им, при необходимости, предупреждать об превышении нормативов при интерпретации результатов и точно отвечать на запросы клиентов о достоверности. Количественная оценка достоверности результата качественного анализа особенно полезна, когда более вероятны ошибочные результаты. Настоящее Руководство предназначено для использования в тех случаях, когда желательна количественная оценка достоверности результата качественного анализа.

Это Руководство основано на опыте нескольких областей аналитики, где часто используется качественный анализ, например, в судебной [15] и клинической областях [16] – [18], а также на обширном общем руководстве [7].

* Предмет данного Руководства выходит далеко за рамки «качественного химического анализа» в традиционном понимании, когда задача сводилась к обнаружению того или иного компонента пробы. В данном случае речь идет об определении *качественных свойств* объекта анализа – группы свойств, которые не могут быть выражены размером и которые в международной метрологической терминологии получили наименование *nominal properties*. В отличие от измерений *величин* определение качественных свойств приводит не к числовому результату, а к утверждению, основанному на определенной классификации объектов (по качественному или количественному критерию). Целью этого Руководства является количественное выражение *достоверности* такой классификации. Это достигается с помощью ряда показателей (показатели эффективности, или характеристики эффективности), которые в конечном итоге отражают вероятность того, что принимаемая классификация верна или ошибочна. Эти характеристики эффективности и служат здесь для выражения «неопределенности результатов». Только в таком понимании можно говорить об «оценке и представлении неопределенности в качественном анализе». Оценивание неопределенности в строгом смысле, следуя принципам основополагающего *Руководства по выражению неопределенности в измерениях* (GUM), в качественном анализе невозможно. (Прим. редактора)

Область применения

Настоящее Руководство предназначено для помощи лабораториям в разработке и внедрении соответствующей методологии для оценки эффективности методов качественного анализа и оценки неопределенностей в качественном химическом анализе.

В настоящем Руководстве качественный анализ определяется как «*Классификация в соответствии с заданными критериями*». Для аналитической химии и смежных дисциплин под «критериями» понимается, как правило, информация, необходимая для определения химического состава, свойств и/или структуры анализируемых объектов.

В настоящем Руководстве рассматриваются следующие типы критериев:

- Количественный критерий, в котором числовой результат используется для отнесения объекта испытаний к заранее установленному классу;
- Качественный критерий, такой как наличие или отсутствие того или иного признака, изменение цвета и т. д.

Настоящее Руководство не является исчерпывающим при описании доступных средств для оценки эффективности методов качественного анализа и неопределенности результатов качественного анализа. Характеристики эффективности, представленные в данном Руководстве, основаны на измеренных или оцененных уровнях ошибочных результатов и не учитывают, например, мер соответствия между качественными методами или трактовку классификации по порядковым шкалам¹, кроме как правильную или неправильную классификацию.

¹ Порядковая шкала — это шкала естественных упорядоченных категорий, где расстояние между категориями неизвестно. Шкала Мооса — это порядковая шкала твердости минералов.

Аббревиатуры и символы

В данном Руководстве используются следующие аббревиатуры и символы. Символы, используемые в этом документе, не гармонизированы во всех областях науки, где они применяются. Например, в медицинской лаборатории *FN* и *FNR* являются аббревиатурами для «числа ложноотрицательных результатов» и «доли ложноотрицательных результатов» соответственно.

<i>A</i>	Интенсивность иона в масс-спектре	$LL_{RR.95}$	Нижний предел 95 % доверительного интервала для доли результатов <i>RR</i> (например, <i>SS</i>)
\bar{A}	Средняя интенсивность иона в масс-спектре		
<i>AR</i>	Отношение интенсивности ионов в масс-спектре	$LL_{RR.95}^{tg}$	Целевое или минимальное значение для $LL_{RR.95}$
<i>c</i>	Измеренная концентрация (или любая другая величина как свойство объекта анализа)	LOD	Предел обнаружения
<i>CI</i>	Доверительный интервал	LOQ	Предел количественного определения
c_{max}	Максимально допустимая концентрация	<i>LR</i>	Отношение правдоподобия
c_{min}	Минимально допустимая концентрация	<i>LR(+)</i>	Отношение правдоподобия положительных результатов
<i>DOR</i>	Диагностическое отношение шансов	<i>LR(-)</i>	Отношение правдоподобия отрицательных результатов
<i>E</i>	Эффективность	<i>n</i>	Число отрицательных результатов
<i>fn</i>	Число ложноотрицательных результатов	<i>nc</i>	Число отрицательных случаев
<i>FN</i>	Доля ложноотрицательных результатов по отношению к положительным случаям	<i>NPV</i>	Отрицательное прогностическое значение
<i>fp</i>	Число ложноположительных результатов	<i>O(·)</i>	Шансы в пользу события, например, <i>O(A)</i> означает шанс события <i>A</i>
<i>FP</i>	Доля ложноположительных результатов по отношению к отрицательным случаям	<i>p</i>	Число положительных результатов
<i>GC-MS</i>	Газовая хроматография с масс-спектрометрическим детектированием	<i>P(·)</i>	Вероятность события; например, <i>P(A)</i> это вероятность события <i>A</i>
<i>GC-MS/MS</i>	Тандемная газовая хромато-масс-спектрометрия (ГХ-МС/МС)	<i>P(+)</i>	Априорная вероятность положительного случая
<i>GUM</i>	Руководство по выражению неопределенности в измерениях	<i>P(-)</i>	Априорная вероятность отрицательного случая
$HL_{RR.95}$	Верхний предел 95 % доверительного интервала для доли результатов <i>RR</i> (например, <i>SS</i>)	<i>pc</i>	Число положительных случаев
<i>LC-MS</i>	Жидкостная хроматография с масс-спектрометрическим детектированием	<i>PN</i>	Апостериорная вероятность отрицательного случая (см. Приложение А)
		<i>PP</i>	Апостериорная вероятность положительного случая (см. Приложение А)
		<i>PPV</i>	Положительное прогностическое значение
		qPCR	Количественная полимеразная цепная реакция

<i>RT-PCR</i>	Полимеразная цепная реакция с обратной транскрипцией	<i>tp</i>	Число истинно положительных результатов
<i>RA</i>	Относительная интенсивность	<i>TP</i>	Доля истинно положительных случаев по отношению к положительным случаям
<i>Rn</i>	Нормализованная интенсивность		
<i>RR</i>	Уровень результата (доля)	<i>t_R</i>	Время удерживания
<i>s_A</i>	Стандартное отклонение интенсивности иона	\bar{t}_{Ri}	Среднее время удерживания
<i>SP</i>	Специфичность	<i>u(c)</i>	Стандартная неопределенность <i>c</i>
<i>SS</i>	Чувствительность	<i>w</i>	Массовая доля
<i>s_{Ri}</i>	Стандартное отклонение времени удерживания	<i>Y</i>	Индекс Юдена
<i>tn</i>	Число истинно отрицательных результатов	ΔRn	Значение нормализованной интенсивности за вычетом фонового сигнала
<i>TN</i>	Доля истинно отрицательных случаев по отношению к отрицательным случаям	ρ	Коэффициент корреляции Спирмена между парами интенсивностей ионов

1 Введение

Многие важные области социально-экономических интересов или интересов отдельных граждан, такие как промышленная производительность и состояние здоровья, зависят от химического анализа. Некоторые из этих анализов являются исключительно качественными или предполагают последующую количественную оценку идентифицированного химического соединения. Интересы, которые должны быть защищены этими анализами, сохраняются только в том случае, если аналитическое качество соответствует предполагаемому использованию.

В некоторых публикациях для «качественного анализа» используются термины «исследование» («examination») [1], «исследование качественных свойств» [19] или «испытание» и «тестирование». Международный стандарт по аккредитации медицинских лабораторий использует термин «исследование» как для количественного, так и для качественного анализа [14]. Поэтому, поскольку среди различных соответствующих международных сообществ не было достигнуто консенсуса относительно этих терминов, в настоящем Руководстве для определения качественных свойств в химическом анализе используется термин «качественный анализ».*

В широком смысле качественный аналитический результат — это простое утверждение или категоризация объекта испытаний или материала, т. е. классификация. Решения неизменно принимаются на основе категоризации; например, следует ли выпустить партию удобрения, пригодна ли вода для питья, есть ли у человека вещество, которое подлежит контролю, или нет, или имеет ли вновь синтезированный материал правильную

структуру, отвечающую определенным требованиям. Неправильная классификация — например, «принятие» продукта, когда он непригоден для использования, — влечет за собой риски для всех сторон. Чтобы контролировать эти риски, профессионалы, занимающиеся анализом, усердно работают над тем, чтобы их методики приводили к приемлемо низким рискам неправильной классификации.

Из этого следует, что на каком-то этапе разработки любой такой методики испытаний необходимо провести оценку риска неправильной классификации. Следовательно, для большинства таких методик разумно ожидать, что лаборатория установит или получит доступ к информации о рисках получения неверных результатов. Важным исключением является использование стандартизированных методик испытаний, установленных группами специалистов вне данной лаборатории и пригодными для предполагаемой цели [20] – [21], [22], [23]. Лаборатория может иметь ограниченный доступ или даже не иметь доступа к данным об эффективности таких методик испытаний. Однако эти методики неизменно определяют испытание с соответствующими деталями, и обычно ожидается, что лаборатория продемонстрирует, что соответствующие факторы, находящиеся под ее контролем, действительно соответствуют требованиям методики испытаний. Это, в свою очередь, может включать в себя демонстрацию того, что неопределенность контролируемых параметров и эффективность адекватны цели испытаний.

* В настоящее время отсутствует единая терминология, относящаяся к определению качественных свойств, хотя на международном уровне рекомендован общий термин *examination of nominal properties*. В различных областях деятельности соответствующие операции называют по-разному: *испытание, тестирование, анализ, исследование*. Это затрудняет выбор подходящих русских эквивалентов при переводе. В англоязычном тексте Руководства широко используются слова *test* и *testing*, при переводе которых мы отдавали предпочтение терминологии испытаний: «объект испытаний» («test item»), «методика испытаний» («test procedure»), «результат испытаний» («test result»). В то же время применительно к тем областям, где принято «выполнение тестов» и «тестирование» (биомедицинские лаборатории), использован именно такой вариант перевода *test* и *testing*. В целом же, поскольку документ так или иначе сконцентрирован на вопросах определения химического (биохимического) состава и свойств, традиционный термин «анализ» часто оказывался наиболее подходящим. (Прим. редактора)

Оценка неопределенностей, связанных с количественными параметрами или результатами анализа, была предметом значительных усилий с момента публикации «Руководства по выражению неопределенности в измерениях» (GUM), которое доступно в виде Руководства ISO 98 [25], а также документа JCGM [24]. С другой стороны, неопределенности результатов в качественном анализе уделялось гораздо меньше внимания. После публикации первого издания ISO/IEC 17025 [26] возрос интерес к неопределенности качественного анализа. Соответственно, проблемы установления неопределенности в качественном анализе, например, «соответствует / не соответствует», анализе идентичности или сравнительной идентичности, привлекли больше внимания, особенно в областях, где влияние ошибочных результатов качественного анализа чрезвычайно важно, например, в судебной экспертизе или анализе на допинг.

Существует множество показателей для выражения неопределенности качественных результатов [7]. Тем не менее, консенсус в отношении того, какие показатели использовать, ограничен. Исключение составляют области эпидемиологии и лабораторной диагностики, где в качестве параметров клинической точности последовательно используются понятия «клиническая чувствительность» и «клиническая специфичность» [27].

Количественный и качественный анализ существенно различаются по способу представления результатов и связанных с ними неопределенностей. В то время как количественные результаты представляют в виде интервала, который включает «истинное значение» измеряемой величины с определенным доверительным уровнем, качественные свойства представляют в виде классификации с показателями, которые выражают вероятность правильной или неправильной классификации. Этот «шанс» может быть описан вероятностью, правдоподобием, шансами или другими показателями, оцененными на основе интерпретации входной информации.

Качество сообщаемых показателей зависит от

количества и разнообразия изученных случаев. Определение этих показателей позволяет идентифицировать случаи, где процедуры должны быть улучшены, чтобы уменьшить вероятность получения ошибочных результатов.

В этом Руководстве описаны общие принципы оценки эффективности качественного анализа для сообщения о неопределенности результатов качественного анализа и представлены примеры применения описанной теории. В Руководстве не обсуждается способность образца для испытаний представлять группу идентичных образцов или более крупный объект; то есть в нем не обсуждается влияние отбора проб на эти оценки.

Порядковые результаты можно свести к бинарным (да/нет) результатам и обработать с помощью методов, описанных в этом руководстве, путем трактовки результатов порядковой классификации как «правильных» или «неправильных». Другие методы обработки порядковых шкал выходят за рамки данного руководства.

2 Виды качественного анализа

Как упоминалось в области применения настоящего Руководства, качественный анализ² определяется как «Классификация в соответствии с заданными критериями» [28]. В Таблице 1 перечислены некоторые примеры. Хотя все эти случаи кажутся очень разными, у них есть одна общая черта; как только критерии определены, эффективность методологии классификации относительно просто описать с точки зрения частоты успеха или неудачи. Эти показатели успеха и неудачи составляют основу большинства показателей эффективности качественного анализа.

Качественный анализ, описанный в основном тексте, делится на две категории на основе различных типов критериев классификации, т. е. качественный или количественный критерий. В Таблице 1 представлены примеры каждого из них. Раздел 3 описывает стратегии оценки эффективности для различных типов критериев классификации. Для качественного анализа, когда число правильных или неправильных ответов зависит от количественного свойства,

такого как наличие запрещенного вещества, обнаружение которого зависит от присутствующего количества, также учитывается предел обнаружения (см. раздел 3.4).

Оценка соответствия значения количественного свойства объекта предельному значению или интервалу иногда может рассматриваться как преобразование результата измерения в качественный результат («соответствует» или «не соответствует»). Использование измеренных значений и их неопределенностей для оценки соответствия подробно рассматривается в другом руководстве Eurachem/CITAC [29] и, соответственно, не рассматривается подробно в настоящем Руководстве. Однако в Приложении В обсуждается, как некоторые показатели, используемые для оценки эффективности или неопределенности качественного анализа, могут быть определены для количественной оценки соответствия.

Таблица 1. Виды качественного анализа на основе разных типов критериев классификации.

Критерий классификации	Пример качественного анализа
Качественный	<ol style="list-style-type: none"> (1) Обнаружение алифатических альдегидов в растворе по изменению окраски после добавления реактива Шиффа. (2) Идентификация кристаллической формы материала путем наблюдения. (3) Идентификация марки и года вина с помощью органолептического анализа. (4) Идентификация биологического вида путем определения или обнаружения конкретной последовательности ДНК. (5) Идентификация группы крови человека путем наблюдения за агглютинацией.
Количественный	<ol style="list-style-type: none"> (1) Идентификация остатков пестицидов во фруктах с использованием измеренных масс фрагментов и относительного содержания фрагментов в ГХ-МС. (2) Определение эквивалентности инфракрасного спектра между новым и ранее принятым промышленным сырьем с использованием критериев длины волны и интенсивности. (3) Идентификация диуретика в моче спортсмена с использованием времени удерживания и измеренной массы фрагментов в GC-MS. (4) Идентификация препарата в крови по времени удерживания и измеренной массе фрагментов в LC-MS. (5) Обнаружение вируса в клиническом образце на основе интенсивности флуоресценции количественной полимеразной цепной реакции в реальном масштабе времени (qPCR).

² «Качественное испытание» в принципе является более широкой областью, чем «качественный анализ» просто потому, что химический анализ, часто называемый аналитической работой, является одним из видов деятельности среди многих областей испытаний. Однако в данном Руководстве для химиков-аналитиков и специалистов смежных дисциплин эти термины используются как синонимы.

3 Оценка эффективности качественного анализа

3.1 Общие соображения

В этом разделе приведены рекомендации по оценке и выражению характеристик эффективности для процедур, предназначенных для обеспечения простой классификации на два класса («бинарная классификация»). «Классы» помечены здесь как «положительные» или «отрицательные» для обозначения принадлежности к интересующему нас классу или «непринадлежности». Эта классификация охватывает большинство практических ситуаций, в том числе «выше предела», «приемлемо», «неприемлемо», «идентичность как» или «присутствие определенных частиц».

Предполагается, что классы являются исчерпывающими и исключительными, чтобы можно было однозначно вычислить долю неправильных ответов. Это означает, что ни один испытуемый объект не может быть отнесен к третьему классу. Как правило, этого можно достичь путем тщательного определения критериев классификации. Однако остается вероятность того, что результат не может обеспечить достаточную уверенность в отнесении к одному из двух классов. В этих обстоятельствах для аналитика вполне разумно сообщить о результате испытания как «неоднозначном» в смысле недостаточно надежном. Неоднозначные результаты требуют дальнейшего изучения, чтобы представить эти результаты как «однозначные». Эти результаты признаются в медицинских лабораториях как принадлежащие к «серой зоне» или «зоне неоднозначности».

