|  |  |  |
| --- | --- | --- |
| **Федеральное агентство**  **по техническому регулированию и метрологии** | | |
| https://rostestsibir.ru/wp-content/uploads/2016/08/image12.jpeg | **национальный**  **стандарт**  **российской**  **федерации** | **ГОСТ Р**  *(проект, первая редакция)* |

**Изделия медицинские для диагностики *in vitro***

**Контроль функциональных характеристик качественных методов исследования**

*Настоящий проект стандарта не подлежит применению до его утверждения*

**Москва**

**Российский институт стандартизации**

**202**

**Предисловие**

1 РАЗРАБОТАН Ассоциацией специалистов и организаций лабораторной службы «Федерация лабораторной медицины» (Ассоциация «ФЛМ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 380 «Клинические лабораторные исследования и диагностические тест-системы ин витро»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от №

4 В настоящем стандарте учтены основные нормативные положения международного документа CLSI EP12 «Изделия медицинские для диагностики in vitro. Контроль функциональных характеристик качественных методов исследования» (CLSI EP12 «Evaluation of Qualitative, Binary Output Examination Performance»)

5 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в* [*статье 26*](https://login.consultant.ru/link/?req=doc&base=LAW&n=200912&date=22.04.2021&dst=100282&fld=134) *Федерального закона от 29 июня 2015 г. N 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок* – *в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования* - *на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (www.rst.gov.ru)*

© Оформление. ФГБУ «Институт стандартизации», 202

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

**Содержание**

|  |  |
| --- | --- |
| 1 Область применения……………………………………………………………………. |  |
| 2 Нормативные ссылки.....………………………………………………………….……. |  |
| 3 Основания ...........………………………………………………………….…….............. |  |
| 4 Стандартные меры предосторожности..........……………….................................... |  |
| 5 Термины, определения, сокращения и символы.................................................... |  |
| 6 Качественные исследования..................................................................................... |  |
| 7 Разработка качественного исследования с бинарным выводом........................... |  |
| 8 Валидация качественного исследования с бинарным выводом............................. |  |
| 9 Лабораторная проверка заявленных характеристик............................................... |  |
| Приложение А (справочное) Примеры исследований................................................ |  |
| Приложение Б (справочное) Связь между распределением внутреннего  непрерывного ответа и пробит- или логит-моделями...................... |  |
| Приложение В (справочное) Качественное бинарное исследование с визуальным чтением................................................................................................ |  |
| Приложение Г (справочное) Оценка точности секвенирования следующего  поколения.......................................................................................... |  |
| Приложение Д (справочное) Определение нижнего предела обнаружения для  качественных исследований, основанных на методах  полимеразной цепной реакции.......................................................... |  |
| Приложение Е (справочное) Схемы прецизионных исследований............................ |  |
| Приложение Ж (справочное) Исследования точности наблюдателей...................... |  |
| Приложение И (справочное) Метод оценки Уилсона для расчета доверительных интервалов........................................................................................... |  |
| Приложение К (справочное) Примеры анализа клинической эффективности......... |  |
| Приложение Л (справочное) Пример полимеразной цепной реакции в реальном времени для ванкомицин-резистентных энтерококков................ |  |
| Библиография............................................................................................................. |  |

**Введение**

Результаты диагностических исследований *in vitro* показывают наличие или характеристики клинического свойства. Количественные исследования (измерения) измеряют и представляют числовое значение свойства клинических образцов. Другие исследования клинических образцов сообщают о характеристиках свойства путем отнесения их к двум бинарным категориям: неупорядоченным (номинальным) и упорядоченным (порядковым). В настоящем стандарте рассматривают качественные исследования, которые дают двоичные (бинарные, например «да»/«нет», «положительный»/«отрицательный») номинальные результаты (примеры см. в приложении A). Данные исследования охватывают широкий спектр специальностей медицинских лабораторий, медицинских целей, технологий измерения и типов представляемых результатов. Бинарный ответ создается на основе:

- внутренней непрерывной реакции устройства и «отсечки» для получения двоичных результатов;

- алгоритмических методов принятия решений, определяющих наличие аналита;

- в некоторых случаях можно ответить «да» или «нет» без помощи приборов.

Эффективность исследований *in vitro* должна быть оценена во время и после их разработки, а их эффективность должна быть подтверждена до того, как результаты исследования будут использованы для принятия клинических решений. Настоящий стандарт рассматривает разработку качественных, бинарных, имеющих результат или «выходных» исследований (в дальнейшем именуемых качественными бинарными исследованиями) и призвано способствовать созданию единообразия в оценке эффективности:

а) разработчиков качественных бинарных исследований:

- для проектирования и разработки исследований;

- установления и проверки результатов исследований на основании того, как они разработаны;

б) лабораторий, проверяющих качественные бинарные исследования перед вводом в эксплуатацию;

в) лабораторий, которые разрабатывают свои собственные качественные бинарные исследования.

Многие количественные исследования предоставляют значения измеряемой величины в единицах плюс «порог принятия решения», который можно применить для получения бинарной интерпретации (см. примеры исследований в приложении A). Хотя эффективность бинарного вывода этих исследований может быть определена с помощью методов, приведенных в настоящем стандарте, оценки эффективности, разработанные для количественных методов, обеспечивают большую гибкость и лучше выявляют различия в эффективности. Поэтому оценка эффективности количественных исследований должна основываться на рекомендациях, приведенных в иных международных руководствах[[1]](#footnote-1)).

В настоящем стандарте описаны категории качественных исследований с бинарным выходом. Каждая категория характеризуется тем, определяет ли она наличие целевого состояния, присутствует ли аналит, и все ли образцы имеют внутренний непрерывный ответ. Основные описанные оценки эффективности включают неточность, которая сосредоточена на оценке C5 и C95, и классификацию, в которой подробно описаны методы определения клинической чувствительности и специфичности или положительное процентное соглашение и отрицательное процентное согласие. Среди других параметров эффективности − стабильность и помехи. Обсуждение этих тем проходит через весь цикл работы с продуктом, начиная с определения типа исследования, установления характеристик, подтверждения утверждений о продукте и, наконец, проверки утверждений в лаборатории.

**НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ИЗДЕЛИЯ МЕДИЦИНСКИЕ ДЛЯ ДИАГНОСТИКИ *IN VITRO***

**Контроль функциональных характеристик качественных методов   
исследования**

Medical devices for *in vitro* diagnostics. Control of functional characteristics of qualitative   
research methods

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Дата введения –**

#### 1 Область применения

Настоящий стандарт устанавливает руководство по проектированию исследований и протоколы для оценки эффективности этапов создания и внедрения этапов модели испытаний (см. [1]). Настоящий стандарт характеризует целевое состояние (ЦС), имеющее только два возможных результата (например, «положительный»/«отрицательный», «присутствует»/«отсутствует», «реактивный»/«нереактивный»). Настоящий стандарт может быть применен как для производителей качественных, бинарных исследований с получением результатов (в дальнейшем именуемых качественными бинарными исследованиями), так и для медицинских лабораторий, создающих лабораторные бинарные исследования (в обоих случаях именуемых разработчиками). Протоколы, содержащиеся в настоящем стандарте, также предназначены для того, чтобы помочь пользователям проверить эффективность исследований в их собственной тестовой среде. Оценка эффективности исследований, которые предоставляют результаты с более чем двумя возможными категориями в неупорядоченном (номинальном) наборе или результатом которых является информация о порядковых категориях, не входит в область применения настоящего стандарта.

#### 2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

#### ГОСТ Р ИСО 5725-1 Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения

#### ГОСТ Р ИСО 5725-2 Точность (правильность и прецизионность) методов и результатов измерений. Часть 2. Основной метод определения повторяемости и воспроизводимости стандартного метода измерений

#### 3 Основания

Часто бывает необходимо предоставить результаты метода исследования, которые имеют двоичный результат (например, «да»/«нет»). Медицинские работники могут назначить исследование, чтобы определить наличие заболевания у пациента, присутствие аналита в образце, беременность женщины или положительные результаты теста на наркотики. Для таких исследований используют две основные оценки: клиническая эффективность (чувствительность и специфичность) и точность.

Оценку клинических показателей при создании и валидации процесса бинарной категоризации проводят путем исследования образцов от субъектов с известной категорией (т.е. «ЦС присутствует» по сравнению с «ЦС отсутствует»). В идеале это исследование включает в себя сравнение бинарных результатов для предполагаемой популяции образцов, полученных в ходе исследования-кандидата, с независимой процедурой, которая обеспечивает наилучшую доступную оценку ЦС. Однако во многих случаях такое сравнение бинарных результатов представляет собой оценку согласия с положительными или отрицательными результатами сравнительного исследования, которое не является идеальной процедурой оценки.

Качественное исследование, определяющее неточность, зависит от представления о значениях в соответствующей шкале, где бинарное исследование признает образец положительным в 5 % случаев (C5) и в 95 % случаев (C95). В отличие от клинических показателей, способы определения C5 и C95 различаются в зависимости от того, как проводят бинарную категоризацию (см. 7.2.4 и 7.3).

Стабильность реагентов и помехи при исследовании − еще два показателя эффективности систем исследования. Хотя обе эти темы рассматриваются в соответствующих руководствах (например помехи рассматриваются в [2], а стабильность – в и [3]), существует несколько положений, уникальных для качественных исследований, которые требуют дополнительного рассмотрения в настоящем стандарте. Более подробная информация по данным положениям приведена в 7.5 (стабильность) и 8.4 (помехи).

#### 4 Стандартные меры предосторожности

Поскольку зачастую невозможно определить, какие изоляты или пробы могут быть зараженными, все пробы пациентов и лабораторные образцы рассматривают как зараженные и обрабатывают в соответствии со стандартными мерами предосторожности. Стандартные меры предосторожности содержат рекомендации, которые сочетают в себе основные черты универсальных мер предосторожности и практики изоляции биологических веществ.

Стандартные меры предосторожности охватывают передачу всех известных инфекционных агентов и, таким образом, являются более всеобъемлющими, чем универсальные меры предосторожности, которые предназначены только для передачи патогенов, передающихся через кровь. Опубликовано руководства, в которых рассматриваются ежедневные операции по диагностике заболеваний человека и животных, а также поощряется культура безопасности в лаборатории (см. [4]). Конкретные меры предосторожности по предотвращению передачи всех известных инфекционных агентов из лабораторных инструментов и материалов, а также рекомендации по действиям при контакте со всеми известными инфекционными заболеваниями приведены в соответствующем руководстве (см. [5]).

#### 5 Термины, определения, сокращения и символы

**5.1 Терминология**

Приоритетом для организаций всего мира, связанных со стандартизацией в области медицины, является достижение глобальной гармонизации во всех возможных случаях. Гармонизация − это процесс признания, понимания и объяснения различий при одновременном принятии мер по достижению единообразия во всем мире. Необходимо отметить, что медицинские конвенции в мировом метрологическом сообществе развивались по-разному в разных странах и регионах, и что юридически обязательное использование терминов, региональное использование и различные сроки консенсуса являются важными факторами в процессе гармонизации. В рамках настоящего стандарта процесс консенсуса сосредоточен на гармонизации терминов для облегчения глобального применения стандартов и руководств. Таблица 1 приведена для разъяснения предполагаемых толкований следующих терминов.

Таблица 1 − Общие термины и фразы с предполагаемыми толкованиями

|  |  |
| --- | --- |
| Термин или фраза | Предполагаемое толкование |
| «Должен» или  «обязан» | Объясняет действие, непосредственно связанное с выполнением нормативных и/или аккредитационных требований, или указывает на необходимый шаг для обеспечения безопасности пациента или надлежащего выполнения процедуры |
| «Требуется» | Представляет собой заявление, которое непосредственно отражает нормативные требования, требования аккредитации, требования к производительности, продукции или организации, либо требования или спецификации, приведенные в стандарте |
| «Следует» | Описывает рекомендацию, приведенную в лабораторных документах, утверждение о надлежащей лабораторной практике или предложение о том, как выполнить требование |

**5.2 Термины и определения**

В настоящем стандарте применены следующие термины с соответствующими определениями. Связанные с ними термины и определения доступны в базе данных гармонизированной терминологии CLSI по адресу <https://htd.clsi.org>.

5.2.1 **согласие:** Доля проб, для которых результаты, полученные с помощью исследования-кандидата, и результаты, полученные с помощью сравнительного исследования, совпадают.

Примечания

1 Согласие также может быть измерено в пределах исследования с использованием одной и той же популяции испытуемых, но с учетом двух факторов, таких как операторы, участки, приборы или партии реагентов.

2 См. отрицательное процентное согласие и положительное процентное согласие.

5.2.2 **аналит:** Компонент, представленный в названии измеряемой величины (см. [6]).

Примечания

1 В типе величины «масса белка в 24-часовой моче» аналит – «белок». В «количестве вещества глюкозы в плазме крови» аналит – «глюкоза». В обоих случаях вся фраза в кавычках обозначает измеряемую величину (см. [6]).

2 В типе величины «каталитическая концентрация изофермента лактатдегидрогеназы 1 в плазме крови» аналитом является «изофермент лактатдегидрогеназы 1». Вся фраза в кавычках обозначает измеряемую величину (см. [7]);

3 В случае качественных исследований с обнаружением аналитов аналит − единица, которую можно подсчитать, например, «копии определенной бактерии», для которой указанная бактерия является аналитом.

5.2.3 **анализатор:** Прибор и/или устройство для обработки и перемещения проб, выполняющее измерения клинически значимых аналитов в пробах пациента.

Примечание – Часть пробы пациента расходуется в процессе исследования.

5.2.4 **оценка** **(целевого состояния [ЦС]):** Наилучший доступный метод оценки целевого состояния.

Примечания

1 Оценку (целевого состояния обозначают как клинический эталонный стандарт в Стандартах отчетности об исследованиях диагностической точности (STARD) (см. [8], [9]).

2 Классификацию данным методом рассматриваютя как истинное бинарное определение.

3 Может быть методом, разработанным на основе клинического консенсуса, который определяет, является ли проба от человека с ЦС или без него.

4 Для исследований, связанных с обнаружением аналитов, оценка может быть выполнена, например, с помощью композитного эталонного метода, включающего несколько исследований и/или методов и алгоритм объединения их результатов в бинарную категорию.

5.2.5 **смещение (измерения):** Оценка систематической ошибки измерения (см. [10]).

Примечание – Для качественного, бинарного исследования смещение обычно определяют на срезе и выражают в терминах внутреннего непрерывного ответа инструмента.

5.2.6 **C5:** Значение в соответствующей шкале (например, внутренний непрерывный ответ), при котором бинарное исследование признает образец положительным в  
5 % случаев.

Примечание – Как и точность, C5 связан с оговоренными условиями точности.

5.2.7 **C95:** Значение в соответствующей шкале (например, внутренний непрерывный ответ), при котором бинарное исследование признает образец положительным в 95 % случаев.

Примечание – Как и точность, C95 связан с оговоренными условиями точности.

5.2.8 **сертифицированный стандартный образец:** Стандартный образец, сопровождаемый документами, выданными авторитетной организацией и обеспечивающий одно или несколько заданных значений свойств с соответствующими неопределенностями и прослеживаемостью, с использованием действительных процедур (см. [6].

5.2.9 **клиническая эффективность:** Способность бинарного исследования различать субъектов с целевым состоянием и без него.

***Пример − Дискриминация между беременными и небеременными женщинами***.

Примечания

1 Клиническую эффективность обычно измеряют с помощью таких параметров, как клиническая чувствительность и клиническая специфичность.

2 На клиническую эффективность могут влиять предвзятость, неточность и способ принятия бинарного решения.

5.2.10 **клиническая чувствительность [чувствительность]:** Способность исследования дать положительный результат для субъектов из предполагаемой популяции, имеющих целевое состояние, на которое они исследуются.

Примечания

1 Это доля субъектов с ЦС, имеющих положительный результат исследования, и часто выражается в процентах.

2 Ранее использовался термин «диагностическая чувствительность».

5.2.11 **клиническая специфичность [специфичность]:** Способность исследования давать отрицательный результат для субъектов из целевой популяции, не имеющих целевого заболевания, на которое они исследуются.

Примечания

1 Это доля субъектов без ЦЗ, имеющих отрицательный результат теста, и часто выражается в процентах.

2 Ранее использовался термин «диагностическая специфичность».

5.2.12 **отсечка:** Внутренний порог непрерывного ответа для бинарного исследования, выше которого результат считается положительным, а ниже – отрицательным.

Примечания

1 Конкурентные исследования ведут себя противоположным образом, результаты при меньших значениях ответа, чем отсечка, считают положительными.

2 Отсечку устанавливают во время разработки исследования; клинические результаты не могут быть оценены до тех пор, пока не будет установлена отсечка.

3 Некоторые бинарные исследования могут иметь эквивокальную зону, для которой результат является неопределенным. Информация приведена в разделе «Эквивокальная зона».

5.2.13 **эквивокальная зона [неопределенная зона]:** Интервал внутреннего непрерывного ответа (ВКР) вокруг среза, в пределах которого окончательный результат не может быть представлен как положительный или отрицательный, а определен как «эквивокальный», с указанием или без указания фактического значения результата.

Примечания

1 В большинстве случаев эквивокальный результат запускает последующие шаги для получения бинарного результата.

2 Многие бинарные исследования включают внутренний контроль, который может указать, когда ВКР ненадежен или результат не может быть представлен. Эквивокальная зона отличается от эквивокального результата.

3 Примеры исследований, приведенные в настоящем стандарте, предполагают наличие всех бинарных результатов (т. е. отсутствие пропущенных или исключенных результатов).

4 Также известна как «серая зона» в медицинской лаборатории и как «защитная полоса» в химической метрологии.

5.2.14 **исследование:** Совокупность операций, целью которых является определение значений или характеристик объекта (см. [11]).

Примечание – Лабораторные исследования, определяющие значения объекта, являются измерительными процедурами; исследования, определяющие характеристики объекта, являются качественными исследованиями.

5.2.15 **исследование (бинарное) [качественное бинарное исследование]:** Исследование, дающее двоичный результат.

Примечания

1 Бинарные исследования могут иметь внутренний непрерывный ответ, который имеет количественные значения. Однако не требуется характеризовать эти значения, за исключением близости к отсечке, используемой для получения бинарного результата.

2 В настоящем стандарте во всех случаях, когда термин «качественное исследование» используют без модификатора, предполагают, что он означает качественное бинарное исследование.

5.2.16 исследование (номинальное): Исследование, которое характеризует выборку по нескольким возможным неупорядоченным категориям.

***Пример − Группа крови.***

5.2.17 **исследование (порядковое):** Исследование, которое характеризует выборку по нескольким возможным упорядоченным категориям.

Примечание – Обычные примеры дают такие выходные данные, как 0, +, ++, +++ и ++++.

5.2.18 **процент попаданий:** Доля положительных результатов исследования в выборке.

5.2.19 **неточность:** Случайная дисперсия набора реплицированных измерений и/или значений, выраженная количественно с помощью статистики, такой как стандартное отклонение или коэффициент вариации.

Примечания

1 Для качественного, бинарного исследования неточность может быть охарактеризована диапазоном интервала от C5 до C95.

2 См. точность (измерений).

5.2.20 **внутренний непрерывный ответ**; *ICR*: Индикация, получаемая от анализатора при исследовании образца.

Примечания

1 В некоторых качественных исследованиях эту индикацию сравнивают со значением отсечки для получения бинарного результата.

2 Внутренний непрерывный ответ также можно назвать сигналом.

5.2.21 **предел холостого хода**; ПХХ: Наивысший результат измерения, который может наблюдаться (с установленной вероятностью [α]) для холостого образца.

Примечание – ПХХ также называют «критическим значением чистой переменной состояния» (см. [12]).

5.2.22 **нижний предел обнаружения**; *LLoD*: Значение измеряемой величины, полученное при данной процедуре исследования, для которого вероятность ложного утверждения об отсутствии исследуемого вещества в материале равна β, при вероятности α ложного утверждения о его присутствии (см. [10]).

Примечания

1 В количественных процедурах − наименьшая концентрация аналита, которая может быть стабильно обнаружена как превышающая предел холостого хода (обычно в более 95 % проб, исследованных в обычных условиях медицинской лаборатории и в определенном типе проб).

2 Также называется «минимально обнаруживаемая концентрация» (или значение), «предел обнаружения» и формально обозначается как «минимально обнаруживаемое значение переменной состояния» (см. [12]).

3 При качественном исследовании с обнаружением аналита *LLoD* определяют с помощью исследования частоты попаданий.

5.2.23 **измеряемая величина:** Величина, предназначенная для измерения (см. [10]).

Примечания

1 Спецификация измеряемой величины для количественного исследования требует знания вида величины, описания состояния явления, тела, или вещество, несущее количество, включая любой соответствующий компонент, и участвующие химические соединения (см. [10]).

2 Хотя качественные исследования могут иметь основной сигнал, который указывается таким же образом, как указано выше для количественных исследований, основная цель качественного исследования состоит в том, чтобы описывать, а не количественно оценивать. Поэтому определение измеряемой величины для качественного исследования требует знания основного сигнала, местоположения отсечки в шкале сигнала и описания двух взаимоисключающих состояний, которые должны быть классифицированы.

5.2.24 **метод измерения:** Общее описание логической организации операций, используемых при измерении (см. [10]).

5.2.25 **процедура измерения:** Подробное описание измерения в соответствии с одним или несколькими принципами измерения и заданным методом измерения, основанное на модели измерения и включающее любые расчеты для получения результата измерения (см. [10]).

Примечание − Обычно используют для описания количественного исследования.

5.2.26 **отрицательное процентное согласие**; *NPA*: Доля отрицательных результатов сравнительного метода, при которых результат метода исследования является отрицательным, если сравнительный метод не является общепризнанным лучшим методом оценки целевого состояния.

5.2.27 **образец пациента:** Одна или несколько частей, взятых из пробы (см. [10]), полученного в условиях клинического лечения, ориентированного на пациента.

***Пример − Порция сыворотки, взятая из пробы свернувшейся крови.***

5.2.28 **положительное процентное согласие**; *PPA*: Доля положительных результатов сравнительного метода, при которых результат метода исследования является положительным, если сравнительный метод не является общепризнанным лучшим методом оценки целевого состояния.

5.2.29 **прецизионность (измерений):** Близость согласия между показаниями или значениями измеряемой величины, полученными при повторных измерениях одного и того же или аналогичных объектов в заданных условиях (см. [10]).

Примечания

1 Прецизионность измерений обычно выражается численно через меры неточности, такие как стандартное отклонение, дисперсия или коэффициент вариации при заданных условиях измерений (см. [10]).

2 Иногда (для образцов с низкими значениями) прецизионность описывается   
2,5-го и 97,5-го перцентилей.

3 «Указанные условия» могут быть, например, условиями повторяемости измерений (в рамках одного и того же цикла), условиями промежуточной точности измерений или условиями воспроизводимости измерений (см. [10]).

4 Точность измерений используют для определения повторяемости измерений, промежуточной точности измерений и воспроизводимости измерений (см. [10]).

5.2.30 **эталонный материал:** Материал, достаточно однородный и стабильный в отношении заданных свойств, который, как было установлено, пригоден для использования по назначению при измерении или исследовании номинальных свойств (см. [10]).

5.2.31 **эталонный метод выполнения измерений:** Метод выполнения измерений, принятый как обеспечивающий результаты измерений, пригодные для использования по назначению при оценке истинности значений измеряемой величины, полученных в результате других методик выполнения измерений величин того же вида, при калибровке или при определении характеристик стандартного образца (см. [6]).

5.2.32 **повтор:** Значение, полученное в результате повторного исследования того же образца или аналогичных объектов, если это необходимо.

5.2.33 **воспроизводимость (измерения):** Точность измерения при воспроизводимых условиях измерения (см. [10]).

Примечания

1 Соответствующие статистические термины приведены в ГОСТ Р ИСО 5725-1 и   
ГОСТ Р ИСО 5725-2 (см. [10]).

2 Для целей иммуногематологического исследования воспроизводимость обычно сводится к проверке способности автоматизированной системы последовательно давать одну и ту же интерпретацию.

3 В более широком смысле определяется как согласие между повторными измерениями одной и той же пробы при различных условиях.

4 Близость согласия между результатами измерений одного и того же измеряемого вещества, проведенных в измененных условиях измерения

5 Измененные условия могут включать принцип или метод измерения, наблюдателя, средство измерения, место, условия использования и время.

6 Оперативные условия должны быть всегда указаны.

7 В рамках настоящего стандарта условия воспроизводимости измерений могут содержать межприборные или межлабораторные данные.

5.2.34 **прогон:** Интервал, в течение которого ожидается стабильность аналитических характеристик испытательной системы, но он не может быть длиннее периода времени, рекомендованного производителем.

Примечания

1 Между аналитическими прогонами могут происходить события, которые делают процесс измерения восприимчивым к значительным изменениям.

2 Последовательность образцов, исследуемых последовательно без перерыва, если только рекомендуемая операция не требует такого перерыва.

3 Прогон или партия − интервал времени или условия измерения, для которых погрешность измерения является наименьшей. Таким образом, точность измерений в пределах партии определяет повторяемость.

5.2.35 **образец:** Одна или несколько частей, взятых из системы и предназначенных для предоставления информации о системе или для того, чтобы служить основой для принятия решения о системе (см. [13]).

***Пример − Объем сыворотки, взятый из большего объема сыворотки.***

Примечание – Для целей настоящего стандарта образец может быть физически или химически изменен по сравнению с исходной пробой пациента (см. проба), например, если в него добавлено потенциально мешающее вещество.

5.2.36 **проба:** Дискретная часть жидкости тела, дыхания, волос или ткани, взятая для осмотра, изучения или анализа одного или нескольких количеств или свойств, предположительно относящихся к целому (см. [11]).

Примечание – Для целей настоящего стандарта проба − компонент, взятый непосредственно из тела, с антикоагулянтами и консервантами или без них, который не подвергался физическим или химическим изменениям, за исключением того, что его могли центрифугировать (т. е. клетки крови были отделены от сыворотки или плазмы).

5.2.37 **целевое состояние**; ЦС: Заболевание, стадия заболевания, состояние здоровья или любое другое идентифицируемое состояние или характеристика, представляющая интерес для субъекта, например, стадирование уже известного заболевания или состояние здоровья, которое может побудить к клиническим действиям, например, к началу, изменению или прекращению лечения.

Примечание − Для исследований с обнаружением аналитов ЦС может быть обобщено как субъект, имеющий аналит (например, вирус).

5.2.38 **ошибка I типа:** Ложное отклонение нулевой гипотезы, когда она верна.

5.2.39 **ошибка II типа:** Ложное принятие нулевой гипотезы, когда она ложна.

5.2.40 **валидация:** Фаза этапа установления модели жизненного цикла исследований; проверка, при которой установленные требования адекватны предполагаемому использованию (см. [10]).

Примечания

1 Подтверждение, посредством предоставления объективных доказательств, что требования для конкретного предполагаемого использования или применения были выполнены (см. [14]).

2 Валидация − установление и/или подтверждение, посредством обширных испытаний, аналитических и/или клинических характеристик процедуры измерения.

3 Оценки характеристик, проводимые разработчиком для установления того, что характеристики адекватны предполагаемому использованию.

5.2.41 **верификация:** Этап стадии реализации модели фаз жизненного цикла испытаний; предоставление объективных доказательств того, что изделие соответствует установленным требованиям (см. [10]).

Примечания

1 Подтверждение, посредством предоставления объективных доказательств, того, что установленные требования были выполнены (см. [14]).

2 Для лаборатории верификация − это подтверждение посредством испытаний того, что рабочие характеристики исследования соответствуют установленным требованиям маркировки. Верификация обычно проводится конечным пользователем до внедрения исследования для рутинного использования в лаборатории.

**5.3 Сокращения**

*ANOVA* − дисперсионный анализ;

*BDS* − двунаправленное секвенирование;

ДИ − доверительный интервал;

*CRISPR* − кластеризованные регулярно перемежающиеся короткие палиндромные повторы;

*CV* − коэффициент вариации;

*CX* − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в *X* % случаев;

ДНК − дезоксирибонуклеиновая кислота;

*FN* – ложноотрицательный результат;

*FP* – ложноположительный результат;

ВИЧ − вирус иммунодефицита человека;

ВПЧ − вирус папилломы человека;

*ICR* − внутренний непрерывный ответ;

*IVD* − диагностика *in vitro*;

*LLoD* − нижний предел обнаружения;

*LoB*  − предел холостого хода;

*LR* − коэффициент правдоподобия;

*NGS* − секвенирование нового поколения;

*NLR* − отрицательное отношение правдоподобия;

*NPA* − отрицательное процентное согласие;

*NPV* − отрицательная прогностическая ценность;

*OR* − отношение шансов;

ПЦР – полимеразная цепная реакция;

*P*FP − вероятность ложноположительного результата;

*PLR* − положительное отношение правдоподобия;

*PPA* − положительное процентное согласие;

*P*POS  − вероятность быть признанным положительным;

*PPV* − положительная прогностическая ценность;

*QC* − контроль качества;

*RFU* − относительная единица флуоресценции (единиц);

РНК − рибонуклеиновая кислота;

*SD* − стандартное отклонение;

*Se* − клиническая чувствительность;

*Se*cand − клиническая чувствительность кандидата;

*Seold* − клиническая чувствительность ранее применяемого исследования;

*Sp* − клиническая специфичность;

*Spcand* − клиническая специфичность кандидата;

*Spold* − клиническая специфичность ранее применяемого исследования;

ЦС − целевое состояние;

*TN* − истинно отрицательный;

*TP* − истинно положительный;

*DF* − степени свободы;

ИФА − иммуноферментный анализ;

*HMW-CAP EIA* – иммуноферментный анализ высокомолекулярных клеточно-ассоциированных белков;

*CT* – порог цикла.

**5.4 Символы**

α − вероятность ошибки первого типа;

∆Se − разница между чувствительностями;

∆Sp − разница между специфичностями;

Φ − кумулятивная стандартная нормальная функция вероятности;

*A* − количество истинно положительных результатов;

*a* − перехват;

*B* − количество ложных срабатываний;

*b* − наклон;

*C* − количество ложноотрицательных результатов;

*c* − небольшое положительное число;

*D* − количество истинно отрицательных результатов;

*l*1, *l*2 − нижние границы двухсторонних 95 %-ных ДИ;

*N* − количество образцов;

*u*1, *u*2 − верхние границы двухсторонних 95 %-ных ДИ;

*X* − концентрация, преобразованная в логарифмы.

#### 6 Качественные исследования

6.1 Цель качественных бинарных исследований − классифицировать исследуемых на когорты с ЦС (т. е. ЦС присутствует) и когорты без ЦС (т.е. ЦС отсутствует). Методы, используемые для проведения качественных исследований, разнообразны, в них применяют технологии от тестов бокового потока с визуальным считыванием, автоматизированного секвенирования нуклеиновых кислот и иммуноанализа. Качественные исследования также используют для широкого спектра клинических целей, включая скрининг, установление диагноза и подтверждение результатов. Даже с учетом различий в методологии и целях существуют общие черты, позволяющие оценить их клиническую эффективность и точность.

**6.2 Качественные параметры и параметры группировки**

Качественные бинарные исследования можно разделить на подкатегории по двум параметрам: назначение и проектирование процедуры исследования.

Два вида целевого использования:

- включает исследования, используемые для определения наличия у пациента, чей образец исследуется, определенного ЦС. Примеры целевых клинических состояний включают беременность, предыдущую вакцинацию или инфицирование вирусом, концентрацию наркотиков, превышающую установленный уровень, и рак;

- исследования на обнаружение аналитов используют для определения наличия в образце целевого аналита. Целью этих исследований является признание образцов, не содержащих целевого аналита, отрицательными, а образцов, содержащих одну или более копий целевого аналита, − положительными. Примеры аналитов включают такие микробы, как *Trichomonas vaginalis* и коронавирус 2 тяжелого острого респираторного синдрома.

Несколько способов разработки процедуры исследования:

- исследования, в которых для всех образцов используется срез по единой шкале внутреннего непрерывного ответа (*ICR*), могут использовать для обоих видов целевого использования;

- исследования, использующие алгоритм принятия решения, который может использовать несколько *ICR* и/или категориальные данные для получения бинарного результата категоризации, обычно применяют только для предполагаемого использования обнаружения аналитов. Примеры включают:

а) исследования, в которых *ICR* равен *X* для образцов без аналита и внутренний ответ отличается от *X* для образцов с аналитом (например, цикл ПЦР, при котором пересекается порог);

б) исследования с несколькими *ICR*, для которых алгоритм использует решения «если/то» для бинарной классификации;

в) исследования, в которых используются многоступенчатые решения о категоризации, исключающие варианты до принятия окончательного решения о бинарной категоризации.

- исследования, которые визуально считываются пользователем без *ICR*, могут использоваться по назначению в обоих случаях. В качестве примера можно привести исследования, основанные на появлении осадка в пробирке с образцом или цветной полоски на тест-полоске, опущенной в образец.

Примечание – Хотя это можно рассматривать как алгоритм, использование отсечения по одному *ICR* настолько распространено, что ему присвоена отдельная категория проектирования, и только более сложные процессы принятия решений считают основанными на алгоритмах.

**6.3 Подкатегории качественных исследований с бинарным результатом**

Используя данные концепции, качественные бинарные исследования можно разделить на следующие категории:

- исследования, которые классифицируют наличие и отсутствие ЦС путем сравнения *ICR* с клиническим значением отсечки. Для таких исследований отсечку по шкале непрерывных ответов обычно выбирают для оптимизации клинической чувствительности и специфичности исследования для целевой популяции с учетом компромисса между пользой и риском или для соответствия заранее установленной точке принятия медицинского решения. Для таких исследований C5 и C95 могут быть измерены с помощью погрешности *ICR*. Более подробная информация о клинической чувствительности и специфичности приведена в соответствующем руководстве (см. [15]).

- исследования для обнаружения аналитов, которые определяют наличие или отсутствие целевого аналита и в которых *ICR* доступен для всех образцов. Например, отсечка ПЦР-исследования может быть установлена с учетом предела шума системы на шкале непрерывного ответа или того, что можно назвать пределом холостого хода (*LoB*). Если распределение *ICR* образцов без аналитов практически не совпадает с распределением образцов с некоторыми аналитами, отсечка может быть установлена таким образом, чтобы оптимизировать частоту ложноположительных (*FP*) и ложноотрицательные результаты (*FN*). При таких исследованиях C95 можно измерить с помощью погрешности *ICR*, но C5 может не поддаваться измерению;

- исследования, основанные на процессе и алгоритме обнаружения уникальной последовательности нуклеиновых кислот целевого аналита. Будь то ПЦР, кластеризованные регулярно перемежающиеся короткие палиндромные повторы (*CRISPR*) или другая технология, такие исследования направлены на обнаружение только целевого аналита. В случае успеха образец с отсутствием целевого аналита будет признан отрицательным почти в 100 % случаев (без *FP*), то есть исследование имеет 100 % клиническую специфичность. Клиническая чувствительность может меняться в зависимости от процесса исследования. Поскольку *ICR* не существует, единственная другая метрика оценки определяет наименьшую концентрацию целевых копий, которая обнаруживается в 95 % случаев в соответствии с пробит-фиттингом для данных частоты попаданий, как описано в соответствующем руководстве (см. [16]). Определение C5 по этой шкале может быть или не быть осуществимым (см. приложение Б);

- визуальное чтение, качественные бинарные исследования (см. приложение В)

На рисунке 1 представлен процесс принятия решений, используемый для классификации качественных бинарных исследований, и описаны методы, используемые для определения ключевых параметров качественных исследований, таких как C5 и C95. Здесь также показан ход событий от разработки исследования до валидации и проверки в лаборатории.



C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев;   
*ICR* − внутренний непрерывный ответ.

Рисунок 1 − Установление, подтверждение и проверка результатов качественного исследования[[2]](#footnote-2))

#### 7 Разработка качественного исследования с бинарным выводом

7.1 Данный раздел содержит рекомендации для разработчиков всех типов качественных бинарных исследований, включая производителей диагностических исследований *in vitro* (*IVD*) и медицинские лаборатории, создающие исследования в качестве лабораторных тестов. По завершении разработки должны быть определены характеристики исследования, необходимые показатели по этим характеристикам, и в процессе работы исследование должно быть оптимизировано для достижения этих желаемых результатов.

Все качественные исследования, основанные на инструментах, генерируют *ICR*. В некоторых исследованиях непосредственно используют сигнал и отсечку для ответа «да» или «нет». Таким ответом может быть внутреннее напряжение (например, от светового детектора). В обычном качественном исследовании, обеспечивающем *ICR* для всех образцов, для определения соотношения сигнала и отсечки используют расчетное соотношение сигнала от образца. Для определения значения отсечки в таких исследованиях используются клинические критерии. В 7.2 приведены этапы разработки данного вида исследований.

Качественные исследования для обнаружения аналитов также могут быть разработаны таким образом, чтобы использовать отсечки по шкале значений *ICR*, доступных для всех образцов. Поскольку системный шум на шкале *ICR* может играть роль в нахождении оптимальной точки разграничения между присутствием и отсутствием аналита, оценка *LoB* на шкале *ICR* может быть полезна при выборе значения отсечки. Эти виды исследований рассматриваются в 7.2.5.

В других случаях обнаружения аналитов комбинация сигналов и/или категориальных данных используется в алгоритме категоризации для получения бинарных результатов: аналит присутствует или отсутствует. Такой подход характерен для качественных исследований, основанных на обнаружении аналитов с помощью ПЦР. Поскольку единая шкала *ICR* не используется для всех потенциальных образцов, основным показателем эффективности обычно является C95, измеряемый в количестве копий на единицу стандартного объема, а не *ICR*. Подробные сведения об этапах разработки данного типа исследований приведены в 7.3.

**7.2 Разработка качественных испытаний с использованием сравнения срезов с внутренним непрерывным ответом**

7.2.1 Для многих типов качественных бинарных исследований *ICR* содержит значения значительно ниже и значительно выше отсечки. Примером такого типа исследований является определение антител к ВИЧ с помощью иммуноферментного анализа, который дает низкие ответы от образцов, полученных от лиц без ВИЧ, и потенциально чрезвычайно высокие ответы от образцов, полученных от лиц с недавно диагностированным ВИЧ. Эти ответы основаны на отношении сигнала прибора от образца пациента к сигналу отсекающего образца. Единственная область, представляющая интерес для этих процедур, находится в районе бесприборного значения отношения отсечки, равного 1,0. Не имеет значения, будет ли отношение сигнала к отсечке 0,1 или 0,8, поскольку результат все равно будет отрицательным, так же как 1,1 и 1000000 являются положительными результатами. Отношение сигнала к отсечке, как функция концентрации аналита, должно демонстрировать монотонность вблизи отсечки. При исследовании на антитела такую монотонность можно продемонстрировать с помощью серии заборов крови, которые показывают увеличение количества антител по мере того, как пациент становится иммунологически реактивным к вирусу. Однако нет необходимости демонстрировать линейность между отношением сигнала к отсечке и концентрацией антител в образце. Та же логика применима к исследованию на обнаружение аналитов, для которого система исследования обеспечивает значение *ICR* даже при отсутствии аналита. Если наблюдаемый результат *ICR* выше значения отсечки, проба считается положительной, если ниже − отрицательной (см. 7.2.5).

**7.2.2 Установление отсечки: ошибки классификации и вариабельность исследования**

Качественные бинарные исследования используют для оценки наличия или отсутствия ЦС, и для получения бинарных результатов им требуется отсечка. Например, результат может быть положительным, если *ICR* больше или равен отсечке, и отрицательным, если *ICR* меньше отсечки. Обычно отсечку выбирают для оптимизации клинической чувствительности и клинической специфичности исследования. Оно основано на наборе клинических данных из популяции предполагаемого использования и учитывает соотношение пользы и риска неправильных результатов. Например, качественное исследование-кандидата для скрининга при исследовании донорской крови на инфекционные заболевания, такие как ВИЧ или вирус гепатита В, отсечка должна быть подобрана таким образом, чтобы обеспечить высокую чувствительность, поэтому вероятность того, что инфицированный донор будет исключен из системы кровоснабжения, соответственно высока. Та же измерительная система может использоваться для проведения качественного исследования с целью диагностики, при этом отсечку выбирают с учетом высокой специфичности, чтобы не применять инвазивные методы лечения у пациентов без ЦС. В других сценариях отсечка может быть установлена таким образом, чтобы повторять заранее установленные медицинские точки принятия решения (например, заранее установленный порог принятия решения о злоупотреблении наркотиками).

Для установления отсечки используют образцы из двух популяций (одна с присутствием ЦС, другая − с отсутствием ЦС). Если внутренние распределения ответов для популяций, которые исследование должно различать, сильно разнятся, отсечку можно установить в широком интервале, не влияя на результаты исследования, как показано на рисунке 2. Кроме того, присущая *ICR* вариабельность не является критичной, поскольку отсутствуют или мало образцов со значениями ответов вблизи отсечки.

Население без целевого состояния (негативное)

Население с целевым заболеванием

(положительное)

Отсечка

**Непрерывный ответ**

Рисунок 2 – Два клинических состояния с широко разнесенными   
распределениями

Если распределения *ICR* из двух популяций близки или перекрываются, как приведено на рисунке 3, выбор отсечки становится вопросом компромисса между частотой *FP* и *FN* и способностью исследования стабильно давать одинаковые результаты вблизи отсечки. По мере сближения этих двух распределений и ограничения частоты ошибочной классификации, допускающей только минимальные показатели *FP* и *FN*, необходимо снижать основную неточность *ICR*. Оптимизация расположения отсечки и максимально возможное снижение неточности обычно выполняют в пространстве непрерывных ответов. Эта оптимизация включает в себя формулировку реагентов, рабочий процесс обработки образцов и производительность компонентов инструментария (например, точность объемов пипетирования).



Население без целевого состояния (негативное)

Население с целевым заболеванием

(положительное)

Отсечка

Настоящие

негативы

Настоящие

позитивы

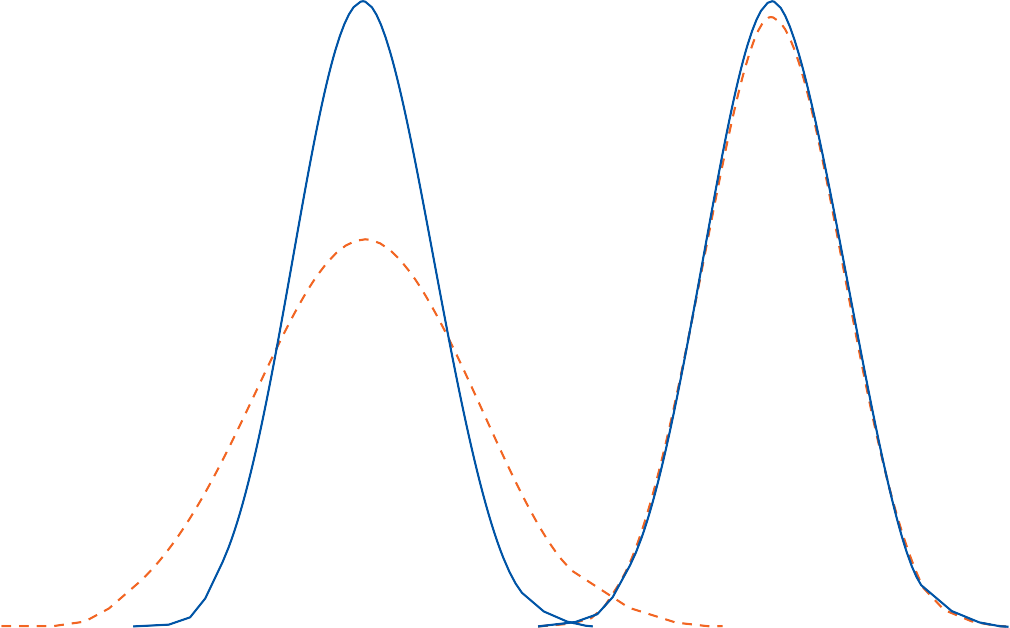
Ложноотрицательные результаты

Ложноположительные результаты

**Непрерывный ответ**

Рисунок 3 – Два клинических состояния с перекрывающимися распределениями

Один из способов представить себе процесс оптимизации − это вообразить, что можно исследовать выборку из каждой особи в популяции предполагаемого использования. На рисунке 4 две сплошные синие линии представляют собой непрерывные распределения ответов, соответствующие особям из популяции предполагаемого использования, которые делятся на две субпопуляции (т. е. одна с присутствующим ЦC, а другая с отсутствующим ЦC). В данном примере 5 %-ная ошибка классификации является приемлемой для субпопуляции отсутствующих ЦС (допускается 5 % *FP*) и 0,04% ошибочной классификации допустимо для субпопуляции присутствующих ЦС (допускается 0,04 % *FN*), при этом два непрерывных распределения вероятности ответа (т. е. сплошные синие линии) пересекаются на приблизительно 1 %. Отсечка (т. е. сплошная желтая вертикальная линия) для исследования была ранее установлена таким образом, чтобы правильно классифицировать как можно больше присутствующих ЦС, но при этом не допустить правильной классификации некоторых отсутствующих ЦС. Чтобы удовлетворить требованию 0,04 % неправильной классификации для присутствующих ЦС (чувствительность около 100 %), учитывая биологическую вариативность и установленную отсечку, вариабельность, присущая исследованию образцов от присутствующих ЦС, должна быть очень низкой (пунктирная оранжевая линия представляет дисперсию как от неточности, так и от индивидуальных ответов субпопуляционных образцов). Такая дисперсия означает, что вариабельность исследования не вносит существенного вклада в дисперсию популяционных результатов. Однако при требовании 5 %-ной ошибочной классификации для лиц, не имеющих ЦС (95 %-ная специфичность), и с учетом биологического распределения и той же отсечки, присущая вариабельность исследования для образцов от лиц, не имеющих ЦС, может быть гораздо выше (пунктирная оранжевая линия распространяется дальше, чем сплошная синяя линия). Пунктирные оранжевые линии представляют собой измерение биологической истинности исследования, то есть включают в себя как биологическую вариативность, так и погрешность и неточность исследования (в данном примере погрешность равна нулю).



Население без целевого состояния (негативное)

Население с целевым заболеванием

(положительное)

Отсечка

Синие сплошные линии представляют

биологическую истину

Оранжевая пунктирная линия

представляет наблюдаемые результаты

**Непрерывный ответ**

Рисунок 4 − Сравнение частоты ошибочной классификации при качественном   
исследовании в субпопуляциях с ЦС и без ЦС

Поскольку невозможно оценить выборку от каждого человека из популяции предполагаемого использования, можно сравнить общую вариабельность процедуры измерения вокруг отсечки с допустимой вариабельностью, которая была рассчитана. Для такого сравнения могут быть полезны понятия C5 и C95, которые описаны в 7.2.3. C5 и C95 отражают вариабельность или *SD* исследования в оговоренных условиях точности (проектирование исследования) и могут быть использованы для определения того, приведет ли вариабельность к недопустимому значению уровня неправильной классификации для данного исследования. Зная допустимую вариабельность для образцов как от людей с ЦС, так и без ЦС (полученную из приемлемой частоты ошибочной классификации), можно рассчитать допустимый интервал С5 − С95 для любой отсечки. Если измеренный интервал C5 − C95 шире этого допустимого интервала, частота ошибочной классификации может быть неприемлемой, и неточность *ICR* должна быть уменьшена для улучшения результатов исследования.

При проектировании исследования следует определить отсечку по шкале *ICR* на выборках, репрезентативных для предполагаемой популяции. Следует изучить диапазон отсечек и выбрать отсечку, удовлетворяющую требованиям клинической эффективности (см. рисунок 3); этот период изучения часто называют «тренировкой». Более подробная информация приведена в соответствующем руководстве (см. [15]). Затем, после фиксации отсечки, клинические характеристики исследования должны быть оценены в отдельном исследовании с другим, обычно большим набором образцов из популяции предполагаемого использования, как описано в 8.3.

Некоторые качественные исследования дают положительный, отрицательный или эквивокальный результат. Такие исследования включают эквивокальную зону, в которой *ICR* действительна, но окончательный результат сообщается как эквивокальный или неопределенный, а не как положительный или отрицательный. Например, если эквивокальная зона, основанная на *ICR*, обозначается [C1, C2], результаты могут быть отрицательными, если *ICR* менее C1, эквивокальными, если *ICR* более или равен C1 или менее или равен C2, и положительными, если *ICR* более C2.

Разработчики исследований могут установить зону равновероятных результатов, чтобы свести к минимуму количество *FP* и *FN* результатов исследований из-за совпадения распределений *ICR* в субпопуляциях с ЦС и без ЦС. Если обнаружена неоднозначная зона, производитель, как правило, рекомендует провести повторное исследование. Рекомендация может включать сбор другого образца в более позднее время и повторное тестирование с использованием той же или другой технологии. Для неточности *ICR* ключевыми факторами при разработке теста являются размер эквивокальной зоны и частота эквивокальных результатов, а также частота результатов *FP* и *FN* в популяции предполагаемого использования. Следует изучить преимущества наличия эквивокальной зоны определенного размера для предотвращения сообщения неверного результата, фиксирования неубедительных результатов, которые могут привести к нежелательным последствиям, или необходимости проведения дополнительного тестирования. Клинические показатели этого типа исследования могут быть охарактеризованы с помощью другого процесса, как описано в 8.3.5.3.

Другие качественные исследования дают бинарный результат и позволяют провести повторное исследование в зоне вокруг отсечки. Первоначальный бинарный результат достоверен, но из-за неточности разработчик рекомендует исследовать тот же образец еще раз с помощью того же теста. Этот сценарий особенно полезен, если используют автоматизированную систему исследования, которая сохраняет образец для повторного исследования. Окончательный положительный или отрицательный результат основан на алгоритме интерпретации с использованием обоих результатов. В этой ситуации клиническую эффективность исследования отражают с помощью окончательного бинарного результата, как описано в 8.3, и дополнительно указывают частоту повторных испытаний.

7.2.3 Взаимосвязь между отсечкой и C5, C50 и C95

C5 и C95 − это конструкции вокруг отсечки при проведении испытания, в которой используется отсечка, применяемое к *ICR*. Образец со значением *ICR* в соответствующей шкале, где бинарное испытание признает образец положительным в 50 % случаев (C50), дает отрицательный результат в 50 % случаев, если множество реплик этого образца выполняется в оговоренных условиях, как показано на рисунке 5. Если непрерывные ответы реплик распределены симметрично (т. е. нормально) для образца на отсечке (уровне), то C50 является оценкой отсечения.

**Относительная**

**частота**

Рисунок 5 − Гипотетическое распределение повторных результатов для выборки с отсечкой

Отсечка

**Непрерывный ответ**

Значения C5 и C95, как показано на рисунке 6, связаны с отсечкой на основе вариабельности реакции, присущей исследованию вблизи отсечки. Если множество реплик образца со значением ответа C5 будет выполнено в оговоренных условиях, 5 % реплик будут выше отсечения. Аналогично, 95 % реплик образца со значением ответа C95 будут иметь значение ответа выше отсечки. Чем больше вариабельность ответа в исследовании, тем дальше значения C5 и C95 от среза. Таким образом, ширина интервала от C5 до C95 является показателем вариабельности ответов в ходе исследования. Узкий интервал свидетельствует о низкой вариабельности, а   
широкий − о высокой вариабельности ответов.

На рисунке 6 приведено распределение результатов непрерывного ответа для одного образца при C5 1,40 и для другого образца при C95 1,68, оба основаны на отсечке 1,5. Протоколы для установления C5 и C95 различаются в зависимости от информации, полученной в результате качественных исследований. Если *ICR* доступен (как правило, производителю), этот ответ следует использовать для оценки вариабельности исследования в этом пространстве, поскольку это наиболее эффективный способ оценки эффективности в районе отсечения. Определение C5 и C95 − важные этапы в процессе разработки качественного исследования, поскольку, являясь мерой неточности, они могут определить необходимость дополнительных улучшений в конструкции и определить, как следует проводить последующие исследования эффективности.

**Относительная частота**

**Непрерывный ответ**

2.1

2

1.9

1.8

1.7

1.6

1.5

1.4

1.3

1.2

1.1

0

0

C95

0

C5

Отсечка

Интервал от C5 до C95

C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев.

Рисунок 6 − Гипотетические распределения повторных результатов для двух   
разных выборок

**7.2.4 Установление C5 и C95 для качественных исследований с помощью внутреннего непрерывного ответа**

Если *ICR* доступен и охватывает широкий интервал вокруг отсечки, C5 и C95 могут быть получены из точного профиля этой реакции. Неточность *ICR* должна быть измерена с помощью четырех или более образцов, которые охватывают интервал от ниже ожидаемого уровня C5 до выше ожидаемого уровня C95. Если это возможно, следует провести внутрилабораторное исследование по соответствующему руководству (см. [17]), которое включает воспроизводимость, и межпрогонные и компоненты междневной дисперсии. В зависимости от их известного влияния на исследование разработчику исследования рекомендуется включить другие факторы изменчивости, которые повлияют на оценку погрешности исследования. Общая вариабельность или *SD* рассчитывают с помощью анализа компонентов дисперсии результатов ответов из каждой выборки.

Предполагаемые уровни C5 и C95 можно оценить, используя оценку *SD* примерно на уровне отсечки, умножив ее на два, а затем добавив и вычтя это значение из значения отсечки. Без первоначальной оценки *SD* в качестве интервала, используемого для построения профиля точности, можно использовать отсечение плюс или минус 20 % от отсечения.

После получения общих *SD* и *CV* из внутрилабораторных данных следует составить профиль точности, чтобы определить, изменяется ли вариабельность по мере увеличения *ICR*. Если *SD* остается постоянным для разных уровней, можно оценить общую вариабельность (т. е. *SD*), объединенную для всех уровней. Если вариабельность изменяется с увеличением уровня внутреннего ответа, *SD* на C5 и C95 следует рассчитать на основе модели профиля точности для данных из нескольких выборок (например, четырех-пяти) вокруг отсечки. Если *CV* оказывается постоянным в интересующем интервале, его оценка в этом интервале может быть использована для расчета C5 и C95 по шкале *ICR*. Создание профиля точности подробно приведено в приложении Г.

Если результаты повторного тестирования или остатки от подгонки модели с компонентами дисперсии распределены нормально, внутрилабораторные C5 и C95 можно оценить непосредственно с помощью отсечки и оцененного общего внутрилабораторного *SD* на C5 и на C95. При нормальном распределении C95 и *SD* − это среднее и *SD* распределения внутреннего ответа, при котором его 5-ный процентиль равен отсечению исследования, полученному при повторном тестировании выборки с соответствующими выходами внутреннего ответа. Аналогично, C5 и *SD* − это среднее и *SD* распределения внутреннего ответа, при котором его 95-ны процентиль равен отсечке, полученной при повторных исследованиях образца с соответствующим значением внутреннего ответа. C5 и C95 могут быть вычислены по формулам (1) и (2) соответственно.

С5 = Отсечка – 1,645*SDC5*, (1)

С95 = Отсечка + 1,645*SDC95*, (2)

где 1,645 − 95-ный процентиль стандартного нормального распределения (Z0,95).

Если изменчивость в области вокруг отсечки имеет приблизительно постоянный *SD*, то *SD*С5 и *SD*С95 - одно и то же значение. Если образцы вокруг отсечения имеют примерно одинаковый *CV*, *SD*С5 можно заменить на *CV*·C5, а *SD*С95 − на *CV*·C95. Тогда C5 и C95 допускается вычислять по формулам (3) и (4) соответственно.

(3)

(4)

Приведенные расчеты предполагают, что распределение результатов для значения *ICR* является нормально распределенным. Если значения не являются нормальными, шкала *ICR* может быть преобразована. В таких случаях обычно используются различные логарифмические преобразования, в том числе предложенные Боксом и Коксом (см. [18]).

Для определения C5 и C95 допускается использовать другие типы оценок состояния неточности (*SD* или *CV*), используя формулы (1) − (4). В некоторых случаях (например, при использовании образцов цельной крови) нестабильность образца не позволяет измерить ежедневную погрешность. В таких случаях допускается использовать исследования, измеряющие только повторяемость, также известную как внутрипробеговая неточность. Также возможно проведение исследований, охватывающих внутрипрогонную и межпрогонную неточность. C5 и C95 также могут быть определены с помощью *SD* или допускается, полученных в результате исследований воспроизводимости, включающих вариабельность между лабораториями и, возможно, другие факторы, такие как партия реагентов. Более полный список потенциальных факторов, которые необходимо учитывать приведен в соответствующем руководстве (см. [17]). Конкретные использованные условия неточности должны быть указаны во всех случаях, когда сообщается о C5 и C95. Рекомендации по исследованию неточности для качественных исследований приведены в 8.2.

На рисунках 7а) и 7б) приведены эффекты изменения неточности между двумя различными формулировками разрабатываемого испытания (также см. приложение Б). Аналогичным образом, формулировки исследований или рабочие процессы для обработки образцов также должны быть оценены на предмет предвзятости, отражающейся в изменении отсечения. На этапе разработки исследования следует оптимизировать аспекты системы, влияющие на ее работу, чтобы уменьшить погрешность. На рисунках 7а) и 7б) показано, как интервал от C5 до C95 может быть использован для такой оптимизации. На рисунке 7б) показана перспектива, связанная с вероятностью признания образца положительным, а на рисунке 7а) − перспектива тех же данных, основанная на неточности исследования. Взаимосвязь между вероятностью положительного результата и неточностью перспектив внутреннего ответа дополнительно описана в приложении В. Как показано на рисунке 7а), общая вариабельность первой версии исследования (первая версия представлена сплошной красной линией на верхнем графике) больше, чем общая вариабельность второй версии исследования (вторая версия представлена сплошной синей линией на нижнем графике), поэтому дизайн, используемый для второй версии, является предпочтительным. Кроме того, область C5 − C95 первой версии шире, чем область C5 − C95 второй версии, когда эти две версии имеют одинаковые отсечки, как показано на рисунке 7б). Если не принимать во внимание другие соображения, такие как скорость *FP* или *FN*, дизайн, используемый в версии два, является предпочтительным.

Примечание – Оценка погрешностей различных качественных исследований путем сравнения интервалов C5 и C95 не имеет смысла, если нет возможности проследить *ICR* к общему эталону. Эта несопоставимость справедлива даже при одинаковом значении отсечки (например, соотношение сигнал/отсечка = 1,0).

**Вероятность положительного результата исследования**

**Относительная частота**



0.5

**Непрерывный ответ**

1

1.5

2

2.5

1.34

0

0.5

1

1.5

2

2.5

3

1.09

C5

1.66

(C95) 1.91

**Непрерывный ответ**

0.8

0.7

**A**

**B**

SD= 0.1

SD= 0.25

Отсечение= 1,5

0.9

C5

1.09

C95

1.91

0.1

SD= 0.25

0.6

C50= 1.5

0.5

C5

1.34

C95

1.66

0.4

0.3

0.2

SD= 0.1

|  |  |
| --- | --- |
| а) Версия 1 (верхний график) имеет более широкое распределение (больше неточности) как на C5, так и на C95, чем версия 2 (нижний график). Версия 2 предпочтительнее. | б) Версия 1 (вверху) с SD 0,25 имеет более широкий интервал от C5 до C95, чем версия 2 с SD 0,1. Версия 2 является более предпочтительной. |

C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C50 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 50 % случаев;   
C95 − значение в соответствующей шкале, при котором бинарное исследование признает   
образец положительным в 95 % случаев; *SD* − стандартное отклонение

Рисунок 7 − Сравнение интервалов от C5 до C95 двух исследуемых составов

**7.2.5 Использование внутреннего непрерывного ответа для исследования обнаружения аналитов**

Если цель исследования − определить, присутствует ли в образце какой-либо аналит, то при измерении образцов без аналита отсечка по шкале *ICR* может учитывать шум системы. *LoB* представляет собой шум системы (определенный в [[16](#_bookmark27)] как 95-ный процентиль распределения нулевых аналитов [«холостых»]). При использовании непрерывного ответа и установке отсечки на уровне *LoB* показатель *FP* составит 5 %. Аналогично, можно установить различные отсечки, чтобы обеспечить различные показатели *FP* (например, 1 % или ниже) при использовании распределения результатов проб с нулевым содержанием аналитов по сравнению с распределением проб с некоторым содержанием аналитов. В зависимости от предполагаемого использования продукта, при определении отсечки необходимо учитывать компромиссные решения для баланса показателей *FP* и *FN*. После того как отсечка установлена, C95 также может быть определена с помощью методов, описанных в 7.2.4. Часто распределение ниже C50 слишком сжато, чтобы можно было использовать эти методы для определения C5.