Некоторые из описанных понятий, в принципе, могут быть распространены на большее количество классов, например, в классификации по порядковой шкале, путем оценки доли правильной и неправильной классификации для всех классов. Полезным расширением является трактовка идентификации структуры или идентичности (формально многоклассовой проблемы) как «правильной» или «неправильной»), и здесь предполагается именно такой подход. Однако подробное рассмотрение проблемы нескольких классов, которая может включать одновременно несколько отнесений или отнесение к нескольким классам, выходит за рамки

настоящего Руководства. Качественный анализ включает в себя различные этапы, а именно: (1) описание задачи, (2) разработку метода и (3) его валидацию, (4) испытание неизвестных образцов и проверку результатов путем контроля качества, и (5) представление результатов (Рис. 1). Однозначная спецификация свойства, которое необходимо определить, и оценка пригодности анализа для предполагаемого использования имеют решающее значение. Представление результата качественного анализа должно опираться на действующие методики и адекватный контроль качества испытаний. Способ представления результатов зависит от цели анализа и получателя отчета. В этом Руководстве не описывается подробно, как должен быть разработан метод испытаний или как должен быть организован контроль качества.

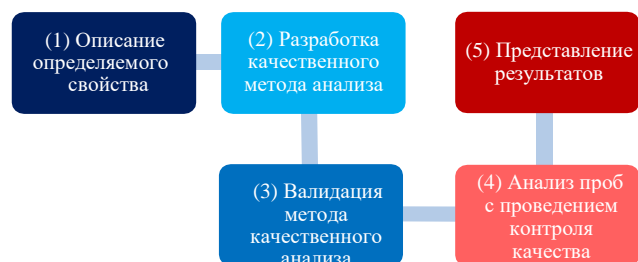


Рисунок 1. Процесс качественного анализа от описания задачи до представления результатов.

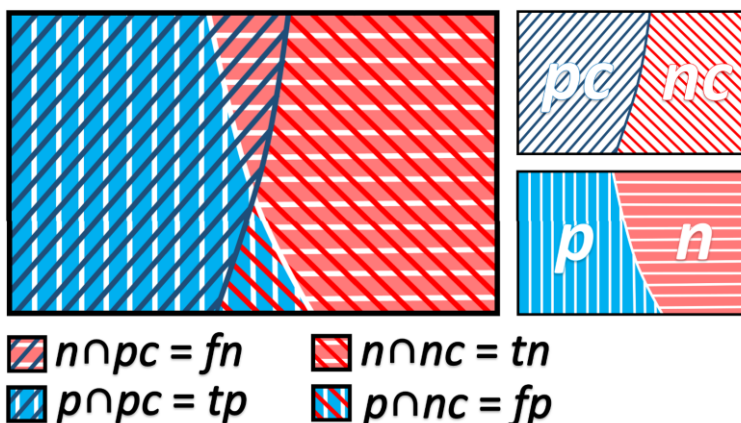
3.2 Количественная оценка эффективности качественного анализа

3.2.1 Определение основных подходов к оценке эффективности

Самый простой способ количественной оценки эффективности метода качественного анализа — это вычисление доли ошибочных результатов. При «положительных» или «отрицательных» результатах полезно сообщать о доле «истинно положительных» и «ложноположительных» или «истинно отрицательных» и «ложноотрицательных» результатов соответственно. Однако доля этих результатов может относиться либо к общему числу случаев или результатов определенного типа, либо к общему числу возможных случаев или результатов.

Например, долю ложноположительных

Рисунок 2. Графическое представление примера перекрытия числа положительных, pc , или отрицательных, nc , случаев с числом положительных, p , или отрицательных, n , результатов. Символ « \cap » обозначает пересечение групп, например, $n \cap pc$ здесь обозначает множество отрицательных результатов из положительных случаев. Множества « $n \cap pc$ », « $p \cap pc$ », « $n \cap nc$ » и « $p \cap nc$ » определяют fn , tp , tn и fp , соответственно.



результатов можно определить, как:

- i) Доля отрицательных случаев, ошибочно признанных положительными (fp/nc), где fp и nc — число ложноположительных результатов и отрицательных случаев соответственно. Рисунок 2 графически представляет перекрытие различных типов случаев и результатов. Доля (fp/nc) представлена отношением площадей пересечения (\cap) положительных результатов, « p », с « nc » ($p \cap nc = fp$) и площадью « nc ». Величина FP в Таблице 2 представляет это определение.
- ii) Доля положительных результатов, ошибочно признанных положительными (fp/p), где p — число положительных результатов. На Рисунке 2 эта доля представлена отношением между площадями ($p \cap nc = fp$) и « p ».
- iii) Доля от общего числа случаев или результатов, ошибочно признанных положительными ($fp/(pc+nc) = fp/(p+n)$), где pc и n представляют число положительных случаев и отрицательных результатов соответственно. На Рис. 2 эта доля представлена отношением площади, отмеченной ($p \cap nc = fp$), к общей площади фигуры.

Разница между этими определениями принципиальна. В случае i) доля не зависит от доли «отрицательных случаев» в совокупности, т.е. $nc/(nc+pc)$, поскольку $FP = fp/nc$. Однако для случаев ii) и iii) доля ложноположительных результатов зависит от $nc/(nc+pc)$, поскольку в совокупностях,

содержащих больше nc , наблюдается больше fp . Следовательно, эти определения по-разному характеризуют выполнение качественного анализа и, следовательно, предполагают различную интерпретацию их значений. Доли истинно положительных, TP , (tp/pc) и истинно отрицательных, TN , (tn/nc), результатов отнесённые к соответствующему числу случаев, известны в клинической химии как «чувствительность» и «специфичность» качественного анализа соответственно [7] (см. Таблицу 2). Определение клинической чувствительности и специфичности требует надлежащего определения изучаемых случаев по окончательному клиническому диагнозу. Для количественного анализа термин «чувствительность» [1] или «аналитическая чувствительность» [30] имеет другое значение.³

Число истинно положительных случаев, отнесённое к числу положительных случаев (tp/p), также известно как «прецизионность» качественного анализа или «положительное прогностическое значение», PPV [30]. Термин «отрицательное прогностическое значение», NPV , используется для числа истинно отрицательных случаев, отнесённого к общему числу отрицательных случаев (т.е. tn/n). Эффективность качественного анализа определяется как доля правильных результатов любого типа от всех результатов (т.е. $(tp+tn)/(p+n)$). Индекс Юдена является альтернативным способом количественной оценки успеха качественного анализа (Таблица 2) [31].

³ Согласно Международному словарю по метрологии [1], «чувствительность измерительной системы» — это «отношение изменения показаний измерительной системы к соответствующему изменению значения величины, которая измеряется».

Таблица 2. Альтернативные характеристики эффективности в качественном анализе.

Характеристики эффективности	Выражение
Доля истинно положительных результатов, <i>TP</i> (Чувствительность, <i>SS</i>)	$tp/pc = tp/(tp+fn) = 1-FN$
Доля ложноположительных результатов, <i>FP</i>	$fp/nc = fp/(tn+fp) = 1-TN$
Доля истинно отрицательных результатов, <i>TN</i> (Специфичность, <i>SP</i>)	$tn/nc = tn/(tn+fp) = 1-FP$
Доля ложноотрицательных результатов, <i>FN</i>	$fn/pc = fn/(tp+fn) = 1-TP$
«Прецизионность» или «Положительное прогностическое значение», <i>PPV</i>	$tp/p = tp/(tp+fp)$
«Отрицательное прогностическое значение», <i>NPV</i>	$tn/n = tn/(tn+fn)$
Эффективность, <i>E</i>	$(tp+tn)/(p+n)$
Индекс Юдена, <i>Y</i>	$SS(\%)+SP(\%)-100$
Отношение правдоподобия положительных результатов, <i>LR(+)</i>	TP/FP
Отношение правдоподобия отрицательных результатов, <i>LR(-)</i>	TN/FN
Апостериорная вероятность	См. Приложение А

tp – число истинно положительных результатов; *fp* – число ложноположительных результатов; *tn* – число истинно отрицательных результатов; *fn* – число ложноотрицательных результатов; *p* – число положительных результатов (*tp + fp*); *n* – число отрицательных результатов (*tn + fn*); *pc* – число положительных случаев и *nc* – число отрицательных случаев.

Хотя показатели, относящиеся к числу положительных или отрицательных случаев, не зависят от преобладания соответствующих типов случаев, сами по себе эти числа не могут обеспечить вероятность того, что конкретный результат является правильным. Чтобы оценить вероятность того, что результат является правильным, необходимо также учитывать соответствующую частоту результатов и преобладание соответствующих случаев. Этот и другие показатели достоверности качественных результатов обсуждаются в разделе 4.

3.2.2 Определение «основы для сравнения» при оценке эффективности

Показатели, используемые для количественной оценки эффективности качественного анализа, могут иметь дополнительные особенности. Положительные и отрицательные случаи могут быть установлены по-разному. Некоторые случаи или пробы, используемые в качестве образцов сравнения, могут быть известны как «положительные» по характеристике исходя из их происхождения или состава. Другие могут быть, как определено АОАС International, случаями, когда результаты «подтверждающего метода и, в дополнение, другого аналитического метода являются положительными» [32]. Некоторыми

примерами адекватного происхождения положительных случаев могут быть пациенты, у которых диагностировано определенное заболевание, или почва, о которой известно, что она загрязнена. Положительные образцы для испытаний могут быть приготовлены путем добавления компонентов, которые подлежат идентификации, в матрицу, эквивалентную анализируемым пробам, например, пестицид в пищевом продукте, который подтвержден или не подтвержден в отношении наличия природных уровней этого пестицида. Если эффективность идентификации значительно зависит от уровня содержания компонента (например, концентрации идентифицируемого или обнаруживаемого вещества), состав образца должен позволять определять этот уровень. Отрицательный случай также может быть истолкован как случай, известный как отрицательный по своему происхождению, приготовлению или определяемый как отрицательный, поскольку «результаты, подтверждающего метода и другого аналитического метода являются отрицательными». Определения положительных и отрицательных случаев в АОАС International имеют более широкое применение, поскольку это единственный подход, применимый к анализу сложных объектов, которые трудно воспроизвести путем приготовления. Однако он зависит от качества

результатов используемых аналитических методов. В некоторых областях сложно искусственно подготовить образцы с исследуемым аналитом и возможными мешающими веществами для проверки эффективности анализа, поскольку матрицы проб неизвестны и непредсказуемы.

Положительные и отрицательные случаи также могут быть предоставлены в качестве справочных данных, например спектр, про который известно, что он относится к конкретному соединению. После определения критериев идентификации можно определить вероятность сообщения о правильном или неправильном отнесении состава при сравнении по этим критериям. Например, в масс-спектрометрии идентификация может быть основана на оценке присутствия или отсутствия и интенсивности характеристических ионов. Вероятность спектроскопического совпадения можно предсказать с помощью биномиальной или гипергеометрической статистики, как показано в примерах E1 и E2.

3.2.3 Представление результатов оценки эффективности метода

3.2.3.1 Таблицы сопряженности

Очень удобный способ представления эффективности качественного метода анализа, который существенно не меняется в пределах области применения, — это таблица сопряженности признаков. В Таблице 3 представлен пример такой таблицы. В этом примере TP , FP , TN и FN составляют 97,8 % (228/233), 0,33 % (1/301), 99,7 % (300/301) и 2,1 % (5/233) соответственно.

Как правило, область применения метода может включать различные уровни

определяемых веществ или свойств и различные матрицы анализируемых проб. Поэтому могут потребоваться отдельные таблицы сопряженности для разных частей области применения.

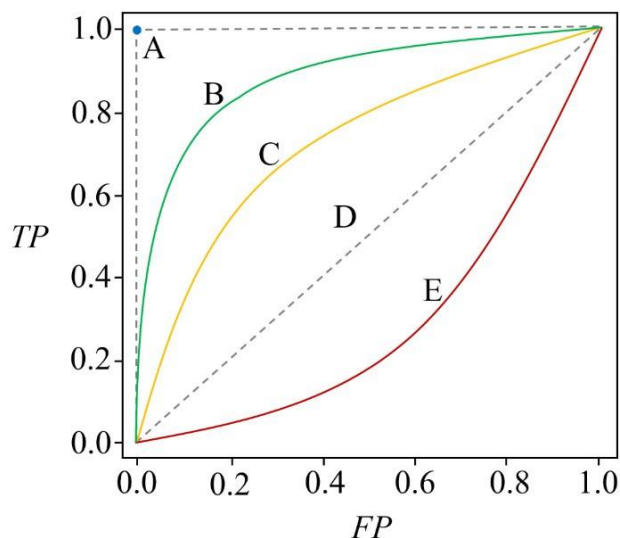
3.2.3.2 Характеристическая кривая обнаружения (ROC-характеристика)

Для качественного анализа, основанного на оценке количественной характеристики, выбор критериев классификации, которые оптимально сочетают доли правильных и ошибочных результатов, обычно TP и FP , может быть выполнен с использованием характеристических кривых обнаружения (ROC, *Receiver Operating Characteristic*), которые отображают пару значений (TP , FP) при варьировании критерия классификации (т. е. порога дискриминации). Эти кривые также можно использовать для сравнения различных методик качественного анализа [31]. Хотя подробное описание этих кривых выходит за рамки данного руководства, на Рисунке 3 представлены пять схематических примеров ROC-характеристик. Каждая кривая показывает, как доля истинно положительных и ложноположительных результатов изменяется по мере того, как критерий идентификации меняется от более строгого к менее строгому выявлению положительных случаев, связанных, соответственно, с низким или высоким TP . Помимо предоставления визуальной иллюстрации эффективности, площадь под кривой (часто сокращенно обозначаемая «AUC» от «*area under the curve*») может использоваться как обобщенный показатель эффективности классификатора [31].

Таблица 3. Конкретный пример таблицы сопряженности признаков, описывающей эффективность качественного аналитического метода, которая предполагается приблизительно постоянной в пределах его области применения.

Результат	Случай		Общее число результатов
	Положительный (pc)	Отрицательный (nc)	
Положительный (p)	$tp = 228$	$fp = 1$	$p = 229$
Отрицательный (n)	$fn = 5$	$tn = 300$	$n = 305$
Общее число случаев	$pc = 233$	$nc = 301$	

Рисунок 3. Пять примеров ROC-характеристик, на которых показано изменение TP и FP при изменении количественных критериев идентификации. Кривая А (синяя точка) представляет идеальный тест, в котором критерии идентификации не влияют на результаты, а TP и FP равны 100 % и 0 % соответственно. Кривые В и С представляют подходящие методы, в которых $TP \geq FP$. Из этих трех методов метод В предпочтительнее метода С. Кривая D представляет собой диагональ шансов, где $TP=FP$ для всех порогов принятия решений, и это не было бы полезным классификатором. Кривая E, на первый взгляд, кажется очень плохим классификатором, она последовательно дает долю ложноположительных результатов выше, чем истинно положительных результатов. Однако простая замена TP на FP в результатах приводит к ROC-характеристике, близкой к кривой С, и тогда классификатор может оказаться полезным.



3.3 Оценивание доли ложноположительных и ложноотрицательных результатов

3.3.1 Область применения метода и детали валидации

Валидация метода качественного анализа включает в себя установление требований к эффективности и проверку их выполнения [30].

Перед этой оценкой эффективности область применения метода должна быть четко определена с точки зрения типа классификации (например, наличие пентахлорфенола выше $1 \text{ мг} \cdot \text{кг}^{-1}$) и анализируемых объектов (например, кожаные изделия). Следует указать метод классификации, а именно метод анализа (например, ГХ-МС/МС), способ использования этого метода (например, подготовка проб и инструментальные условия) и критерии классификации. Критерии классификации должны быть четко описаны, чтобы гарантировать, что собранные данные об эффективности будут применимы к последующему анализу.

В некоторых видах качественного анализа, руководствуясь соображениями эффективности, аналитический метод делится на два этапа: предварительный, более быстрый и дешевый метод скрининга, за которым, когда

это необходимо, следует более трудоемкий и дорогой подтверждающий метод. Подтверждение выполняется, когда первая оценка дает результаты, противоречащие ожидаемым, или когда она может оказать существенное влияние на частные или групповые интересы людей. Тем не менее, важно оценить число ложноотрицательных и положительных результатов для всей процедуры, включающей скрининг и подтверждающие тесты. Например, когда подтверждению подлежат только положительные результаты, важно проверить, достаточно ли низок уровень ложноотрицательных результатов на этапе скрининга.

Что касается деталей валидации метода, для методов, применимых к разнообразным объектам (например, различным пищевым продуктам), эффективность должна быть проверена для репрезентативного набора типов образцов. Типы и количество анализируемых образцов зависят от влияния матрицы проб на эффективность. В некоторых случаях понимание принципов классификации может позволить предвидеть группы объектов, демонстрирующих одинаковую эффективность качественного анализа, из которых можно выбрать представительный образец и исследовать его. Эффективность, оцененную при анализе такого образца, затем можно

экстраполировать на группу объектов, характеризующихся одинаковой эффективностью качественного анализа. Если эффективность метода классификации позволяет, может быть принято решение об изучении эффективности анализа образцов и/или значений свойств, где доля ошибочных результатов достигает самых высоких значений. Лаборатория должна управлять тщательностью оценки эффективности, не забывая при этом о наличии времени и ресурсов для этой оценки.