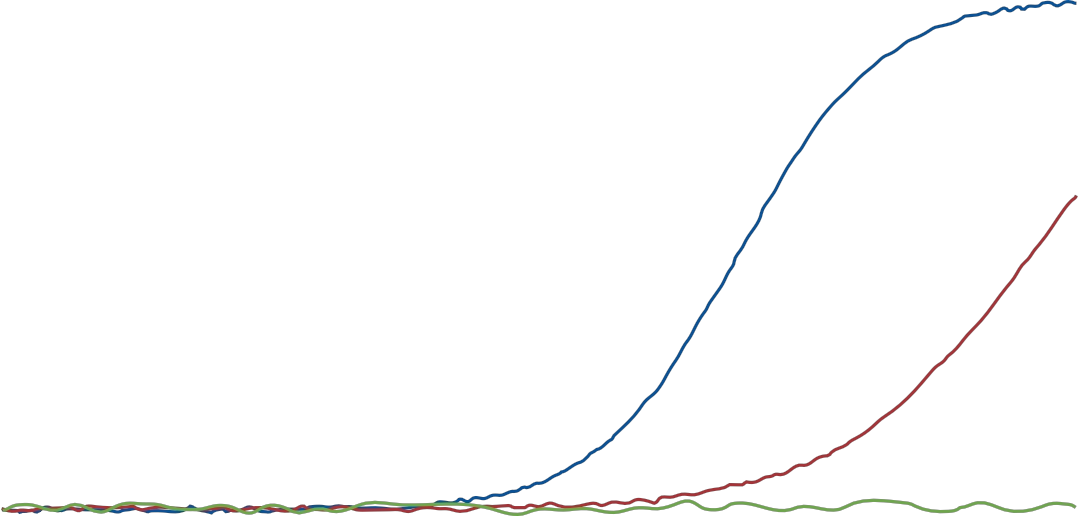
**7.3 Разработка методов обнаружения аналитов, качественные исследования**

Значительное и растущее число качественных исследований, которые можно классифицировать как исследования по обнаружению аналитов, выполняют с помощью ПЦР. Для разработки любого ПЦР-исследования необходимо выполнить множество шагов, включая выбор оптимальных методов выделения, амплификации и измерения присутствия целевого аналита (например, нуклеиновой кислоты, уникальной для конкретных вирусов или бактерий, аллелей или мутаций). Если выбраны правильные методы и последовательности целевых нуклеиновых кислот, можно получить данные, подобные тем, что приведены на рисунке 8.

ПЦР-исследования включают в себя цикл репликации, при котором количество копий нуклеиновой кислоты увеличивается примерно вдвое с каждым циклом. По мере увеличения числа копий флуоресцентный сигнал возрастает. Для этого типа качественного исследования производитель может установить отсечение с помощью ИКР, например, относительных единиц флуоресценции (*RFU*) на 45-м цикле, как показано на рисунке 8. В этом примере *RFU*, определяющий порог цикла, потенциально будет хорошим вариантом для отсечения. Для таких случаев непосредственно применимы процессы разработки, описанные в 7.2.5. Отсечка, C95 и, возможно, C5 могут быть определены с помощью этой шкалы.

ПЦР – полимеразная цепная реакция

**Относительные флуоресцентные единицы**



0

**Количество циклов ПЦР**

45

40

35

30

25

20

15

10

5

0

Ноль копий

25

Порог

50

Один

экземпляр

75

1000

экземпляров

100

200

175

150

125

Рисунок 8 − Исследование ПЦР

Далее предполагают, что алгоритм использует несколько *ICR* и категориальные данные для бинарной классификации. Единой шкалы *ICR* не существует, поэтому для определения эффективности исследования по выявлению аналитов используют не точность ICR, а число копий. Хотя ПЦР используется в качестве примера такого типа качественного исследования, допускается использовать и другие методики (например, CRISPR), если они дают качественные бинарные результаты.

Для многих качественных исследований, основанных на ПЦР, алгоритм обнаружения таков: образец является положительным, если флуоресценция пересекает заранее установленный порог до определенного количества циклов ПЦР, или отрицательным. В примере, приведенном на рисунке 8, исследование с использованием этого алгоритма в течение 45 циклов ПЦР может, вероятно, обнаружить присутствие одной копии целевой последовательности нуклеиновой кислоты.

Дополнительный учет *ICR* или системного шума может не потребоваться, поскольку клинические показатели соответствуют необходимым качественным критериям эффективности исследования с использованием данного алгоритма. Такие исследования основаны на нацеливании на определенные последовательности ДНК или РНК. Они не обнаружат эту последовательность, если она отсутствует. Если целевая последовательность нуклеиновой кислоты уникальна для целевого аналита, исследование не даст результата *FP*, что эквивалентно 100 % клинической специфичности. Тестирование популяции отрицательных образцов и оценка помех при исследовании помогут продемонстрировать эту надежную клиническую специфичность. Весь процесс выделения, амплификации, измерения и алгоритмической классификации должен быть тщательно продуман, чтобы обнаружить аналит и в тех случаях, когда он присутствует. Цель состоит в том, чтобы иметь возможность обнаружить одну копию целевого аналита, если он присутствует в исследуемом образце. При наиболее эффективных качественных исследованиях этот процесс может обеспечить клиническую чувствительность, близкую к 100 %.

Более сложные алгоритмы могут объединять более одной последовательности целевой ДНК, чтобы дать ответ «да» или «нет» на вопрос «Обнаружен ли целевой организм?». В других алгоритмах используют несколько этапов процесса, которые поочередно устраняют все возможности нецелевого анализа, оставляя только образцы с целевым анализом, признанные положительными. Производитель обязан создать алгоритм, обеспечивающий наилучшую клиническую чувствительность и клиническую специфичность, близкую к 100 %.

После того как процесс принятия решения «да» или «нет» установлен, можно количественно оценить способность этого процесса обнаруживать целевой аналит. Производительность качественного исследования, основанного на алгоритме обнаружения аналита, измеряется как C95 путем нахождения зависимости между частотой попаданий для нескольких образцов и их известной концентрацией аналита. Этот анализ может быть таким же, как описанный в [16] для измерения нижнего предела обнаружения (*LLoD*) с помощью пробит или другого метода. Ключевыми моментами для создания наилучшего процесса измерения C95 являются:

- использование достаточного количества и соответствующего распределения образцов;

- точное определение концентрации аналита (копий в см3) в каждом образце;

- наличие достаточного количества реплик для каждого образца для получения разумной оценки частоты попаданий;

- нахождение хорошего соответствия данных соответствующей обобщенной линейной модели (например, пробит, логит).

На рисунке 9 приведен пример графика набора данных, полученных в ходе качественного исследования с превосходной клинической чувствительностью. На рисунке представлены образцы в пяти концентрациях, охватывающих интервал от 0 до   
7 кол/см3. Самая высокая концентрация включена, чтобы обеспечить 100 % попадание. Эти образцы исследуются по 25 раз каждый, и отмечается доля положительных определений (попаданий). К этим данным подгоняется кривая пробит, и концентрация, при которой эта кривая пересекает пунктирную линию 95 % вероятности (C95), по расчетам, составляет примерно 3 кол/см3.

**Концентрация вируса, кол-во/см3**

7

6

5

4

3

2

1

0

0.0

C95

0.2

0.4

0.6

0.8

1.0

**Вероятность обнаружения**

C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев

Рисунок 9 − Использование подгонки под кривую пробит для оценки C95

В отличие от предыдущих описаний измерения C95 в этой главе, шкала, используемая для определения C95, основана на внешней единице измерения (кол-во/см3), которая отличается от любой внутренней шкалы; она не измеряется с помощью интервала погрешности *ICR* выше отсечки. Тем не менее, она измеряет точку (по шкале количество/см3), при которой исследование признает образец положительным в 95 % случаев. Если концентрация C95 низкая благодаря хорошей разработке процедуры исследования, больше образцов с низкой концентрацией аналитов могут быть признаны положительными, что повышает клиническую чувствительность исследования. Таким образом, C95 является полезной характеристикой, которую следует использовать в процессе разработки исследования. В приложении Д описаны аспекты измерений для исследований с низкой концентрацией C95, как показано на рисунке 9.

Аналогично, C5 не измеряется с помощью интервала погрешности *ICR* ниже отсечки. Если его можно достоверно определить, он все равно измеряет точку (по шкале счет/см3), в которой исследование признает образец положительным в 5 % случаев. Как упоминалось ранее, если последовательность целевой нуклеиновой кислоты уникальна для целевого аналита, клиническая специфичность будет близка к 100 %. C5 коррелирует с клинической специфичностью, но в процессе разработки разработчику будет полезнее измерить множество отрицательных образцов и убедиться в отсутствии интерференции в отрицательных образцах, чем пытаться измерить C5 с помощью исследования частоты попаданий. По этой причине C5 не измеряют.

Для качественных исследований, основанных на алгоритмах обнаружения аналитов, C95 является ключевым атрибутом, который часто используют для определения образцов, которые следует использовать в других методах проверки эффективности (например, образцы с концентрацией копий в 1,5−3 раза больше C95). В требованиях к конструкции может быть указан максимально допустимый C95. Если это требование не выполняется, необходимо продолжать разработку.

C95 используется в настоящем стандарте в качестве показателя результативности качественных исследований. *LLoD* также является показателем эффективности количественных исследований. В случае качественных исследований с обнаружением аналитов эти две величины могут быть измерены одинаково. Часто эти два термина используют как взаимозаменяемые для данного типа исследований.

Для измерения эффективности исследования необходимо использовать образцы, для которых известно среднее количество копий последовательности целевой ДНК на объем образца. Точное создание образцов для тестирования в низких концентрациях, таких как показано на рисунке 9, может быть затруднено, поскольку эталонные материалы часто доступны только в высоких концентрациях (например, 106 копий/см3). Многократные серийные разведения, необходимые для снижения концентрации на шесть порядков и более, могут внести неизбежные ошибки, которые могут привести к занижению или завышению C95. В качестве альтернативы образцы с более низкой концентрацией допускается измерять с помощью эталонной процедуры измерения, такой как цифровая ПЦР (см. [19]).

В примере, приведенном на рисунке 9, каждый образец исследуется 25 раз, чтобы оценить пропорциональную частоту, при которой этот образец будет признан положительным (т. е. частоту попаданий). В этом случае разрешение измерения каждого зарегистрированного составляет 1/25 или 4 %. Использование нескольких образцов обеспечивает некоторое снижение вариабельности общей линейной модели и, следовательно, оценки C95, но некоторая остаточная вариабельность все же ожидается из-за такого разрешения измерений. По этой причине для данного типа исследований рекомендуется использовать не менее 20 реплик.

Для исследований на основе алгоритма обнаружения аналита вероятность обнаружения аналита в положительном образце зависит от среднего количества копий в тестовом объеме и минимального количества копий, необходимого для обнаружения. Фактическое количество копий аналита в одной реплике образца может варьироваться в зависимости от распределения Пуассона. Этот источник неопределенности может потенциально скрывать истинные характеристики тестируемого образца. Более подробная информация приведена в приложении Д. Для уменьшения этого влияния можно использовать более высокие объемы измерений и, соответственно, большее количество реплик из исследуемых образцов. На рисунке 9 показана подгонка под кривую пробит для нетрансформированных данных о количестве копий; часто такая подгонка лучше (т. е. меньше остаточная ошибка) после преобразования подсчетов с помощью функции на основе логарифма, которая измеряется при нулевой концентрации (см. [18]). Такая модель может обеспечить лучшую подгонку для C5, если это необходимо. Разработчики должны изучить методы, модели и шкалы, чтобы найти оптимальное соответствие модели.

Цель должна заключаться в том, чтобы найти модель, которая лучше всего соответствует наблюдаемым результатам по количеству попаданий, но при этом подчеркивает соответствие вокруг расчетного C95.

Образцы, признанные положительными по алгоритму порога цикла, могут также давать числовой результат в виде цикла, в котором пересекается порог цикла флуоресценции. Это значение может быть использовано в качестве *ICR* для реплик, признанных положительными. Однако репликаты, считающиеся отрицательными, предоставляют цензурированные данные, поскольку сообщается только максимальное число циклов. Поэтому для измерения неточности с помощью дисперсионного анализа (*ANOVA*) можно использовать только образцы с таким значением в соответствующей шкале, при котором бинарное исследование признает образец положительным в 100 % случаев (C100). Для такого *ANOVA*, как правило, необходимо использовать только образцы со значением C95 в 1,5 раза выше, что ограничивает полезность такого исследования неточности.

**7.4 Контрольные материалы**

7.4.1 В данном подразделе приведены правила использования контроля правильности, материалов внешнего контроля качества и материалов внутреннего контроля.

**7.4.2 Контроль истинности**

Многие качественные исследования сравнивают ИКР образца с материалом, имеющим *ICR* на уровне отсечки. Распространенным типом таких исследований являются исследования соотношения сигнал/отсечка. Отсекающий материал метрологически прослеживается либо до внутреннего эталонного материала производителя, установленного при разработке исследования, либо до внешнего эталонного материала, используемого в процессе начальной настройки отсечки. Его заданное значение может быть использовано для калибровки системы перед использованием или для обеспечения поддержания правильности системы. Требования к метрологической прослеживаемости и неопределенности для калибраторов и контроля правильности описаны в международных стандартах (см. [6]).

**7.4.3 Материалы внешнего контроля качества**

Использование соответствующих внешних материалов для контроля качества обеспечивает стабильное качество бинарных исследований. Эти внешние контроли проводятся так же, как и образцы пациента, как будто они являются неизвестными образцами. Основное отличие заключается в том, что они имеют установленный интервал результатов реакции, поэтому их результат можно сравнить с ожидаемым результатом. Если лаборанту предоставляется *ICR* для контрольного материала, то руководство для лабораторий, представленное в [20] для количественных исследований, может быть использовано с этими значениями для создания процесса мониторинга качества. Если предоставлен только бинарный результат, допускается следовать этим общим рекомендациям, но при этом следует учитывать процесс принятия решений, описанный ниже.

Внешние материалы *QC* должны включать как минимум один отрицательный образец с ожидаемым результатом и как минимум один положительный образец с ожидаемым результатом. В некоторых ПЦР-исследованиях используется отрицательный образец *QC*, в котором не обнаруживается последовательность целевой ДНК на фоне нормальной ДНК, а также контроль «без шаблона» (без содержания ДНК). Для многих качественных исследований рекомендуется тестировать отрицательный и положительный контроль ежедневно и/или на каждую партию; однако для некоторых качественных исследований производитель рекомендует использовать разные частоты тестирования контроля. Если возможно, образцы материала внешнего *QC* должны обрабатываться так же, как и образцы пациента (например, если требуется экстракция или разведение), или должно быть указано, какой аспект рабочего процесса исследования представляет данный контроль.

Необходимо тщательно продумать положение контрольных материалов на шкале внутренних ответов. Положительные и отрицательные контроли, расположенные вблизи отсечки, более чувствительны к обнаружению ошибок процесса, чем очень высокие или очень низкие отсечки; однако контроли, расположенные вблизи отсечки, могут привести к отклонению пробных тестов на основании случайного стечения обстоятельств (для C5 и C95 такое отклонение происходит в 5 % случаев). Уровни контроля должны быть выбраны таким образом, чтобы обеспечить ожидаемый результат при правильном функционировании теста, что предполагает, что процесс тестирования должен контролироваться на уровнях, более далеких, чем C5 и C95 от отсечения.

Как описано в [20], для достижения целей лаборатории в области качества необходимо учитывать как пределы допуска для результатов *QC* материалов, так и процесс обеспечения качества. Кроме того, следует учитывать риски для пациентов, связанные с изменением функционирования системы. Конкретный пример, в котором результат *FP* представляет наибольший риск, приведен в таблице 2.

Таблица 2 − Примеры риска ошибок при тестировании и материалов *QC* для качественных исследований

|  |  |  |
| --- | --- | --- |
| Сдвиг в функционировании системы | Направление смены | Риск для пациента |
| Ошибка в процессе качественного исследования или отказ реагентов приводят к изменению результатов у пациента | Негатива | Высокий риск:  Ошибочное медицинское решение не лечитьb |
| Позитивс | Риск среднего уровня:  Ненужное или неадекватное решение о лечении |
| Низкий риск:  Подтверждающий тест может дать правильный отрицательный результат |
| aНизкоположительный материал *QC* редко дает отрицательные результаты, когда процедура находится под контролем, но вероятность получения *FN*-результатов увеличивается при изменении режимов отказа, вызывая отрицательный сдвиг.  bОтрицательный сдвиг (*FN*-результат) может привести к ошибочному решению не лечить пациента или не отбраковывать инфицированный образец крови. Примером положительного сдвига (*FP*-результат) является серология иммуноглобулина G против краснухи, для которой положительный результат говорит медицинскому работнику, что беременная женщина защищена, хотя это не так, что представляет собой более высокий медицинский риск, чем *FN*-результат.  cВысокоотрицательный материал *QC* редко дает положительные результаты, когда процедура находится под контролем, но вероятность получения *FP*-результатов возрастает при наличии режимов отказа, вызывающих положительный сдвиг.  Примечание – КК − контроль качества. | | |

**7.4.4 Внутренние контрольные материалы**

Внутренние контрольные материалы (также называемые контролем процесса, контролем комплекта или бортовым контролем) контролируют этапы или функции в процессе исследования. Контрольные материалы часто выполняются перед каждым образцом пациента или вместе с ним. Одним из примеров является тест на беременность, в котором одна цветная линия на тест-полоске для мочи показывает, присутствует ли аналит выше пороговой концентрации, а вторая линия показывает, достаточно ли мочи впитала тест-полоска, что свидетельствует о достоверности теста. Положительный или отрицательный результат не может быть определен, если этот внутренний контроль не указывает на достоверность теста.

В примере ПЦР внутренний контрольный материал может содержать последовательность ДНК/РНК неродственного гена, которая выполняется перед каждым образцом пациента или вместе с ним, чтобы убедиться, что извлечение и амплификация соответствуют ожиданиям. В качестве альтернативы можно исследовать «домашний» ген, который присутствует в каждом образце пациента, или экзоны самого целевого гена. Ожидаемое значение *ICR* известно заранее. Подобно внешнему контролю качества, этот внутренний контроль проходит через большинство (в идеале − через все) этапов процесса, чтобы получить результаты для пациента. Сравнение *ICR* с ожидаемым ответом показывает, активны ли реагенты и правильно ли они работают, был ли добавлен образец и правильно ли протекали растворы через устройство.

**7.5 Определение стабильности реагентов, используемых в качественных исследованиях**

Реагенты, используемые в качественных бинарных исследованиях, могут включать материалы, используемые в ходе процедуры исследования, в том числе используемые для выделения исследуемого аналита или для создания сигнала прибора или отсекающего образца, используемого для получения непрерывного ответа по соотношению сигнал/отсекающий образец. Во всех случаях, если *ICR* доступен производителю, ответы исследуемых образцов могут быть использованы для оценки тенденций с помощью методов регрессии, как описано в [3] чтобы определить, были ли достигнуты цели стабильности. В некоторых случаях производитель может не использовать *ICR* для качественного исследования (например, при визуальном исследовании). Производителям настоятельно рекомендуется искать способы создания непрерывной реакции, чтобы можно было использовать протоколы, приведенные в [3]. Например, результаты визуального исследования мочи с помощью дипстика могут быть количественно оценены с помощью колориметрического считывателя полосок. При исследовании на основе ПЦР для обнаружения аналитов может использоваться цикл (порог) для количественной оценки характеристик стабильности с использованием суррогатных образцов, которые имитируют образцы от людей с присутствующим аналитом, как показано на рисунке 8. Однако если *ICR* недоступна, для исследования стабильности можно использовать схему исследования, описанную в 9.4.

Срок годности и стабильность при использовании, при которых реагенты деградировали и, как следствие, изменили наблюдаемое значение тестового образца, могут привести к смещению отсечения и, таким образом, негативно повлиять на все параметры эффективности, включая C5, C95, клиническую чувствительность и клиническую специфичность по сравнению с результатами, полученными при использовании свежего материала. Поэтому разработчик обязан начать оценку стабильности реагента как можно раньше, а затем приступить к оценке стабильности производственной партии после того, как будет разработан дизайн процедуры исследования. Чтобы убедиться, что все применимые аспекты стабильности учтены, следует обратиться к [3].

Любые образцы, используемые в исследованиях стабильности реагентов, должны храниться в условиях, гарантирующих, что они не изменятся в ходе исследования. Эти исследования проверяют эффективность реагентов, но не образцов, и включают в себя исследования срока годности для определения того, как долго реагент или набор может храниться в рекомендованных условиях до использования, и исследования в процессе использования для определения того, как долго реагент или набор стабилен после того, как его начали использовать. Если пробы, полученные от пациентов, не являются стабильными (например, образцы цельной крови), разработчикам, возможно, придется изучить возможность использования пустых или суррогатных образцов с добавлением шипов для исследований стабильности.

#### 8 Валидация качественного исследования с бинарным выводом

8.1 Сбор данных для валидации качественного бинарного исследования включает оценку аппаратного и программного обеспечения, а также реагентов (включая «расходные материалы») в конфигурации, находящейся в «производственном» (т. е. окончательном) состоянии. Все данные собирают с использованием тех же инструкций по применению, которые приведены во вкладыше к упаковке продукта (в случае распространяемого набора) или в окончательных стандартных операционных процедурах (в случае разработанного лабораторией теста).

В данном разделе рассматривают валидационные испытания, которые делятся на три основные категории:

- неточность (исследования точности и частоты попаданий):

а) внутрилабораторное исследование;

б) исследование воспроизводимости.

- классификация (клиническая эффективность или согласие);

- мешающие вещества.

**8.2 Точные исследования**

8.2.1 Независимо от типа качественного исследования, для оценки неточности обычно проводят два типа испытаний: внутрилабораторные испытания неточности (внутри производителя) и испытания на воспроизводимость, включающие лаборатории вне производителя (в случае распространяемого набора) или референс-лаборатории. Для описания требований к испытаниям прецизионности качественные бинарные исследования допускается группировать в следующие категории (описание приведено в 6.3):

- исследования, которые классифицируют наличие и отсутствие ЦС путем сравнения *ICR* с клинически установленным значением отсечения;

- исследования по обнаружению аналитов, которые определяют наличие или отсутствие целевого аналита с помощью отсечки по единой шкале *ICR*, доступной для всех образцов;

- исследования по обнаружению аналитов, основанные на процессе и алгоритме, который обнаруживает уникальную последовательность нуклеиновых кислот целевого аналита (может использовать несколько *ICR*);

- визуальное чтение, качественные, бинарные исследования (см. приложение В).

Дизайн прецизионных испытаний включает два основных аспекта: оцениваемые условия прецизионности (источники вариабельности) и панель образцов. Исследование (оборудование, программное обеспечение, реагенты и рабочий процесс) определяет логистику дизайна и влияет на то, сколько источников изменчивости можно оценить одновременно. Ранее определенные оценки параметров исследования, такие как C95, влияют на выбор тестовых образцов, используемых для исследования. Различия в обоих аспектах (логистика и оценки параметров) и то, как они влияют на дизайн исследований, проведено ниже. Более подробная информация о дизайне прецизионных исследований содержится в приложении Е.

**8.2.2 Испытания для исследований с отсечкой по внутреннему непрерывному ответу**

8.2.2.1 В данном подразделе рассматриваются первые два типа исследований, упомянутые выше: наличие или отсутствие ЦС на основании клинически установленного среза *ICR* и обнаружение аналитов (ноль против одной или более копий) на основании одного среза *ICR* шкалы. Поскольку *ICR* предоставляет числовые значения для всех результатов анализа образцов, для таких исследований допускается использовать процедуры сбора и анализа данных исследований прецизионности, описанные в [17]. Сюда входят внутрилабораторные исследования и исследования воспроизводимости на разных объектах. В [17] приведены стратегии, которые допускается использовать для определения влияния таких компонентов дисперсии, как повторяемость, прогон, день, прибор, оператор, участок и партия реагента. Разработчику могут быть известны другие источники вариабельности, которые следует учитывать с учетом технологии, проб (включая их сбор и хранение до обработки), системы исследования (аппаратное и программное обеспечение, реагенты и другие расходные материалы), методы предварительных исследований, таких как выделение аналитов, и предполагаемого использования продукта.

8.2.2.2 Точные исследования

Панель прецизионности, используемая в внутрилабораторных исследованиях и исследованиях воспроизводимости, должна включать:

- заведомо отрицательный образец (образец с нулевым содержанием аналитов для исследований на определение аналитов);

- высокоотрицательный образец (около C5 для клинически установленной отсечки исследований *ICR*);

- образец выше отсечения (около C95);

- умеренно положительный образец (выше C95, но в идеале − ближе к C100);

- высокоположительная проба.

Примечание – Для исследований с клинически установленной отсечкой *ICR* высокоотрицательная проба должна быть направлена на значение C5. Для исследований с отсечкой *ICR*, установленной на пределе системного шума или вблизи него (т. е. *LoB*), высокоотрицательная проба не требуется. Варианты высоконегативной пробы для исследований без отсечки, с обнаружением аналитов, см. в 8.2.3.1.

В идеале все образцы должны быть получены из немодифицированных проб пациента. Вопросы, связанные с сравнением необработанных и обработанных образцов, рассматриваются в [21], варианты использования образцов панели точности ниже отсечки для исследований с определением аналитов описаны в 8.2.3.1. Поскольку *ICR* доступен для всех образцов для данного типа исследований, *ANOVA* следует проводить для каждого результата и/или уровня прецизионной пробы отдельно, если предположения *ANOVA* уместны. Сильно асимметричные распределения могут потребовать преобразования результатов *ICR* перед анализом, а бимодальное распределение может исключить проведение *ANOVA*. Для получения более подробной информации об анализе результатов неточности с помощью *ANOVA* допускается обратиться к [17]. Если для образца невозможно провести корректный *ANOVA*, вместо него следует указать 2,5 и 97,5 перцентили результатов *ICR*, а также медиану.

C5 и C95 (если имеются данные по панели точности) должны быть рассчитаны с использованием формул (1) и (2) и представлены на основе внутрилабораторной погрешности. Кроме того, необходимо представить описательную сводку, должны быть представлены бинарные результаты. Для каждого образца в панели прецизионности следует рассчитать процент положительных, отрицательных и недействительных результатов. Пример такой сводки для внутрилабораторного исследования точности приведен в 8.2.4. Аналогичные результаты должны быть представлены для испытаний воспроизводимости.

8.2.2.3 Исследования частоты попаданий для определения C95 по количественной шкале

Если существует эталонный материал с известным количественным значением или доступна эталонная процедура измерения, позволяющая получить образцы с количественным значением, C95 следует оценивать с помощью этих образцов с количественным значением. Эту оценку проводят с помощью исследования количественных показателей. В идеале исследование должно состоять:

- из одного исследования на партию реагентов с использованием не менее двух партий реагентов с 20 или более повторами на разведение в каждой партии в течение не менее трех дней;

- не менее шести уровней, включающих образец с одним заведомо отрицательным (или нулевым) результатом, образцы с тремя уровнями и процентом попадания от 10 % до 90 %, и один образец с процентом попадания более 95 %, но в идеале близким к 100 %;

- одного образца при 100 % попадании.

Этот набор образцов обычно создают на уровне, близком к желаемому, но не всегда точно соответствующем ему, путем разбавления. Этот дизайн исследования имитирует дизайн, описанный в [16] для оценки *LLoD* с помощью исследования частоты попаданий. Анализ этих данных для каждой партии реагентов, более подробно описанный в 8.2.3.2, позволяет определение ожидаемой частоты попаданий в интервал образцов. Наиболее распространенным параметром, определяемым таким образом, является значение C95 (определяется по эталону количественной шкалы). Согласно [16], если отсечкой является *LoB*, то это значение часто называют *LLoD*.

**8.2.3 Исследования по обнаружению аналитов с использованием алгоритма вместо внутренней непрерывной реакции с отсечкой**

8.2.3.1 Точные исследования

Хотя эти исследования дают только результаты «да» или «нет» для образцов, все равно следует проводить те же внутрилабораторные испытания и испытания воспроизводимости, которые описаны в 8.2.2. Часто невозможно провести *ANOVA*, а C5 и C95 не могут быть измерены по шкале *ICR*. Для некоторых исследований *ICR* может быть доступен для образцов в высоких концентрациях. В таких случаях следует руководствоваться правилами *ANOVA*, описанными в 8.2.2.2. Для каждого типа исследования должны быть представлены бинарные отчеты, как показано в 8.2.4. Для панели прецизионности исследования обнаружения аналита уровень образца ниже отсечки может быть определен с помощью результатов кандидата и компаратора.

Если более 10 % всех испытуемых из целевой популяции, признанных компаратором положительными, имеют наблюдаемое значение меньше C95 в исследовании-кандидате, высокоотрицательный образец не включается в панель точности. Он заменяется образцом в интервале, в котором бинарное исследование признает образец положительным от 20 % до 80 % времени, как показано на рисунке 10.

Если менее 10 % всех испытуемых из популяции предполагаемого использования, признанных компаратором положительными, имеют наблюдаемое значение меньше C95 в исследовании-кандидате, высокоотрицательный образец исключается из прецизионной панели.

**Скорость попадания**



**Копии/см3**

60

50

40

30

20

10

0

0.2

0.1

0.0

C20

0.3

C80

0.4

80%

20%

0.5

Пробит

1.0

0.9

0.8

0.7

0.6

C20 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 20 % случаев; C80 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 80 % случаев

Рисунок 10 − Прецизионные уровни панели

8.2.3.2 Исследования частоты попаданий для определения нижнего предела обнаружения

В исследованиях с коэффициентом попадания значения образцов получают путем их измерения с помощью процедуры количественного эталонного измерения или создания путем разбавления эталонного материала. Схема исследования коэффициента попадания такая же, как описана в 8.2.2.3 для оценки значения образца, при котором 95 % реплик будут признаны положительными (т. е. *LLoD*).