В некоторых случаях может быть приемлемой стратегия постоянной валидации, при которой каждый раз, когда анализу подвергается объект, который является новым для лаборатории, выполняются дополнительные и специальные проверки качества анализа.

3.3.2 Использование информации из литературы

Для широко используемых методик качественного анализа можно ожидать, что информация об эффективности находится в открытом доступе. Прежде чем приступить к изучению эффективности хорошо зарекомендовавшей себя аналитической методики, следует провести соответствующее изучение подходящей литературы в этой области для сбора независимой информации о пригодности методики для предполагаемого использования. Однако опубликованные данные об уровне неправильных ответов следует использовать с осторожностью; они могли быть получены с использованием специального оборудования, реагентов и персонала и относиться к конкретным матрицам проб и характерным уровням содержаний, поэтому аналитик должен учитывать, является ли его/ее ситуация эквивалентной описанной. Например, если объекты, исследуемые в литературе, имеют уровни содержаний, далекие от пороговых

значений, используемых для различения классов, и если их матрицы относительно свободны от мешающих компонентов, найденные характеристики идентификации могут быть слишком оптимистичными по сравнению с «реальными» аналитическими проблемами, с которыми сталкивается лаборатория. Таким образом, частота правильных и ошибочных результатов сильно зависит от доступных данных.

В некоторых случаях можно предвидеть, будут ли характеристики, наблюдаемые в литературе, лучше или хуже, чем характеристики, наблюдаемые при качественном анализе в лаборатории. Если делается вывод, что методика качественного анализа эффективна для наихудших сценариев, т. е. может давать результаты, соответствующие ее назначению, эта методика может быть использована для анализа неизвестных объектов без ограничений.

В разделе 4 обсуждается, как можно установить критерии для принятия решения о том, подходит ли анализ для предполагаемого использования.

3.3.3 Оценивание исключительно экспериментальным путем

Независимо от вида качественного анализа, которые упомянуты в разделе 2, значения *FP* и *FN* можно оценить непосредственно по числу ошибочных результатов набора анализов. В качественном анализе, основанном исключительно на качественных критериях классификации, это единственный способ оценки неопределенности качественного анализа. Однако, если неправильные ответы маловероятны, этот подход требует большого числа испытаний.

Таблица 4. Минимальное число испытаний, необходимое для обнаружения одного или нескольких ошибочных (положительных или отрицательных) результатов.

Доля ошибочных результатов	Доверительный уровень	
	95 %	99 %
0,5 %	598	919
1 %	299	459
5 %	59	90

Учитывая, что количество неправильных ответов в идеале должно быть небольшим, возникает вопрос: сколько образцов нужно проанализировать, чтобы быть достаточно уверенным в обнаружении ненулевого числа неправильных ответов?

Из опубликованной информации (см., например, Ferrara et al. [33]) видно, что доля ложноположительных или ложноотрицательных результатов может составлять всего 0,5 %, а в некоторых случаях даже ниже [6, 8, 9]. Для диапазона вероятностей ошибочных результатов в Таблице 4 показано число проб, которое необходимо проанализировать, чтобы быть уверенным, по крайней мере в пределах указанных доверительных уровней, в обнаружении одного или нескольких ошибочных результатов. Таблица 4 использует биномиальное распределение и показывает, что для 95% вероятности обнаружения одного или нескольких ошибочных результатов число испытаний, которые необходимо выполнить, будет в три раза больше, чем число испытаний, дающих в среднем один ошибочный результат. Например, для метода с показателем ложноположительных результатов 1 % установлено, что (в среднем) на каждые 100 анализов с отрицательными случаями наблюдается один положительный результат. Однако, чтобы быть «на 95 % уверенным» в том, что наблюдается ложноположительный результат, необходимо провести 299 (около 3×100) испытаний с отрицательными результатами.

Значений Таблицы 4 недостаточно для хорошей оценки доли ошибочных результатов или для сравнения различных методов. Даже для приблизительной оценки обычно требуется в пять-десять раз больше минимального количества наблюдений, приведенного в Таблице 4. Эту таблицу также можно интерпретировать как минимальное число анализов, необходимых для проверки соответствия различным допустимым уровням ошибочных результатов, как обсуждается ниже.

Пытаясь определить частоту ошибочных результатов непосредственно в ходе экспериментов с новой методикой, аналитик часто сталкивается с дилеммой. С одной стороны, для данной методики процент ошибочных ответов неизвестен, и поэтому любые проведенные классификации могут быть недостоверными.

С другой стороны, простой анализ до появления первого ошибочного ответа не обязательно даст истинную картину частоты ошибочных ответов. Чтобы решить эту проблему, аналитику предлагается заранее определить допустимые уровни для двух значений доли ошибочных ответов. Для выбранного доверительного уровня можно использовать биномиальное распределение для оценки числа экспериментов, необходимых для обнаружения одного или нескольких ошибочных ответов с достаточной достоверностью. Этот подход не гарантирует получения точного значения частоты ошибочных ответов, но он укажет её границу. Например, предположим, что аналитик решает, что значение FP 5 % является приемлемым, и после выполнения 59 экспериментов (Таблица 4), охватывающих вероятный диапазон матриц, не обнаружено ни одного неправильного ответа. В этом случае можно сделать вывод, что FP не превышает 5 %. В качестве меры контроля качества этой валидированной методики также рекомендуется чередовать анализируемые пробы с холостыми и со стандартными образцами, содержащими целевую характеристику (например, аналит) на соответствующих уровнях. Всегда следует помнить, что частота ошибочных результатов очень сильно зависит от случайностей и/или особенностей совокупности проб, из которых отбирается выборка, а также от стратегии пробоотбора.

Таблица 4 показывает, что при низкой доле неправильных ответов может быть нецелесообразно анализировать большое число проб для обнаружения неправильного отклика. Соответственно, если анализ недорогой и/или предназначен для широкого использования, например, в качестве скринингового теста на наркотики, может быть приемлемым сначала установить экспериментальным путем, что доля неправильных ответов не превышает верхнего предела, скажем, 5 %, а затем уточнить эту цифру в свете опыта с другими образцами.

В тех случаях, когда число проб, вероятно, будет относительно небольшим и/или анализ является дорогостоящим, все тесты следует проводить параллельно с подтверждающим тестом, и время от времени следует пересчитывать долю неправильных ответов.

Математическая обработка имеющейся информации может быть использована для преодоления некоторых ограничений экспериментального определения частоты неправильных ответов (см. разделы 3.3.4 и 3.3.5).

Краткое руководство 1 – Пример моделирования сигнала

Если для идентификации дельтаметрина в оливковом масле с помощью ГХ-МС оцененное стандартное отклонение повторяемости времени удерживания, s_{tRi} , составляет 0,022 мин при $\nu = 32$ степенях свободы, допустимое время удерживания для идентификации этого соединения в пробе может быть: $(t_R \pm t \cdot s_{tRi}) = (t_R \pm 2,04 \times 0,022) = (t_R \pm 0,045)$ мин; где t_R — время удерживания, наблюдаемое при однократной ежедневной инъекции стандартного раствора, а t — двустороннее 95 % критическое значение для t -распределения с 32 степенями свободы. Следовательно, для t_R равного 36,055 мин, допустимый интервал времени удерживания будет составлять $(36,055 \pm 0,045)$ мин. Если предположить, что мешающее вещество имеет время удерживания на 0,05 мин меньше, чем для дельтаметрина, и точность обоих времен удерживания эквивалентна, то вероятность того, что мешающее вещество будет иметь время удерживания в пределах этого допустимого интервала, составит 1,5 %. Это значение оценивается кумулятивным t -распределением для t -значения $(-0,05/0,022)$ и ν (формула MS-Excel: СТЬЮДЕНТ.РАСП(-0,05/0,022;32;ИСТИНА)).

3.3.4 Оценивание на основе базы данных

Альтернативой определению доли ошибочных результатов экспериментальным путем являются исследования случайных несоответствий в справочных базах данных, таких как базы данных масс-спектров или инфракрасных спектров. В некоторых случаях это эквивалентно многим тысячам экспериментов. Однако, хотя они информативны и мощны, в настоящее время существует ограничение, заключающееся в том, что такие базы данных часто совершенно непредставительны в отношении исследуемой совокупности объектов; например, в то время как распространенность различных материалов в общем использовании широко варьируется, типичная справочная база данных будет содержать только один вид каждого материала. Это может привести к значительному смещению оценок вероятности; опять же, полученные значения вряд ли будут лучше, чем оценки порядка величины. Примеры E1 и E2 иллюстрируют использование этой методологии для оценки эффективности качественного анализа.

3.3.5 Оценивание на основе моделирования количественных данных

Оценка эффективности высокоселективного и трудоемкого и/или дорогостоящего качественного анализа исключительно на основе экспериментов, проведенных в одной лаборатории, невозможна.

В качественном анализе, основанном на классификации количественных результатов (например, инструментальный анализ), модели рассеяния результатов могут использоваться для оценки доли правильных и ошибочных результатов. Приложение В содержит

дополнительную информацию. Например, если соответствующий инструментальный сигнал, такой как время удерживания аналита в хроматографическом методе, имеет нормальное распределение, можно предсказать вероятность того, что мешающий компонент будет иметь время удерживания в пределах допустимого интервала времени удерживания для данного аналита (см. Краткое руководство 1, стр. 15).

Однако моделирование зависит от достоверности предположенной модели и значений входных переменных. Например, поскольку относительное время удерживания не может быть распределено нормально, предположение о нормальности может привести к недооценке частоты ошибочных результатов. Моделирование инструментальных сигналов методом Монте-Карло является удобным способом оценки FP и FN исходя из параметров, не имеющих нормального распределения [8, 9].

Пример E5 иллюстрирует моделирование рассеяния инструментального сигнала для оценки частоты ошибочных результатов высокоселективной идентификации методом ГХ-МС/МС.

3.3.6 Оценивание эффективности качественного анализа, когда она зависит от непрерывной переменной

Многие подтверждающие или выявляющие тесты показывают сильную зависимость вероятности обнаружения или доли неправильных ответов от некоторой непрерывной переменной. Например, эффективность детектирования часто зависит от концентрации или числа частиц искомого вещества. Может оказаться полезным смоделировать зависимость доли неправильных ответов от непрерывных (или других) переменных.

Логистическая регрессия и пробит-регрессия [34, 35] обычно применяются к таким задачам и были предложены (с примерами) для оценки эффективности качественных методов анализа [36]. Применение логистической регрессии было продемонстрировано при обнаружении ДНК на уровне низкого числа копий [37]. Процедура хорошо описана в учебниках и доступна практически во всех пакетах статистического программного обеспечения, поэтому здесь она подробно не представлена. Простая логистическая регрессия моделирует вероятность бинарного ответа как функцию некоторой непрерывной переменной. Модель:

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

или эквивалентная:

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

где p — представляющая интерес вероятность (например, вероятность положительного результата), x — непрерывная переменная (обычно концентрация анализируемого вещества) а b_0 и b_1 — коэффициенты регрессии. Большинство статистических пакетов предоставляют метод подбора либо на основе исходных данных (пары концентрация/качественный результат), либо на основе долей, рассчитанных исходя из числа результатов. Обратите внимание, что для первого подхода требуется только последовательность значений да/нет (или 1/0); он не требует вычисления доли. Это позволяет применять метод к ряду проб с различными (известными или независимо измеренными) концентрациями, каждая из которых подвергается качественному анализу только один раз.

Как только взаимосвязь установлена, становится возможным оценить пределы обнаружения (см. ниже) на основе подобранной зависимости между концентрацией и вероятностью обнаружения, просто выбрав соответствующий предел для вероятности обнаружения, который соответствует используемому определению способности обнаружения («detection capability»).

В примере E4 представлен практический пример логистической регрессии.

3.3.7 Экспертное суждение

Когда нет данных об эффективности аналитического метода от третьей стороны и невозможно оценить эффективность исключительно на основе экспериментов (раздел 3.3.3) или моделирования (разделы 3.3.4 и 3.3.5), аналитик может использовать свой практический опыт в методе классификации для изучаемых или аналогичных объектов, чтобы решить, подходит ли метод для предполагаемого использования.

Когда это возможно, решение о пригодности метода для предполагаемого использования должно быть подтверждено объективными доказательствами.

Процесс формулирования экспертных суждений явился предметом нескольких исследований. На суждения влияет множество факторов, приводящих к соответствующим оценкам неопределенности результата [38].

3.4 Предел обнаружения и селективность

3.4.1 Предел обнаружения

Предел обнаружения (LOD) обычно описывает самую низкую концентрацию вещества, которая приводит к надежному обнаружению. Для тестов, в которых классификация включает оценку количественной характеристики, и значение этой характеристики влияет на качественные результаты, «предел обнаружения» (LOD) и/или «предел количественного определения» (LOQ), учитываемые в качественном и /или количественном анализе, должны быть проверены с точки зрения эффективности качественного анализа [30]. Результат качественного анализа должен соответствовать предполагаемому использованию на этом уровне (уровнях).

ПРИМЕЧАНИЕ: Постановления Комиссии (ЕС) № 589/2014 [39] и № 152/2009 [40] определяют LOQ как «наименьшее содержание аналита, которое может быть измерено с достаточной статистической достоверностью, отвечающее критериям идентификации, как описано в международно-признанных стандартах» [41].

Для исключительно качественного анализа LOD можно найти, применяя методику анализа к образцам, содержащим постепенно уменьшающиеся уровни определяемой характеристики, пока вероятность получения ошибочных результатов не достигнет заранее установленного критерия. В этом типе оценки также можно использовать логистическую и пробит-регрессию (раздел 3.3.4).

3.4.2 Селективность

Селективность в том смысле, в котором этот термин обычно используется в аналитической химии, относится к «степени, в которой конкретный метод может быть использован для определения аналитов в заданных условиях в присутствии других компонентов с аналогичным поведением» [42]. Международный словарь по метрологии (VIM) определяет этот термин как свойство измерительной системы [1].

ПРИМЕЧАНИЕ. Термин «специфичность» в контексте количественного анализа используется для совершенно селективного анализа [42, 43], что может быть заявлено в химии лишь в очень редких случаях. Однако существует явное альтернативное использование термина «специфичность» в контексте качественного анализа (см. Таблицу 2). В настоящем Руководстве термин «селективность» используется в общем смысле, а термин «специфичность» зарезервирован для целей, указанных в таблице 2.

Селективность можно оценить путем анализа одного или нескольких анализируемых образцов, имеющих известные или вероятные мешающие характеристики, т. е. характеристики, которые не являются целевыми для данного анализа, но которые можно считать вероятной причиной отклика.

Иногда можно идентифицировать мешающие компоненты или ситуации, которые с наибольшей вероятностью могут давать ложноположительные результаты. Например, можно с достаточным основанием ожидать, что тесты на аммиак реагируют на первичные амины, а тесты на конкретные бактериальные штаммы могут реагировать на любые бактерии того же общего вида.

Если качественный анализ работает относительно хорошо в наихудших ситуациях, можно сделать вывод, что методика эффективна для анализируемых объектов всех типов.

И хотя частота неправильных ответов может быть измерена для каждого отдельного вещества или каждого присутствующего мешающего компонента, маловероятно, что исследования селективности приведут к одному определенному значению селективности. Это связано с тем, что ответ зависит от потенциально перекрестно реагирующих частиц, включенных в исследование, и от уровня содержания этих частиц. Таким образом, исследования селективности лучше всего рассматривать как дающие общее представление об адекватности метода качественного анализа при столкновении с различными проблемами.

4 Выражение достоверности в качественном анализе

4.1 Общие соображения

В то время как установление неопределенности измерения в количественном анализе обычно приводит к диапазону значений, например, к интервалу расширенной неопределенности или минимальной степени чистоты, установление классификации обычно невозможно связать с диапазоном. В целом, нельзя заключить, что на 90 % материал «соответствует» требованиям или что с вероятностью 99 % аналит присутствует или что структурные элементы соединения находятся в пределах некоторой последовательности. Вместо этого типичная информация о неопределенности носит вероятностный характер. То есть указывается вероятность того, что данная классификация верна, или типичные вероятности неправильной классификации объектов, правильный класс которых известен.

Показатели эффективности, которые могут быть получены в результате валидационных исследований, могут быть представлены вместе с результатом качественного анализа. Однако в целом они редко дают прямую информацию (например) о вероятности того, что качественный результат верен. В этом

разделе описаны два вида показателей, которые были предложены для этой цели, чтобы помочь понять и улучшить состояние дел в области выражения неопределенности результатов качественного анализа. Представленные здесь показатели используют варианты правила Байеса [4] (см. Приложение А). Их можно использовать, чтобы дать: а) указание на силу доказательств, обеспечиваемых одним или несколькими качественными результатами, в пользу одной возможной классификации по сравнению с другой; б) в сочетании с достоверной информацией о вероятности обнаружения разных (истинных) значений качественных характеристик - указание вероятности того, что та или иная классификация верна при данном конкретном результате качественного анализа.

4.2 Отношение правдоподобия

Наиболее знакомая и широко используемая форма представления эффективности качественного анализа — это доля ошибочных результатов, в частности, FP и FN или дополняющие их доли, TN и TP , соответственно (например, $TN = 1 - FP$). Оба из этих отношений можно удобно объединить в одну и ту же

Краткое руководство 2 – Интерпретация отношения правдоподобия

Если сообщается о положительном результате, вероятность того, что случай действительно положительный, PP , рассчитывается по уравнению. (Q2.1) (см. ниже). Это уравнение представляет собой хорошо известную теорему Байеса (Приложение А), заменяющую доли истинно положительных и ложноположительных результатов условными вероятностями.

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

где $P(+)$ — вероятность того, что случай будет положительным, известная до испытания. Это также может быть выражено в форме «шансов» (см. Приложение А):

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

В Q2.2 отношение $P(+)/P(-)$ представляет вероятность положительного результата до применения качественного теста; то есть «априорные шансы». Отношение TP/FP — это рассчитанное отношение правдоподобия $LR(+)$.

Таким образом, отношение правдоподобия описывает, как вероятность (представленная шансами) изменяется после положительного результата теста; его можно рассматривать как меру дополнительной информации, предоставленной тестом.

В частном случае, когда $P(+)=P(-)=0,5$, так что априорные шансы равны 1,0, $LR(+)$ представляет отношение апостериорных вероятностей положительного или отрицательного случая. Например, при равной распространенности (или предполагаемой распространенности) положительных и отрицательных случаев положительный результат с $LR(+)$ 1000 будет означать, что апостериорная вероятность того, что случай действительно положительный, в 1000 раз больше, чем вероятность того, что случай отрицательный.

характеристику: отношение правдоподобия LR .

Если сообщается положительный результат, $LR(+)$ оценивается по уравнению (2):

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

$LR(+)$ представляет собой отношение двух вероятностей; вероятность сообщения о положительном результате, если случай положительный, деленная на вероятность сообщения о положительном результате, если случай отрицательный. В широком смысле отношение правдоподобия дает меру изменения вероятности того, что проба действительно положительная после получения положительного результата теста.

Математически отношение правдоподобия представляет собой изменение вероятности, выраженной как «шанс» (см. Приложение А). Высокое значение отношения правдоподобия в качественном анализе указывает на то, что испытуемый образец с большей вероятностью будет положительным, чем можно было бы сказать до выполнения анализа.

Иногда это интерпретируется как «весомость доказательств», основанных на положительном результате теста, в пользу того, что испытуемый образец действительно положительный.

В частном случае, когда как положительные, так и отрицательные случаи равновероятны ($P(+)=P(-)=0,5$; где $P(+)$ и $P(-)$ — распространенность положительных и отрицательных случаев соответственно), $LR(+)$ можно понимать, как указание на то, насколько более вероятно, что сообщаемый положительный результат будет правильным, а не ошибочным.

Например, если положительные и отрицательные случаи считаются равновероятными перед испытанием, положительный результат, связанный с $LR(+)$ равным 7300 означает, что положительный результат в 7300 раз более вероятно будет правильным, чем ошибочным.

Если сообщается отрицательный результат, $LR(-)$ представляет собой отношение:

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

В частном случае, когда до проведения исследования положительный и отрицательный исходы равновероятны, $LR(-)$ показывает, насколько более вероятно, что отрицательный результат является правильным.

Некоторые авторы объединяют оба отношения правдоподобия в параметре «диагностическое отношение шансов», DOR ($DOR = LR(+)/LR(-)$) [30].

Одна из наиболее полезных функций отношения правдоподобия ($LR(+)$ или $LR(-)$) состоит в том, что если классификация зависит от двух независимых доказательств (т. е. результат сообщается только тогда, когда его подтверждают два независимых анализа, проведенных по независимым методикам), $LR_{(1\&2)}$ результата обоих анализов оценивается путем перемножения LR , которые количественно выражают неопределенности каждого доказательства ($LR_{(1)}$ и $LR_{(2)}$):

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

Например, если присутствие загрязнителя в пищевом продукте, определенное с помощью ГХ-МС, основано на времени удерживания с

Краткое руководство 3 – Вероятность при рассмотрении нескольких доказательств, которые не являются независимыми

Вероятность наступления двух независимых событий А и В, $P(A \cap B)$, оценивается по формуле (Q3.1).

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

где $P(A)$ и $P(B)$ — вероятности наступления событий А и В по отдельности, например, при положительном результате анализа положительного случая (т. е. TP).

Однако, если $P(A)$ и $P(B)$ взаимосвязаны, вероятность совпадения обоих событий определяется уравнением (Q3.2), которое включает условную вероятность события В в предположении, что событие А уже произошло:

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

Или, что то же самое:

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

Для прямой или обратной корреляции, связанной с $r_{AB} > 0$ или $r_{AB} < 0$, соответственно, $P(A \cap B)$ будет соответственно больше или меньше, чем для случая, когда А и В независимы.

$LR(+)$ 99,9 и данных масс-спектра с $LR(+)$ 490, $LR(+)$ идентификации, основанной на обоих этих подходах, становится $4,9 \times 10^4$ (т. е. $99,9 \times 490$). Уравнение (4) следует из того, что вероятность сходимости двух независимых результатов оценивается путем умножения соответствующих индивидуальных вероятностей. Если рассматриваются m независимых доказательств (от $i = 1$ до m), необходимых для сообщения о положительном или отрицательном результате, т. е. результат сообщается только в том случае, если на него указывают m свидетельств, LR оценивается из объединенных доказательств по уравнению (5).

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

где Π обозначает произведение последовательности переменных, а $LR_{(i)}$ – отношение правдоподобия i -го качественного анализа ($LR_{(i)}(+)$ или $LR_{(i)}(-)$).

Когда отдельные доказательства не являются независимыми, уравнение (5) будет недооценивать или переоценивать совместную вероятность. В Кратком руководстве 3 показано, как можно объединять вероятности, которые не являются независимыми.

Отношения правдоподобия могут быть сложными для интерпретации, особенно для неспециалистов. Для применения в области судебной экспертизы была рекомендована [44] шкала в Таблице 5, чтобы дать словесную интерпретацию весомости доказательств. Согласно этой таблице, собранные доказательства считаются «чрезвычайно сильными» только в том случае, если LR

больше 10^6 . В принципе, такой подход можно адаптировать для других обстоятельств, если требуется общее указание на весомость доказательств.

Например, для определения типа полимера микропластика, собранного из отложений при мониторинге окружающей среды, критерии, представленные в Таблице 5, являются слишком строгими. Для этих анализов результаты, связанные с $LR(+)$ больше 19, должны быть достаточными (т. е. с $TP \geq 95\%$ и $FP \leq 5\%$), поскольку загрязнение определяется после идентификации многих частиц из нескольких проб [45].

Хотя определение бинарного свойства может дать только один из двух результатов, если наиболее вероятный результат (например, либо «да», либо «нет») связан с низким LR , может быть принято решение сообщить о результате как о неоднозначном, а не использовать словесный эквивалент из Таблицы 5. Например, может быть принято решение сообщать положительный или отрицательный результат, если соответствующий LR больше, чем (например) 10^5 и результат с более низким LR считается неоднозначным. Эта «серая зона» для значения LR может быть установлена ниже 10^5 , 19, или любого другого значения в зависимости от цели анализа. Сообщение о результате как о неоднозначном полезно, если и ложноположительные, и ложноотрицательные результаты оказывают существенное влияние. При тестировании на допинговые вещества в моче спортсмена ложноположительные

Таблица 5. Интерпретация отношений правдоподобия, предложенная в области судебной экспертизы Европейской сетью судебно-экспертных учреждений (ENFSI) [44].

Значение отношения правдоподобия	Словесный эквивалент
1	Свидетельства не поддерживают ни одно из утверждений по сравнению с другим
2 - 10	Слабая поддержка первого утверждения относительно альтернативного
10 - 100	Умеренная поддержка первого утверждения, по сравнению с альтернативным
100 - 1000	Умеренно сильная поддержка первого утверждения, по сравнению с альтернативным
1000 - 10 000	Сильная поддержка первого утверждения по сравнению с альтернативным
10 000 - 1 000 000	Очень сильная поддержка первого утверждения по сравнению с альтернативным
> 1 000 000	Чрезвычайно сильная поддержка первого утверждения по сравнению с альтернативным

результаты являются гораздо более серьезными, чем ложноотрицательные результаты, из чего можно заключить, что, если не наблюдается никаких свидетельств наличия допинга, результат может быть указан как отрицательный (т.е. свидетельствующий об отсутствии подтверждений допинга) [9, 46]. Однако, как ложноположительные, так и ложноотрицательные результаты могут быть проблемой при установлении материнства или отцовства, что даёт основания заключить, что положительное совпадение с низким LR не следует представлять как окончательный вывод об «отсутствии совпадений» [47].

4.3 Апостериорная вероятность

Если имеется надежная информация о преобладании определенного признака (например, популяция с хорошо задокументированной распространенностью определенного заболевания), $LR(+)$, связанный с результатом теста, может быть преобразован в вероятность PP того, что проверяемый объект действительно положительный, имея ввиду положительный результат теста. Эта вероятность называется апостериорной и оценивается с помощью теоремы Байеса (Приложение А). Одна из форм этого преобразования, использующая отношение правдоподобия, выглядит следующим образом:

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

Здесь $P(+)$ и $P(-)$ — априорные вероятности, т.е. информация, доступная до теста, а PP и PN (см. далее) — апостериорные вероятности.

Взяв предыдущий пример анализа загрязнителя в пищевом продукте с помощью ГХ-МС, где положительный результат связан с $LR(+)$ $4,9 \times 10^4$, предполагая, что $P(+)$ = $P(-)$ = 0,5, получаем, что PP равно 99,998 % ($PP = 4,9 \times 10^4 / (4,9 \times 10^4 + 1)$).

Если сообщается об отрицательном результате, апостериорная вероятность того, что образец действительно отрицательный, PN , оценивается по формуле:

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

Это уравнение подобно уравнению (6). $P(+)$ и $P(-)$ выражают преобладание положительных или отрицательных случаев.

В широком смысле, поскольку апостериорная вероятность относится к представленной как результат классификации, апостериорную вероятность можно рассматривать как меру вероятности того, что представленный результат является правильным.

Апостериорные вероятности могут быть трудно применимы на практике. Иногда достаточно релевантные и надежные априорные вероятности недоступны. Хотя некоторые авторы полагают, что эту проблему можно преодолеть, если предположить, что положительные и отрицательные результаты равновероятны, так что $P(+)/P(-) = 1$, это не всегда разумно. Иногда, особенно в судебной экспертизе, может быть неуместно делать выводы об априорных вероятностях для конкретного случая на основании знаний о несвязанных событиях. В таких случаях отношение правдоподобия (раздел 4.2) может предоставить полезную информацию о достоверности, обеспечиваемой результатом теста, без необходимости определения априорных вероятностей.

В некоторых областях, таких как медицинские науки, рассмотрение преобладания какого-то состояния или характеристики при принятии решений на основе качественных аналитических результатов может иметь решающее значение для диагностики. Диагноз заболевания или клинической ситуации, основанный на результатах клинического анализа, также будет основываться на дополнительной информации, такой как цвет слизистой оболочки, локализация и интенсивность болей, возраст и пол, поведение, сопряженное с риском, и т. д. То, как эта информация влияет на окончательное решение относительно наблюдаемого результата, можно проиллюстрировать путем вычисления PP или PN , хотя клиницисты обычно не выполняют эти расчеты; скорее ожидается, что они будут знать об общей важности распространенности при постановке диагноза на основе результата теста.

Более подробная информация об этих показателях представлена в библиографии [4, 8, 9].

4.4 Надежность показателей

Надежность вычисленных значений LR , PP или PN зависит от надежности рассматриваемой доли результатов и, для апостериорных вероятностей, от надежности любой используемой априорной вероятности. В Таблице 4 представлено число испытаний, необходимое для надежного обнаружения одного или нескольких неправильных ответов при различных вероятностях неправильного отклика. Для надежного *количественного* определения доли неправильных ответов обычно требуется гораздо большее число испытаний (см. раздел 3.3.3). Возможно, потребуется ещё большее увеличение числа изучаемых случаев, чтобы охватить всю область применения метода испытаний. Например, при испытании пищевой продукции может потребоваться изучить несколько различных матриц. Моделирование инструментального сигнала, рассматриваемого в качественном анализе, может сделать возможным количественное определение небольшого числа ошибочных результатов, но успех зависит от адекватности входных данных и алгоритма моделирования.

Для определения PP или PN исходя из сильно различающихся $P(+)$ и $P(-)$ можно использовать Таблицу 4 для определения количества случаев (взятых из целевой совокупности), которые необходимо изучить.

Качество входных данных для оценки этих показателей становится еще более важным, когда объединяются различные доказательства и рассчитываются показатели, количественно определяющие силу суммарной информации.

Таким образом, представленные показатели следует использовать с осторожностью, помня о соответствующих особенностях входных данных, о том, как используется полученный результат, и о соответствующих последствиях. Излишне благоприятная интерпретация данных об эффективности качественного анализа может быть столь же вредной, как игнорирование ограничений конкретного качественного анализа.

4.5 Неопределенность долей

Статистическое качество оцениваемой частоты результатов, которая зависит от числа испытаний, использованных для ее определения, может быть выражено в виде доверительного интервала, CI . Этот доверительный интервал также известен как «условная неопределенность» (4.4.6 в [18]), и обычно рассчитывается для 95 % доверительного уровня (95 % CI).

Например, широкий 95 % CI для чувствительности SS указывает на то, что «истинное» значение SS может сильно отличаться от оцененного. Та же логика может быть применена к другим долям, характеризующим, например, специфичность SP . Поскольку искомые доли результатов не оцениваются на основе какой-либо априорной информации о генеральной совокупности, эти интервалы характеризуют только оцениваемое качество аналитической работы.

Интерпретация 95 % CI , в некоторой степени, аналогична той, что имеет место в случае расширенной неопределенности измерений [1]. Для 95 % CI существует 5 % вероятность того, что «истинное» значение доли результатов выходит за рамки CI . Аналогичным образом, 95 % CI для экспериментально определенной доли дает статистическую неопределенность для вычисленной доли.

Например, если способность метода качественного анализа правильно идентифицировать положительные случаи проверяется на основе анализа 400 таких случаев, и все 400 результатов являются положительными, оцененная чувствительность, SS , на уровне 100 %, связана с 95 % CI , ограниченным между 99 % и 100 %; то есть истинное значение SS варьируется от 99 % до 100 % с 95 % достоверностью. Если метод проверяется только с 5 положительными случаями, 95 % CI для SS будет ограничен 57 % и 100 %. 95 % CI позволяет выразить качество параметров аналитического метода, что требуется для их осмысленной интерпретации. В приведенных выше примерах оба значения SS оценены как 100 %, но оценка SS гораздо более надежна в первом случае. Расчет 95 % CI для SS и SP является стандартной практикой в клинической лаборатории (10.1.3 в [27]).

Несколько моделей для вычисления CI были опубликованы [48, 49, 50, 51, 52, 53, 54]. Интервал оценки Уилсона [54] использовался

для простоты применительно к небольшим выборкам. Уравнения 8 и 9 могут использоваться для расчета нижнего, $LL_{SS.95}$ и верхнего, $HL_{SS.95}$, пределов 95 % CI для SS или TP .

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

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

где:

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

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

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

Уравнения (10) и (11) используются для расчёта нижнего, $LL_{SP.95}$, и верхнего, $HL_{SP.95}$, пределов 95 % CI для SP и TN .

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

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

где:

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

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

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

Целевое или минимальное значение $LL_{SS.95}$ или $LL_{SP.95}$ (то есть $LL_{SS.95}^{lg}$ или $LL_{SP.95}^{lg}$) должно быть определено в соответствии с целью анализа. Цель особенно важна, когда влияние ошибочных результатов имеет решающее значение. Например, для компонентов крови, используемых при переливании, скрининг на инфекционные заболевания должен проводиться с помощью тестов, имеющих $LL_{SS.95}$ близкое к 100 %, которые могут быть подтверждены, если во время валидации было протестировано много положительных случаев.

Когда найденная доля случаев сравнивается с минимальным целевым значением, или когда изучается либо увеличение, либо уменьшение этого параметра, должна проводиться односторонняя оценка. Для теста с 95 % достоверностью коэффициент 1,96 должен быть изменен на 1,64.

5 Представление результата качественного анализа

В настоящее время аккредитованные лаборатории не обязаны представлять результаты качественного анализа с неопределенностью. Примеры в этом разделе соответственно предназначены для того, чтобы предложить возможные подходы к отчетности, когда лаборатория решит сделать это, чтобы помочь клиенту.

Положительный результат можно сообщить с *TP* и *FP*, *LR(+)* или *PP* и отрицательный результат с *TN* и *FN*, *LR(-)* или *PN*. Другие показатели, представленные в Таблице 2, также могут быть использованы для сообщений о доверии к результату.