Для исследований, связанных с обнаружением аналитов, цель состоит в том, чтобы образцы без аналитов имели 0 % положительных (обнаруженных) результатов. Для оценки *LLoD* следует использовать пробит-анализ (или альтернативный подходящий анализ, например, с использованием обобщенной линейной модели) частоты попаданий. Более подробная информация приведена в приложениях Б и В. Стандартную пробит-модель, использующую логарифмически преобразованную концентрацию *X*, вычисляют по формуле

(5)

где  *X* − концентрация аналита (для *X* > 0);

*Ppos(X)* − вероятность получения положительного результата исследований в зависимости от концентрации *X*;

*Φ* − кумулятивная стандартная нормальная функция вероятности;

*a и b* − перехват и наклон пробита, соответственно;

*log* − логарифм с основанием 10. Допускается использовать и другие основания; допускается использовать и другие преобразования, например,   
*log от X'* = *X+ c*, для которого *c* − небольшое положительное число (см. [18]).

На рисунке 11 приведена подгонка к преобразованному в log10 набору данных при *c* = 0,01.

**Скорость попадания**

**Log, копий/см3**

2

1

0

-1

-2

0.1

Проба без

аналитов

0.3

0.2

Probit 95%

C95

0.4

Подгонка

по методу

1.0

0.9

0.8

0.7

0.6

0.5

C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев

Рисунок 11 − Стандартная кривая пробит

Многие пакеты статистического программного обеспечения обеспечивают нелинейную подгонку этой пробит-модели после ввода экспериментальных данных. Более подробное объяснение того, как эта модель подходит к данным о количестве попаданий, независимо от того, преобразована ли ось *x*, приведено в приложении В. Если вариабельность присуща исследованиям по обнаружению аналитов из-за системного шума или несоответствия образцов, образцы без аналитов могут иметь результаты *FP*. В таких случаях стандартная пробит-модель, описанная выше, должна быть расширена, поскольку немодифицированная модель предсказывает частоту попаданий для образца без аналита как 0 %. В расширенной модели образцы с отсутствием аналита будут иметь α процент обнаруженных результатов, где   
α − вероятность ошибки I типа (*P* = вероятность *FP*). Расширенная пробит-модель (см. [22]) представлена в уравнении (рассчитывается по формуле

(6)

где все общие параметры те же, что и в стандартной пробит-модели, *X* > 0, а *P* − вероятность *FP*. Замена *X* на *X*' = *X* + *c*, причем *c* − небольшое положительное число, означает, что модель больше не ограничена нулем. Статистическое программное обеспечение также может быть использовано в этом случае для нелинейной подгонки экспериментальных данных.

Из формулы (6) следует, что *Ppos(X)* приближается к *PFP* по мере приближения *X* к нулю (или *X*' к *c*) (см. рисунок 12).

Расширенная пробит-модель обычно дает меньшую оценку *LLoD*, чем стандартная пробит-модель.

**Доля положительных результатов**

1.0

0.95

0.75

0.5

Проба без

аналитов

0.25

LLoD

0.1

0

**Концентрация**

*LLoD* − нижний предел обнаружения; *PFP* − вероятность ложноположительного результата

Рисунок 12 − Расширенная кривая пробита с *PFP* 5 %

**8.2.4 Пример бинарного отчета**

Во всех типах качественных исследований результаты исследования могут быть представлены в виде частоты, с которой образец признан положительным, отрицательным или недействительным. Например, в схеме исследования, в котором оцениваются три партии реагентов с двумя операторами и двумя приборами в течение пяти дней (см. таблицу 3 для ежедневной логистики), бинарные результаты за все дни для образцов на разных уровнях (например, C5, C95 и C100) обобщены в таблице 4. Хотя это не показано в таблице 4, цель исследования на обнаружение аналитов для образца без аналитов − никогда не признавать его положительным. Примеры дизайна внутрилабораторных и воспроизводимых исследований приведены в приложении Е.

Таблица 3 − Количество реплик для внутрилабораторного прецизионного дизайна исследования

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | День 1 | | | | | | | | | | | |
| Набор реагентов 1 | | | | Набор реагентов 2 | | | | Набор реагентов 3 | | | |
| Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | |
| Инст 1 | Инст 2 | Инст 1 | Инст 2 | Инст 1 | Инст 2 | Инст 1 | Инст 2 | Инст 1 | Инст 2 | Инст 1 | Инст 2 |
| Реплики | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Примечание – Инст – прибор. | | | | | | | | | | | | |

Таблица 4 − Пример подсчета результатов бинарного исследования

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Источник | Номер | Уровень 1, C5 | | | Уровень 2, C95 | | | Уровень 3, C100 | | |
| Положительный | Отрицательный | Недействительный | Положительный | Отрицательный | Недействительный | Положительный | Отрицательный | Недействительный |
| Комбинированный | 120 | 6 | 114 | 0 | 113 | 7 | 0 | 119 | 0 | 1 |
| День 1 | 24 | 2 | 22 | 0 | 23 | 1 | 0 | 23 | 0 | 1 |
| День 2 | 24 | 1 | 23 | 0 | 22 | 2 | 0 | 24 | 0 | 0 |
| День 3 | 24 | 1 | 23 | 0 | 22 | 2 | 0 | 24 | 0 | 0 |
| День 4 | 24 | 1 | 23 | 0 | 23 | 1 | 0 | 24 | 0 | 0 |
| День 5 | 24 | 1 | 23 | 0 | 23 | 1 | 0 | 24 | 0 | 0 |
| Партия реагентов  1 | 40 | 2 | 38 | 0 | 37 | 3 | 0 | 40 | 0 | 0 |
| Партия реагентов  2 | 40 | 2 | 38 | 0 | 38 | 2 | 0 | 40 | 0 | 0 |
| Партия реагентов  3 | 40 | 2 | 38 | 0 | 38 | 2 | 0 | 39 | 0 | 1 |
| Оператор 1 | 60 | 3 | 57 | 0 | 57 | 3 | 0 | 60 | 0 | 0 |
| Оператор 2 | 60 | 3 | 57 | 0 | 56 | 4 | 0 | 59 | 0 | 1 |
| Инструмент 1 | 60 | 3 | 57 | 0 | 56 | 4 | 0 | 60 | 0 | 0 |
| Инструмент 2 | 60 | 3 | 57 | 0 | 57 | 3 | 0 | 59 | 0 | 1 |
| Примечание – Обозначения: C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; C100 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 100 % случаев; N − количество общих реплик для каждого уровня. | | | | | | | | | | |

Кроме вычисления пропорций или процентов для каждой ячейки таблицы, никакого анализа таких данных не требуется. Следует изучить и объяснить сильно разнящиеся показатели попаданий. Сложив доли определения каждого уровня в таблице может помочь прояснить, были ли получены ожидаемые результаты и какие факторы требуют изучения. Например, если у одного оператора доля положительных результатов на С95 значительно выше, чем у другого, это может указывать на недостаточный уровень подготовки по использованию исследования. Аналогичным образом, очень разные результаты по партиям реагентов могут указывать на несоответствующий процесс производства реагентов. В приложении Г приведен пример анализа дляисследования методом секвенирования следующего поколения (*NGS*).

**8.2.5 Положения, касающиеся прецизионных исследований**

Для исследований с визуальным чтением в двоичном формате *ICR* не существует. Как указано в подразделе 8.2.3.1 и таблице 3, оба внутрилабораторных исследования точности и воспроизводимости допускается проводить с использованием бинарного отчета, поскольку это единственный доступный анализ неточности. Если исследование предназначено в первую очередь для использования вне лаборатории, в испытании воспроизводимости следует задействовать не менее трех внешних мест. Более подробное описание оценки исследования, учитывающего вариативность читателя, приведено в приложении Ж.

Если разработчик хочет заявить об использовании нескольких типов образцов для исследования, при создании дизайна исследования следует ссылаться на [23]. Любой основной или преобладающий тип образца должен быть тщательно валидирован в соответствии с описанными выше схемами прецизионных испытаний. Согласно [23], если было показано, что другие типы образцов имеют эквивалентные характеристики, может быть проведена менее строгая проверка результатов на таких типах образцов.

Настоящий стандарт не содержит руководства по оценке эффективности для мультиплексных тестов амплификации нуклеиновых кислот (см. [24]) или для микрочипов (см. [25]).

**8.3 Оценка клинической деятельности**

8.3.1 Качественные исследования используют для скрининга, выявления, прогноза, оценки риска и других целей. Требования к клиническим характеристикам могут различаться в зависимости от клинической необходимости. Клиническая эффективность кандидата на проведение качественного бинарного исследования может быть охарактеризована его способностью классифицировать образцы или субъектов из предполагаемой популяции в их правильные бинарные классы (ЦС присутствует или ЦС отсутствует), как определено с помощью компаратора. В качестве компаратора может выступать наилучший доступный метод определения наличия или отсутствия ЦС или другое исследование.

**8.3.2 Сравнение с известным целевым состоянием**

Если компаратор является наилучшей доступной оценкой ЦС, то компаратор рассматривают как оценку ЦС, которая обеспечивает известное ЦС (присутствующее или отсутствующее). В этих случаях можно оценить клиническую эффективность исследования-кандидате. Наиболее распространенной парой показателей клинической эффективности является клиническая чувствительность и клиническая специфичность. Клиническая чувствительность − это доля испытуемых с ЦС, признанных положительными по результатам исследований-кандидатов. Клиническая специфичность − это доля лиц без ЦС, у которых исследование-кандидат выявило отрицательный результат.

**8.3.3 Сравнение с альтернативным исследованием**

Если исследование-кандидате оценивают путем сравнения со сравнительным исследованием, которое не является общепризнанным лучшим методом оценки ЦС, клиническая чувствительность и клиническая специфичность не могут быть легко оценены.

Вместо этого исследование-кандидата характеризуется способностью классифицировать образцы или предметы так же, как сравнительное исследование классифицирует предметы, не зная, правильна ли эта классификация. В такой ситуации оценки эффективности рассчитываются так же, как чувствительность и специфичность, но оценки называются положительным процентным согласием (*PPA*) и отрицательным процентным согласием (*NPA*), чтобы отразить, что эти оценки не дают количественной оценки клинической эффективности, а скорее определяют согласие исследования-кандидата со сравнительным исследованием. Кроме того, такие величины, как положительная прогностическая ценность (*PPV*), отрицательная прогностическая ценность (*NPV*), а также положительное и отрицательное отношение правдоподобия (*PLR* и *NLR*), не могут быть рассчитаны, поскольку состояние испытуемого неизвестно.

**8.3.4 Положения об исследовании эффективности классификации**

8.3.4.1 Положения данного пункта содержат общие рекомендации по выбору компаратора и отбору образцов для различных ситуаций. Для исследований, предназначенных для оценки клинической эффективности, необходимо включить следующие два ключевых компонента, которые более подробно рассматриваются в [15]:

- стандартное или сравнительное исследование, или наилучший доступный метод оценки ЦС;

- субъекты с распределением (спектром) состояний здоровья, репрезентативным для предполагаемой популяции пользователей.

8.3.4.2 Определение критериев приемлемости результатов исследования

Критерии приемлемости результатов исследования устанавливают на основе анализа соотношения пользы и риска с точки зрения допустимой ошибочной классификации с учетом предполагаемого использования исследования-кандидата. Критерии должны быть сформулированы в терминах оценок и статистической достоверности, необходимых для оценки эффективности исследования. Поскольку результаты исследования характеризуются парой показателей эффективности (например, чувствительность и специфичность или *PPA* и *NPA*), для каждой пары показателей необходимы критерии приемлемости. Критерии приемлемости обычно основаны на критериях границ доверительных интервалов (ДИ) для оценок эффективности (например, критерии нижних границ ДИ). Одних только точечных оценок недостаточно для оценки эффективности исследования, поскольку они не отражают вариабельность оценок. Исследование должно быть достаточного размера, чтобы оценить исследование-кандидата на основе желаемого уровня неопределенности (т. е. ДИ) и целевой эффективности, что влияет на размер исследования, как описано в 8.3.4.4б).

8.3.4.3 Выбор компаратора

Выбор компаратора должен определяться типом необходимой характеристики. Если целью является характеристика клинической чувствительности и специфичности, а также соответствующих отношений правдоподобия (ОВ) и прогностических значений, следует использовать сравнительный метод, который является лучшим из доступных методов оценки ЦС. Таким стандартом может быть сочетание наилучших доступных аналитических методов и медицинских критериев принятия решения для определения присутствия или отсутствия интересующего состояния. Например, для кандидата на проведение теста на ДНК вируса папилломы человека (ВПЧ), предназначенного для помощи в диагностике рака шейки матки, оценка ЦС может представлять собой предварительно определенный алгоритм, включающий критерии, основанные на гистологии и клиническом наблюдении. Методы оценки ЦС развиваются с течением времени по мере расширения знаний и совершенствования аналитических систем. Поэтому клиническая чувствительность и специфичность всегда должны быть представлены и интерпретированы при четком описании использованного компаратора. Информация о том, как определить результаты исследований при сравнении с оценкой ЦС, приведены в 8.3.5.

В некоторых случаях оценка ЦС основана исключительно на клинической картине у испытуемых, например, признаках и симптомах предполагаемой инфекции или потере регуляторной функции (неконтролируемый рост клеток). Кроме того, бывают и другие ситуации, когда наличие или отсутствие ЦС напрямую совпадает с наличием или отсутствием одного аналита. В качестве примера можно привести использование теста на наркотики для анализа образца на наличие или отсутствие метадона. В этой ситуации оценка ЦС может быть аналитической референсной процедурой измерения аналита (т. е. метадона).

В приведенном выше примере с ВПЧ другое качественное бинарное исследование может быть использовано в аналитическом исследовании согласия (исследование, оценивающее согласие результатов исследования целевых аналитов), в то время как процедура биопсии шейки матки под контролем кольпоскопии необходима для определения наличия или отсутствия рака шейки матки в исследовании клинической эффективности. Каждое исследование имеет свою цель, и исследование-кандидат должно оцениваться отдельно путем сравнения с подходящими компараторами.

Для лаборатории или разработчика кандидата на проведение качественного бинарного исследования, который хочет заменить существующий метод, целью может быть демонстрация способности кандидата на проведение анализа в достаточной степени соответствовать сравнительной экспертизе, используемой в настоящее время. Для этой цели сравнительным методом может быть используемое в настоящее время исследование. Однако если новое исследование лучше (например, более чувствительна), чем ранее применяемый метод, эти два метода могут не совпадать. Эта относительная характеристика может также применяться к сравнительному исследованию, которая использовалась для оценки ЦС. Обсуждение расхождения результатов приведено в 8.3.6.3.

8.3.4.4 Отбор испытуемых и образцов

Образцы, включенные в исследование клинической эффективности, должны представлять популяцию предполагаемого использования. Если участвующие в исследовании субъекты не являются репрезентативными для предполагаемой популяции, оценки клинической эффективности подвержены эффекту спектра (см. [26], [27]). Например, если в выборку попадают только субъекты из крайних точек ЦС (например, либо здоровые субъекты, либо субъекты с поздней стадией заболевания), эффективность может казаться лучше, чем она есть на самом деле. Этот эффект возникает потому, что испытуемые с результатами, близкими к базовой границе исследования, которые не попадают в выборку, обычно труднее поставить правильный диагноз. Когда испытуемые находятся ближе к базовому срезу, присущая исследованию вариабельность может увеличить вероятность постановки неправильного диагноза, если исследование не было правильно разработано. Важность правильного спектра испытуемых, отражающего предполагаемую популяцию, рассмотрено в различных источниках (см. [15], [28]−[30]).

а) процесс отбора

Перспективное планирование помогает гарантировать, что исследуемые субъекты и образцы обеспечат доказательства, подтверждающие предполагаемое использование, и минимизируют возможность предвзятого отношения к результатам исследования. Рекомендуется использовать протокол, в котором указаны процесс набора и отбора испытуемых, критерии включения или исключения испытуемых и образцов, а также тип мест проведения испытаний (объем и опыт). Наилучшим сценарием для минимизации потенциальной предвзятости является использование протокола, для которого верны следующие утверждения:

- протоколы отбора испытуемых устанавливают до набора и тестирования (чтобы результаты теста не влияли на выбор испытуемых или клиническое исследование);

- пробы собирают и исследуют немедленно (в отличие от хранения и последующего тестирования).

Если для оценки используют хранящиеся или архивные образцы, разработчик исследования должен быть уверен в стабильности таких образцов в установленных условиях хранения, включая надлежащее использование процедур замораживания−оттаивания. Если для оценки используются образцы, ранее собранные для другой цели, репрезентативность этих образцов для текущего предполагаемого использования должна быть оценена, насколько это возможно, путем понимания первоначальных критериев отбора и того, как образцы с определенными характеристиками были исключены из коллекции.

б) количество испытуемых или образцов

Общее количество испытуемых или образцов, необходимых для проведения исследования, определяется исходя из желаемой вероятности успеха (мощности) для достижения заявленных целей, при этом учитывают более широкие последствия исследования. Чтобы определить размер исследования, разработчик может принять во внимание:

- знание истинной эффективности методов-кандидатов и методов-компараторов и уверенность в этом знании;

- риск, связанный с невыполнением критериев приемки производительности;

- рабочий процесс или технологические требования;

- необходимость сохранения объема образца для будущих исследований.

Размер исследования рассматривается отдельно для субъектов или образцов с интересующим ЦС и без него, поскольку эффективность включает чувствительность и специфичность или *PPA* и *NPA*. Как правило, на одного испытуемого приходится один образец, так что размер исследования отражает количество испытуемых. Если на одного испытуемого приходится более одного образца, необходимо рассмотреть статистические методы оценки эффективности, отличные от представленных в настоящем стандарте (см. методы статистического анализа для кластеризованных данных в [31]). Дополнительное рассмотрение данной ситуации не входит в область применения настоящего стандарта.

**8.3.5 Сравнение бинарного исследования с оценкой состояния цели**

Клиническую эффективность качественного бинарного исследования обычно оценивают с точки зрения правильности классификации (клиническая чувствительность и специфичность) и прогностической ценности (положительная и отрицательная). Таблица 5 представляет собой таблицу случайностей 2×2, в которой сравниваются бинарные результаты качественного исследования и результаты оценки ЦС, используемые для определения наличия или отсутствия ЦС. Запись в каждой ячейке таблицы представляет собой количество образцов (или субъектов, предполагая один образец на субъекта), соответствующих заголовкам в таблице.

Таблица 5 − Качественное исследование в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС Отсутствует |
| Положительный | Количество *TP* = *A* | Количество *FP* = *B* | *A* + *B* |
| Отрицательный | Количество *FN* = *C* | Количество *TN* = *D* | *C* + *D* |
| Всего | *N*1 = *A* + *C* | *N*0 = *B* + *D* | *N* = *N*1 + *N*0 = *A* + *B* + *C* + *D* |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначения: *FN* − ложноотрицательный; *FP* − ложноположительный; *N* − количество образцов; ЦC − целевое состояние; *TN* − истинно отрицательный; *TP* − истинно положительный. | | | |

Формулу (7) используют для расчета клинической чувствительности, а формулу (8) − для расчета клинической специфичности.

Клиническая чувствительность (*S*e) = (7)

Клиническая специфичность (Sp) = (8)

Примечание – Все коэффициенты, относящиеся к клиническим показателям, включая клиническую чувствительность и специфичность, можно выразить в процентах, умножив коэффициент на 100 %.

Если частота *TP* (чувствительность) равна частоте *FP* (1 − специфичность), то исследование не дает никакой информации о ЦС. С другой стороны, если чувствительность и специфичность близки к 100 %, исследование имеет отличные клинические показатели.

*PLR*, который может быть рассчитан с помощью формулы (9), определяет, насколько велика вероятность положительного результата обследования у субъекта с ЦС по сравнению с субъектом без ЦС. Аналогично *NLR*, который может быть рассчитан с помощью формулы (10), определяет, насколько велика вероятность отрицательного результата обследования у субъекта с ЦС по сравнению с субъектом без ЦС.

(9)

(10)

*LR*, равный единице, указывает на то, что исследование не информативно, т. е. результат исследования не дает никакой дополнительной информации, кроме преобладания информации о наличии у испытуемого ЦС. Исследование с *PLR* > 1 и *NLR* < 1 считаются информативными, и *LR*, далекий от единицы, более желателен, чем *LR*, близкий к единице. Большее значение *PLR* и меньшее значение *NLR* указывают на большую диагностическую способность исследования и являются предпочтительными. Положительные и отрицательные *LR* могут быть объединены для получения отношения шансов (*OR*), как показано в формуле (11) (см. [32]).

(11)

Распространенность − частота встречаемости ЦС в определенной группе или популяции, выраженная в виде пропорции. Если оценку исследования проводят на случайной выборке пациентов пользователя, распространенность для оценки и распространенность, ожидаемая при типичном использовании исследования, должны быть похожи. Измеренная распространенность может отличаться от ожидаемой, если, например, оценка включает повышенное число положительных пациентов. Оценку распространенности вычисляют по формуле

(12)

Другая распространенная пара показателей эффективности − *PPV* и *NPV* исследования на ЦС. *PPV* и *NPV* всегда следует представлять и рассматривать как пару показателей. *PPV* − доля испытуемых, признанных положительными, у которых есть ЦС согласно оценке ЦС [см. формулу (13)]. *NPV* − доля испытуемых, признанных отрицательными, у которых нет ЦС согласно оценке ЦС [см. формулу (14)]. *PPV* и *NPV* важны для характеристики, поскольку клинические решения по ведению и лечению испытуемых часто основываются, по крайней мере частично, на результатах исследования. *PPV* и *NPV* отражают ситуацию с отдельным пациентом, поскольку они определяют условную вероятность наличия или отсутствия ЦС при положительном или отрицательном результате исследования. Однако *PPV* и *NPV* не являются показателями внутренней клинической эффективности метода, как клиническая чувствительность и клиническая специфичность, и не должны рассчитываться на основе исследования, предназначенного для автоматического определения клинической чувствительности и специфичности. Если их рассчитывают, то необходимо учитывать распространенность заболевания в исследуемой популяции. Однако расчеты *PPV* и *NPV* зависят от распространенности ЦС, которая может варьироваться в зависимости от географического региона, учреждения или других факторов. Для интерпретации *PPV* и *NPV* необходимо учитывать, что наблюдаемые значения должны быть установлены в соотношении с распространенностью и «1 минус распространенность», соответственно, для интерпретации. Так, тест с очевидно низким *PPV* (например, 15 %) может оказаться очень полезным, если распространенность заболевания составляет 1 %. А высокая *NPV* (например, 99 %) может оказаться бесполезной в случае низкой распространенности (например, 1 %). Более того, при оценке клинической эффективности распространенность ЦС может быть искажена. Например, в перспективном исследовании редкого ЦС субъекты с ЦС могут быть намеренно завышены по сравнению с их реальной распространенностью, чтобы обеспечить размер исследования, позволяющий снизить неопределенность оценок эффективности для субъектов с ЦС.

Прогностические значения бинарного исследования объединяют распространенность ЦС с чувствительностью и специфичностью. Распространенность для оценки важна для определения *PPV* и *NPV* исследования. Прогностические значения применимы только к популяциям, которые имеют сходную распространенность и спектр ЦС с популяциями, использованными для оценки прогностического значения.

*PPV*, %, и *NPV*, %,вычисляют по формулам

(13)

(14)

Поскольку *PPV* и *NPV* зависят от распространенности, иногда бывает полезно определить *PPV* и *NPV* при фиксированной распространенности по формулам

(15)

(16)

*PLR* и *NLR* могут быть полезны из-за их связи с *PPV* и *NPV*. В отличие от *PPV* и *NPV*, *PLR* и *NLR* не зависят от распространенности, поскольку они являются функциями чувствительности и специфичности, которые представляют собой вероятности, обусловленные наличием у субъекта ЦС. Тем не менее, *PLR* и *NLR* тесно связаны с *PPV* и *NPV*. *PPV* увеличивается с увеличением *PLR*. Таким образом, тест с более высоким *PLR*, чем другой тест для того же ЦС и предполагаемой популяции, имеет более высокий *PPV*. Более точно, *PLR* − относительное увеличение шансов на ЦС, связанное с положительным результатом исследования. Например, если *PLR* равен 3,0, вероятность наличия ЦС в три раза выше для субъекта с положительным результатом исследования, чем для нетестированного субъекта, случайно выбранного из популяции предполагаемого использования. Формула (17) описывает взаимосвязь между *PPV* и *PLR*.

(17)

Аналогичным образом *NPV* увеличивается с уменьшением *NLR*. Таким образом, исследование с более низким *NLR*, чем другое исследование для того же ЦС и предполагаемой популяции, имеет более высокую *NPV*. Более точно, *NLR* − относительное снижение вероятности ЦС, связанное с отрицательным результатом исследования. Например, если *NLR* = 1/3, то вероятность наличия ЦС в три раза меньше для субъекта, получившего отрицательный результат, чем для нетестированного (случайно выбранного) субъекта. Формула (18) описывает связь между *NPV* и *NLR*.

(18)

Основополагающим результатом является то, что *PPV* монотонно возрастает при *PLR*, а *NPV* монотонно убывает при *NLR*. Если *PLR* больше, а *NLR* меньше для одного исследования по сравнению со вторым исследованием, то первое исследование имеет лучшие *PPV* и *NPV* для той же распространенности. Таким образом, *LR* можно использовать для определения порядка прогностических значений для двух исследований без знания распространенности ЦС, что может иметь место в некоторых исследованиях, распространяющихся на ЦС. КИ для LR и прогностических значений могут быть основаны на ДИ отношения двух независимых биномиальных пропорций (см. [31], [33] и [34]). Благодаря математической взаимосвязи между параметрами *NLR*, *NPV*, *PLR* и *PPV*, оценки из исследования клинической эффективности могут быть объединены с установленной распространенностью для прогнозирования значений *PPV* и *NPV* для исследования, что, в свою очередь, может дать представление о том, как интерпретировать результаты исследования и, возможно, обосновать клинические решения.

Показатели эффективности, рассчитанные с помощью формул (7) − (18), являются лишь оценками истинной эффективности, поскольку они основаны только на подмножестве испытуемых из целевой популяции. Если будет протестировано другое подмножество испытуемых или если те же самые испытуемые будут протестированы в другое время (в которое ЦС или результат исследования может измениться для некоторых испытуемых), оценки эффективности будут численно отличаться. Оценочные показатели эффективности могут быть подвержены смещению. К аспектам исследования, которые могут внести погрешность, относятся набор испытуемых и процесс отбора, не соответствующий предполагаемому использованию, отсутствие слепого контроля за клиническими результатами со стороны коллекторов, а также процедура анализа данных. Потенциальные источники предвзятости должны быть известны, чтобы их можно было избежать или свести к минимуму. Простое увеличение общего числа испытуемых в исследовании не уменьшит смещение. Информацию по отбору субъектов для исследования приведена в 8.3.4.4. Всестороннее рассмотрение предвзятости в исследованиях диагностических устройств приведено в [15], [31], [35], [36].

8.3.5.2 Доверительные интервалы для чувствительности и специфичности

Доверительные границы чувствительности и специфичности могут быть рассчитаны на основе подходов к расчету ДИ для биномиальной пропорции. Простой метод заключается в том, что ДИ основывается на том, что доля выборки распределена приблизительно нормально в больших выборках. Однако предположение о нормальности не всегда подходит для небольших выборок или для пропорций, близких к нулю или 100 % (см. [37]); поэтому методы расчета ДИ, основанные на нормальном приближении, не рекомендуются. Существует несколько методов, которые можно получить из многих пакетов программ или из опубликованных таблиц ( см. [38] − [40]). Рекомендуется использовать метод баллов Уилсона, поскольку он обладает хорошими статистическими свойствами и может быть рассчитан напрямую (см. [37] − [39], [41], [42]). Информация о получении необходимых расчетных значений приведена в приложении И (формула И.1). Доверительные границы баллов Уилсона имеют тенденцию давать более узкие ДИ, чем ДИ Клоппера-Пирсона, что приводит к увеличению нижней доверительной границы. Например, при *n* = 30 образцов и 29/30 = 96,7 % нижняя граница двустороннего 95 %-ного ДИ составляет 83,3 %, в то время как нижняя граница Клоппера-Пирсона равна 82,8 %. В настоящем стандарте рассматриваются только двусторонние ДИ.

8.3.5.3 Положения об эквивокальной зоне

Эквивокальная зона часто вводится для повышения чувствительности или специфичности исследования, если в паре с клиническим процессом рекомендуется провести повторное исследование. Однако проведение такого повторного тестирования во время оценки результатов качественного бинарного исследования кандидата в клиническом исследовании часто нецелесообразна. Включение результатов последующего исследования в оценку результативности исследования-кандидата может быть не обобщающим и не полезным, если последующее исследование и интерпретация результатов не указаны. Поэтому при такой оценке результаты исследования-кандидата являются положительными, двусмысленными или отрицательными, а результаты исследования могут быть представлены в виде таблицы случайностей 3×2, как показано в таблице 6.

Таблица 6 − Качественное исследование с неоднозначными результатами в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС Отсутствует |
| Положительное | *A*1 | *B*1 | *A*1 + *B*1 |
| Эквивокальное | *A*2 | *B*2 | *A*2 + *B*2 |
| Отрицательное | *A*3 | *B*3 | *A*3 + *B*3 |
| Всего | *N*1 = *A*1 + *A*2 + *A*3 | *N*0 = *B*1 + *B*2 + *B*3 | *N* |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначения: *N* − количество образцов; ЦС − целевое состояние. | | | |

Для оценки эффективности допускается рассматривать наихудший сценарий. В этом сценарии, если чувствительность рассчитывают по формуле (19), равновероятные результаты (*A*2) считаются отрицательными. Если специфичность рассчитывают по формуле (20), то эквивокальные результаты (*B*2) считаются положительными. Таким образом, доля эквивокальных результатов вычисляется по формуле (21)

(19)

(20)

(21)

Эквивокальный результат можно рассматривать как третью категорию при оценке таблицы случайностей 3×2. Тогда *LR* эквивокального результата может быть рассчитан и интерпретирован как изменение шансов на ЦС, которые дает эквивокальный результат, по отношению к случайно выбранному (не тестированному) субъекту. Если исследование-кандаидат должен быть повторен, если первый результат оказался неоднозначным, исследование-кандаидат может быть спланировано таким образом, чтобы включить результаты повторных исследований. Если повторные исследования всегда приводят к тому, что испытуемые с первоначальными неоднозначными результатами получают либо положительный, либо отрицательный результат исследования, то полученная таблица 2×2 условных единиц может служить основой для оценки эффективности, а показатель повторного исследования указывается отдельно. Если все неоднозначные результаты не могут быть разрешены при повторном исследовании, итоговая таблица 3×2 случайностей, полученная после повторных исследований, оценивается или сжимается в таблицу 2×2 случайностей с наихудшим сценарием, описанным выше, если это необходимо.