Эти показатели обычно предоставляют информацию об отдельном результате испытаний. Однако для случаев, когда значение показателя является постоянным для всей области применения, такие параметры можно интерпретировать как характеризующие аналитический метод.

Следующие четыре примера показывают, как можно представить качественные результаты с соответствующей эффективностью или неопределенностью.

Пример 1 (*текст курсивом говорит о неопределенности качественного анализа*):

Г-жа А. Б. заражена вирусом SARS-CoV-2.

(тест с чувствительностью 90 % и специфичностью 99 %)

Пример 2 (*текст курсивом говорит о неопределенности качественного анализа*):

Моча г-на С. Д. содержит остатки канренона

(идентификация с отношением правдоподобия 4.9×10^4)

Пример 3 (*текст курсивом говорит о неопределенности качественного анализа*):

Кокаин присутствует в образце 123

(идентификация с отношением правдоподобия 4.9×10^4 , что считается «очень сильным» доказательством присутствия аналита)

Пример 4 (*текст курсивом говорит о неопределенности качественного анализа*):

Остатки бензина были идентифицированы в пробе с кодом 456, взятой из завала после пожара

(идентификация с апостериорной вероятностью 99,998 %, оцененной путем моделирования сигнала в предположении, что присутствие или отсутствие аналита одинаково вероятно)

6 Выводы и рекомендации

Для лабораторий важно проверить, по крайней мере, наиболее критический уровень неверных откликов. Для некоторых показателей необходимо установить как ложноположительные, так и ложно-отрицательные уровни.

Вполне возможно ожидать, что большинство лабораторий имеют соответствующие параметры своих методик качественного анализа (когда условия анализа находятся под адекватным контролем. Доказательства этого обычно включают:

- четкие доказательства адекватной метрологической прослеживаемости значений параметров, находящихся под контролем ввиду их значимости для испытания;
- доказательства того, что неопределенности этих параметров достаточно малы для цели анализа.

Разумно ожидать, что лаборатории будут следовать опубликованным стандартам наилучшей практики в качественном анализе, где они доступны, включая использование соответствующих справочных данных и стандартных образцов.

Количественные (то есть численные) оценки неопределенности в качественных результатах испытаний должны быть представлены таким образом, чтобы избежать неправильного толкования.

Всякий раз, когда делается вывод, что полученный аналитический результат связан со слишком низким и слишком высоким уровнем правильных или, соответственно, неправильных результатов, совершенно разумно сообщать о результатах испытания как неоднозначных в смысле недостаточно надежных.

7 Примеры

Примеры описаны после указания их области применения.

7.1 E1: Идентификация соединений с помощью масс-спектрометрии низкого разрешения с использованием поиска в базе данных или наличия характеристических ионов

7.1.1 Введение

Этот пример разделен на Случаи А и В, где используются разные процедуры для идентификации соединений в сложных матрицах с помощью масс-спектрометрии с низким разрешением. Параллельное изложение этих двух случаев подчеркивает альтернативный характер вариантов идентификации.

Обратите внимание, что на практике для идентификации обычно используют несколько критериев, таких как совпадение масс-спектров и хроматографического времени удерживания. Этот пример фокусируется только на компоненте масс-спектрометрии. Пример E5 дает пример использования нескольких критериев.

Область:

Тип качественного анализа: Анализ на основе количественных критериев

Образец/матрица: А) Мясные продукты и В) криминалистические или экологические пробы

Параметр/аналит: А) Диэтилстильбоэстрол, DES (запрещенный гормон роста для говядины и мяса птицы) или В) Героин, ДЭС и дихлордифенилтрихлорэтан (ДДТ)

Тип критерия классификации: 1) Идентификация на основе допусков относительной интенсивности (relative abundance, RA) конкретных ионов масс-спектра; 2) Идентификация по наличию специфических ионов масс-спектра вне зависимости от значений RA.

Метод/оборудование: Газовая хроматография с масс-спектрометрией низкого разрешения с использованием ионизации электронным ударом (ГХ-МС)

Форма представления результатов: Отношение правдоподобия

В этом примере описывается оценка неопределенности идентификации соединений с помощью ГХ-МС с использованием различных критериев идентификации (разделы 7.1.1 и 7.1.2). В примерах представлены результаты идентификации трех соединений (т. е. DES, героина и ДДТ) в двух типах проб (мясные продукты и пробы для судебной экспертизы или анализа окружающей среды).

Масс-спектрометрия, особенно в сочетании со стадией хроматографического разделения, является мощным инструментом, помогающим идентифицировать неизвестные соединения. В большинстве случаев масс-спектрометрия с низким разрешением с использованием ионизации электронным ударом (ЭУ) является предпочтительным методом, когда требуется идентификация, а не количественная оценка. Масс-спектр может содержать много ионов, не все из которых полезны для диагностических целей. В связи с этим возникает вопрос, существует ли минимальное число ионов, достаточное для обеспечения однозначной идентификации.

ПРИМЕЧАНИЕ: В некоторых аналитических областях минимальное необходимое число ионов определено для идентификации соединений [20] [21][22][23].

7.1.2 Идентификация на основе относительной интенсивности характеристических ионов

Srphon [55] исследовал минимальное число ионов, которое необходимо контролировать для однозначной идентификации диэтилстильбоэстрола⁴ (DES) в мясных продуктах. Данные, относящиеся к последующему исследованию [56], основанному на коммерческой библиотеке масс-спектров, содержащей около 270 000 записей, представлены в Таблице E1.1.

⁴ DES использовался в качестве гормона роста для крупного рогатого скота и птицы и впоследствии был запрещен после того, как были доказаны его канцерогенные свойства.

Таблица E1.1. Количество спектров библиотеки Wiley с 270 000 записей, соответствующих определенным критериям для относительной интенсивности некоторых ионов.

Критерий идентификации			Совпадения
#	Ион (m/z)	RA (%) допустимый интервал	
1	268	1 – 100	9 995
2	268	1 – 100	5 536
	239	1 – 100	
3	268	90 – 100	46
	239	10 – 90	
4	268	90 – 100	9
	239	50 – 70	
5	268	90 – 100	15
	239	50 – 90	
	145	5 – 90	
6	268	90 – 100	1 (DES) ^a
	239	50 – 70	
	145	45 – 65	

RA: относительная интенсивность (процент по отношению к сигналу с наибольшей интенсивностью, основной пик)
^a Одно совпадение соответствует масс-спектру DES

В Таблице E1.1 показано количество спектров в использованной библиотеке, соответствующих заданным критериям для относительной интенсивности (RA) одного или нескольких ионов. RA оценивается путем деления интенсивности исследуемого иона на интенсивность иона с наибольшей интенсивностью (т. е. основного пика). Эта нормализация направлена на получение параметра идентификации, менее зависящего от уровня аналита (например, концентрации). Из Таблицы E1.1 видно, что при рассмотрении большего числа ионов и более узких диапазонов интенсивности число совпадений в базе данных резко сокращается. Набор критериев идентификации #6 выделяет масс-спектр DES, что приводит к единственному совпадению.

Для сравнения с альтернативной библиотекой в Таблице E1.2 представлено количество совпадений из общедоступной справочной библиотеки, содержащей на тот момент 62 235 спектров, с учетом допусков относительной интенсивности (RA) одного или нескольких ионов [57]. По мере того, как допуски, связанные с RA большего числа ионов, становятся уже, выделяются масс-спектры меньшего числа соединений. В Таблице E1.2 представлено число спектров, соответствующих трем различным целевым соединениям, а именно DES, героину и ДДТ. Героин и ДДТ имеют отношение к анализу некоторых образцов для судебной экспертизы и образцов из окружающей среды, соответственно.

Сравнение выделения масс-спектра DES в обеих библиотеках, как это описано в таблицах E1.1 и E1.2, позволяет сделать вывод, что, как и следовало ожидать, количество совпадений зависит от количества спектров в библиотеке (см. критерии идентификации #1 и #2 в таблицах E1.1 и E1.2). Если количество совпадений разделить на общее количество записей, различия, наблюдаемые в таблицах E1.1 и E1.2, уменьшатся.

В Таблице E1.3 обобщается информация, собранная в таблицах E1.1 и E1.2 по наиболее избирательным идентификациям. В Таблице E1.3 собранная информация преобразуется в TP и FP , которые далее объединяются в $LR(+)$, которое оценивает неопределенность положительного результата (т. е. сообщает о наличии аналита).

Оценка TP (т. е. приблизительно 100 %) предполагает, что определенные допуски для RA ионов учитывают их изменчивость. В идеале допуски должны быть установлены на основе моделей изменчивости сигнала, построенных на основе повторно снятых спектров растворов с соответствующими концентрациями аналита [8, 9] (пример E5).

Значение FP , представленное в Таблице E1.3, предполагает, что возможные мешающие вещества — это все соединения, масс-спектр которых содержится в используемой библиотеке. Многие соединения

присутствующие в пробах, не могут быть обнаружены с помощью ГХ-МС или удалены при подготовке проб. С другой стороны, многие соединения, спектр которых есть в библиотеке, вряд ли будут присутствовать в анализируемых пробах из-за химической несовместимости, или независимости от источников или происхождения. Для наихудшего случая оценивается FP , эквивалентный единице отнесенной к общему количеству спектров N минус 1 ($FP = 1/(N - 1)$), поскольку известно, что FP не будет равен нулю. Описанные ограничения того, как оценивался FP , следует учитывать при использовании этого значения. Как вариант, FP можно оценить из моделей шума сигнала, как обсуждается в Примере Е5 [8, 9].

Из данных Таблицы Е1.3 видно, что при использовании одной и той же библиотеки все аналиты с одним совпадением имеют одинаковые TP , FP и $LR(+)$ в зависимости только от количества записей в библиотеке. В соответствии с критериями, установленными Европейской сетью судебно-экспертных учреждений (Таблица 5), идентификация аналита с помощью масс-спектрометрии с использованием описанной процедуры идентификации дает «Очень убедительные» доказательства присутствия аналита: ($LR(+)$ между 10^4 и 10^6). Предположим, что идентификация также подтверждается временем удерживания аналита в хроматографической системе (т. е. в ГХ), и диапазон значений удерживания выбран правильно. В таком случае $LR(+)$ идентификации может увеличиться (Пример Е5).

Таблица Е1.2. Количество спектров в общедоступной библиотеке с 62 235 записями, соответствующих определенным критериям для относительной интенсивности некоторых ионов.

#	Критерий идентификации		Совпадения
	Ион (m/z)	RA (%) допустимый интервал	
1	268	1 – 100	3597
2	268	1 – 100	1597
	239	1 – 100	
3	268	55 – 95	83
4	268	55 – 95	4
	239	30 – 70	
5	268	55 – 95	1 (DES)
	239	30 – 70	
	145	60 – 100	
6	369	1 – 100	1672
	327	1 – 100	
7	369	1 – 100	526
	327	1 – 100	
8	369	45 – 85	43
	327	60 – 100	
9	369	45 – 85	1 (Heroin)
	327	60 – 100	
10	352	1 – 100	1242
	235	1 – 100	
11	352	1 – 40	1140
	235	60 – 100	
12	352	1 – 40	1 (DDT)
	235	60 – 100	
13	352	1 – 40	7
	235	1 – 100	
	237	48 – 88	
14	352	1 – 40	1 (DDT)
	235	60 – 100	
	237	48 – 88	

RA : относительная интенсивность (процент по отношению к сигналу с наибольшей интенсивностью, основной пик)

^a Одно совпадение соответствует масс-спектру DES, героина или ДДТ

7.1.3 Идентификация на основе присутствия характеристических ионов – расчетные вероятности случайного совпадения

Предположим, что вместо идентификации аналита с использованием допусков для *RA* конкретных ионов идентификация основана на простом присутствии трех селективных ионов. В этом случае для грубой оценки неопределенности идентификации можно использовать следующие математические вычисления. Предположим, что используется масс-спектрометр с низким разрешением, который может различать только единицы m/z (т. е. единицы отношения массы к заряду), а ионы должны иметь значения m/z от 180 m/z до 480 m/z . В этом случае в спектре можно наблюдать примерно 300 возможных значений m/z . Следовательно, поскольку количество комбинаций из 300 предметов, взятых по три, равно $300! / [3! \cdot (300 - 3)!] = 4\,455\,100$, для трех пиков вероятность того, что три случайных пика случайно совпадут с тремя выбранными ионами, при условии, что все отношения m/z равновероятны, будет равна $1/4,6 \times 10^6$, или примерно $2,2 \times 10^{-7}$. Это, однако, не учитывает того факта, что большинство масс-спектров обычно имеют более трех ионов в интересующей области.

Таблица Е1.3. Неопределенность идентификации нескольких аналитов в разных матрицах с помощью ГХ-МС, оцениваемая по количеству совпадений спектров из используемой библиотеки с учетом допусков относительной интенсивностью двух или трех ионов.

Аналит	Анализируемый объект	Кол-во спектров в библиотеке, <i>N</i>	Кол-во совпадений ^a	<i>TP</i> (%) ^b	<i>FP</i> (%) ^c	<i>LR(+)</i> (<i>TP/FP</i>)
DES	Мясные продукты	270 000	1 в <i>N</i>	~100	3.7×10^{-4}	2.7×10^5
DES	F&E	62 235	1 в <i>N</i>	~100	1.6×10^{-3}	6.2×10^4
Heroin	F&E	62 235	1 в <i>N</i>	~100	1.6×10^{-3}	6.2×10^4
DDT	F&E	62 235	1 в <i>N</i>	~100	1.6×10^{-3}	6.2×10^4

DES: диэтилстилбоэстрол; F&E: криминалистические и экологические пробы; TP: доля истинно положительных результатов; FP: доля ложноположительных результатов; LR(+): отношение правдоподобия (TP/FP) (см. Таблицу 2).

^a Количество совпадений с учетом установленных допусков для *RA* конкретных ионов.

^b Оптимистическая оценка TP (примерно 100 %).

^c Оценено как $1/(N-1)$ (наихудший сценарий на основе собранной информации).

Это увеличивает вероятность того, что масс-спектр, содержащий *n* ионов, может соответствовать выбранным трем в число раз, равное $n!/[3! \cdot (n - 3)!]$. Принимая десять за типичное количество пиков, вероятность случайного совпадения увеличивается на $10!/3! \times 7! = 120$ раз. Расчетная вероятность случайного совпадения составляет соответственно $120 \times 2,2 \times 10^{-7}$ или приблизительно $2,6 \times 10^{-5}$. Это показывает ориентировочно частоту неправильных ответов, *FP*.

Предполагая, что уровни загрязняющих веществ достаточно высоки для получения достоверных масс-спектров примесей, можно считать, что *TP* составляет приблизительно 1, или 100 %.

Оцененные выше *TP* и *FP* можно объединить, чтобы получить *LR(+)* $1/2,6 \times 10^{-5} = 3,8 \times 10^5$, что указывает верхнюю границу отношения правдоподобия для идентификации, основанной на описанной процедуре.

ПРИМЕЧАНИЕ. Эта оценка предполагает, что все комбинации ионов равновероятны и равновозможны и что их появления независимы; известно, что это грубое приближение (см. пример 7.2). Таким образом, действительная вероятность случайного совпадения вероятно будет намного выше рассчитанного значения, что приведет к более низкому отношению правдоподобия.

7.1.4 Заключительные замечания

Подходы к оценке *LR(+)* при идентификации с помощью ГХ-МС, представленные в этом примере, имеют тенденцию быть чрезмерно оптимистичными в отношении достоверности качественных аналитических результатов (см. пример 7.2 для сравнения с фактическими вероятностями совпадения). Поэтому эти расчеты следует использовать только в качестве первоначальной оценки достоверности идентификации. В примере Е5 обсуждаются альтернативные и более реалистичные оценки неопределенности идентификации с помощью ГХ-МС/МС.

Хотя представленные здесь подходы к оценке вероятностей случайных совпадений и отношений правдоподобия могут быть оптимистичными, все же можно заключить, что, например, идентификация

героина в пробах с места преступления на основе только присутствия ионов с m/z 369 и 327 будет неадекватной, поскольку она связана с $LR(+)$ равным 118 ($118 = TP/\{FP\} = 100/\{[526/(62235-1)] \cdot 100\}$) (см. раздел 7.1.1: случай 7 Таблицы E1.2) или $4,5 \times 10^4$ ($4,5 \times 10^4 = 100/\{[1/(44850-1)] \cdot 100\}$; где $44850 = 300!/[2! \cdot (300 - 2)!]$) (см. раздел 7.1.2) в зависимости от подхода, используемого для оценки достоверности результатов. На практике это будет означать, что для обеспечения адекватной достоверности потребуются дополнительные критерии или дополнительные подтверждающие тесты.

Точно так же, хотя Sphon [55] и другие предположили, что остатки ветеринарных препаратов у крупного рогатого скота можно идентифицировать с помощью трех ионов масс-спектра, для официального мониторинга запрещенных веществ Европейский Союз (EU) требует сбора дополнительных доказательств присутствия этих соединений [20]. Например, если с помощью ГХ-МС можно контролировать только два характеристических иона при соответствующих уровнях анализа, необходимо рассмотреть два независимых хроматографических цикла, основанных на электронном ударе или химической ионизации, чтобы подтвердить присутствие этого анализа [20].

7.2 E2: Идентификация очищенных соединений с помощью инфракрасной спектроскопии

Область:

Тип качественного анализа: Анализ на основе количественных критериев

Образец/матрица: Очищенное химическое соединение

Параметр/аналит: Соединение из доступной базы данных инфракрасных спектров

Тип критерия классификации: Совпадение волнового числа трех или шести полос инфракрасного спектра в интервале [500, 1800] см⁻¹

Метод/оборудование: Инфракрасная спектроскопия

Форма представления результатов: Отношение правдоподобия

Несколько авторов исследовали использование статистики баз данных для оценки критериев качественного анализа. De Ruig и др. [58] предложили критерии для выявления остатков ветеринарных препаратов в мясных продуктах (см. раздел 7.1.2). Авторы дают ориентировочные значения вероятностей случайных совпадений на основе простой биномиальной модели. Эллисон и др. показали, что гипергеометрическое распределение является более подходящей моделью для случайных совпадений в спектрах, поскольку оно допускает небольшое количество совпадающих пиков между двумя спектрами, каждый из которых содержит большее количество пиков [5]. Последние авторы сосредоточились на случайных совпадениях, когда инфракрасный спектр сравнивается со спектральной библиотекой.

Эллисон и др. [5] изучали достоверность идентификации очищенных соединений путем сравнения полученного инфракрасного спектра со спектрами из библиотеки. Библиотека, используемая Эллисоном с соавторами была библиотека Sadtler, содержащая спектры 59 626 различных веществ. Из этой библиотеки выбирали случайную подгруппу из тридцати соединений и для каждого соединения отмечали количество полос m в интервале [500, 1800] см⁻¹. Было определено, что

Таблица E2.1. Случайные совпадения с шестью полосами в инфракрасной базе данных.

Вещество	Число полос m в спектральном интервале	Вероятность случайного совпадения ^a	Прогнозируемое число совпадений ^b	Наблюдаемое число совпадений	$LR(+)$
1-хлор-3-(1-нафтилокси)-2-пропанол	23	$3,19 \times 10^{-4}$	19	192	311
α -Цианокоричная кислота, метиловый эфир	17	$5,03 \times 10^{-5}$	3	29	2056
Фенил ν -триазоло-[1,5- α]-пиридин-3-ил кетон	24	$4,19 \times 10^{-4}$	25	190	314
Бензо- β -тиофен-6-акриловая кислота	20	$1,34 \times 10^{-4}$	8	52	1147
3-((Дипропиламино)метил)1-5-нитроиндол	17	$5,03 \times 10^{-5}$	3	29	2056
2-Мезитил-5-фенил-оксазол	22	$2,52 \times 10^{-4}$	15	99	602
p -Гидроксibenзойная кислота	18	$6,71 \times 10^{-5}$	4	44	1355
Капроновая кислота, изобутиловый эфир	8	$1,36 \times 10^{-7}$	0	1	59626
1-бромадамантан	10	$9,4 \times 10^{-7}$	0	1	59626
Фенилпропиловый эфир	17	$5,03 \times 10^{-5}$	3	47	1269

^a Вероятность случайного совпадения — это вероятность того, что $n = 6$ пиков случайно совпадут в двух спектрах, состоящих из m полос, при условии случайного попадания полос в спектр. Вероятность была рассчитана с использованием гипергеометрического распределения (ссылка [5]).

^b Прогнозируемое число совпадений — это вероятность случайного совпадения, умноженная на количество спектров в базе данных

среднее число полос в спектре в интервале $[500, 1800]$ см^{-1} , M , равно 16. Имеющееся спектральное разрешение составляло 4 см^{-1} и это означало существование $1300/4 = 325$, N , дискретных положений пиков в интервале $[500, 1800]$ см^{-1} . Для каждого отдельного спектра в выбранном поднаборе всю базу данных просматривали дважды: сначала для поиска минимум $n = 3$ совпадающих пиков, а второй раз для поиска минимум $n = 6$ совпадающих пиков.

При $n \geq 3$ число наблюдаемых совпадений примерно в два раза превышает число, предсказанное гипергеометрическим распределением. При $n \geq 6$, хотя число совпадений было значительно меньше, как и следовало ожидать, наблюдаемые совпадения превышали предсказанные совпадения в десять раз. Часть данных для шести совпадающих пиков представлена в Таблице E2.1.

Рассчитанные вероятности случайных совпадений в случае шести пиков находились в интервале $[10^{-7}, 10^{-5}]$. Вероятность случайного совпадения для соединения, умноженная на количество записей в базе данных, оценивает количество соединений, соответствующих критериям поиска. В случае двух соединений в таблице E2.1, таких как изобутиловый эфир капроновой кислоты и 1-бромадамантан, критерии поиска дают одно совпадение и, следовательно, будут адекватными, если подозреваются именно эти соединения. Для остальных соединений находится гораздо больше совпадений, что указывает на необходимость более строгих критериев

Предполагая, что TP равно приблизительно 100 % (поскольку ИК-спектры чистых соединений достоверно совпадают с их собственными эталонными спектрами) и принимая FP как отношение числа наблюдаемых совпадений к общему количеству спектров библиотеки (т. е. 59 626), можно оценить $LR(+)$ идентификации. В последнем столбце Таблицы E2.1 представлены найденные значения $LR(+)$ (вычисленные как TP/FP , где $TP = 1,0$). Получаемое $LR(+)$ намного ниже минимального значения 10^6 , которое считается по классификации доказательств как «чрезвычайно сильное» (Таблица 5), что позволяет предположить, что простое совпадение шести пиков только по длине волны может не обеспечить достаточной достоверности без дополнительных критериев. Для обеспечения однозначной идентификации могут потребоваться дополнительные критерии (такие как дополнительное совпадение интенсивности пиков, отсутствие пиков, отсутствующих в целевом соединении, совпадение хроматографического времени удерживания или требование точного визуального совпадения для полного спектра).

В этом примере подчеркивается, что справочные базы данных, одним из типов которых являются спектральные библиотеки, могут дать только ориентировочную информацию об уровне неправильных ответов. Уровни таких ответов действительно надежны только для совокупностей проб, аналогичных ожидаемым. Аналитик также должен решить, какое из набора совпадений, если таковые имеются, соответствует целевому соединению.

7.3 E3: Идентификация наркотических веществ в моче с помощью метода иммуноферментного анализа (EMIT) и альтернативного метода

Область:

Тип качественного анализа: анализ, основанный на количественных критериях (полученных с использованием качественной информации)

Образец/матрица: Моча

Параметр/аналит: Кокаин, метадон или опиаты

Тип критерия классификации: Не указано

Метод/оборудование: Метод иммуноферментного анализа с ферментативным усилением, EMIT и альтернативный собственный метод.

Форма представления результатов: Отношение правдоподобия и вероятность того, что положительный результат является правильным.

Хотя определения EMIT включают обработку инструментального сигнала, в этом примере оценивается эффективность этого качественного анализа на основе экспериментально определенных показателей ложноположительных и ложноотрицательных результатов. Таким образом, этот пример иллюстрирует определение эффективности качественного анализа на основе качественной информации.

В литературе сообщалось об использовании баз данных выборочных результатов для получения соответствующих вероятностей для байесовской оценки эффективности качественного анализа. Для проверки наличия наркотиков в моче, Ferrara et al. [33] собрали базу данных, содержащую информацию о типах наркотиков, аналитических методах, частоте неправильных ответов для этих методов и распространенности наркотиков. Для лаборатории цитируемых авторов Таблица E3.1 обобщает часть этих данных, относящуюся к EMIT. В таблице также представлена расчетная апостериорная вероятность того, что испытуемая проба будет действительно положительной, PP , как описано в уравнении (6). Для расчетов принимается, что распространенность отрицательных результатов равна $(1 - P(+))$, $LR(+) = TP/FP$ и $TP = 1 - FN$.

Таблица E3.1. Вероятность обнаружения с помощью EMIT наркотиков в моче [33].

Значения характеристик эффективности для разных веществ или классов веществ

Вероятность	Опиаты	Метадон	Кокаин
$P(+)$	0,44	0,26	0,20
FP	0,028	0,004	0,09
FN	0,069	0,18	0,056
PP	0,963	0,988	0,963

$P(+)$ – Распространенность положительных результатов.

Например, для идентификации метадона PP оценивается по уравнению (E3.1) как 0,988 (PP — апостериорная вероятность, определяемая по теореме Байеса). Другими словами, аналитик может быть более чем на 98 % уверен, что положительный ответ на метадон действительно указывает на присутствие этого наркотика. Однако обратите внимание, что это частично зависит от наблюдаемой высокой распространенности в конкретной совокупности отобранных проб. PP может быть намного ниже в генеральной совокупности, т. е. в совокупности, где медицинские препараты употребляются не так часто.

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

В Таблице E3.2 приведены аналогичные данные для другого, неиммунохимического метода. Обратите внимание, что FP для кокаина по этому методу определяется как ноль. Тем не менее, вопрос о том, действительно ли частота неверных ответов для таких скрининговых тестов может быть равна нулю, остается спорным. В данном случае ложноположительных результатов обнаружено не было, но при анализе большего числа образцов мог появиться один или несколько ложноположительных ответов. Соответственно, при расчете апостериорной вероятности использовалась оцененная частота неправильных ответов.

Таблица E3.2. Вероятность обнаружения наркотических средств в моче по собственной методике [33].

Значения характеристик эффективности для разных веществ или классов веществ			
Вероятность	Опиаты	Метадон	Кокаин
$P(+)$	0,44	0,26	0,20
FP	0,038	0,012	0,000
FN	0,276	0,179	0,247
PP	0,937	0,960	0,95 §

§ Вероятность PP рассчитана с использованием оцененного наихудшего значения FP , равного 0,001 (1 на 1000 тестов).

Снова рассматривая метадон, видим, что, PP составляет 0,960. Это достаточно высокая вероятность, хотя и несколько менее убедительная, чем та, которую дает тест ЕМІТ.

Чтобы проиллюстрировать эффект объединения данных, предположим, что были выполнены оба скрининговых теста. Если в каждом случае получен положительный ответ, то объединенная оценка PP становится равной 0,9998 (см. уравнение (E3.2), основанное на объединении уравнений (5) и (6).

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

В этом примере доступны надежные значения долей (т. е. априорные вероятности). Если бы их не было под рукой или если бы аналитик предпочел их не использовать, вместо них можно было бы использовать отношения правдоподобия; соответствующие значения равны 246 (ЕМІТ) и ~68 (собственный метод), что дает объединенное отношение правдоподобия приблизительно 17 000 (уравнение (E3.3)), что, согласно Таблице 5, соответствует «очень сильному» доказательству наличия метадона.

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

Во всех случаях ГХ-МС использовалась в качестве референтного метода для определения доли ошибочных результатов. Конкретная база данных, упомянутая здесь, является достаточно полной для изученных аналитов и была разработана для предоставления репрезентативной коллекции, позволяющей проводить байесовский анализ данных. Неизбежно есть некоторые отсутствующие значения, но по мере добавления новых данных их количество должно уменьшаться, а точность прогнозов еще более повышаться.

Еще одним преимуществом базы данных, настроенной для записи репрезентативных данных для нескольких различных методов, является информация, которую она предоставляет, чтобы позволить оптимизировать аналитическую эффективность. Например, выбор метода скрининга с низким уровнем ложноположительных результатов должен минимизировать затраты на дорогостоящие подтверждающие анализы. Однако необходимо учитывать и другие факторы, такие как предел обнаружения метода, частота ложноотрицательных результатов, а также скорость и стоимость анализа.

7.4 E4: Идентификация гена SRY человека в биологическом материале с помощью количественной ПЦР (qPCR)*

Область:

Тип качественного анализа: Анализ, основанный на количественных критериях (превышение порога флуоресценции)

Образец/матрица: Биологический материал

Параметр/аналит: Ген SRY (локализованный в определяющей пол области хромосомы Y)

Тип критерия классификации: Не указано

Метод/оборудование: Количественная полимеразная цепная реакция (qPCR)

Форма представления результатов: Доля истинно положительных ответов.

Хотя интерпретация qPCR включает обработку инструментального сигнала, в этом примере эффективность оценивается путем определения TP при различных концентрациях ДНК.

На рис. E4.1 показаны некоторые экспериментальные данные исследования обнаружения гена SRY человека с помощью количественной ПЦР (qPCR) в биологическом материале [37]. Данные были получены из исследования с 3 планшетами с использованием анализа 5'-нуклеазной активности с TaqMan зондом, содержащим двойную флуоресцентную метку, соответствующую области SRY генома человека, специфичного для мужчин. Результат считался «положительным», если наблюдаемая флуоресценция превышала заданный порог после 55 циклов амплификации. Хотя в рамках исследования по валидации использовались разные методы подготовки пробы, статистические тесты не выявили существенных различий, поэтому данные рассматривались как единый набор данных.

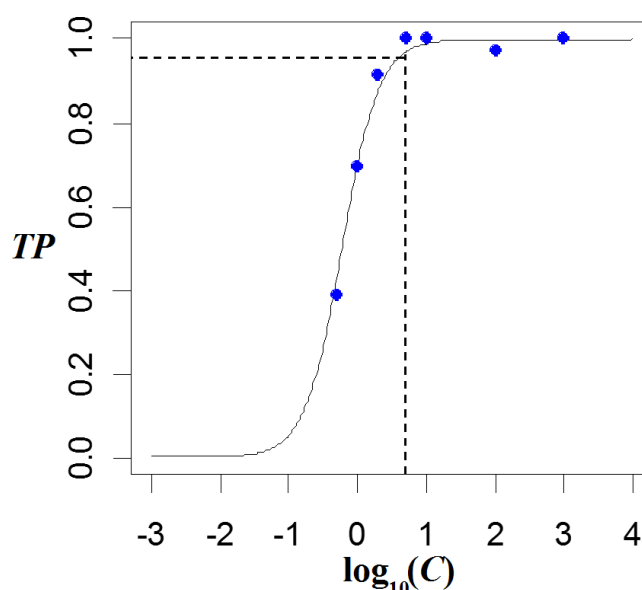


Рисунок E4.1. Данные по обнаружению ДНК, обработанные методом логистической регрессии. Величина TP построена как функция от $\log_{10}(C)$, где C - число копий гена SRY, в исследовании возможностей обнаружения и классификации ДНК. Точки показывают долю положительных результатов из 36 повторов для каждого числа копий последовательности. Сплошная линия показывает подгонку логистической регрессии с $b_0 = 0,85$ и $b_1 = 3,75$ (см. Раздел 3.3.6). Пунктирная линия показывает значение $TP = 0,95$, соответствующее логарифму числа копий последовательности $\log_{10}(C) = 0,56$ для $C = 3,6$ копий.

* В данном случае имеется в виду метод, известный как *полимеразная цепная реакция в реальном времени* (Real-Time PCR). Под количественной ПЦР (qPCR) обычно понимают ПЦР в реальном времени, выполняемую с применением калибраторов с целью измерения содержания аналита. (Прим. редактора)

Обратите внимание, что точка с $\log_{10}(C) = 2$ не показывает 100 % *TP*, хотя точки меньше и больше 2 показывают, что затрудняет точную оценку возможности классификации. Однако кривая регрессии позволяет достаточно точно определить действительный предел обнаружения, как показано на рисунке; использование модели непрерывной функции эффективно сглаживает случайные данные. Это значительное преимущество, которое позволяет аналитику изучать большое число концентраций с относительно небольшим числом повторов на каждом уровне вместо того, чтобы требовать много повторов на небольшом числе уровней и при этом получать относительно надежные оценки вероятности.

Этот пример иллюстрирует важную оговорку в моделировании. Данные о концентрации наносятся на график и моделируются в логарифмическом масштабе – это обычная практика работы с экспериментами по измерению концентрации ДНК или подсчету микробов. Для зависимой переменной преобразование часто диктуется распределением ошибок. Однако в этом случае нет веских причин выбирать логарифмическое преобразование для независимой переменной, выбор по сути произвольный. То же самое можно сказать и о выборе логистической модели: другие модели также могут достаточно хорошо согласовываться с данными. Там, где разные, но гладкие модели сходным образом соответствуют данным, результат интерполяции не сильно зависит от выбора модели. Однако вероятности на краях диапазона действительно могут быть очень чувствительны к выбору модели. Из этого следует, что даже если модель обеспечивает хорошее описание данных и относительно надежные вероятности и пределы обнаружения в этом диапазоне, очень небезопасно экстраполировать оценки вероятности далеко за пределы изучаемого диапазона без существенных доказательств пригодности модели.

7.5 E5: Идентификация остатков пестицидов в пищевых продуктах с помощью ГХ-МС/МС на основе времени удерживания и отношения интенсивностей ионов

Область:

Тип качественного анализа: Анализ на основе количественных критериев

Образец/матрица: Пищевые продукты растительного происхождения

Параметр/аналит: Хлорпирифос-метил и малатион

Тип критерия классификации: интервалы времени удерживания и отношения интенсивностей характеристических ионов.