**8.3.6 Согласие качественного исследования с бинарным выходом**

8.3.6.1 В распространенной ситуации, когда компаратор не является общепринятой оценкой ЦС, клиническая чувствительность и специфичность не могут быть легко оценены. Вместо этого проводятся те же численные расчеты, но оценки называются *PPA* и *NPA*, а не клинической чувствительностью и специфичностью, поскольку речь идет не о клинических показателях, а о согласии исследования-кандидата со сравнительным исследованием. Кроме того, такие величины, как *PPV*, *NPV*, а также положительный и отрицательный *LR*, не могут быть рассчитаны, поскольку состояние обследуемого неизвестно.

8.3.6.2 Оценка согласия между исследованием-кандидатом и сравнительным исследованием

Если метод исследования-кандидата сравнивают с чем-то другим, а не с наилучшей доступной оценкой ЦС, полезным может оказаться использование таблицы результатов 2×2 и отчета о том, насколько часто совпадают результаты исследования-кандидата и сравнительного исследования. Пример представления результатов приведен в таблице 7.

Таблица 7 − 2×2 таблица случайностей для сравнительного исследования

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Сравнительное исследование | | |
| Положительное | Отрицательное | Всего |
| Положительное | *A* | *B* | *A* + *B* |
| Отрицательное | *C* | *D* | *C* + *D* |
| Всего | *A* + *C* | *B* + *D* | *N* = *A* + *B* + *C* + *D* |
| Примечание − *N* − количество образцов. | | | |

В отличие от таблицы 5, которая представляет собой сравнение с оценкой ЦС, столбцы таблицы 7 являются результатом сравнительного исследования. Таким образом, хотя оба исследования могут быть согласованы, они могут быть неверными по сравнению с реальным ЦС. Поэтому записи в таблице 7 (*A*, *B*, *C*, *D*) не являются такими же определяемыми, как *TP*, *FP*, *FN* и *TN*, соответственно, и не могут быть использованы для оценки клинической чувствительности или клинической специфичности; они могут быть использованы только для оценки *PPA* и *NPA*.

*PPA*, %, и *NPA*, %, вычисляют по формулам

(22)

(23)

Если компаратором является исследование, ДИ биномиальной пропорции (клиническая чувствительность и специфичность) такие же, как и при оценке ЦС, но выражены в процентах. Примеры вычислений приведены в 8.3.2 и приложении К.

8.3.6.3 Разрешение несоответствий

Расхождения между исследованием-кандидатом и сравнительным исследованием могут возникнуть из-за ошибок в одном из них. Кроме того, с развитием новых технологий возможно, что исследование-кандидат имеет лучшие клинические характеристики, чем лучшая из доступных в настоящее время оценок ЦC, используемых для сравнения. Одним из примеров такой лучшей производительности является сравнение *NGS* с двунаправленным секвенированием Сангера (*BDS*) для выявления специфических генетических мутаций. Если C95 для BDS выше, чем C95 для *NGS*, могут возникнуть ситуации, когда *BDS* не обнаруживает мутацию, а *NGS* обнаруживает, в результате чего результат *NGS* будет ошибочно обозначен как «*FP*» при наличии мутации. В таких случаях можно провести дополнительные исследования, чтобы оценить расхождение результатов.

В примерах, подобных приведенному выше, вероятность получения неверных результатов можно определить, оценив сначала характеристики эффективности (C95) как исследования-кандидата, так и сравнительного исследования, используя эталонные материалы (в идеале − сертифицированные эталонные материалы) для оценки каждого C95. Образцы вблизи отсечки и С95 по-прежнему следует оценивать с помощью образцов пациентов и их разведений соответствующим разбавителем (обычно образцы пациентов с нулевым содержанием аналитов), но критерии приемки должны учитывать различия в С95. В целом, для тех исследований, в которых используется *ICR*, список числовых результатов исследования-кандидата и сравнительного исследования изучают, близки ли расхождения в результатах к отсечению для обоих исследований или расхождения могут быть обусловлены различиями в C95. Если C95 ближе к отсечению для исследования-кандидата, чем для компаратора, может потребоваться большее количество образцов, чтобы продемонстрировать, что два метода исследований сопоставимы.

В других ситуациях, когда возникают расхождения в результатах, расходящиеся образцы могут быть дополнительно оценены такой «судейской» оценкой, если есть другой способ дополнительной оценки образцов с наилучшим доступным в настоящее время методом оценки ЦС или клиническим наблюдением за испытуемым. Дополнительная оценка дает информацию о том, согласно ли исследование-кандидат или сравнительное исследование в большей степени с экспертом, но сама по себе эта информация не может быть использована для прямой оценки эффективности исследования (см. [36]). Один из возможных подходов заключается в том, что все оригинальные результаты остаются в оригинальных таблицах данных 2×2, а диссонирующие образцы снабжаются сносками с результатами исследования, или результаты исследования представляют в отдельном списке, в котором указывают результат исследования-кандидата, результат сравнительного исследования и результат реферирования для каждой дискордантной выборки. Чтобы рассчитать оценку эффективности исследования-кандидата по методу судьи, все или репрезентативное подмножество совпадающих результатов также должно быть оценено судьей, а оценки эффективности должны быть рассчитаны на основе результатов исследования как для совпадающих, так и для несовпадающих образцов.

8.3.7 Сравнение клинической чувствительности и специфичности двух исследований

8.3.7.1 Пользователь может захотеть оценить различия в чувствительности и специфичности между двумя обследованиями. Допускается использовать формулу (И.1) приложения И может быть использовано для оценки ДИ клинической чувствительности и специфичности для каждого исследования. Как указано в 8.3.5, такие расчетные показатели эффективности являются обобщенными (несмещенными) для ожидаемой эффективности лабораторных исследований только в той степени, в которой образцы для оценки являются типичными для образцов, анализируемых в лаборатории. В 8.3.4.4 описано, как выбрать репрезентативных субъектов исследования.

При сравнении двух исследований клиническая чувствительность и специфичность должны сравниваться одновременно, а не по отдельности. Совместное сравнение может показать, что и чувствительность, и специфичность для одного исследования выше, чем для другого, что указывает на превосходство одного исследования над другим. Однако если один из показателей (например, чувствительность) одного исследования лучше, чем другого, а другой парный показатель (например, специфичность) хуже, может быть неочевидно, какое исследование больше подходит для предполагаемого использования. В такой ситуации для сравнения исследований допускается использовать положительный и отрицательный *LR* (или *PPV* и *NPV*) (см. [32]).

8.3.7.2 Сравнение парных клинических чувствительностей и специфичностей

Схема исследования трехстороннего сравнения между исследованием-кандидатом, ранее применяемым исследованием и оценкой ЦС представлена в таблице 8. Данные считаются парными, так как одни и те же образцы исследуются и исследованием-кандидатом, и ранее применяемым исследованием. Результаты исследования-кандидата сравнивают с результатами ранее применяемого исследования при наличии оценки ЦС (для сравнения чувствительности) и отдельно, если оценка ЦС отсутствует (для сравнения специфичности). Когда сравнивают чувствительность и специфичность двух исследований, план исследования должен гарантировать, что порядок тестирования образцов не повлияет на результат. Формулы (25) − (28) используют для расчета разницы в чувствительности между исследованием-кандидатом и ранее применяемым исследованием. Детали того, как эти расчеты могут быть модифицированы для аналогичных расчетов специфичности, описаны в формулах (29) − (31).

Таблица 8 − Трехстороннее сравнение между исследованием-кандидатом, ранее применяемым исследованием и оценкой ЦС

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Оценка ЦСа | | | | Оценка ЦСа | | | |
| ЦС присутствует | | | | ЦС отсутствует | | | |
| Исследование-кандидат | Ранее применяемое исследование | | Всего | Исследование-кандидат | Ранее применяемое исследование | | Всего |
| Поло-житель-ное | Отрицательное | Поло-житель-ное | Отрицательное |
| Положительное | *А*1 | *В*1 | *A*1 + *B*1 | Положительное | *А*0 | *В*0 | *A*0 + *B*0 |
| Отрицательное | *С*1 | *D*1 | *C*1 + *D*1 | Отрицательное | *С*0 | *D*0 | *C*0 + *D*0 |
| Всего | *А*1 + *С*1 | *B*1 + *D*1 | *N*1 | Всего | *А*0 + *С*0 | *B*0 + *D*0 | *N*0 |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначения: *N* − количество образцов; ЦС − целевое состояние. | | | | | | | |

В таблице 8:

*N*1 = *A*1+ *B*1+ *C*1 + *D*1, (24)

*N*0 = *A*0 + *B*0 + *C*0 + *D*0. (25)

По данным таблицы 8 можно построить две таблицы в форме таблицы 5 (*TP A*, *FP B*, *FN C* и *TN D* в сравнении с оценкой ЦC), одну для исследования-кандидата (см. таблицу 9) и одну для ранее применяемого исследования (см. таблицу 10). Однако обратное утверждение не верно. То есть пользователь не может построить таблицу 8, если известны только данные, приведенные в таблице 5, для каждого исследования (исследование-кандидат против оценки ЦС и ранее применяемое исследование против оценки ЦС). Чтобы построить таблицу 8, необходим трехсторонний вывод данных по исследованию-кандидату, ранее применяемому исследованию и оценке ЦС для каждой отдельной выборки.

Таблица 9 – Исследование-кандидат в сравнении с оценкой ТС

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС отсутствует |
| Положительное | *A*1 + *B*1 | *A*0 + *B*0 | *A*1 + *B*1 + *A*0 + *B*0 |
| Отрицательное | *C*1 + *D*1 | *C*0 + *D*0 | *C*1 + *D*1 + *C*0 + *D*0 |
| Всего | *N*1 | *N*0 | *N* |
| aНаилучшая имеющаяся оценка ТС.  Примечание – Обозначения: *N* − количество образцов; ЦС − целевое состояние. | | | |

Таблица 10 – Ранее применяемое исследование в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Ранее применяемое исследование | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС отсутствует |
| Положительное | *A*1 + *C*1 | *A*0 + *C*0 | *A*1 + *C*1 + *A*0 + *C*0 |
| Отрицательное | *B*1 + *D*1 | *B*0 + *D*0 | *B*1 + *D*1 + *B*0 + *D*0 |
| Всего | *N*1 | *N*0 | *N* |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначения: *N* − количество образцов; ЦС − целевое состояние. | | | |

Оцененная клиническая чувствительность исследования-кандидата (*Se*cand) и ранее применяемого исследования (*Se*old), а также их разница могут быть вычислены по формулам

(26)

(27)

(28)

Оценки клинической специфичности двух исследований могут быть вычислены по формулам

(29)

(30)

(31)

ДИ для парных различий чувствительности и специфичности можно рассчитать с помощью формул (И.8) и (И.18), соответственно.

8.3.7.3 Сравнение непарных (независимых) клинических чувствительностей и специфичностей

Чувствительность и специфичность исследований-кандидатов и ранее применяемых исследований также можно сравнить, если для каждого исследования использовать независимые наборы образцов (непарные). Каждый набор состоит из одного обследования и оценки ЦС с помощью таблицы 2× 2, подобной таблице 5. Для любого такого сравнения для двух наборов образцов необходимо использовать одинаковый дизайн (например, спектр популяции пациентов). Трехстороннее сравнение результатов, как, например, результаты, представленные в таблице 8, неприменимо. В [33] рекомендуется метод, основанный на оценке ДИ для двух отдельных пропорций, который описан ниже.

Пусть ∆ обозначает разницу между двумя оцененными чувствительностями,  
∆*Se* = *Se*cand − *Se*old. Пусть *l*1, *u*1, *l*2 и *u*2 обозначают нижнюю и верхнюю границы двухсторонних 95 %-ных ДИ для *Se*cand и *Se*old, соответственно (получены с помощью формулы И.8). Двусторонний 95%-ный ДИ для разницы между непарными чувствительностями, ∆, может быть вычислен по формуле

. (32)

Аналогичным образом можно вычислить 95 %-ный ДИ для разницы между непарными специфичностями. Рабочий пример этих расчетов приведен в приложении К.

**8.4 Испытание на наличие мешающих веществ при проведении качественных исследований**

Разработчик должен изучить влияние мешающих веществ на результаты качественного бинарного исследования так же, как и при количественном исследовании. Такая оценка исследования позволит разработчику понять ошибки, связанные с отдельными образцами, которые не учитываются при определении точности или клинических показателей. Если *ICR* доступны для всех образцов, для разработки исследования можно использовать [2]. Если *ICR* оценить невозможно, разработчик должен провести испытание на интерференцию, описанное в 8.5.

Мешающие вещества классифицируют на эндогенные и экзогенные. Эндогенные мешающие вещества − вещества, источником которых является проба самого пациента, например метаболиты, образующиеся при патологических состояниях или колонизирующие микроорганизмы. Экзогенными интерферентами могут быть вещества, введенные во время лечения пациента, вещества, проглоченные пациентом, или вещества, намеренно введенные в пробу во время подготовки пробы или случайно введенные во время сбора или обработки пробы (например, гемолиз, элюирующий буфер для выделения нуклеиновых кислот). Выбор потенциальных мешающих веществ должен основываться на типе образца (например, цельная кровь, плазма, сыворотка, моча, образцы дыхательных путей, стул, генитальные мазки), предполагаемой популяции и аналите, на котором проводится исследование.

**8.5 Отбор образцов для испытания на наличие мешающих веществ**

При отборе проб учитывают как уровень аналита, выявляемого в ходе исследования, так и концентрация исследуемого интерферента. Как указано в [2], два различных уровня аналитов должны проходить испытание, соответствующее лицам с ЦС и без ЦС. В [43] приведены руководства по выбору уровней интерференционных тестов.

Для качественных исследований с обнаружением аналитов информация о C95 должна использоваться при подготовке образцов для интерференционного исследования. Образец для интерференционного исследования должен содержать наименьшее количество аналита, при котором кандидат в эксперты с очень высокой вероятностью получит положительный результат (C100). Обычно рекомендуется рассматривать образец с 1,5−2-кратным превышением C95. В этом случае отрицательный результат указывает на потенциальную интерференцию. Когда проводят подобные исследования на перекрестную реактивность (например, исследование на ВИЧ, в которое в качестве интерферента добавлен вирус гепатита С), используют отрицательный образец, а положительный результат указывает на перекрестную реактивность.

Для исследований-кандидатов с отсечкой C50 по шкале *ICR* следует оценить отрицательные и положительные погрешности, вызванные потенциальными помехами. Для этой оценки необходимо использовать информацию о C5 и C95. Если C5 известен, расстояние от отсечки до C5, умноженное на 1,5−2, должно использоваться в качестве целевого значения для образца, который, как ожидается, будет признан отрицательным [т. е. C50−2(C50−C5)]. Целевым уровнем ответа для образцов, которые будут признаны положительными, должно быть расстояние от отсечения до C95, умноженное на 1,5−2 [т. е. C50+2(C95− C50)]. Этот результат является разумной оценкой C100.

Иногда можно найти образцы с обоими желаемыми целевыми значениями. Однако часто бывает трудно получить образец естественного происхождения с одним из целевых значений концентрации аналита, и тогда для достижения соответствующего значения допускается использовать образец от человека, не имеющего ЦС, образец сыворотки, образец мочи и т. д. Первая аликвота любого из образцов называется контрольным образцом. Чтобы проверить интерференцию, во второй «контрольный образец» следует добавить тестируемый интерферант. В общем случае, если микроорганизмы проверяются на интерференцию в исследовании, бактерии добавляются на уровне примерно 1×106 колониеобразующих единиц/см3, а вирусы − на уровне   
1× 105 бляшкообразующих единиц/см3. Информация предпочтительных концентрациях многих целевых аналитов-интерферентов приведена в [2].

Оба образца сравнивают с соответствующим контрольным образцом, чтобы определить, оказывает ли интерференция вредное влияние на исследование. Одной реплики может быть достаточно, так как контрольная концентрация известна, но можно рассмотреть и большее количество реплик, если повторяемость исследования или концентрация образца менее определенна. Такое снижение неопределенности особенно актуально, если невозможно оценить *ICR*. Далее, каждый исследуемый образец должен иметь не менее трех-пяти реплик для каждого исследования или контрольного сравнения. Если любая из этих реплик дает неожиданный результат (отличающийся от результата контрольного образца), значит, интерференция проявилась и может потребоваться дополнительное тестирование для количественной оценки уровня интерференции (например, различные уровни интерферирующего вещества и большее количество реплик). Если имеется *ICR*, интерференцию следует оценивать с помощью числовых значений ICR в соответствии с [2].

**8.6 Отчеты о результатах исследований**

8.6.1 Отчеты о результатах исследований по утверждению эффективности должны включать описание дизайна, достаточное для повторения, чтобы облегчить понимание результатов.

**8.6.2 Точность**

Во вкладыше к качественному исследованию должно быть описано, какие источники вариабельности оценивают и как каждый из них используют в исследовании точности. Следует также провести различие между внутрилабораторными и мультилабораторными исследованиями прецизионности. Если возможно, следует указать значения *ICR* и, возможно, концентрации, установленные для значений в соответствующей шкале, при которых бинарное исследование признает образец положительным в 0 % случаев (C0), C5, C95 и C100, чтобы облегчить пользователям проверку заявлений производителя. Количество операторов, партий, реплик, запусков в день и общее количество дней должно быть представлено в соответствии с [17]. Если проводят *ANOVA*, для прецизионных результатов следует придерживаться форматов отчетов о результатах, предложенных в [17]. Для каждого исследования неточности должна быть представлена таблица бинарных отчетов, аналогичная таблице 4. Для исследований коэффициента попадания следует представить описание всех факторов, связанных с каждым исследованием (включая эталонный материал или эталонную процедуру измерения, использованную для образцов исследования), а также рассчитанный *LLoD*.

**8.6.3 Клинические показатели, связанные с классификацией**

Клиническая чувствительность и специфичность (или положительное и отрицательное согласие) вместе с их двусторонними 95 %-ными ДИ должны быть указаны во вкладыше к препарату. Кроме того, следует указать, использовалась ли в качестве компаратора наилучшая доступная оценка ЦС или другое качественное исследование. Должен быть назван компарато и должно быть представлено подробное описание, включающее соответствующие ссылки. При использовании наилучшей доступной оценки ЦС расчет *PPV* и *NPV* должен быть скорректирован, если распространенность в исследовании не соответствует распространенности в популяции предполагаемого использования. Следует указать распространенность ЦС, использованную для определения *PPV* и *NPV*, и способ определения этой распространенности (например, из исследуемой популяции или из справочных документов). Примеры таблиц и расчетных результатов для различных дизайнов исследований клинической эффективности приведены в приложении К.

**8.6.4 Другие требования к производительности**

Отчеты по другим показателям должны соответствовать форматам, предложенным в [2] для мешающих веществ и [3] для стабильности реагентов.

#### 9 Лабораторная проверка заявленных характеристик

9.1 Когда лаборатория решает предложить какое-либо *IVD*-исследование, заявления о точности и клинической эффективности или относительном согласии, содержащиеся в упаковке, должны быть проверены до получения результатов от пациентов. Проверка определяет, правильно ли выполняют исследование. Если имеется подходящая спецификация характеристик для использования по назначению и показаниям к применению, необходимо убедиться в лаборатории, что новое исследование соответствует этой спецификации. Лаборатория должна проанализировать практические рекомендации и выяснить, используется ли исследование в клинических целях для диагностики, прогнозирования или предсказания. Кроме того, при поступлении новых партий реагентов следует проверить их соответствие используемым в настоящее время партиям реагентов. Эти рекомендации предназначены для продуктов, уже одобренных регулирующими органами, но если в исследование вносятся какие-либо изменения, включая использование типов образцов, не перечисленных производителем во вкладыше к упаковке, к этому новому показанию к применению предъявляются дополнительные требования по валидации, как описано в разделе 8.

Процесс разработки, рассмотренный в разделе 7, и процесс аналитической валидации, описанный в разделе 8, могут быть использованы лабораторной организацией, разрабатывающей качественные исследования. Таким образом, процесс, включенный в настоящий раздел, может быть использован для определения того, проверена ли эффективность экспертизы при стандартном использовании, обычно в нескольких лабораториях в рамках данной организации.

Эти рекомендации должны рассматриваться как минимально необходимые для проверки и даны с целью обеспечения баланса между практичностью и статистической строгостью. Они предполагают, что производитель выполнил требования по валидации, необходимые для выпуска качественного продукта. Степень согласия между результатами лабораторного исследования *IVD* и заявлениями производителя на упаковке можно оценить более глубоко, используя методы, описанные в разделе 8. Примеры лабораторной проверки утверждений о продукте приведены в приложении Л.

Если качественное исследование имеет *ICR*, доступную для лаборатории конечного пользователя, то подход к проверке точности является более простым процессом, который может потребовать меньшего количества образцов и генерировать результаты, содержащие больше информации. Однако, в случае клинических показателей или согласия, разница в рекомендуемых процессах невелика, за исключением того, что исследование с *ICR* может помочь лаборантам понять любые различия в бинарных результатах.

Перед проведением качественного исследования необходимо изучить факторы, влияющие на этапы до, после и во время испытания, а также на процесс, для которого планируется проведение верификационного исследования. Необходимо изучить вкладыш в упаковку, чтобы определить характеристики, заявленные производителем, а также понять принцип проведения исследования, методологию, ограничения и рекомендации по контролю качества. Этот обзор необходим для определения характеристик эффективности исследования и того, при каких условиях они достигаются для планирования валидации верификационного исследования и установления критериев приемки. По завершении проверки допускается подготовить протокол проверки, рабочую документацию, средства контроля и образцы для проверки и приступить к проверке.

**9.2 Проверка точности**

9.2.1 Прецизионность проверяют в условиях, аналогичных тем, которые указаны во вкладыше к упаковке производителя. В некоторых случаях испытания по проверке точности, например, проверка точности нескольких приборов, когда в лаборатории имеется только один, могут оказаться нецелесообразными. И наоборот, хотя партии реагентов могут быть значительным источником вариаций, производитель мог не учесть этот фактор при составлении заявления во вкладыше к продукту. Поэтому лаборатория должна стремиться к тому, чтобы условия проверки соответствовали условиям, использованным производителем при утверждении результатов исследования. Верификационное исследование должно проводиться операторами, которые выполняют рутинное тестирование пациентов.

Например, в некоторых ситуациях производитель реагентов может не предоставить данные о прецизионности именно для концентраций измеряемого вещества C5 или C95, поэтому доступны только данные о прецизионности образцов донора или пациента с ЦС или без него. Медицинской лаборатории может потребоваться секвенирование или получение образцов из разведений. Пример того, как медицинская лаборатория может провести такую оценку, приведен в приложении Л.

Если поставщик предлагает провести или оказать помощь в проведении верификационного исследования, это может быть приемлемо только для полностью автоматизированных испытательных систем, в которых межоператорская вариативность не является проблемой, а верификационное исследование проводится на территории лаборатории в месте, где будет проводиться исследование. В конечном итоге, ответственность за выполнение нормативных требований несет испытательная лаборатория.

**9.2.2 Элементы точности**

9.2.2.1 Хотя те же факторы, которые влияют на количественные исследования, могут влиять на точность качественных бинарных исследований, результаты неточности при оценке качественных исследований выражают по-другому. Коэффициенты вариации заменяют пропорциями положительных или отрицательных результатов. Из-за недостаточной разрешающей способности, присущей бинарным результатам, отдельные и независимые оценки повторяемости и воспроизводимости недостижимы, если только объем выборки не очень велик.

Точность качественных исследований, не имеющих зарегистрированных *ICR*, лучше всего оценивать с использованием образцов с ответами, достаточно близкими к отсечению, чтобы поставить под сомнение способность исследования правильно признать образцы положительными или отрицательными (например, C5 и C95).

Лаборатория несет ответственность за проверку возможности многократного исследования одних и тех же образцов в один и тот же день и в разные дни и получения воспроизводимых результатов. Эти результаты должны быть получены независимо от того, кто из операторов проводит исследование. В идеале в этой оценке должны участвовать несколько лаборантов, чтобы определить вариабельность лаборатории. Для полностью автоматизированных тест-систем, которые не зависят от оператора, изменчивость оператора не должна влиять на точность исследования, и может не потребоваться участие нескольких операторов в проверочном исследовании.

При необходимости следует провести исследования точности для всего обследования. Если исследование будет проводиться в нескольких местах, то в исследование репликации следует включить межлабораторный компонент.

9.2.2.2 Повторяемость

Повторяемость − тесная связь между воспроизведенными результатами, полученными из одного и того же образца при одинаковых условиях работы (например, в рамках одного испытания). Однако для наилучшей оценки этого параметра в рамках внутрилабораторного исследования точности оценки повторяемости обычно проводятся в течение многих серий или дней. Такой подход необходим, поскольку однократные измерения повторяемости могут не отражать тонких, неконтролируемых, временных изменений в методе испытания и/или работе прибора. В рамках каждого прогона условия (например, оператор, партия реагентов, прибор) остаются постоянными, тестируется несколько реплик и оценивается воспроизводимость при различных концентрациях измеряемого вещества.

9.2.2.3 Воспроизводимость

Воспроизводимость − тесная связь между результатами измерений при изменении условий эксплуатации. Важной переменной при проверке утверждений о неточности является время, которое измеряется путем выполнения исследований в отдельные дни. Рекомендуется проводить исследования, охватывающие пять или более рабочих дней (последовательных или непоследовательных), чтобы учесть потенциально значительную повседневную погрешность и обеспечить адекватное покрытие этого компонента погрешности. Если единственными неповторяющимися факторами, учитываемыми при исследовании, являются день за днем или прогон за прогоном, то можно сказать, что исследование воспроизводимости измеряет внутрилабораторную погрешность. Другие типы исследований воспроизводимости могут измерять другие факторы, такие как оператор, прибор и/или место проведения исследования.

**9.2.3 Точная верификация без внутреннего непрерывного ответа**

9.2.3.1 Рекомендуемый протокол

Для проверки прецизионности без *ICR* следует подготовить три образца. Если производитель заявляет, что C5, C50 и C95 соответствуют определенным концентрациям аналита, можно рассчитать отношение C5 и C95 к C50 (например, C95 = 115 % от C50). Три образца должны быть подготовлены в соответствии с этими концентрациями. Если образец готовят на уровне С50 с использованием образцов, заявленных как положительные, более практичным способом установления разведения образца С50 для проверки может быть поэтапный подход с серийными разведениями, исследуемыми с меньшим количеством реплик на разведение. Эталонные образцы с указанными количествами аналита могут быть предпочтительнее, поскольку для установления C50 им требуется меньше первоначальных испытаний (при условии, что любые эффекты матрицы были оценены до использования).

Три образца должны быть протестированы не менее чем в 20 повторах. Если производитель использует другой набор образцов для подтверждения исследования, лаборатория может пожелать выбрать сопоставимые образцы не на уровне C5, C50 и C95. Протокол должен охватывать как можно больше известных источников вариабельности. Например, если оператор является важной переменной, то можно собрать не менее двух реплик на каждого из двух операторов в течение пяти дней (последовательных или непоследовательных). Если исследование полностью автоматизировано и используют несколько приборов, можно использовать два прибора вместо двух операторов. Ключевым моментом во всех случаях является обеспечение исследования со сбалансированной выборкой (т. е. экспериментальной схемой, в которой все комбинации лечения имеют одинаковое количество наблюдений) важных компонентов дисперсии. Информация о сбалансированной выборке представлена в [17].

Если производитель не предоставил абсолютные или относительные концентрации (процентное разбавление от C50) в C5 и C95, лаборатория должна определить процентное разбавление, которое, по ее мнению, должно привести к значительному изменению положительных или отрицательных пропорций. Например, лаборатория ожидает, что 20 %-ное разбавление C50 приведет к значительному снижению положительных результатов в пробе C50 (см. таблицу 11). Если исследуют 20 копий обоих образцов, а количество положительных результатов уменьшается с восьми положительных результатов в образце С50 до шести положительных результатов в 20 %-ном разведении, лаборатория может сделать вывод, что качественное исследование неприемлемо, поскольку оно не обладает необходимой точностью. Хотя разбавление ± 20 % может быть разумным разрешением, ожидаемым для некоторых исследований, которые основаны на реакциях с иммунологическими реагентами, качественные исследования, основанные на ПЦР в реальном времени, могут давать такие же ответы при разведениях, превышающих 100 %.

Таблица 11 − ДИ вокруг ожидаемого числа положительных результатов (см. [44])

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Истинная доля выборки | 95 % ДИ | *N* | | |
| 20 | 30 | 40 |
| C5 | Наименьшее количество положительных результатов | 0 | 0 | 0 |
| Наибольшее количество положительных результатов | 5 | 6 | 7 |
| C25 | Наименьшее количество положительных результатов | 2 | 3 | 5 |
| Наибольшее количество положительных результатов | 10 | 13 | 17 |
| C50 | Наименьшее количество положительных результатов | 5 | 9 | 14 |
| Наибольшее количество положительных результатов | 15 | 21 | 26 |
| C75 | Наименьшее количество положительных результатов | 10 | 17 | 23 |
| Наибольшее количество положительных результатов | 18 | 27 | 35 |
| C95 | Наименьшее количество положительных результатов | 15 | 24 | 33 |
| Наибольшее количество положительных результатов | 20 | 30 | 40 |
| Примечание – Обозначения: C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C25 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 25 % случаев; C50 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C75 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 50 % случаев; C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; ДИ − доверительный интервал; *N* − количество всех исследований. | | | | |

9.2.3.2 Анализ данных

Хотя исследование верификации по умолчанию должно проводиться как минимум в течение пяти дней, количество реплик будет недостаточным для количественной оценки компонента погрешности между днями в доле процента положительных или отрицательных результатов. Поэтому анализ заключается в сравнении общего количества положительных или отрицательных результатов с ожидаемыми значениями.