Метод/оборудование: ГХ-МС/МС

Форма представления результатов: отношение правдоподобия и вероятность того, что положительный результат достоверен.

В этом примере обсуждается оценка FP высокоселективных определений остатков пестицидов в пищевых продуктах с помощью ГХ-МС/МС посредством моделирования инструментального сигнала с использованием метода Монте-Карло. Моделирование методом Монте-Карло было выполнено в электронной таблице MS-Excel.

Идентификация аналитов основана на времени удерживания t_R в хроматографической системе и на отношении интенсивностей $AR = A_1/A_2$ двух характеристических ионов в масс-спектре аналита. Хотя t_R распределено приблизительно нормально, распределение AR может значительно отклоняться от нормального. Известно, что отношение коррелированных переменных не имеет нормального распределения, особенно если в знаменателе стоит переменная с меньшей точностью (т. е. с большим разбросом значений) [8].

Разработка и валидация методики идентификации следовых количеств аналитов в пищевых продуктах с помощью ГХ-МС/МС начинается с определения метода качественного анализа, включая подготовку проб и условия ГХ-МС/МС. Спецификация условий ГХ-МС/МС включает выбор характеристических ионов в масс-спектре аналита (ионы хлорпирифос-метила: 208 m/z и 271 m/z ; ионы малатиона: 99 m/z и 127 m/z). Затем проводят повторные вводы исходных растворов аналитов и пищевых экстрактов. Вводы исходных растворов аналита используются для изучения дисперсии t_R и интенсивностей ионов, A_1 и A_2 , а также корреляции интенсивностей ионов каждого аналита. Данные об эффективности были собраны при различных концентрациях аналита, поскольку значения и дисперсии A_1 и A_2 изменяются в зависимости от концентрации. В Таблице E5.1 представлены сводные данные по параметрам эффективности t_R , A_1 и A_2 . Повторный анализ экстрактов без детектируемых содержаний аналита был использован для определения моделей дисперсии шумов сигнала в диапазоне значений времени удерживания (Таблица E5.1). Изучены сигналы экстрактов пищевых продуктов, представляющих схемы питания пищевыми продуктами растительного происхождения с высоким содержанием воды.

На основе данных Таблицы E5.1 были разработаны модели изменчивости t_R и AR . Модели t_R были построены из доверительных интервалов на основе t -распределения Сьюдента ($\bar{t}_{Ri} \pm t_{SiRi}$; где \bar{t}_{Ri} и s_{iRi} - среднее значение и стандартное отклонение t_R , а t есть значение t -распределения для заданного доверительного уровня и степеней свободы \bar{t}_{Ri} и s_{iRi}). Модели AR были построены на основе моделирования методом Монте-Карло. На основе наблюдаемой дисперсии различных значений A_1 и A_2 аналита моделировали отношения коррелированных A_1 и A_2 (т.е. AR). На холостых экстрактах моделировали шум сигнала, а затем и AR в холостых пробах. Шум сигнала моделировали как нормальное распределение, усеченное до нуля, поскольку хроматографические пики не имеют отрицательных значений площади. В Таблице E5.2 представлены расчетные значения дисперсии t_R и AR . В этой таблице также представлены формулы MS-Excel, использованные при моделировании A_1 и A_2 . Доверительные интервалы для t_R и AR были установлены для доверительного уровня 99,9 % или 98 % соответственно, что соответствует TP .

Значение FP при идентификациях, основанных на t_R , было установлено на уровне 10% на основе опыта аналитика. Это значение FP представляет собой вероятность того, что пик, неподтвержденный как относящийся к аналиту, будет наблюдаться в пределах определенного диапазона значений времени удерживания этого аналита. Значение FP при идентификациях, основанных на AR , было оценено

Таблица Е5.1. Параметры эффективности, относящиеся к идентификации хлорпирифос-метила и малатиона в экстрактах растительного происхождения методом ГХ-МС/МС. Все параметры оценивались с 11 степенями свободы.

Аналит: Хлорпирифос-метил								
		Интенсивность						
		Время удерживания, t_R			Ион: 208 m/z		Ион: 271 m/z	
w	t_{Ri}	s_{tRi}	\bar{A}	s_A	\bar{A}	s_A	ρ	
Экстракт	(мг кг ⁻¹)	(мин)	(мин)	(а.у.)	(а.у.)	(а.у.)	(а.у.)	
Е §	3,33	17,24	0,024	105668	13,3	138678	6,14	0,9956
Е §	0,33	17,24	0,024	10163	10,5	15025	8,10	0,6151
Е §	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	-

Аналит: Малатион								
		Интенсивность						
		Время удерживания, t_R			Ион: 99 m/z		Ион: 127 m/z	
w	t_{Ri}	s_{tRi}	\bar{A}	s_A	\bar{A}	s_A	ρ	
Экстракт	(мг кг ⁻¹)	(мин)	(мин)	(а.у.)	(а.у.)	(а.у.)	(а.у.)	
Е §	3,33	19,45	0,070	226592	7,85	226765	10,3	0,9988
Е §	0,33	19,45	0,070	22354	17,4	22969	15,6	0,9672
Е §	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	-

§ – Дисперсия интенсивности иона оценивалась путем объединения сигналов от анализируемого вещества в чистом растворителе с сигналами от экстракта. Матрица экстракта: G – имбирь, O – зеленый лук, I – водоросли ирландского мха, E – неуказанная матрица.
 w – массовая доля анализируемого вещества (мг/кг); t_{Ri} – среднее время удерживания (мин) (этот параметр может меняться в зависимости от дня ввода); s_{tRi} – стандартное отклонение времени удерживания, оцененное в условиях повторяемости (мин); \bar{A} – среднее значение интенсивности иона (условные единицы измерения, а.у.); s_A – стандартное отклонение интенсивности иона; ρ – коэффициент корреляции Спирмена.

на основе моделирования шума сигнала методом Монте-Карло и определения того, сколько раз смоделированный шум даст отношение AR в пределах допустимого интервала для этого параметра. Поскольку FP может быть чрезвычайно большим для низких уровней аналита, его определяли при различных массовых долях аналита путем определения минимальной интенсивности каждого иона. В Таблице Е5.3 представлены расчетные значения TP , FP и их комбинации в $LR(+)$. В последнем столбце таблицы приведены данные по эффективности идентификации на основе как t_R , так и AR . В таблице Е5.3 также представлена апостериорная вероятность PP того, что анализируемый образец является положительным при условии, что положительные и отрицательные результаты равновероятны (т. е. $P(+)=P(-)=0,5$).

В соответствии с данными Таблицы Е5.3 идентификация «очень сильно» подтверждается доказательствами (т. е. $10^5 < LR(+) < 10^6$) только тогда, когда она выполняется на уровне или выше предела количественного определения (0,14 мг/кг и 0,38 мг/кг для хлорпирифос-метила и малатиона, соответственно). Все характеристики эффективности в Таблице Е5.3 (TP и FP , $LR(+)$ и PP), являются допустимыми альтернативами для представления эффективности или неопределенности качественного анализа при различных массовых долях аналита. Представление $LR(+)$ имеет то преимущество, что объединяет TP и FP в одном показателе и не требует предположения о преобладании определяемого пестицида в анализируемых пробах.

Таблица E5.2. Допустимые интервалы для времени удерживания и отношения интенсивностей ионов.

Аналит	Экстракт	Интервал массовой доли, w (мг кг ⁻¹)	Макс. разница t_{Ri} (мин) (с.л. 99.9 %) §	Интервал AR (с.л. 98 %) †
Хлорпирифос-метил	Е	0,04 – 3,33	0,18	0,439 – 1,18
Малатион	Е	0,11 – 3,33	0,54	0,467 – 1,54

Е – неуказанная матрица; с.л. – доверительный уровень;

§ – Максимальная разница между временем удерживания анализируемого вещества в стандартном растворе и в анализируемом образце ($\sqrt{2}t_{SIRi}$);

† – Формула MS-Excel, используемая при моделировании: Первый ион: $A_1 = \bar{A}_1 + s_{A1} *$

СТБЮДРАСПОБР (R_1, ν_1) и второй ион: $A_2 = \bar{A}_2 + s_{A2} * (СТБЮДРАСПОБР (R_1, \nu_2) * \rho +$

СТБЮДРАСПОБР(R_2, ν_2) * $(1 - \rho^2)^{0,5}$), где ν_i это степени свободы, связанные с A_i и s_{Ai} , а R_1 и

R_2 два независимых генератора случайных величин U(0,1) (Формула Excel СЛЧИС())

Таблица Е5.3. Характеристики эффективности идентификации хлорпирифос-метила и малатиона методом ГХ-МС/МС.

Аналит	<i>w</i> (мг/кг)	Характеристики эффективности для разных уровней аналита, <i>w</i>				
		<i>t_R</i>	<i>AR</i>	<i>t_R & AR</i>		
Хлорпирифос-метил	<i>TP</i> (%)	≥ 0,04	99,9	98	97,9	
	<i>FP</i> (%)	0,04 §	10	30,2	3,03	
		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	98000	979020	
	<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	
	Малатион	<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	3,29	32,9	
		0,23	9,99	98000	979020	
		0,38 †	9,99	98000	979020	
		0,77	9,99	98000	979020	
<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	

§ – Предел обнаружения; † – предел количественного определения; *t_R* – время удерживания анализируемого вещества; *AR* – отношение интенсивностей двух характеристических ионов масс-спектра аналита.

Этот пример иллюстрирует, как моделирование сигналов методом Монте-Карло может преодолеть трудности экспериментального определения уровня ложноположительных результатов высокоселективной идентификации.

7.6 Е6: Идентификация РНК SARS-CoV-2 методом амплификации нуклеиновых кислот

Область:

Тип качественного анализа: Анализ на основе количественного критерия

Образец/матрица: Мазки из носа, мазки из носоглотки и ротоглотки

Параметр/аналит: РНК SARS-CoV-2

Тип критерия классификации: Пороговый цикл, C_t . Значения, равные или ниже уровня отсечки пороговой линии C_t , классифицируются как положительные; более высокие значения классифицируются как отрицательные.

Метод/оборудование: Полимеразная цепная реакция с обратной транскрипцией (RT-PCR)

Форма представления результатов: Доля истинно положительных ответов и доля истинно отрицательных ответов.

Коронавирус 2-го типа, вызывающий острый респираторный синдром (SARS-CoV-2) — это вирус, вызывающий заболевание COVID-19, ответственное за пандемию COVID-19. Один из скрининговых тестов на наличие этого вируса в мазках из носа включает полимеразную цепную реакцию с обратной транскрипцией (RT-PCR), метод тестирования, основанный на амплификации нуклеиновых кислот (NAAT). «Внутрилабораторная» валидация этого теста включает определение уровня отсечки пороговой линии, предела обнаружения, LOD, (клинической) чувствительности (SS) и специфичности (SP) (таблица 2) с другими параметрами эффективности (в данном примере не представлены). В клиническом анализе значения SS и SP используются как оценки клинической точности. Клиническая точность ограничена значениями FN и FP , а также эпидемиологической распространенностью, типами и подтипами возбудителей, мутациями и другими биологическими факторами.

7.6.1 Пороговый цикл

«Количество циклов, необходимое для того, чтобы сигнал амплификации был обнаружен на уровне выше фона», определяется как пороговый цикл (C_t) [59] — количество циклов, необходимое для амплификации вирусной РНК до достижения обнаруживаемого уровня. Нужно оценить влияние некоторых переменных на уровень отсечки пороговой линии (C_t cut-off). R_n , нормализованное значение сигнала, — это величина сигнала, генерируемого в ПЦР при данном наборе условий. ΔR_n (Рисунок Е6.1) представляет собой нормализованное значение минус фоновый отклик. Порог — это уровень сигнала, отражающий статистически значимое увеличение по сравнению с рассчитанным фоновым сигналом. Эта линия принятия решения устанавливается для того, чтобы отличать соответствующие сигналы амплификации от фона. В этом примере программное обеспечение устанавливает пороговое значение, равное 10-кратному стандартному отклонению значения фоновой флуоресценции. Предел определяется в области, связанной с экспоненциальным ростом продукта ПЦР.

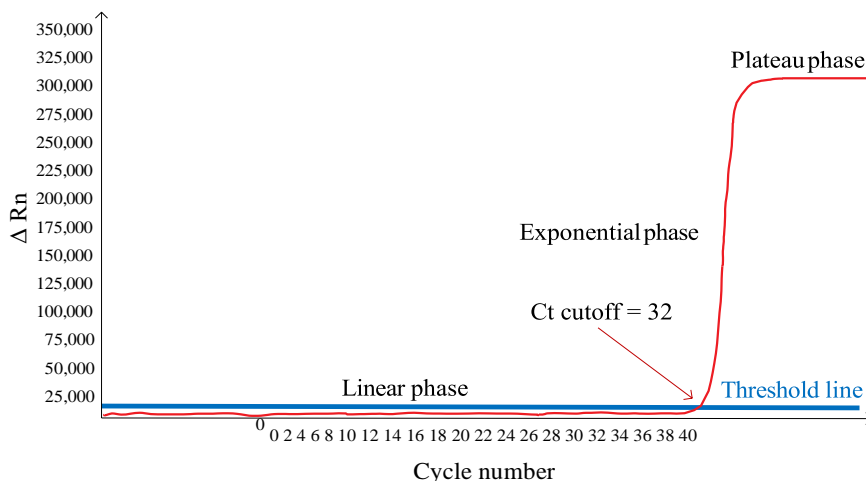


Рисунок Е6.1. Обнаружение РНК SARS-CoV-2 в истинно положительной пробе.

Рисунок Е6.1 иллюстрирует положительную классификацию по числу циклов (меньше или равно уровню отсечки Ct cut-off, равному 32) для пробы, взятой у человека. Флуоресцентные кривые с числом циклов выше 32 классифицируются как отрицательные. Обратите внимание, что только сигмоидальная форма кривых амплификации указывает на истинную амплификацию (см. Рисунок Е6.1).

7.6.2 Оценка предела обнаружения (LOD)

Возможность обнаружения вблизи уровня отсечки (Ct cut-off) оценивается с использованием предела обнаружения (LOD). LOD определяется как величина, кратная концентрации (выраженная, например, в числе копий последовательности на миллилитр), связанная с TP 95 % ($LOD_{95\%}$). $LOD_{95\%}$ оценивается путем моделирования зависимости TP от концентрации и определения концентрации, при которой TP составляет 95 %, с использованием пробит-регрессии [60] – [63]. В данном примере $LOD_{95\%}$, оцененный в серии из семи разведений образца с концентрацией копий 500 мл^{-1} , составляет 114 мл^{-1} . Более подробная информация о представленной здесь процедуре определения $LOD_{95\%}$ доступна в библиографии (п. 5.5 в [63]).

7.6.3 Клиническая точность

Оценка клинической точности включает определение целевых или минимальных значений нижних пределов 95 % CI для SS и SP , $LL_{SS,95}^{tg}$ и $LL_{SP,95}^{tg}$, и проверку того, отвечают ли оцениваемые нижние пределы соответствующим целевым значениям (т.е. $LL_{SS,95} \geq LL_{SS,95}^{tg}$ и $LL_{SS,95} \geq LL_{SP,95}^{tg}$). Для цели этого теста $LL_{SS,95}^{tg} = 95\%$ и $LL_{SP,95}^{tg} = 90\%$, то есть нижний предел чувствительности должен быть больше 95 %, а нижний предел специфичности должен быть больше 90 %.

Для валидации метода было проанализировано 200 проб мазков из носа, носоглотки и ротоглотки: 100 от лиц, о которых известно, что они инфицированы SARS-CoV-2, и 100 от лиц, у которых подтверждено отсутствие инфекции этим вирусом. В Таблице Е6.1 показана таблица сопряженности, полученная в результате 200 испытаний. Эта таблица показывает, что не было зарегистрировано ни одного ложноотрицательного результата и было три ложноположительных результата, поэтому клинические значения SS и SP составляют 100 % и 97 % соответственно. Пределы 95% CI для SS и SP , рассчитанные по формулам (8)–(11) равны [96,3 100] и [91,6, 99,0] соответственно (Таблица Е6.2). Поскольку $LL_{SS,95}$ и $LL_{SP,95}$ выше 95 % и 90 %, соответственно, аналитический метод считается обоснованным. SS дополняется проверкой сероконверсии [64].

Таблица Е6.1. Таблица сопряженности, в которой обобщаются характеристики метода обнаружения РНК SARS-Cov-2 в мазках из носа, носоглотки и ротоглотки.