9.2.3.3 Критерии приемлемости

Проверка заявленной производителем точности должна основываться на том, соответствует ли число положительных результатов этому заявлению. В таблице 11 приведено ожидаемое количество положительных результатов при истинной базовой концентрации и количестве всех исследований (*N*). Нижнее и верхнее значения представляют собой 95 % доверительные границы, или пределы положительных результатов, которые пользователь должен ожидать в 95 % случаев при одинаковом количестве исследований, проведенных на выборке. Эти ДИ количества положительных результатов рассчитывают с использованием оценки Уилсона и округлением до ближайшего целого числа для нижнего предела и округлением до ближайшего целого числа для верхнего предела (см. приложение И для расчета ДИ оценки).

В примере качественного бинарного обследования без *ICR* наблюдатель предоставляет результаты исследования для получения бинарного результата. В приложении Ж представлено обсуждение прецизионных исследований, предназначенных для оценки различий в бинарных результатах между наблюдателями, с примером работы.

**9.2.4 Точная верификация при наличии внутреннего непрерывного ответа**

Аналогично процедурам количественных исследований, при качественном исследовании необходимо проверить неточность. Отсечка для качественного исследования − порог, установленный производителем для *ICR*, при превышении которого результат считается положительным, а при снижении − отрицательным или наоборот. Как уже было рассмотрено в 8.2, исследования прецизионности для качественных исследований должны давать оценку неточности исследования при концентрациях аналита вблизи отсечки. Учитывая, что необходимый образец (образцы) имеется в наличии и что *ICR* доступен для предоставления результатов исследования, пользователь может обратиться к [45] для получения информации о проверке прецизионности.

**9.3 Клинические показатели или согласие**

**9.3.1 Клинические показатели**

Для оценки клинических показателей лаборатории должны иметь практически равное количество образцов от субъектов, у которых, как известно, присутствуют ЦС, и субъектов, для которых известно, что ЦС отсутствуют. Такой сценарий является наиболее желательным типом сравнения. Подходящие показатели для описания согласованности включают оценки клинической чувствительности, клинической специфичности, *PPA* и *NPA*. Более подробная информация приведена в 8.3. Эти образцы могут быть получены из коммерческих источников или других медицинских лабораторий или лабораторий общественного здравоохранения, если они были охарактеризованы с помощью надежного метода для оценки ЦС.

Необходимо приложить усилия, чтобы выборка (выборки) была репрезентативной для предполагаемой популяции. Однако некоторые образцы должны быть несколько близки к отсечению (например, более C5 и менее C95), чтобы эффективность экспертизы была под вопросом. Этого легче добиться, если в результате исследования будет получен *ICR*. При условии отсутствия ограничений, указанных в инструкции по применению, набор образцов может также включать некоторые старые сохранившиеся образцы (например, пятна крови и замороженные образцы).

Хотя референсные панели могут быть выгодны для эффективного и рационального использования ресурсов и времени эксперта, такие панели могут быть недоступны. Кроме того, эталонные панели могут не представлять клиническую популяцию лаборатории или типичную распространенность и спектр интересующего заболевания. Однако референс- панели и образцы квалификации, используемые в сочетании с обычным исследованием клинических образцов, могут обеспечить дополнительную уверенность в результатах процесса оценки. В 8.3 рассматривается анализ данных для тех видов сравнений, которые лаборатория, скорее всего, будет проводить не только для проверки клинических показателей или согласия, но и для оценки величины очевидных различий в чувствительности и специфичности.

**9.3.2 Согласие**

В менее идеальном случае сравнение может характеризоваться только согласием кандидата с другим тщательно описанным методом. Например, в качестве компаратора может выступать существующее в лаборатории исследование, которое не считается наилучшей доступной оценкой ЦС. В этом случае термины «чувствительность» и «специфичность» больше не подходят для описания сравнительных результатов. Вместо этого актуальной целью для многих лабораторий может быть демонстрация степени согласия со сравнительным исследованием. Критерии согласия зависят от сравниваемых исследований. Например, хотя одно или меньшее количество дискордантных отрицательных или дискордантных положительных результатов из 20 пар результатов может свидетельствовать о достаточном согласии при сравнении исследований с аналогичными возможностями, для исследования, которое, как ожидается, будет более чувствительным, может потребоваться большее количество положительных результатов.

Расхождения между исследование-кандидатом и сравнительным исследованием могут возникнуть из-за ошибок в одном из них. Если сравнительное исследование не дает стопроцентного результата, образцы с расхождениями между результатами исследования-кандидата и сравнительного исследования могут быть протестированы с учетом клинического состояния пациента, если оно известно, для получения результатов, которые могут быть полезны для устранения расхождений. В случаях, когда оператор может повлиять на результаты, он должен продемонстрировать мастерство выполнения исследования до его проведения. Также можно изучить *ICR* результатов, чтобы определить, находятся ли расходящиеся результаты вблизи отсечки для обоих исследований, в этом случае расхождение можно объяснить незначительными различиями в C50 или неточностью в исследованиях. Следует проанализировать клинические диагнозы и другую клиническую информацию об исследуемой популяции, чтобы определить, есть ли среди образцов с расходящимися результатами преобладающее клиническое состояние, которое следует дополнительно исследовать.

**9.3.3 Количество образцов для проверки клинических показателей или согласия**

Количество образцов, необходимых для каждой части проверочного исследования, может быть разным. Необходимо протестировать не менее 20 образцов с целью примерно равного распределения людей с ЦС и без ЦС. Однако, если *PPA* или чувствительность считаются значительно более важными, чем *NPA* или специфичность, соотношение положительных и отрицательных образцов может быть разумно увеличено. Соответствующее количество образцов зависит от многих факторов, включая точность и сложность исследования, распространенность цели (целей) в указанной популяции, клинические показатели сравнительного метода, стоимость и целесообразность метода, используемого для анализа данных, а также уровень приемлемой статистической достоверности. При наличии *ICR* по результатам обследования компаратора количество образцов, необходимых для верификации, может быть меньше, чем при его отсутствии. Такая ситуация возникает потому, что результаты выборки могут быть подобраны таким образом, чтобы легче было проверить результаты исследования-кандидата вблизи отрезков.

Количество отобранных образцов и пропорции положительных и отрицательных результатов будут сильно влиять на ширину 95 %-ного ДИ оценок чувствительности и специфичности. Поэтому минимальный размер выборки в 20 образцов определяет только то, соответствуют ли результаты процедуры-кандидата заявлениям производителя. Даже при 100 %-ном согласии результатов исследования равного количества положительных и отрицательных образцов (по 10) 95 %-ный ДИ будет составлять от 72 % до 100 % как для чувствительности, так и для специфичности. Если относительные пропорции меняются или обнаруживается дискордантная проба (пробы), ДИ увеличивается еще больше. Несколько примеров приведены в таблице 12.

Таблица 12 − Примеры ДИ по шкале Уилсона

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Результаты | Всего образцов | 20 | 20 | 20 | 40 | 40 |
| Настоящие положительные | A | 10 | 9 | 15 | 20 | 18 |
| Ложные срабатывания | B | 0 | 1 | 0 | 0 | 2 |
| Ложные отрицательные | C | 0 | 1 | 0 | 0 | 210 |
| Настоящие отрицательные | D | 10 | 9 | 5 | 20 | 18 |
| Чувствительность, 95 % ДИ | Среднее | 1,00 | 0,90 | 1,00 | 1,00 | 0,90 |
| Низкий | 0,72 | 0,60 | 0,80 | 0,84 | 0,70 |
| Высокий | 1,00 | 0,98 | 1,00 | 1,00 | 0,97 |
| Специфичность, 95 % ДИ | Среднее | 1,00 | 0,90 | 1,00 | 1,00 | 0,90 |
| Низкий | 0,72 | 0,60 | 0,57 | 0,84 | 0,70 |
| Высокий | 1,00 | 0,98 | 1,00 | 1,00 | 0,97 |
| Примечание – Обозначение: ДИ − доверительный интервал. | | | | | | |

**9.3.4 Анализ данных**

Во многих случаях лаборатории оценивают исследование-кандидата только по сравнительному исследованию (см. таблицу 7); таким образом, в качестве выводов используются *PPA* и *NPA*, как показано в формулах (22) и (23). Хотя это в первую очередь обязанность производителя, лаборатория может оценить клиническую эффективность исследования, если для образцов имеется клинический диагноз, полученный в результате наилучшей доступной оценки ЦС (см. таблицу 5). В этих случаях клиническая чувствительность и специфичность могут быть оценены по формулам (7) и (8) (см. 8.3.5) следующим образом: определяют, соответствуют ли полученные результаты заявлениям производителя или клиническим требованиям. Необходимо заранее установить критерии, по которым эти параметры будут приниматься при проведении исследования-кандидата.

Если оцениваются *PPA* и *NPA* или чувствительность и специфичность, то разница в рекомендуемых этапах исследования данных незначительна. В любом случае, если при качественном бинарном исследовании вместе с бинарным результатом также предоставляется результат *ICR*, такая количественная оценка может помочь объяснить, как лежащий в основе *ICR* приводит к любым различиям между исследованиее-кандидатом и компаратором.

**9.4 Исследование стабильности реагентов**

Если лаборатории должны проверять стабильность реагентов, указанную во вкладыше производителя, следует обратиться к [3]. Если имеется *ICR*, его следует использовать в контексте протоколов исследования (см. [3]), разработанных для таких лабораторных проверок. Как минимум, должны использоваться образцы как присутствующих, так и отсутствующих ЦС, чтобы можно было выявить изменения в ту или иную сторону с течением времени. Как правило, для такого выявления необходимо отобрать образец с реакцией на C5 (или образец, не содержащий ЦС, если C5 не представляется возможным) и C95.

Если *ICR* недоступен, для отбора образцов для исследования стабильности реагента можно использовать соображения, изложенные в 9.2.3. В каждой выбранной временной точке можно использовать размер выборки и количество положительных результатов из таблицы 11, чтобы определить, является ли реагент более нестабильным (т. е. отличается от исходного уровня). Если образцы с известной концентрацией отобрать невозможно, то для проверки стабильности в каждой временной точке допускается использовать сравнение тест/контроль, описанное в 8.5 для проверки на интерференцию. В любом случае регрессия не используется в анализе стабильности.

**9.5 Смена партий реагентов**

Когда в лабораторию поступают новые реагенты, перед их использованием в отчетах о результатах пациентов следует проверить ожидаемые характеристики. Более подробная информация о методах, которым следует следовать при наличии *ICR*, приведена в [46]. Если *ICR* недоступна, следует провести 20 повторов образца с реакциями, близкими к отсечению (C50), используя новую и старую партии реагентов (общее количество тестов равно 40), и получить сопоставимую долю положительных результатов (см. таблицу 11 для получения дополнительной информации). Результаты можно оценить с помощью точного теста Фишера для двух пропорций. В противном случае, если разница между количеством положительных результатов из каждой партии реагентов превышает семь, эти две партии следует считать достоверно различными   
(α = 0,05). Партия с меньшим числом положительных результатов должна считаться менее чувствительной. В этом случае для получения истинного C50 и применения критерия отказа более 7 количество положительных результатов в образцах C50 из старшей партии реагентов должно составлять от восьми до 12.

### Приложение A

### (справочное)

### Примеры исследований

В таблице A.1 приведены три различных типа исследований с примерами. Первый тип − качественные бинарные, отчитывающиеся о результатах исследования (качественные, бинарные исследования), которые рассматриваются в настоящем стандарте. Второй тип − небинарные исследования с порядковыми и номинальными результатами с более чем двумя категоризациями. Третий тип − процедуры количественных измерений. Руководство по использованию справочных документов (документов CLSI) для оценки процедур количественных измерений приведены в соответствующих руководствах[[3]](#footnote-3)).

Таблица A.1 − Примеры качественных бинарных исследований, небинарных исследований категоризации и количественных процедур измерения

|  |  |  |
| --- | --- | --- |
| Качественные бинарные исследования | | |
| Отчетный результат (характеристика) | Основание результата  исследования | Пример технологии |
| Положительный или отрицательный (беременна или не беременна) | Концентрация общего βhCG в моче | Тест на беременность на основе агглютинации или преципитации |
| Положительный или отрицательный (беременна или не беременна) | Концентрация общего βhCG в моче | Картридж для теста на беременность в моче |
| Наличие или отсутствие белка | Полоса на агарозном геле, уникальная для конкретного белка | Белковая полоса определенной молекулярной массы в матрице, распознаваемая специфическим антителом |
| Положительный или отрицательный результат на антитела | Отношение сигнала к срезу  < или > 1,0 | Иммуноферментный анализ на ВИЧ, сравнивающий сигнал образца пациента с сигналом образца-отсекателя |
| Обнаружена или не обнаружена ДНК | Превышение или отсутствие порога копирования в течение заданного количества циклов ПЦР | Сверхспецифическое, автоматизированное ПЦР-обнаружение *MRSA* |
| Обнаружена или не обнаружена РНК | Превышение или отсутствие порога копирования в течение заданного количества циклов ПЦР | Сверхспецифическое, автоматизированное ПЦР-обнаружение вируса гепатита С |

*Продолжение таблицы А.1*

|  |  |  |
| --- | --- | --- |
| Качественные бинарные исследования | | |
| Отчетный результат (характеристика) | Основание результата  исследования | Пример технологии |
| Присутствует или отсутствует | Волчаночные антикоагулянты в плазме крови, направленные против фосфолипидно-белковых комплексов | Соотношение времени свертывания испытуемого образца и нормальной плазмы |
| Положительные или отрицательные | Пики ионов в масс-спектрометрии с характерным для данного аналита соотношением масс и зарядов | Тандемное масс- спектрометрическое измерение содержания указанного вещества в плазме крови |
| Положительный или отрицательный результат на мутацию или аллель | Разница в измерениях ПЦР между образцом и образцом эталонного гена | *NAAT* определение ∆C между тестовыми и контрольным образцом |
| Исследования по небинарной категоризации | | |
| Отчетный результат (характеристика) | Основание результата  исследования | Пример технологии |
| Отрицательный, +, ++, +++,++++ Полоса или пятно с различной интенсивностью (порядковый номер) | Концентрация целевого аналита | Анализ мочи на билирубин, глюкозу или кетоновые тела |
| Конечный титр при 1:40, 1:80, 1:160 и т.д. (порядковый) | Антитела, специфичные для конкретного анализатора | Непрямой иммунофлуоресцентный анализ *ANA* на клетках HEp-2 |
| O+, O-, A+, A-, AB+, AB- (номинальный) | Группа крови | Серологическое тестирование |
| Геномное типирование (номинальное) | Генотипирование | Полиморфизм длины рестрикционных фрагментов |
| Процедуры количественных измерений | | |
| Отчетный результат  (значение) | Основание результата  исследования | Пример технологии |
| Концентрация, пг/см3 | Концентрация в пг/см3, но с порогом клинического решения | Автоматизированный иммуноферментный анализ тропонина в сыворотке крови |
| Концентрация, мкг/см3 | Концентрация в мкг/см3, но с клиническим отсечением | Автоматизированное иммунотурбидиметрическое определение димера D в плазме крови |

*Окончание таблицы А.1*

|  |  |  |
| --- | --- | --- |
| Процедуры количественных измерений | | |
| Отчетный результат  (значение) | Основание результата  исследования | Пример технологии |
| Концентрация, нг/см3 | Концентрация в нг/см3, но порог обнаружения 150 нг/см3 | Иммуноферментный анализ для выявления наличия метаболитов кокаина в моче |
| Количество окрашенных клеток, на мм2 | Количество окрашенных клеток на мм2 участка опухоли | Автоматизированное сканирование изображений *IHC* |
| Процентное содержание общего белка в сыворотке крови, г/дл | Перевод интенсивности линий или пятен в проценты от общего количества, поддающегося количественному измерению | Представление фракций белков сыворотки крови (альбумин и α1-, α2-, β-, и γ-глобулин) |
| Примечание – Обозначения: βhCG − хорионический гонадотропин человека, β-субъединица; ∆C − изменение порога цикла; *ANA* − антинуклеарные антитела; ДНК − дезоксирибонуклеиновая кислота;  IHC − иммуногистохимический; MRSA − метициллин (оксациллин) − устойчивый золотистый стафилококк; NAAT − тест амплификации нуклеиновых кислот; ПЦР − полимеразная цепная реакция; РНК − рибонуклеиновая кислота. | | |

**Приложение Б**

### (справочное)

### Связь между распределением внутреннего непрерывного ответа и пробит- или логит-моделями

Сокращения, применяемые в настоящем приложении:

*CX* − внутреннее значение непрерывного ответа в соответствующей шкале, где при бинарном исследовании образец признают положительным в *X* % случаев;

*ICR* − внутренний непрерывный ответ;

*P*pos −вероятность быть признанным положительным;

*Pr* – вероятность.

Символы, применяемые в настоящем приложении:

Φ − кумулятивная нормальная функция распределения;

µ − среднее;

σ − стандартное отклонение;

*a* − перехват линейной модели для преобразованной вероятности положительного результата;

*b* − наклон линейной модели для преобразованной вероятности положительного результата;

*e* − ошибки;

*R* − значение по внутренней шкале непрерывных ответов;

*X* − независимый (по оси *x*) масштаб;

*Y* − бинарный результат (1 = положительный, 0 = отрицательный).

Значение внутреннего непрерывного ответа (*ICR*) в соответствующей шкале, где бинарное исследование признает образец положительным в 5 % случаев (C5), и значение *ICR* в соответствующей шкале, где бинарное исследование признает образец положительным в   
95 % случаев (C95), можно оценить с помощью измерений непрерывного ответа по слепку или путем моделирования (подгонки кривой) доли положительных результатов исследования (процент попаданий) по сравнению с истинным значением ответа. Модели пробит и логит обычно используют для характеристики этой зависимости с использованием среднего значения нескольких определений в оговоренных условиях как неизвестное истинное значение ответа. На рисунке Б.1 приведен гипотетический пример зависимости между долей положительных результатов исследования (ось *y*) и средним значением множественных определений для пяти различных образцов с уровнями около отсечки. Доля положительных результатов (результаты выше отсечения на уровне 1,5) определяют по площади заштрихованной области как долю всей площади под кривой. Среднее значение нескольких гипотетических определений является центром каждой кривой. В данном примере повторные результаты для образца распределены нормально, а вариабельность постоянна для разных образцов. Кривая на нижней панели рисунка Б.1 соответствует пробит-модели.

**Вероятность положительного результата**

**Непрерывный ответ**

2.4

2.2

2

1.8

1.6

1.4

1.2

1

0.8

0.6

1.0

0.8

0.6

0.4

0.2

*P*= 0

*P*= 0.04

*P*= 0.11

*P*= 0.82

*P*= 1

Рисунок Б.1 − Взаимосвязь между долей положительных результатов исследования и средним значением множественных гипотетических определений

Пробит-модель для одного количественного предиктора описывает формула

(Б.1)

где *Ppos*(*X*) − вероятность положительного результата обследования на уровне *X*;

Φ − кумулятивная функция нормального распределения;

*a* и *b* − неизвестные параметры, подлежащие оценке[[4]](#footnote-4)).

Преобразованная вероятность положительного результата линейна согласно формуле

(Б.2)

где Φ−1 − обратное кумулятивное нормальное распределение.

Пробит-модель может быть мотивирована нормальной моделью латентных переменных или пороговой моделью[[5]](#footnote-5)). Нормальная модель латентных переменных предполагает наличие истинного ненаблюдаемого непрерывного отклика *R*, как показано в выражениях

*Y* = 0 (отрицательный), если *R* ≤ Отсечка, (Б.3)

*Y* = 1 (положительный), если *R* ≥ Отсечка. (Б.4)

Если *R* = *a* + *b*(*X* + *e*), а ошибки *e* независимы и нормально распределены со средним µ и равной дисперсией σ2), то вероятность того, что *Y* = 1, учитывая *X*, равна   
*Pr*(*Y* = 1|X) = *Ppos*(*X*) = Φ(*a* + *bX*), как показано в формуле (Б.1), при этом µ = −*a*/*b* = C50 и   
σ = 1/|b|. Пробит-модель − обобщенная линейная модель, которая имеет вид   
Ф−1[*Pr*(*Y* = 1|*X*)] = *a* + *bX* с обратной нормальной функцией связи Ф−1. Если вместо нормального распределения ошибок используют стандартное логистическое, а не нормальное распределение, то вместо пробит-модели используют логистическая модель1).

C5, C50 и C95 можно оценить напрямую с помощью формул (1) − (4) и косвенно на основе оценок *a* и *b* в пробит-регрессии с помощью формул (Б.5) − (Б.7) (см. рисунок Б.2)

(Б.5)

(Б.6)

(Б.7)

0.8

**Непрерывный ответ**

0.6

50%

0.4

0.2

5%

0.0

0.8

1.0

1.2

C5

1.4

Отсечка

1.6 C95

1.8

2.0

1.0

95%

2.0

1.8

1.6

1.4

1.2

1.0

0.8

1.8 2.0

Процент положительных = 95 %

1.6

1.4

1.2

1.0

0.8

Процент

положительных= 5 %

**Вероятность положительного результата**

C5 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; *ICR* − внутренний непрерывный ответ

Рисунок Б.2 − Взаимосвязь между распределениями *ICR* и пробит-моделью для получения C5 и C95

Если ошибки распределены нормально, но дисперсии неравны по уровням, стандартная пробит-регрессионная модель не может правильно отразить эту непостоянную изменчивость, и может потребоваться подход обобщенного линейного моделирования с другой структурой дисперсии. В качестве альтернативы можно использовать преобразования для стабилизации дисперсии.

Если базовый ответ доступен (не латентен), допускается его исследовать и использовать. Если ошибки следуют стандартному логистическому, а не нормальному распределению, можно использовать логистическую модель. На практике может быть трудно отличить нормальное распределение от стандартного логистического распределения. Если базовый ответ доступен, его следует использовать для определения наиболее подходящей модели для данных и для оценки C5 и C95, как описано в 7.2.4.

**Приложение В**

### (справочное)

### Качественное бинарное исследование с визуальным чтением

В.1 Сокращения, применяемые в настоящем приложении:

*CX* − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в *X* % случаев;

*ICR* − внутренний непрерывный ответ.

В.2 Качественное исследование с визуальным чтением − исследование, при котором категоризация производится на основе наблюдения пользователя, а не с помощью приборов. Этот тип бинарного исследования встречается довольно часто и не очень легко вписывается в схему категоризации так же, как другие качественные бинарные исследования, описанные в настоящем стандарте. Такие исследования не имеют внутреннего непрерывного ответа (*ICR*), поскольку их бинарные результаты не зависят от инструментария. Однако ответ «да» или «нет» может имитировать отсечение количественного показателя. В связи с этими различиями в данном приложении рассматриваются уникальные особенности визуальных качественных исследований, которые могут быть использованы для оценки их клинической эффективности в соответствии с 8.3 и 9.2.

Примерами исследований с визуальным чтением являются картриджи для теста на беременность, чашки для скрининга мочи на наркотики и линейные зондовые анализы для обнаружения антител к гепатиту С, в которых положительная реакция проявляется в виде линии на сплошном фоне. Если исследование выполняется и результаты определяются только наблюдением пользователя, без *ICR*, обеспечиваемого приборами, вариабельность, вызванная интерпретацией пользователя, может быть большой по сравнению со считыванием на основе прибора. В этих случаях вариабельность в значительной степени зависит от способности пользователя определить наличие осадка или увидеть цветную линию, на что влияют такие параметры, как условия освещенности, цвет фона (для контраста), а также индивидуальное суждение или острота зрения.

**В.3 Разработка качественных исследований с визуальным чтением**

Целью разработчиков пользовательско-зависимых тестов обычно является создание простого устройства, которое различает положительные и отрицательные результаты при четко определенной пороговой концентрации. Эффективность должна быть определена с использованием образцов с известными значениями измеряемых веществ. В качестве примера можно привести тест-полоску для определения наркотиков в моче, на которой появляется линия при наличии определенного наркотика в заданной концентрации.

Если производительность встроена в физическую функцию тест-полоски, поведение исследования можно охарактеризовать, построив график вероятности признания образца положительным в зависимости от концентрации измеряемого вещества в образце, определенной с помощью отдельной процедуры, что показано на рисунке В.1. В данном примере подготовлено семь образцов, и для каждого образца проведено 20 исследований. Частота попаданий, или доля случаев, когда результат пробы признается положительным (ось *y*), должна быть построена в сравнении со значением измеряемого вещества для каждой пробы (ось *x*), и кривая должна быть добавлена, если она проходит через точки.

Необходимо определить точку пересечения результирующей кривой и, по крайней мере, линий 5 %, 50 % и 95 % измеряемого значения. Следует позаботиться о том, чтобы значение измеряемой величины для каждого образца было известно с минимальной погрешностью, что может быть достигнуто путем:

- выполнения измерений с помощью процедуры количественных эталонных измерений;

- усреднения нескольких определений, которые охватывают соответствующие компоненты дисперсии из выбранной процедуры количественного измерения;

- создания серии образцов с шипами с использованием стандартного материала[[6]](#footnote-6)) (если есть возможность, следует использовать сертифицированный стандартный материал);

- внесения шипов в один высокий образец и создание независимых разведений до известных количеств для создания других суррогатных образцов.

На рисунке В.1 показан пример, в котором производитель разрабатывает полоски для анализа мочи на наркотики с вероятностью 50 % при значении измеряемого вещества 20 единиц, однако значение измеряемого вещества никогда не показывают пользователю. Пользователь может определить только наличие или отсутствие цветной полоски на тест-полоске. Всегда существует некоторая неопределенность в результатах из-за различных источников вариабельности, например, вариабельности производства и различной способности пользователей видеть цветную полоску.

Эти результаты, показанные на рисунке В.1, получены путем подгонки пробит-кривой к полученным данным. Также определяют линии 5 %, 50 % и 95 % измеряемых значений. Преобразования данных о концентрации (например, арифметическое и логарифмическое) допускается выполнять по всем точкам данных (низкий остаток) для достижения наиболее близкого соответствия.

Рисунок В.1 − Пример качественного исследования с визуальным чтением

**Вероятность признания положительным**

0.1

0.0

0

5

10

15

20

25

30

35

40

**Концентрация препарата,**

**единицы**

50% значение

5% значение

0.2

95% значение

0.3

положительная 5 %

Точки данных

0.4

0.5

положительная 50 %

0.6

Подгонка по пробиоту

положительная 95 %

1.0

0.9

0.8

0.7

В некоторых случаях цели эффективности устанавливаются регулирующими организациями (например, цели обнаружения наркотиков). Производителю может потребоваться разработать тест-полоску таким образом, чтобы все образцы (или более или равно 95 % образцов) с определенным значением измеряемого количества наркотика были признаны положительными. В других случаях производитель может разработать тест-полоску таким образом, чтобы ни один образец (или менее или равно 5 % образцов) не был признан положительным, если его содержание ниже определенного значения измеряемой концентрации. В большинстве случаев предельная концентрация является критическим фактором. Поэтому значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 50 % случаев (C50), должно быть оценено и сравнено с заранее установленным отсечением, чтобы определить, насколько успешно система принимает бинарное решение при данной конкретной концентрации.

Как отмечалось выше, при визуальных качественных исследованиях допускается использовать существующую процедуру количественных измерений для присвоения значений тестовым образцам или создавать тестовые образцы с известными значениями, используя эталонные материалы. В процессе разработки такие образцы могут использоваться для определения характеристик тестовой полоски. После определения свойств экспертизы можно провести исследование с использованием аналогичных образцов, чтобы определить, была ли достигнута цель установления желаемых характеристик.

Следует использовать как минимум пять уровней, три из которых должны иметь частоту попаданий от 5 % до 95 %. Один образец ниже 5 % (обычно на значении, при котором бинарное исследование с большой вероятностью признает образец положительным в 0 % случаев [C0]) и один образец выше 95 % (обычно на значении в соответствующей шкале, при котором бинарное исследование с большой вероятностью признает образец положительным в 100 % случаев) обеспечивают необходимые пять образцов, но большее количество образцов на этих крайних уровнях не дает лучшего общего соответствия. Если необходимо больше уровней, их следует добавить между линиями вероятности 5 % и 95 %.

Источники вариабельности следует учитывать, когда лаборатория определяет, как следует собирать наблюдения на каждом уровне. Как правило, вариабельность включает в себя операторов и партии приборов для исследования (например, тест-полосок). Метод измерения вариабельности между наблюдателями приведен в приложении Ж. Несколько партий прибора должны использоваться несколькими разными лаборантами, которые представляют «типичных пользователей» (или «читателей»). Количество необходимых лаборантов и партий должно основываться на любых ранних оценках того, насколько вариабельными они будут в процессе исследования.

**В.4 Иные положения**

Одно из основных отличий бинарных исследований с визуальным чтением − отсутствие *ICR*. Эта особенность влияет на некоторые оценки эффективности. Исследования точности для таких исследований рассматривают в 8.2.5. На оценку клинических характеристик, описанную в 8.3, отсутствие *ICR* не влияет. Измерение стабильности реагентов для исследований с визуальным считыванием описано в 7.5, который ссылается на 9.4. Наконец, измерение влияния мешающих веществ или перекрестных реактивов, а также определение влияния других влияющих факторов для таких исследований рассматривают в 8.5. Определение влияния влияющих факторов особенно полезно в тех случаях, когда создание образцов, соответствующих таким уровням, как значение, при котором бинарное исследование с большой вероятностью признает образец положительным в 5 % или 95 % случаев (C5 или C95), не представляется возможным.

Визуальное исследование бинарного исследования может быть таким же простым, как определение появления цветной полоски на щупе или определение преципитата. Также может потребоваться обнаружение индикатора достоверности исследования с помощью контроля процесса. Пример приведен в 7.4.4. Некоторые визуальные исследования, разработанные для лабораторий, стали инструментальными исследованиями с использованием оптических считывающих устройств, которые количественно определяют интенсивность цвета щупа (например, щуп для анализа мочи на наркотики). В этих случаях исследование больше не зависит от пользователя, поскольку теперь имеется *ICR*. Разработка данного типа инструментальных исследований рассматривают в 7.2.

#### Приложение Г

#### (справочное)

#### Оценка точности секвенирования следующего поколения

Г.1 Сокращения, применяемые в настоящем приложении:

ДИ − доверительный интервал;

*CX* − внутрилабораторное значение в соответствующей шкале, когда при бинарном исследовании образец признают положительным в *X* % случаев;

*QC* − контроль качества;

*SD* − стандартное отклонение;

*VAF* − частота аллелей вариантов.

**Г.2 Пример исследования**

В данном примере моделируемого исследования анализ предназначен для поиска мутации в экзоне 2 гена *KRAS*, расположение c.35G.T (эталонный нуклеотид в этом месте − G, альтернативный или мутационный − T в этом месте). Исследование вычисляет процентную частоту аллелей мутационных вариантов (*VAF*) и сравнивает результат с отсечкой 15 % *VAF*. Если рассчитанная *VAF* более или равна 15, система сообщает конечному пользователю «мутация обнаружена». Если значение менее 15, система сообщит конечному пользователю «мутация не обнаружена». Значения *VAF* не сообщают конечному пользователю, но они доступны производителю.

Цель данного исследования − определить общую вариабельность системы для оценки внутрилабораторного значения в соответствующей шкале, где бинарное исследование признает образец положительным в 5 % случаев (C5), и внутрилабораторного значения в соответствующей шкале, где бинарное исследование признает образец положительным в 95 % случаев (C95) для конечного пользователя.

В исследование включены пять образцов с различными уровнями *VAF* для мутационного варианта *KRAS*. Эти уровни включают 2 %, 10 %, 14 %, 18 % и 25 % *VAF*. Каждый образец выполняется в трех экземплярах двумя операторами с четырьмя приборами. Каждая комбинация оператор-инструмент выполняется в три разных дня начала секвенирования (далее – «дни»).

Одна из возможных схем исследования представлена в таблице Г.1. Этот план исследования охватывает 24 стартовых дня и считается перекрестной и вложенной моделью, поскольку инструменты и операторы перекрещиваются, а день полностью вложен в каждую комбинацию инструментов и операторов.

Таблица Г.1 − Дизайн исследования: перекрестная и вложенная модель с днями, полностью вложенными в прибор и оператора

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Прибор 1 | | | | | | Прибор 2 | | | | | | Прибор 3 | | | | | | Прибор 4 | | | | | |
| Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | |
| День | | | День | | | День | | | День | | | День | | | День | | | День | | | День | | |
| 1 | 9 | 13 | 2 | 5 | 17 | 6 | 10 | 21 | 14 | 18 | 22 | 7 | 19 | 23 | 11 | 15 | 24 | 3 | 16 | 20 | 4 | 8 | 12 |
| Повторы | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

На практике может оказаться более эффективным провести такое исследование за меньшее количество дней. Такого темпа можно достичь, если проводить некоторые комбинации оператор-прибор в одни и те же дни. Пример такой схемы приведен в таблице Г.2. В этом примере операторы могут запускать только два прибора в день. Поскольку системе требуется несколько дней для получения результата, на каждый прибор приходится по одному запуску в день.

Таблица Г.2 − План проведения исследования за шесть дней вместо 24 дней

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Прибор 1 | | | | | | Прибор 2 | | | | | | Прибор 3 | | | | | | Прибор 4 | | | | | |
| Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | |
| День | | | День | | | День | | | День | | | День | | | День | | | День | | | День | | |
| 1 | 3 | 4 | 2 | 5 | 6 | 1 | 2 | 5 | 3 | 4 | 6 | 2 | 3 | 6 | 1 | 4 | 5 | 4 | 5 | 6 | 1 | 2 | 3 |
| Повторы | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

Использование различных партий реагентов в данном примере исследования невозможно. Поскольку определение общей внутрилабораторной вариабельности системы должно включать данные о вариабельности между изготовленными партиями реагентов, вариабельность от партии к партии получена отдельно в ходе внутреннего исследования.

Исходя из вышеуказанных ограничений, логистика сбора необходимых данных требует, чтобы каждый оператор запускал два прибора в день с тремя повторами каждого образца на каждый прибор. Для этого плана необходимо 72 аликвоты каждого уровня образца в исследовании. Таблица Г.3 описывает логистику плана прогона на основе таблицы Г.2.

Таблица Г.3 − План выполнения исследования с шестью не следующими друг за другом стартовыми днями, отсортированный по стартовым дням

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | День начала | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | 2 | | | | 3 | | | | 4 | | | | 5 | | | | 6 | | | |
| Неделя 1 – понедель-  ник | | | | Неделя 1 −  среда | | | | Неделя 1 – пятница | | | | Неделя 2 – понедель-  ник | | | | Неделя 2 −  среда | | | | Неделя 2 − пятница | | | |
| Оператор | | | | Оператор | | | | Оператор | | | | Оператор | | | | Оператор | | | | Оператор | | | |
| 1 | | 2 | | 1 | | 2 | | 1 | | 2 | | 1 | | 2 | | 1 | | 2 | | 1 | | 2 | |
| Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | | Прибор | |
| 1 | 2 | 3 | 4 | 2 | 3 | 1 | 4 | 1 | 3 | 2 | 4 | 1 | 4 | 2 | 3 | 2 | 4 | 1 | 3 | 3 | 4 | 1 | 2 |
| Повторы | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

В таблицах Г.1 – Г.3 представлены различные стратегии для одного и того же дизайна исследования, и для каждой стратегии используется одна и та же модель анализа. Для анализа данных приборы и операторы считаются пересекающимися, а день полностью вложен в прибор и оператора. При таком дизайне для каждой выборки используется модель со случайными эффектами согласно формуле

*yijk* = μ + α*i* + β*j* + γ*ij* + δ*k(ij)* + ε*l(ijk)*, (Г.1)

где µ − сумма истинной оценки;

α*i* – ошибки, связанные с прибором;

β*j* – ошибки, связанные с оператором;

γ*ij* – ошибки, связанные с взаимодействием между оператором и прибором;

δ*k(ij)* – ошибки, связанные с днем, включая оператора и прибора;

ε*l(ijk)* – ошибки, связанные с любой остаточной вариацией, включая повторы каждая из которых нормально распределена со средним нулем и дисперсией σ2, описанной ниже.

,

,

,

,

.

Схема исследования имеет 48 степеней свободы (*DF*) для оценки вариабельности повторов, которая считается достаточной, как приведено в таблице Г.4.

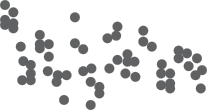
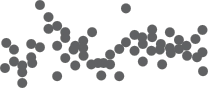
Таблица Г.4 − Факторы исследования и *DF*

|  |  |  |
| --- | --- | --- |
| Фактор | Расчет *DF* | Уравнение *DF* |
| Прибор | Количество приборов − 1 | 4 – 1 = 3 |
| Оператор | Количество операторов − 1 | 2 – 1 = 1 |
| Прибор · оператор | (Количество приборов − 1) · (Количество операторов − 1) | 3 · 1 = 3 |
| День (прибор · оператор) | Количество приборов · количество операторов · (количество дней, включая приборы и оператора − 1) | 4 · 2(3 − 1) = 16 |
| Повторы | Количество приборов · количество операторов · количество дней, включая приборы и оператора · (количество повторов − 1) | 4 · 2 · 3(3 − 1) = 48 |
| Всего | Количество приборов · количество операторов · количество дней, включая приборы и оператора · количество повторов − 1 | 4 · 2 · 3 · 3 – 1 = 71 |

**Г.3 Анализ данных исследования и оценка C5 и C95 с использованием данных непрерывного ответа**

Анализ компонентов дисперсии *VAF* (проценты от нуля до 100) проводят для оценки компонентов прецизионности (*SD*) для каждого источника вариации (прибор, оператор, прибор-оператор, день [прибор, оператор], репликация) и в целом. Оценки получены с помощью метода ограниченного максимального правдоподобия. В аналитическом наборе 72 результата *VAF*. Графики наблюдаемых *VAF* для каждого уникального образца представлены на рисунке Г.1. Разброс (изменчивость) между наблюдениями меньше для образцов с более низким уровнем *VAF* и увеличивается для образцов с более высоким уровнем *VAF[[7]](#footnote-7)).*

*VAF*, %



1

2

Образец

3

4

5

30

25

20

15

10

5

0

1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

День/Оператор/Прибор

*VAF* − частота аллелей вариантов

Рисунок Г.1 − Точечные графики наблюдаемых *VAF* для каждого образца

В таблице Г.5 представлены результаты дисперсионного компонентного анализа результатов *VAF* для каждого источника вариации и общей вариации для каждого образца. В большинстве случаев изменчивость повторов вносит наибольший вклад в общую изменчивость, за исключением образца *S*5 с высоким уровнем и образца *S*3, находящегося вблизи отсечки. В *S*5 наибольший вклад в общую вариабельность вносит вариабельность оператора, а в *S*3 вариабельность оператора аналогична вариабельности повторов.

Таблица Г.5 − Оценки компонентов вариации для отдельных выборок

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Иденти  фика тор обра  зца | *N* | Сред-нее | Прибор | | Оператор | | Прибор − оператор | | День, прибор, оператор | | Повтор | | Итогоа | |
| *SD* | *CV*,% | *SD* | *CV*,% | *SD* | *CV*,% | *SD* | *CV*,% | *SD* | *CV*,% | *SD* | *CV*,% |
| *S*1 | 72 | 2,43 | 0,03 | 1,0 | 0,00 | 0,0 | 0,03 | 1,1 | 0,05 | 2,1 | 0,05 | 2,0 | 0,08 | 3,3 |
| *S*2 | 72 | 9,93 | 0,00 | 0,0 | 0,18 | 1,8 | 0,13 | 1,3 | 0,14 | 1,5 | 0,24 | 2,4 | 0,35 | 3,6 |
| *S*3 | 72 | 14,12 | 0,00 | 0,0 | 0,45 | 3,2 | 0,00 | 0,0 | 0,15 | 1,1 | 0,40 | 2,8 | 0,62 | 4,4 |
| *S*4 | 72 | 17,38 | 0,00 | 0,0 | 0,04 | 0,2 | 0,00 | 0,0 | 0,30 | 1,7 | 0,53 | 3,0 | 0,61 | 3,5 |
| *S*5 | 72 | 25,66 | 0,00 | 0,0 | 0,87 | 3,4 | 0,25 | 1,0 | 0,29 | 1,1 | 0,41 | 1,6 | 1,04 | 4,0 |
| aВключает прибор, оператора, прибор − оператор, день (прибор, оператор) и повтор.  Примечание – Обозначения: *CV* − коэффициент вариации, %; *N* − количество образцов; *SD* − стандартное отклонение. | | | | | | | | | | | | | | |

Внутрилабораторные значения C5 и C95 оценивают по формулам

С5 = Отсечка исследования – 1,645*SD*C5, (Г.2)

С95 = Отсечка исследования + 1,645*SD*C95, (Г.3)

где ±1,645 − 5-й и 95-й процентили нормального распределения, соответственно.

*SD*C5 и *SD*C95 − внутрилабораторные *SD* на уровнях *VAF* C5 и C95, соответственно. Поскольку определение внутрилабораторной вариабельности системы должно включать данные о вариабельности между партиями реагентов, вариабельность между партиями получена отдельно в результате внутреннего исследования и объединена с общей *SD*, как показано в таблице Г.5. *SD* отдельной партии составляет 0,005 (постоянный для всех уровней *VAF*), а внутрилабораторную изменчивость рассчитывают как квадратный корень из (0,0052 + общий *SD*2) для каждого образца; результаты приведены в таблице Г.6. В данном примере, поскольку вариабельность от партии к партии настолько мала, нет никаких видимых изменений в оценках *SD*, представленных в таблице Г.6, поскольку любая разница между общим и внутрилабораторным *SD* видна только в четвертом десятичном знаке.

Таблица Г.6 − Внутрилабораторный *SD*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Идентификатор образца | *N* | Среднее | Итогоа | | Внутрилабораторныеb | |
| *SD* | *CV*,% | *SD* | *CV*, % |
| *S*1 | 72 | 2,43 | 0,08 | 3,3 | 0,08 | 3,3 |
| *S*2 | 72 | 9,93 | 0,35 | 3,6 | 0,35 | 3,6 |
| *S*3 | 72 | 14,12 | 0,62 | 4,4 | 0,62 | 4,4 |
| *S*4 | 72 | 17,38 | 0,61 | 3,5 | 0,61 | 3,5 |

*Окончание таблицы Г.6*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Идентификатор образца | *N* | Среднее | Итогоа | | Внутрилабораторныеb | |
| *SD* | *CV*,% | *SD* | *CV*, % |
| *S*5 | 72 | 25,66 | 1,04 | 4,0 | 1,04 | 4,0 |
| aВключает прибор, оператора, прибор − оператор, день (прибор, оператор) и повтор.  bВключает прибор, оператора, прибор − оператор, день (прибор, оператор), партию реагентов и повтор.  Примечание – Обозначения: *CV* − коэффициент вариации, %; *N* − количество образцов; *SD* − стандартное отклонение. | | | | | | |

Взаимосвязь между внутрилабораторным *SD* и средним *VAF* изучают путем построения графика зависимости соответствующего внутрилабораторного *SD* от среднего *VAF* для каждого образца. Внутрилабораторный *SD* (и *SD* повторяемости) не является постоянным для всех оцениваемых уровней, а линейно увеличивается по мере увеличения уровня образца (процент мутационных *VAF*). Оценки профиля точности представлены в таблице Г.7, а график − на рисунке Г.2.

Таблица Г.7 − Оценки профиля точности

|  |  |  |
| --- | --- | --- |
| Срок | Смета | Стандартная ошибка |
| Перехват | − 0,02633 | 0,057 |
| Наклон | 0,04066 | 0,004 |

Внутрилабораторный *SD*

**Средний *VAF***

30

25

20

Отсечка

10

5

0

0.0

0.2

0.4

0.6

0.8

1.0

*Y* = − 0,02633 + 0,04066 − среднее

1.2

*SD* − стандартное отклонение; *VAF* − частота аллелей вариантов

Рисунок Г.2 − Профиль точности внутрилабораторного SD (оператор, прибор, день начала секвенирования, партия и повтор) в сравнении со средним процентом *VAF* с наложением линейной модели

Оценку *SD* на C5 и C95 получают из профиля точности, где профиль точности оценивается по линейной регрессии внутрилабораторного *SD* на средний *VAF* по образцам, как показано в формуле

*SD*средний *VAF* = Средний *VAF* · Наклон + Перехват. (Г.4)

Поскольку профиль точности является линейным, C95 можно оценить непосредственно с помощью формулы (Г.5), которую получают путем подстановки формулы (Г.4) со средним *VAF* на C95 в формулу (Г.3) и последующего решения для C95.

. (Г.5)

Аналогичным образом C5 рассчитывают по формуле (Г.6), которую получают путем подстановки формулы (Г.4) со средним значением *VAF* в точке C5 в формулу (Г.2) и последующего решения для C5.

. (Г.6)

При наклоне 0,04066 и перехвате − 0,02633 значения для C5 и C95 рассчитывают как

Для этого исследования отсечение составляет 15 % *VAF*, поэтому интервал от C5 до C95 составляет 14,1 % − 16,0 %. Расчетный *SD* при C5 равен 0,547 *VAF*, а *SD* при C95 равен 0,635 *VAF*, что следует из формулы (Г.4). Эти данные представлены на рисунке Г.3. Для каждого протестированного уровня *SD* данных на этом уровне был построен график в сравнении со средним значением уровня. Были определены отсечка исследования, допустимые значения C5 и C95, а также интервал от C5 до C95. На рисунке Г.3 показано увеличение вариабельности (*SD*) по мере увеличения *VAF*, а также модель, подогнанная к данным.

Внутрилабораторный *SD*

Средний *VAF*

30

25

20

C5 ОтсечкаC95

10

5

0

0.0

0.2

0.4

0.6

0.8

1.0

*Y* = −0.02633 + 0,04066 − среднее

1.2

C5 − внутрилабораторное значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5% случаев; C95 - внутрилабораторное значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; *SD* − стандартное отклонение; *VAF* − частота аллелей вариантов.

Рисунок Г.3 − Анализ данных исследования и оценка C5 и C95 с использованием данных   
бинарного или качественного ответа

Те же данные, которые использовались выше, оценивают на основе качественного «звонка», сообщаемого исследованием на основе мутационного статуса образца. Результат считают положительным (или мутация обнаружена), если наблюдаемый *VAF* ≥ 15 % (отсечение). Для образцов со средним *VAF* значительно ниже 15 результаты ожидают отрицательными; для образцов со средним *VAF* значительно выше 15 результаты ожидают положительными; а для образцов со средним *VAF* около 15 ожидают некоторые зарегистрированные результаты для пересечения границы, при этом наблюдают как положительные, так и отрицательные результаты. В таблице Г.8 приведены результаты с соответствующими 95 %-ными доверительными интервалами по всем тестам.

Таблица Г.8 − Процент положительных и процент отрицательных результатов по выборке

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Идентификатор образца | Средний  *VAF* | Ожидаемые результаты | Количест-во попыток | *QC*  неуда-чи | Количество  положительных результатов из достоверных пробa | Процент положитель-ных результатов (95 % ДИb) | Процент отрицатель-ных результатов (95 % ДИb) |
| *S*1 | 2,43 | Отрицательные | 72 | 0/72 | 0/72 | 0  (0, 5,1) | 100  (94,9, 100) |
| *S*2 | 9,93 | Отрицательные | 72 | 0/72 | 0/72 | 0  (0, 5,1) | 100  (94,9, 100) |
| *S*3 | 14,12 | Большинство отрицатель-ных; некоторые положительные | 72 | 0/72 | 4/72 | 5,6  (2,2, 13,4) | 94,4  (86,6, 97,8) |
| *S*4 | 17,38 | Положительные | 72 | 0/72 | 72/72 | 100  (94,9, 100) | 0  (0, 5,1) |
| *S*5 | 25,66 | Положительные | 72 | 0/72 | 72/72 | 100  (94,9, 100) | 0  (0, 5,1) |
| aВ исследовании используют только образцы, прошедшие контроль качества анализа («валидные»).  bДвусторонние 95 %-ные ДИ рассчитаны по методу оценки Уилсона.  Примечание – Обозначения: ДИ − доверительный интервал; *QC* − контроль качества; *VAF* − частота аллелей вариантов. | | | | | | | |

Для каждого образца источники вариабельности, такие как приборы, операторы и дни, и результаты исследования могут быть обобщены с помощью свернутой таблицы, как показано в таблице Г.9. Обычно таблица начинается с наименьшей комбинации источников (один прибор на оператора в день) и затем «поднимается» вверх. В приведенном примере нет проблем в те дни, когда проводилось исследование, поэтому эти дни были свернуты. Если есть вопросы, требующие дополнительного обсуждения, в таблице можно разбить дни на части, как это было сделано для других оцениваемых факторов.

Таблица Г.9 − Свернутая таблица для образца 3 (*S*3) в таблице Г.8

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Прибор | Опера-тор | День | Количест-во попыток | Количество отказов *QC* | Количество положительных результатов из достоверных пробa | Количество отрицательных результатов из достоверных пробa | Процент положи-  тельных результатов | Процент отрица-тельных результатов |
| 1 | 1 | Все | 9 | 0 | 0/9 | 9/9 | 0 | 100 |
| 1 | 2 | Все | 9 | 0 | 0/9 | 9/9 | 0 | 100 |
| 2 | 1 | Все | 9 | 0 | 0/9 | 9/9 | 0 | 100 |
| 2 | 2 | Все | 9 | 0 | 1/9 | 8/9 | 11,1 | 88,9 |
| 3 | 1 | Все | 9 | 0 | 0/9 | 9/9 | 0 | 100 |
| 3 | 2 | Все | 9 | 0 | 1/9 | 8/9 | 11,1 | 88,9 |
| 4 | 1 | Все | 9 | 0 | 0/9 | 9/9 | 0 | 100 |
| 4 | 2 | Все | 9 | 0 | 2/9 | 7/9 | 22,2 | 77,8 |
| 1 | Все | Все | 18 | 0 | 0/18 | 18/18 | 0 | 100 |
| 2 | Все | Все | 18 | 0 | 1/18 | 17/18 | 5,6 | 94,4 |
| 3 | Все | Все | 18 | 0 | 1/18 | 17/18 | 5,6 | 94,4 |
| 4 | Все | Все | 18 | 0 | 2/18 | 16/18 | 11,1 | 88,9 |
| Все | 1 | Все | 36 | 0 | 0/36 | 36/36 | 0 | 100 |
| Все | 2 | Все | 36 | 0 | 4/36 | 32/36 | 11,1 | 88,9 |
| Все | Все | Все | 72 | 0 | 4/72 | 68/72 | 5,6 | 94,4 |
| aДля включения в исследование образцы должны пройти контроль процесса исследования (если применимо) и контроль качества.  Примечание – Обозначение: *QC* − контроль качества | | | | | | | | |

Критерии приемлемости могут быть установлены для отдельных образцов или групп образцов, представляющих аналогичные области пригодного для использования диапазона.

#### Приложение Д

#### (справочное)

#### Определение нижнего предела обнаружения для качественных исследований, основанных на методах полимеразной цепной реакции

Сокращения, применяемые в настоящем приложении:

ДНК − дезоксирибонуклеиновая кислота;

*LLoD* − нижний предел обнаружения (согласно методу частоты попаданий[[8]](#footnote-8)));

ПЦР − полимеразная цепная реакция;

*Pr* − вероятность.

Символы, применяемые в настоящем приложении:

µ − ожидаемое количество копий;

*b* − параметр логита;

*c* − параметр логита;

*e* − Уравнение Пуассона.

При идеальном качественном исследовании, основанном на методах ПЦР, любой образец, содержащий хотя бы одну копию целевого аналита, будет признан положительным. Если посмотреть на исследование частоты попаданий для такого идеального исследования, то можно ожидать, что частота попаданий будет равна нулю для образцов от пациентов без аналита и 100 % для образцов от пациентов, у которых есть аналит (т. е. инфицированных). Однако такого совершенства никогда не наблюдается при исследовании частоты попаданий из-за эффекта Пуассона, который определяет наличие копии аналита в исследуемом образце. Клиническая чувствительность исследования связана с вероятностью обнаружения аналита при его наличии [*Pr*(Обнаружение|Аналит присутствует)]. Однако общая вероятность обнаружения аналита для пациента [*Pr*(Обнаружение)] также зависит от вероятности присутствия копии аналита в исследуемом образце пациента [*Pr*(Аналит присутствует)] в соответствии с формулой

*Pr*(Обнаружение) = *Pr*(Обнаружение|Аналит присутствует) · *Pr*(Аналит присутствует) (Д.1)

*Pr*(Аналит присутствует) из формулы (Д.1) определяют как 1 − уравнение Пуассона для вероятности отсутствия аналита в образце в соответствии с формулой[[9]](#footnote-9))

*Pr*(Аналит присутствует) = 1 − *e*−μ, (Д.2)

где µ − ожидаемое количество копий в исследуемом объеме.

Для многих ПЦР-исследований µ может быть определено путем умножения средней концентрации в образце (например, копий вируса/см3) на объем образца (например, см3) исследования. Например, µ = 4 копии вируса/см3 − 0,5 см3/образец = 2 нити целевой ДНК/образец.

*P*r(Обнаружение|Аналит присутствует) из формулы (Д.1) может быть определен с помощью логит-уравнения, представленного в виде формулы (Д.3).

, (Д.3)

где *P*r(Обнаружение|Аналит присутствует) − вероятность обнаружения при условии, что µ больше нуля;

*c* и *b* − параметры логита, которые варьируются для достижения наилучшего соответствия результатам исследования частоты попаданий.

При желании эту вероятность обнаружения в зависимости от µ можно подогнать под пробит-функцию с помощью оценки максимального правдоподобия.

Формулы (Д.1)−(Д.3) используют для построения графиков на рисунках Д.1 и Д.2. В этих примерах нижний предел обнаружения (*LLoD*) определяют как число копий образца, для которого объединенная кривая обнаружения [*Pr*(Обнаружение)] пересекает 95 % уровень обнаружения. В этих примерах число копий аналита в соответствующей шкале, где бинарное исследование признает образец положительным в 95 % случаев (C95), дает то же значение, что и *LLoD*.

На рисунке Д.1 *LLoD* определено на уровне 30. Модель Пуассона предполагает, что для исследования используется образец объемом 1 см3; следовательно, измеряемой величиной является концентрация в копиях/см3 целевого вируса в образце. *LLoD* измеряеют при гораздо более высоком значении, чем то, которое может дать пуассоновский расчет. Поэтому предполагается, что комбинированное соответствие обусловлено почти исключительно клинической чувствительностью исследования.

Вероятность обнаружения

Среднее количество копий вируса на см3 пробы

60

50

40

30

20

10

0

0.0

C95

0.2

0.1

0.3

95 %

C95

0.4

Пуассон

Комбинированный

0.5

логит

1.0

0.9

0.8

0.7

0.6

C95 − число копий аналита в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; логит − *Pr*(Обнаружение|Аналит присутствует); Пуассон − *Pr*(Аналит присутствует)

Рисунок Д.1 − Умеренно чувствительное обнаружение *NAAT* (тест амплификации нуклеиновых кислот)

На рисунке Д.2 приведено исследование, которое является высокочувствительным для обнаружения аналита. При тех же допущениях, что и на рисунке Д.E1, измерение *LLoD* будет зависеть почти исключительно от того, присутствует ли в образце нить целевой нуклеиновой кислоты (т. е. от распределения Пуассона). Таким образом, клиническая чувствительность исследования не может быть измерена напрямую. Пример такого исследования приведен в 7.3.

Вероятность обнаружения

Среднее количество копий вируса на см3 пробы

5

4

3

2

1

0

0.0

C95

0.2

0.1

0.3

95%

C95

0.4

Пуассон

Комбинированный

0.5

логит

1.0

0.9

0.8

0.7

0.6

C95 − число копий аналита в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; логит − *Pr*(Обнаружение|Аналит присутствует); Пуассон − *Pr*(Аналит присутствует)

Рисунок Д.2 − Высокочувствительная обнаружение *NAAT* (тест амплификации нуклеиновых кислот)

#### Приложение Е

#### (справочное)

#### Схемы прецизионных исследований

Е.1 Сокращение, применяемое в настоящем приложении:

*DF* − степени свободы.

Е.2 Схемы исследований, приведенные в настоящем приложении, являются примерами. Возможны и другие схемы исследований, если учитывать характеристики исследования[[10]](#footnote-10)), а также логистику и возможности исследования. Степени свободы (*DF*) − число, которое можно рассматривать как показатель количества данных, лежащих в основе оценки точности[[11]](#footnote-11)).

**Е.3 Внутрилабораторные исследования**

Полностью перекрестный дизайн можно рассматривать, если оба оператора могут запускать оба прибора и все партии реагентов в течение каждого дня исследования. Другими словами, оба оператора выполняют одну и ту же работу каждый день в разные дни. В примере, приведенном в таблице Е.1, есть две реплики для каждой из 12 комбинаций из трех партий реагентов, двух приборов и двух операторов, оцениваемых в один день. Вся схема исследования повторяется в пять разных дней для каждого оцениваемого образца и уровня.

Таблица Е.1 − Однодневная логистика примера внутрилабораторного прецизионного исследования

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Каждый день | | | | | | | | | | | |
| Набор реагентов 1 | | | | Набор реагентов 2 | | | | Набор реагентов 3 | | | |
| Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | |
| Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 |
| Повторы | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

Для каждого образца будет получено 120 результатов (два прибора − два оператора − три партии реагентов − пять дней − два повтора). Чтобы понять, какой объем данных лежит в основе различных оценок компонентов дисперсии в полностью насыщенной модели для этого дизайна, можно определить *DF*, как показано в таблице Е.2.

Таблица Е.2 − Влияние *DF* на пример дизайна внутрилабораторного прецизионного исследования

|  |  |  |
| --- | --- | --- |
| Фактор | Расчет *DF* | Уравнение *DF* |
| Прибор | Количество приборов − 1 | 2 − 1 = 1 |
| Оператор | Количество операторов − 1 | 2 – 1 = 1 |
| Партия реагентов | Количество партий реагентов − 1 | 3 – 1 = 2 |
| День | Количество дней − 1 | 5 – 1 = 4 |
| Прибор · оператора | (Количество приборов − 1) · (количество операторов − 1) | 1 · 1= 1 |
| Прибор · партия реагентова | (Количество приборов − 1) · (количество партий реагентов − 1) | 1· 2 = 2 |
| Прибор · деньа | (Количество приборов − 1) · (количество дней − 1) | 1 · 4 = 4 |
| Оператор · партия реагентова | (Количество операторов − 1) · (количество партий реагентов − 1) | 1 · 2 = 2 |
| Оператор · деньа | (Количество операторов − 1) · (количество дней − 1) | 1 · 4 = 4 |
| Партия реагентов · деньа | (Количество партий реагентов− 1) · (количество дней − 1) | 2 · 4 = 8 |
| Прибор · оператор · партия реагентова | (Количество приборов − 1) · (количество операторов − 1) · (количество партий реагентов − 1) | 1 · 1 · 2 = 2 |
| Прибор · оператор · деньа | (Количество приборов − 1) · (количество операторов − 1) · (количество дней − 1) | 1 · 1 · 4 = 4 |
| Прибор · партия реагентов · деньа | (Количество приборов − 1) · (количество партий реагентов − 1) · (количество дней − 1) | 1 · 2 · 4 = 8 |
| Оператор · партия реагентов · деньа | (Количество операторов − 1) · (количество партий реагентов − 1) · (количество дней − 1) | 1 · 2 · 4 = 8 |
| Прибор · оператор · партия реагентов · деньа | (Количество приборов − 1) · (количество операторов − 1) · (количество партий реагентов − 1) · (количество дней − 1) | 1 · 1 · 2 · 4 = 8 |
| Повторы | Количество приборов · количество операторов · количество партий реагентов · количество дней · (количество повторов − 1) | 2 · 2 · 3 · 5(2 − 1) = 60 |
| Всего | Количество приборов · количество операторов · количество партий реагентов · количество дней · количество повторов − 1 | 2 · 2 · 3 · 5 · 2 − 1= 119 |
| aУсловия взаимодействия должны быть изучены на стадии разработки и, как ожидается, будут близки к нулю. На стадии валидации условия взаимодействия не включают[[12]](#footnote-12)).  Примечание – Обозначение: *DF* − степени свободы. | | |

Если невозможно, чтобы каждый оператор запускал один и тот же прибор или партию реагентов в один и тот же день (в течение каждого дня исследования), можно создать альтернативные схемы, как показано в таблицах Е.3 и Е.4. В этом примере каждый оператор запускает разные приборы и партии реагентов в разные дни (с двумя репликами каждый раз). Следовательно, для проведения исследования потребуется больше дней, чтобы получить необходимое количество наблюдений в одних и тех же условиях и сгенерировать необходимую *DF* для исследования. Логистика приведена в таблице Е.4.