		Случай		
		Положительный (pc)	Отрицательный (nc)	Общее кол-во результатов
Результат	Положительный (p)	$tp = 100$	$fp = 3$	$p = 103$
	Отрицательный (n)	$fn = 0$	$tn = 97$	$n = 97$
Общее количество случаев		$pc = 100$	$nc = 100$	

Таблица Е.6.2. Клиническая точность метода выявления РНК SARS-Cov-2 в сыворотке или плазме крови человека

Клиническая чувствительность		
$SS = 100\%$	$LL_{SS,95} = 96.3\%$	$HL_{SS,95} = 100\%$
Клиническая специфичность		
$SP = 97\%$	$LL_{SP,95} = 91.6\%$	$HL_{SP,95} = 99.0\%$

INTENTIONALLY BLANK

Приложение А – Теорема Байеса, шансы и отношение правдоподобия

А.1 Теорема Байеса

Теорема Байеса описывает, как вероятность события A (например, вероятность того, что испытуемый образец является действительно положительным) меняется с новой информацией E , например, с положительным результатом теста. Теорема Байеса чаще всего записывается для двух событий A и E в виде:

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

Здесь $P(A)$ и $P(E)$ — вероятности того, что события A и E произойдут по отдельности, $P(A/E)$ — вероятность события A при условии, что E произошло, а $P(E/A)$ — вероятность события E при условии, что A произошло. В статистике $P(A/E)$ и $P(E/A)$ обычно называют «условными вероятностями»; например, $P(E/A)$ можно назвать условной вероятностью события E при условии, что событие A произошло.

В контексте качественного анализа, взяв в качестве примера положительный случай, $P(A)$ можно понимать, как вероятность того, что случайно выбранный образец для испытаний является действительно положительным — до того, как будут выполнены какие-либо тесты. $P(E/A)$ — вероятность того, что действительно положительный испытуемый образец даст положительный результат при испытании — это доля истинно положительных результатов, TP , в Таблице 2. $P(E)$ — вероятность положительного результата испытания независимо от природы образца. Наконец, $P(A/E)$ — это вероятность того, что испытуемый образец является действительно положительным после рассмотрения информации, добавленной положительным результатом испытания. Поскольку вероятность рассчитывается после того, как появляется доказательство наступления события E , $P(A/E)$ обычно называют «апостериорной вероятностью» для A . Оцененная апостериорная вероятность дает прямое выражение достоверности, которую можно принять в классификации.

Важно помнить, что (продолжая рассматривать положительный случай) вероятность $P(E)$ включает как истинно положительные результаты, так и ложноположительные результаты, а также то, что $P(E)$ применяется ко всей совокупности испытуемых образцов. Это означает, что $P(E)$ чувствителен как к доле истинно положительных и ложноположительных результатов, так и к доле действительно положительных и отрицательных образцов. Количественно для двух случаев A и $\neg A$ (обозначая «Не- A » действительно отрицательный образец) вероятность $P(E)$ можно записать в виде взвешенной суммы:

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

Рассматривая E как положительный результат испытания, уравнение А.2 говорит, что полная вероятность события E равна уровню истинно положительных результатов, умноженному на долю действительно положительных образцов, плюс уровень ложноположительных результатов, умноженный на долю действительно отрицательных образцов. Вот почему высокий уровень ложноположительных результатов снижает уверенность в положительных результатах теста. Ссылаясь на уравнение А.1, если существует высокая вероятность положительных результатов, полученных от действительно отрицательных образцов, вероятность $P(A/E)$ уменьшается, потому что $P(E)$ увеличивается. Это соответствует интуиции; каким бы высоким ни был процент истинно положительных результатов, вероятность большого числа ложноположительных ответов должна сделать нас менее уверенными в том, что положительный результат указывает на действительно положительный испытуемый образец.

А.2 Вероятность и шансы

Вероятность P обычно выражается числом от 0 до 1. Однако ее также можно выразить в виде «шансов»,⁵ термина, возможно, более известного по спортивным ставкам. Если вероятность события A равна $P(A)$, а альтернативная возможность просто «Не A », шансы $O(A)$ в пользу A можно рассчитать

⁵ Слово «шанс» используется во множественном числе.

используя:

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

В отличие от вероятностей шансы могут принимать любые неотрицательные значения; возможны шансы 10^6 или «миллион к одному». Шансы можно преобразовать обратно в вероятности, перегруппировав А.3 так, чтобы получить:

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

А.3 Форма шансов теоремы Байеса и отношение правдоподобия

Если есть только две альтернативные и дополнительные по отношению друг к другу гипотезы, A и $\neg A$ (то есть «Не A »), и некоторое доказательство события E (например, положительный результат испытания в пользу A) используется для уточнения вероятности каждой из двух гипотез, теорема Байеса дает апостериорные вероятности как:

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

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

Тогда отношение этих вероятностей равно:

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

или, разделив сомножители справа,

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

Поскольку есть только две гипотезы, A и $\neg A$, сумма априорных и апостериорных вероятностей должна равняться 1, то есть $P(\neg A) = 1 - P(A)$ и $P(\neg A|E) = 1 - P(A|E)$. Это означает, что левая часть в уравнении (А.6) равна $P(A|E)/[1 - P(A|E)]$. По сравнению с выражением (А.3), это просто шансы в пользу A , когда событие E произошло, т.е. $O(A|E)$, «апостериорные шансы» в пользу гипотезы A . Точно так же априорные шансы $O(A)$ появляются в правой части А.6 как $P(A)/P(\neg A) = P(A)/[1 - P(A)] = O(A)$. Оставшееся отношение, $P(E|A)/P(E|\neg A)$ известно как «отношение правдоподобия». Для случая качественного анализа, когда $\neg A$ соответствует действительно отрицательному испытываемому образцу, в Таблице 5 дано (оцененное) отношение правдоподобия TP/FP .

Следовательно, формула шансов теоремы Байеса может быть записана как

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

или, схематически:

$$\text{Апостериорные шансы} = \text{Априорные шансы} \times \text{Отношение правдоподобия}$$

Таким образом, отношение правдоподобия можно количественно интерпретировать как изменение шансов в пользу конкретной гипотезы.

Приложение В – Качественный анализ, связанный с оценкой соответствия количественному пределу

В.1 Оценка соответствия как качественный анализ

Оценку соответствия значения количественного параметра анализируемого объекта предельному значению или интервалу можно считать качественным анализом по одному количественному критерию (раздел 2) с результатами «соответствует» или «не соответствует». В таблице В.1 представлены некоторые примеры таких типов анализа.

Таблица В.1. Примеры качественного анализа, основанного на оценке соответствия значения количественного параметра анализируемого объекта предельному значению или интервалу.

-
- (1) Оценка цвета сырья путем сравнения измерений поглощения с пороговым значением.
 - (2) Оценка соответствия сплава минимальному содержанию основного компонента.
 - (3) Оценка соответствия лекарственного средства установленному интервалу концентрации действующего вещества.
 - (4) Оценка соответствия остаточного содержания пестицида во фруктах с учетом максимального уровня остаточного содержания.
 - (5) Оценка состояния здоровья человека путем сравнения измеренного компонента крови с интервалом значений у здоровых людей.
-

Использование правил принятия решений и неопределенности измерений при оценке соответствия подробно обсуждается в руководстве Eurachem/CITAC «Использование информации о неопределенности при оценке соответствия» [29] («Руководство по соответствию»). Однако для полноты в этом Приложении обсуждается, как информация о неопределенности или эффективности для количественного анализа может быть использована для получения некоторых показателей в Таблице 2. Затем их можно использовать для характеристики эффективности процедур качественного анализа, полностью или частично основанных на сравнении результатов измерений с установленным пределом или спецификацией.

Когда анализ включает в себя оценку того, находится ли измеренное свойство выше, ниже или внутри пределов или интервала спецификации, неопределенность измерения может использоваться для количественного выражения надежности оценки соответствия.

ПРИМЕЧАНИЕ. В настоящем Руководстве не обсуждается, как следует оценивать неопределенность измерений. Оценка неопределенности измерений подробно описана в руководстве Eurachem/CITAC «Количественное описание неопределенности в аналитических измерениях» [65].

Использование неопределенности измерения для решений о соответствии, описанное в руководстве по соответствию Eurachem/CITAC [29], включает установление критерия для принятия решения о соответствии или несоответствии объекта с максимальной вероятностью ошибочных решений о соответствии x %. В Руководстве по соответствию различаются «специфические» и «глобальные» риски. «Специфический риск» количественно определяет вероятность ошибочного решения о соответствии конкретного образца; он основан исключительно на распределении, связанном с результатом измерения для этого образца. Напротив, «глобальный риск» количественно определяет вероятность ошибочных решений о соответствии случайно выбранного будущего образца [66]. Глобальный риск учитывает распределение возможных значений для измеряемых образцов, например, распределение значений в изделиях, полученных на производственной линии, или объектах окружающей среды. Например, для расчета глобального риска производителя требуется вероятность того, что производственная линия будет производить продукцию со значением определяющего параметра, близким к предельному значению, так что она может быть ошибочно признана несоответствующей. Следовательно, для определения глобального риска необходимо хорошо охарактеризовать распределение значений совокупности образцов.

Чаще всего аналитиков интересует оценка соответствия конкретного объекта анализа. Как в таких случаях можно определить показатели, используемые для количественной оценки надежности других типов качественного анализа? Разбор конкретного случая и формулы, используемые для определения этих параметров, представлены ниже.

В.2 Положительные и отрицательные результаты оценки соответствия

Если соответствие требованиям рассматривается как «положительный результат», распределение, описывающее неопределенность измерения, может быть использовано для получения либо отношения правдоподобия $LR(+)$ в пользу соответствия, либо, при некоторых обстоятельствах, апостериорной

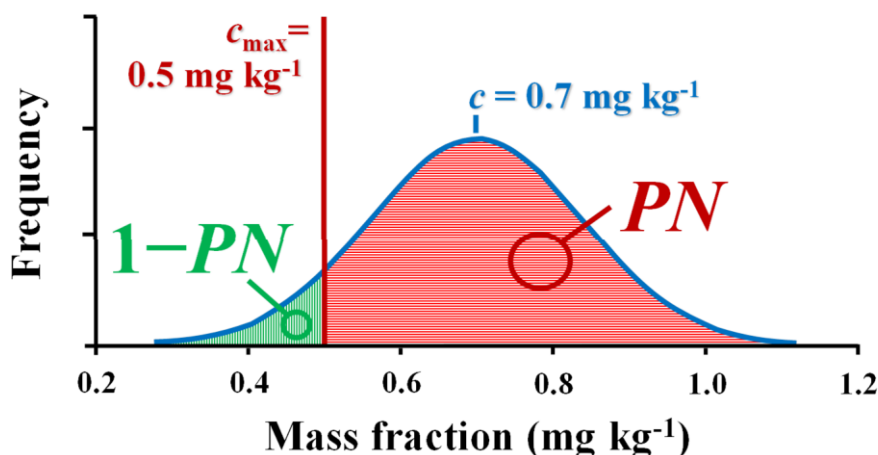


Рисунок В.1. Графическое представление вероятности PN того, что анализируемый образец не соответствует максимальному пределу, c_{max} , на основе измеренного значения, c , и связанной с ним стандартной неопределенностью, $u(c) = 0.14 \text{ мг кг}^{-1}$, и соответствующей вероятности $1 - PN$ того, что соответствует.

вероятности PP того, что объект действительно соответствует («положительный случай»), и вероятности $(1 - PP)$ того, что объект фактически не соответствует («отрицательный случай»). Эквивалентно, если результат сравнения с пределом или интервалом сообщается как «отрицательный», т.е. означающий «несоответствие», неопределенность измерения может быть использована для получения соответствующего отношения правдоподобия $LR(-)$ или апостериорных вероятностей PN и $(1 - PN)$. В приведенном ниже примере рассматриваются апостериорные вероятности.

ПРИМЕЧАНИЕ. Определение «положительных результатов» и «отрицательных результатов» как «соответствующих» и «несоответствующих требованиям», соответственно, является произвольным; можно следовать противоположному соглашению.

В.3 Пример – Оценка соответствия для остатков пестицидов во фруктах

Предположим, что соответствие образца винограда оценивается по максимальному уровню остаточного содержания, c_{max} , $0,5 \text{ мг кг}^{-1}$ ацетамиприда [67], а измеренная массовая доля, c , в образце составляет $0,70 \text{ мг кг}^{-1}$.⁶ Результат измерения имеет нормальное распределение со стандартной неопределенностью, $u(c)$, $0,14 \text{ мг кг}^{-1}$ оцененной с очень большим числом степеней свободы. Поскольку $c > c_{max}$, наиболее вероятным выводом о соответствии является «несоответствие» винограда («отрицательный результат»).

Формально апостериорная вероятность, такая как PP или PN , требует знания априорной вероятности. В данном случае информация об общем распределении ацетамиприда в винограде отсутствует. Однако в случаях подозрения на загрязнение часто разумно предположить, что распределение настолько широкое, что в области результата измерения оно практически неинформативно. Если это так,

⁶ Символ «с» для обозначения концентрации используется в настоящем Руководстве в случаях, относящихся к различным величинам, таким как массовая концентрация, массовая доля и рН.

неопределенность измерения может быть принята как аппроксимация апостериорного распределения. При таком подходе апостериорная вероятность отрицательного (не соответствующего требованиям) анализируемого образца, PN , представляет собой площадь под функцией плотности вероятности результата измерения справа от c_{max} , представленную на рисунке В.1 красным цветом. Эта площадь соответствует верхнему хвосту нормального распределения со средним значением c и стандартным отклонением, равным $u(c)$. Его можно вычислить с помощью электронной таблицы или статистического пакета; в Microsoft Excel, например, необходимая формула: $1-NORM.PАСП(c_{max}=0,5, c=0,7, u(c)=0,14, ИСТИНА)$ (см. Таблицу В.2). Для этого примера площадь равна 0,923, или 92,3 %. Соответствующая вероятность PP того, что образец является положительным (соответствующим требованиям), равна $(1-PN)$ или 7,6 %.

Там, где доступна информативная априорная информация и ее целесообразно использовать, расчеты включают интегрирование по априорному распределению. Результаты интегрирования нормально распределенной априорной вероятности и неопределенность измерений приведены, например, в JCGM 106 [66] вместе с рекомендациями по другим распределениям.

ПРИМЕЧАНИЕ На удивление трудно установить действительно информативное априорное распределение.

Например, простое равномерное распределение в диапазоне, показанном на рисунке В.1 (0,2–1,2 мг/кг), будет соответствовать 30 % априорной вероятности того, что испытуемый образец соответствует пределу, просто потому, что только 30 % этого диапазона находится ниже предела 0,5 мг/кг. Таким образом, полный байесовский анализ будет включать проверку, обычно включающую альтернативные варианты априорного распределения, чтобы убедиться, что вывод не слишком чувствителен к выбору априорного распределения.

В.4 Табличные формулы для вероятностей оценки соответствия

В Таблице В.2 представлены формулы MS-Excel, которые следует использовать, когда рассматриваются различные пределы или интервалы соответствия, а измеренное значение находится ниже, выше, внутри или вне этих пределов.

Если стандартная неопределенность, $u(c)$, оценивается с небольшим числом степеней свободы, v , вместо описания дисперсии оценки измеряемой величины нормальным распределением следует рассматривать t -распределение Стьюдента. В этом случае общая формула, используемая в таблице В.2, $NORM.PАСП(C, c, u(c), ИСТИНА)$, должна быть заменена на $СТЮДЕНТ.PАСП(ABS(C - c), u(c), v, ИСТИНА)$.

Таблица В.2. Формулы MS Excel, используемые для оценки вероятности соответствия, PP , или несоответствия, PN , при принятии решения по конкретному анализируемому образцу. Эти формулы можно использовать для вычисления $LR(+) = PP/(1 - PP)$ и $LR(-) = PN/(1 - PN)$

	Соответствие образца (тип результата)	Критерий надежности соответствия	Формула MS-Excel (на основе совокупного нормального распределения)
S Предел	Сценарий		
1	Мах Соответствие («положительный результат») $c \leq c_{max}$	PP	$NORM.PАСП(C_{max}, c, u(c), ИСТИНА)$
2	Мах Несоответствие («отрицательный результат») $c \geq c_{max}$	PN	$1-NORM.PАСП(C_{max}, c, u(c), ИСТИНА)$
3	Мин Соответствие («положительный результат») $c \geq c_{min}$	PP	$1-NORM.PАСП(C_{min}, c, u(c), ИСТИНА)$
4	Мин Несоответствие («отрицательный результат») $c \leq c_{min}$	PN	$NORM.PАСП(C_{min}, c, u(c), ИСТИНА)$
5	Inter Соответствие («положительный результат») $c_{min} \leq c \leq c_{max}$	PP	$NORM.PАСП(C_{max}, c, u(c), ИСТИНА) - NORM.PАСП(C_{min}, c, u(c), ИСТИНА)$
6	Inter Несоответствие («отрицательный результат») $c \geq c_{max}$ ИЛИ $c \leq c_{min}$	PN	$1-NORM.PАСП(C_{max}, c, u(c), ИСТИНА) + NORM.PАСП(C_{min}, c, u(c), ИСТИНА)$

S – сценарий; Inter., Max. или Min. – Интервал, максимальный или минимальный предел; Положительный или отрицательный результат – соответствующий или несоответствующий результат; c и $u(c)$ – измеренная концентрация и связанная с ней стандартная неопределенность; c_{max} или c_{min} – максимально или минимально допустимая концентрация.

INTENTIONALLY BLANK

Библиография

- [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. Pennecchi, "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.

INTENTIONALLY BLANK