Таблица Е.3 − Альтернативный пример дизайна внутрилабораторного прецизионного исследования

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Набор реагентов 1 | | | | Набор реагентов 2 | | | | Набор реагентов 3 | | | |
| Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | |
| Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 |
| 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день | 2  дня · 2  повтора в  день |

Таблица Е.4 − Логистика исследования по дням для альтернативного примера внутрилабораторного дизайна точного исследования

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Набор реагентов 1 | | | | Набор реагентов 2 | | | | Набор реагентов 3 | | | |
| Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | | Оператор 1 | | Оператор 2 | |
| Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 | Прибор 1 | Прибор 2 |
| Дни | 1 и  7 | 4 и  10 | 4 и  10 | 1 и  7 | 2 и  8 | 5 и  11 | 2 и  8 | 5 и  11 | 3 и  9 | 6 и  12 | 3 и  9 | 6 и  12 |

Для каждого образца имеется 48 результатов (два прибора · два оператора · три партии реагентов · два дня · два повтора). *DF* рассчитывают в соответствии с таблицей Е.5. Попытку отделить эффект календарного дня в исследовании не делают.

Таблица Е.5 − Влияние *DF* на альтернативный пример внутрилабораторного дизайна точного исследования

|  |  |  |
| --- | --- | --- |
| Фактор | Расчет *DF* | Уравнение *DF* |
| Прибор | Количество приборов − 1 | 2 – 1 = 1 |
| Оператор | Количество операторов − 1 | 2 – 1 = 1 |
| Партия реагентов | Количество партий реагентов − 1 | 3 – 1 = 2 |
| Прибор · оператора | (Количество инструментов − 1) · (количество операторов − 1) | 1 · 1 = 1 |
| Прибор · партия реагентова | (Количество приборов − 1) · (количество партий реагентов − 1) | 1 · 2 = 2 |
| Оператор · партия реагентова | (Количество операторов − 1) · (количество партий реагентов − 1) | 1 · 2 = 2 |
| Прибор · оператор · партия реагентова | (Количество приборов − 1) · (количество операторов − 1) · (количество партий реагентов − 1) | 1 · 1 · 2 = 2 |
| День (прибор, оператор, партия реагентов) | Количество приборов · количество операторов · количество партий реагентов · (количество дней в пределах прибора, оператора и партии реагентов − 1) | 2 · 2 · 3(2 − 1) = 12 |
| Повторы | Количество приборов · количество операторов · количество партий реагентов · количество дней внутри прибора и операторов · (количество реплик − 1) | 2 · 2 · 3 · 2(2 − 1) = 24 |
| Всего | Количество приборов · количество операторов · количество партий реагентов · количество дней в пределах приборов, операторов и партий реагентов · количество реплик − 1 | 2 · 2 · 3 · 2 · 2 − 1 = 47 |
| aУсловия взаимодействия должны быть изучены на стадии разработки и, как ожидается, будут близки к нулю. На стадии валидации условия взаимодействия не включают[[13]](#footnote-13)).  Примечание – Обозначение: *DF* − степени свободы. | | |

**Е.4 Воспроизводимость**

В случае дизайна воспроизводимых исследований место проведения становится источником вариабельности и должно быть частью логистики, а также модели исследования. В таблицах Е.6 и Е.7 приведены примеры дизайна с учетом двух операторов на участок, одного прибора на участок и, в данном случае, шести повторов на уровень за прогон.

Таблица Е.6 − Логистика дизайна исследования воспроизводимости

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Участок 1 | | | | | | Участок 2 | | | | | | Участок 3 | | | | | |
| Прибор 1 | | | | | | Прибор 2 | | | | | | Прибор 3 | | | | | |
| Набор реагентов 1 | | | | | | Набор реагентов 1 | | | | | | Набор реагентов 1 | | | | | |
| Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | | Оператор 1 | | | Оператор 2 | | |
| День 1 | День 3 | День 5 | День 2 | День 4 | День 6 | День 1 | День 3 | День 5 | День 2 | День 4 | День 6 | День 1 | День 3 | День 5 | День 2 | День 4 | День 6 |
| Повто-ры | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

Для каждого образца имеется 108 результатов (три участка · два оператора · три дня · шесть повторов). *DF* рассчитывают в соответствии с таблицей Е.7.

Таблица Е.7 − Влияние *DF* на дизайн исследования воспроизводимости

|  |  |  |
| --- | --- | --- |
| Фактор | Расчет *DF* | Уравнение *DF* |
| Участок | Количество участков − 1 | 3 – 1 = 2 |
| Оператор · участок | Количество участков · (количество операторов − 1) | 3(2 − 1) = 3 |
| День, оператор · участок | Количество участков · количество операторов · (количество дней − 1) | 3 · 2(3 − 1) = 12 |
| Повторы | Количество участков · количество операторов · количество дней · (количество реплик − 1) | 3 · 2 · 3(6 − 1) = 90 |
| Всего | Количество участков - количество операторов - количество дней - количество реплик− 1 | 3 · 2 · 3 · 6 – 1 = 107 |
| Примечание – Обозначение: *DF* − степени свободы. | | |

*DF* допускается использовать для оценки или сравнения различных схем исследований с точки зрения количества данных, лежащих в основе различных оценок компонентов дисперсии. Например, рекомендуется[[14]](#footnote-14)), чтобы план исследования воспроизводимости обеспечивал не менее 20 *DF* для повторяемости. Схема, представленная в таблице Е.7, обеспечивает 90 *DF* для изменчивости реплик (повторяемости). Оценка компонентов дисперсии приведена в соответствующих рекомендациях[[15]](#footnote-15)).

#### Приложение Ж

#### (справочное)

#### Исследования точности наблюдателей

Сокращения, применяемые в настоящем приложении:

*ANA* − среднее отрицательное согласие;

*APA* − среднее положительное согласие;

ДИ − доверительный интервал;

*IHC* − иммуногистохимический;

*NPA* − отрицательное процентное согласие;

*PPA* − положительное процентное согласие.

Символы, применяемые в настоящем приложении:

*k* − каппа-статистика;

*i* − наблюдатель 1;

*j* − наблюдатель 2;

*L* − нижний;

*p* − особая пропорция;

*U* − верхний;

*w* − вес;

*y* − количество образцов.

Для некоторых качественных исследований наблюдатель оценивает результаты исследования, чтобы получить бинарный результат. Например, патологоанатом может оценить срез образца ткани, окрашенного иммуногистохимическим (*IHC*) анализом, на предмет процентного содержания окрашенных клеток или интенсивности окрашивания и применить алгоритм (например, пороговые значения) к этим оценкам для получения бинарного результата. Аналогично, для некоторых полуколичественных исследований наблюдатель оценивает результаты анализа, чтобы получить порядковый результат (например, оценки 0 +, 1 +, 2 + или 3 + сверхэкспрессии *HER*2 в образцах ткани опухоли молочной железы или желудка).

Прецизионное исследование может быть разработано специально для оценки вариабельности качественного результата между наблюдателями или для одного наблюдателя, при этом все остальные факторы (условия) остаются неизменными (насколько это возможно). Для оценки согласия между двумя наблюдателями допускается рассматривать положительное процентное согласие (*PPA*) и отрицательное процентное согласие (*NPA*), но это зависит от того, какой наблюдатель оценивается для *PPA* и *NPA* и какой наблюдатель служит в качестве компаратора для оценки, как объясняется в 8.3.6. Желательно использовать меры согласия между двумя наблюдателями, которые не зависят от того, кто из наблюдателей выступает в качестве компаратора для оценки другого. Одной из таких пар мер является среднее отрицательное согласие (*ANA*) и среднее положительное согласие (*APA*), определяемые по формулам (Ж.1) и (Ж.2), соответственно

(Ж.1)

(Ж.2)

где *yij* − количество выборок, на которых бинарные результаты для наблюдателей 1 и 2 равны *i* и *j*, с обозначением отрицательных и положительных результатов, соответственно, *yi∙* = *yi*1 + *yi*2 и *y∙j*= *y*1*j* + y2*j* (см. таблицу Ж.1).

Таблица Ж.1 − Условные обозначения для подсчета парных бинарных результатов *yij* от наблюдателей 1 и 2

|  |  |  |  |
| --- | --- | --- | --- |
| Наблюдатель 1 | Наблюдатель 2 | | Итогоa,b |
| − | + |
| − | *y*11 | *y*12 | *y*1∙ |
| + | *y*21 | *y*22 | *y*2∙ |
| Итогоb,c | *y*∙1 | *y*∙2 | *y*∙∙ |
| a*yi∙ = yi*1 *+ yi*2*, i* = 1,2.  b*y∙∙= y*1∙ *+ y*2∙ *= y*∙1 *+ y*∙2*.*  c*y∙j = y*1*j + y*2*j, j* = 1,2. | | | |

*ANA* допускается пересчитывать как средневзвешенное значение согласно формуле

*ANA = w−NPA*12 *+*  (1 − *w−*) *NPA*21, (Ж.3)

где *NPA*12 = *y*11/*y*∙1 – *NPA* наблюдателя 1 с наблюдателем 2 (выражают как долю отрицательных образцов наблюдателя 2, которые наблюдатель 1 также считает отрицательными);

*NPA*21 = *y*11/*y*1∙ – *NPA* наблюдателя 2 с наблюдателем 1;

*w−* = *y*∙1/(*y*1∙ + *y*∙1) − вес, предельное общее количество отрицательных образцов *y*∙1 для наблюдателя 2 над суммой предельных общих количеств отрицательных образцов для наблюдателей 1 и 2.

Аналогичным образом *APA* допускается пересчитывать как средневзвешенное значение по формуле

*APA = w+PPA*12 *+*  (1 – *w+*) *PPA*21, (Ж.4)

где *PPA*12 = *y*22/y∙2 − *PPA* наблюдателя 1 с наблюдателем 2;

*PPA*21 = *y*22/*y*2∙ − *PPA* второго наблюдателя 2 с наблюдателем 1;

*w*+ = *y*2∙/(*y*2∙ + *y*∙2) − вес, основанный на предельных суммах положительных выборок для двух наблюдателей.

В качестве альтернативы, *ANA* и *APA* допускается интерпретировать как *NPA* и *PPA* для оценки случайно выбранного наблюдателя против другого[[16]](#footnote-16)) (с вероятностями *w−* для *NPA* и *w*+ для *PPA*). Парные данные из 90 образцов приведены в таблице Ж.2.

Таблица Ж.2 − Парные данные из 90 образцов

|  |  |  |  |
| --- | --- | --- | --- |
| Наблюдатель 1 | Наблюдатель 2 | | Всего |
| − | + |
| − | 70 | 2 | 72 |
| + | 5 | 13 | 18 |
| Всего | 75 | 15 | 90 |

Для этих данных *ANA* и *APA* вычисляют как

Доверительный интервал (ДИ) для *ANA* может быть получен, если учесть, что *ANA* = 2*p*−/(1 + *p*−), где *p*−= *y*11/(*y*12 + *y*∙1) − специальная пропорция, для которой ДИ [*p*−*L*, *p*−*U*] может быть основан на оценке Уилсона или методе Клоппера-Пирсона. Поскольку *ANA* является монотонной функцией *p*−, ДИ на *ANA* − это [2*p*−*L*/(1 + *p*−*L*) , 2*p*−*U*/(1 + *p*−*U*)].

Аналогично, ДИ на *APA* можно получить, отметив, что *APA* = 2*p*+/(1 + *p*+), где   
*p*+ = *y*22/(*y*21 + *y*∙2) − специальная пропорция, на которой можно получить ДИ [*p*+*L*, *p*+*U*]. Поэтому ДИ на *APA* − это [2*p*+*L*/(1 + *p*+*L*) , 2*p*+*U*/(1 + *p*+*U*)].

Для примера, уравнение *p*− = 70/(2 + 5 + 70)= 0,9091 используют для данных, представленных выше. Двусторонний 95 %-ный ДИ для *p*− по шкале Уилсона составляет [0,8240, 0,9553]. Таким образом, двусторонний 95 %-ный ДИ для истинного *ANA* составляет   
[2(0,8240)/(1 + 0,8240), 2(0,9553)/(1 + 0,9553)] =[0,9035, 0,9771] или [90,4 %, 97,7 %].

Аналогично, для приведенных выше данных используют уравнение   
*p*+ = 13 /(5 + 2 + 13)= 0,65. Точный двусторонний 95 %-ный ДИ Клоппера-Пирсона для *p*+ составляет [0,4078, 0,8461]. Таким образом, точный двусторонний 95 %-ный ДИ для истинного значения *APA* составляет [2(0,4078)/(1 + 0,4078), 2(0,8461)/(1 + 0,8461)] = [0,5794, 0,9166] или [57,9 %, 91,7 %].

*APA* был первоначально предложен Дайсом1), который назвал его индексом совпадения. В радиологии его обычно называют коэффициентом сходства Дайса[[17]](#footnote-17)). В обзоре индексов согласия между двумя наблюдателями Флейсс[[18]](#footnote-18)) отметил, что коррекция *ANA* или *APA* на случайное согласие дает каппа-статистику (*k*). Чтобы разрешить кажущиеся парадоксы с общим согласием и *k*, Чиккетти и Файнштейн[[19]](#footnote-19)) предложили называть *ANA* и *APA* *p*− и *p*+, соответственно. Исследователи спорят о клиническом значении *p*− и *p*+ и следует ли учитывать их выборочные распределения[[20]](#footnote-20)). Были предложены приблизительные ДИ для *p*− и *p*+[[21]](#footnote-21)), но предпочтительным является метод ДИ, описанный выше. Были предложены общие рекомендации по представлению результатов исследований надежности и согласия[[22]](#footnote-22)).

(-

Подсчет *ANA* и *APA* между наблюдателями может быть использован для оценки неточности по другим факторам, таким как партия и участок.

#### Приложение И

#### (справочное)

#### Метод оценки Уилсона для расчета доверительных интервалов

И.1 Сокращения, применяемые в настоящем приложении:

ДИ − доверительный интервал;

*Secand*− клиническая чувствительность кандидата;

*Seold* − клиническая чувствительность ранее применяемого исследования;

*Spcand* − клиническая специфичность кандидата;

*Spold* − клиническая специфичность ранее применяемого исследования.

И.2 Символы, применяемые в настоящем приложении:

∆Se − разница между чувствительностями;

∆Sp − разница между специфичностями;

*A* − количество истинно положительных результатов;

*B* − количество ложных срабатываний;

*C* − количество ложноотрицательных результатов;

*D* − количество истинно отрицательных результатов;

*l*1, *l*2 − нижние границы двухсторонних 95 %-ных ДИ;

*N* − количество образцов;

*Q* − определения;

*u*1, *u*2 − верхние границы двухсторонних 95 %-ных ДИ.

**И.3 Доверительный интервал пропорции**

Примером использования метода баллов является расчет доверительного интервала (ДИ) для оценки клинической чувствительности с использованием данных из таблицы И.1 (взятой из таблицы 4 в 8.3.5).

Таблица И.1 − Качественная экспертиза в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС отсутствует |
| Положительное | *A* | *B* | *A + B* |
| Отрицательное | *C* | *D* | *C + D* |
| Всего | *A + C* | *B + D* | *A + B + C + D* |
| aНаилучшая имеющаяся оценка ЦС. | | | |

Двухсторонний 95 %-ный ДИ для чувствительности или специфичности может быть вычислен по формуле

, (И.1)

где *Q*1, *Q*2 и *Q*3 вычисляют по данным таблицы И.1 по формулам (И.2) – (И.4).

Для чувствительности,

*Q*1 = 2*A* + 1,962 = 2*А* + 3,84, (И.2)

(И.3)

*Q*3 = 2(*A* + *C* + 1,962) = 2(*А* + *C* + 3,84). (И.4)

В формулах (И.2) – (И.4) 1,96 − квантиль стандартного нормального распределения, который соответствует двустороннему доверительному уровню 95 %. Если требуется другой уровень доверия, вместо 1,96 следует использовать соответствующий квантиль.

95 %-ный ДИ для специфичности [формула (И.1)] рассчитывают аналогичным образом, заменяя *A* на *D* и *C* на *B* в формулах (И.2) – (И.4). Пример таких расчетов приведен в   
приложении К.

В более обобщенном варианте применения ДИ количества экземпляров, удовлетворяющих выбранному условию (например, положительных образцов), может быть рассчитан с помощью формул (И.2) – (И.4), если *A* − количество экземпляров, а *N* = *A* + *C* − общее количество исследованных образцов (см. 8.3.5).

**И.4 Доверительный интервал пропорциональной разницы**

Если сравнить исследование-кандидат и ранее применяемое исследование с оценкой ЦС, можно получить следующие таблицы, как показано в таблице 8.

Таблица И.2 − Трехстороннее сравнение между исследованием-кандидатом, ранее применяемым исследованием и оценкой ЦС

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Оценка ЦСа | | | | Оценка ЦСа | | | |
| ЦС присутствует | | | | ЦС отсутствует | | | |
| Исследование-кандидат | Ранее применяемое исследование | | Всего | Исследование-кандидат | Ранее применяемое исследование | | Всего |
| Поло-житель-ное | Отрицательное | Поло-житель-ное | Отрицательное |
| Положительное | *А*1 | *В*1 | *A*1 + *B*1 | Положительное | *А*0 | *В*0 | *A*0 + *B*0 |
| Отрицательное | *С*1 | *D*1 | *C*1 + *D*1 | Отрицательное | *С*0 | *D*0 | *C*0 + *D*0 |
| Всего | *А*1 + *С*1 | *B*1 + *D*1 | *N*1 | Всего | *А*0 + *С*0 | *B*0 + *D*0 | *N*0 |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначения: *N* − количество образцов; ЦС − целевое состояние. | | | | | | | |

Таблица И.3 − Исследование-кандидат в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС отсутствует |
| Положительное | *A*1 + *B*1 | *A*0 + *B*0 | *A*1 + *B*1 + *A*0 + *B*0 |
| Отрицательное | *C*1 + *D*1 | *C*0 + *D*0 | *C*1 + *D*1 + *C*0 + *D*0 |
| Всего | *N*1 | *N*0 | *N* |
| aНаилучшая имеющаяся оценка ТС.  Примечание – Обозначение: ЦС − целевое состояние. | | | |

Таблица И.4 – Ранее применяемое исследование в сравнении с оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Ранее применяемое исследование | Оценка ЦСа | | Всего |
| ЦС присутствует | ЦС отсутствует |
| Положительное | *A*1 + *C*1 | *A*0 + *C*0 | *A*1 + *C*1 + *A*0 + *C*0 |
| Отрицательное | *B*1 + *D*1 | *B*0 + *D*0 | *B*1 + *D*1 + *B*0 + *D*0 |
| Всего | *N*1 | *N*0 | *N* |
| aНаилучшая имеющаяся оценка ЦС.  Примечание – Обозначение: ЦС − целевое состояние. | | | |

Предполагаемая клиническая чувствительность исследования-кандидата (*Secand*) и ранее применяемого исследования (*Seold*), их разность и ДИ этой разности могут быть рассчитаны с помощью формул (И.5) − (И.8). Такие же расчеты для клинической специфичности выполняют по формулам (И.15) − (И.24).

(И.5)

(И.6)

(И.7)

Двусторонний 95 %-ный ДИ для разницы между чувствительностью исследования-кандидата и ранее применяемого исследования рассчитывают по формуле (И.8), чтобы увидеть, есть ли перекрытие с нулевой разницей

(И.8)

где *Q*5 и *Q*6 вычисляют из данных с помощью формул (И.9) − (И.14). Сначала вычисляют отдельные балльные ДИ для *Secand*и *Seold* с помощью формулы (И.1), а затем вычисляются величины *Q*1 – *Q*6 с помощью формул (И.9) − (И.14), где:

- *l*1 − нижняя граница двухстороннего 95 %-ного ДИ для *Secand*;

- *u*1 − верхняя граница двухстороннего 95 %-ного ДИ для *Secand*;

- *l*2 − нижний предел двухстороннего 95 %-ного ДИ для *Seold*;

*- u*2 − верхняя граница двухстороннего 95 %-ного ДИ для *Seold*.

*Q*1 = (*A*1 + *B*1)(*C*1 + *D*1)(*A*1 + *C*1)(*B*1 + *D*1), (И.9)

*Q*2 = (*A*1*D*1) − (*B*1*C*1), (И.10)

, (И.11)

(И.12)

(И.13)

(И.14)

Двусторонний 95 %-ный ДИ для разницы между парными специфичностями приведен в формулах (И.15) − (И.24)

(И.15)

(И.16)

(И.17)

(И.18)

где:

- *l*1 − нижняя граница двустороннего 95 %-ного ДИ для *Spcand*;

- *u*1 − верхняя граница двустороннего 95 %-ного ДИ для *Spcand*;

- *l*2 − нижняя граница двухстороннего 95 %-ного ДИ для *Spold*;

*- u*2 − верхняя граница двустороннего 95 %-ного ДИ для *Spold*.

*Q*1 = (*A*0 + *B*0)(*C*0 + *D*0)(*A*0 + *C*0)(*B*0 + *D*0), (И.19)

*Q*2 = (*A*0*D*0) − (*B*0*C*0), (И.20)

, (И.21)

(И.22)

(И.23)

(И.24)

#### Приложение К

#### (справочное)

#### Примеры анализа клинической эффективности

К.1 Сокращения, применяемые в настоящем приложении:

ДИ − доверительный интервал;

ИФА − иммуноферментный анализ;

*FN* – ложноотрицательный результат;

*FP* – ложноположительный результат;

*HMW-CAP EIA* – иммуноферментный анализ высокомолекулярных клеточно-ассоциированных белков;

*NPA* − отрицательное процентное согласие;

*NPV* − отрицательная прогностическая ценность;

*PPA* − положительное процентное согласие;

*PPV* − положительная прогностическая ценность;

*Secand*− клиническая чувствительность кандидата;

*Seold* − клиническая чувствительность ранее применяемого исследования;

*Spcand* − клиническая специфичность кандидата;

*Spold* − клиническая специфичность ранее применяемого исследования;

ЦС − целевое состояние;

*TP* − истинно положительный;

*TN* − истинно отрицательный.

К.2 Символы, применяемые в настоящем приложении:

∆Se − разница между чувствительностями;

∆Sp − разница между специфичностями;

*l*1, *l*2 − нижние границы двухсторонних 95 %-ных ДИ;

*N* − количество образцов;

*Q* − определения;

*u*1, *u*2 − верхние границы двухсторонних 95 %-ных ДИ.

**К.3** **Пример − Если имеется оценка состояния цели**

Представлены данные, которые позволят сравнить исследование-кандидата и ранее применяемое исследование по отдельности с независимо полученным диагнозом. В примере из справочных документов восемь коммерческих ИФА использовались для тестирования сывороток, полученных от 102 пациентов, у которых определялся статус инфекции *Helicobacter pylori[[23]](#footnote-23))*. В этом примере все коэффициенты, относящиеся к клиническим показателям, выражены в процентах.

Показатели исследования-кандидата и ранее применяемого исследования определяют в сравнении с наилучшей доступной оценкой результата целевого состояния (ЦС) для каждого образца пациента (т. е. *Z*-стандарт). Таблица случайных чисел 2×2 результатов исследования-кандидата и ранее применяемого исследования приведены в таблицах К.1 и К.2. Для оценки клинической чувствительности и специфичности используются формулами (7) и (8), приведенными в 8.3.5. Обе эти таблицы получены из объединенной таблицы К.3.

Таблица К.1 − 2×2 таблица зависимостей между результатами исследования-кандидата и *Z*-стандартом[[24]](#footnote-24))

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат | *Z*-стандарт для *Helicobacter pylori* | | Всего |
| Положительное | Отрицательное |
| Положительное | *TP* = 57 | *FP* = 2 | 59 |
| Отрицательное | *FN* = 4 | *TN* = 39 | 43 |
| Всего | 61 | 41 | 102 |
| Примечание – Обозначения: *FN* − ложноотрицательный результат; *FP* − ложноположительный результат; *TN* − истинно отрицательный; *TP* − истинно положительный. | | | |

Распространенность заболевания, %, вычисляют по формуле

(К.1)

В данном примере %

Положительная прогностическая ценность *PPV*, %, положительного результата теста вычисляют по формуле

(К.2)

В данном примере %

Отрицательная прогностическая ценность *NPV*, %, отрицательного результата теста вычисляют по формуле

(К.3)

В данном примере

Оценочную чувствительность для кандидатского экзамена *Secand*, %, вычисляют по формуле

(К.4)

В данном примере

95 %-ные ДИ вычисляют по формуле (И.1) приложения И, как показано ниже

*Q*1 = 2 ∙ 57 + 1,962 = 117,84,

*Q*3 = 2(57 + 4 + 1,962) = 129,68

Нижний предел, %, вычисляют по формуле

(К.5)

В данном примере

Верхний предел %, вычисляют по формуле

(К.6)

В данном примере %

Расчетную специфичность для исследования-кандидата *Spcand*, %, вычисляют по формуле

(К.7)

В данном примере

95 %-ные ДИ вычисляют по формуле (И.1) приложения И, как показано ниже

*Q*1 = 2 ∙ 39 + 1,962 = 81,84,

*Q*3 = 2(39 + 2 + 1,962) = 89,68

Нижний и верхний пределы в данном примере вычисляют по формулам (К.5) и (К.6)

Таблица К.2 − 2× 2 таблица зависимостей между ранее применяемым исследованием и оценкой ЦС

|  |  |  |  |
| --- | --- | --- | --- |
| Ранее применяемое  исследование | *Z*-стандарт для *H. pylori*a | | Всего |
| Положительное | Отрицательное |
| Положительное | *TP* = 54 | *FP* = 7 | 61 |
| Отрицательное | *FN* = 7 | *TN* = 34 | 41 |
| Всего | 61 | 41 | 102 |
| a*Z*-стандарт должен быть подробно описан как наилучшая доступная оценка ЦС.  Примечание – Обозначения: *FN* − ложноотрицательный результат; *FP* − ложноположительный результат; ЦC − целевое состояние;*TN* − истинно отрицательный; *TP* − истинно положительный. | | | |

Распространенность заболевания, *PPV* и *NPV* вычисляются по формулам

%,

%,

Формулу (К.8) допускается использовать для оценки чувствительности ранее применяемого исследования

(К.8)

В данном примере

95 %-ные ДИ вычисляют по формуле (И.1) приложения И, как показано ниже

*Q*1 = 2 ∙ 54 + 1,962 = 111,84,

*Q*3 = 2(54 + 7 + 1,962) = 129,68

Нижний и верхний пределы в данном примере вычисляют по формулам (К.5) и (К.6)

Расчетную специфичность для ранее применяемого исследования *Spold*, %, вычисляют по формуле

(К.9)

В данном примере

95 %-ные ДИ вычисляют по формуле (И.1) приложения И, как показано ниже

*Q*1 = 2 ∙ 34 + 1,962 = 71,84,

*Q*3 = 2(34 + 7 + 1,962) = 89,68

Нижний и верхний пределы в данном примере вычисляют по формулам (К.5) и (К.6)

В данном случае имеет смысл проведение совместного статистического сравнения пар чувствительность/специфичность, поскольку предполагаемая чувствительность и специфичность для исследования-кандидата больше, чем предполагаемая чувствительность и специфичность для ранее применяемого исследования. Чтобы сравнить чувствительность и специфичность двух экзаменов, необходимо провести трехстороннее сравнение между ранее применяемым исследованием, исследованием-кандидатом и результатами оценки ЦС. Такие сравнительные результаты косвенно представлены в справочных документах[[25]](#footnote-25)).

Таблица К.3 − Трехстороннее сравнение между исследованием-кандидатом, ранее применяемым исследованием и оценкой ЦС

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Z-с*тандарт: ЦС присутствуетa | | | |  | *Z-с*тандарт: ЦC Отсутствуетa | | | |
| Исследование-кандидат | Ранее применяемое  исследование | | Всего | Исследование-кандидат | Ранее применяемое  исследование | | Всего |
| Положи-тельное | Отрица-тельное | Положи-тельное | Отрица-тельное |
| Положительное | 53 | 4 | 57 | Положительное | 2 | 0 | 2 |
| Отрицательное | 1 | 3 | 4 | Отрицательное | 5 | 34 | 39 |
| Всего | 54 | 7 | 61 | Всего | 7 | 34 | 41 |
| a*Z*-стандарт должен быть подробно описан как наилучшая доступная оценка ЦС. | | | | | | | | |

Сравнение чувствительности ∆Se, %, вычисляют по формуле

(К.10)

В данном примере

или

95 %-ные ДИ для исследований-кандидатов и ранее применяемых исследований могут быть рассчитаны для чувствительности по формуле (И.1) приложения И. Однако более целесообразно рассчитывать 95 %-ные ДИ для разницы между ними по формулам (И.9) − (И.14) приложения И и формулам (К.11) и (К.12), приведенным ниже, где:

- *l*1 = 84,3 % (от ДИ оценки для *Secand*);

- *u*1 = 97,4 %;

- *l*2 = 78,2 % (от ДИ оценки для *Seold*);

- *u*2 = 94,3 %.

*Q*1 = (53 + 4)(1 + 3)(53 + 1)(4 + 3) = 57 ∙ 4 ∙ 54 ∙ 7 = 86 184,

*Q*2 = 53 ∙ 3 – 4 ∙ 1 = 155.

В приведенных ниже примерах *n*1 − общее количество образцов от субъектов с ЦС, а   
*n*0 − общее количество субъектов без ЦС.

Потому, что ,

,

*Q*5 = (93,4 – 84,3)2 – 2 ∙ 0,4241(93,4 – 84,3)(94,3 – 88,5) + (94,3 – 88,5)2 = 71,68 %,

*Q*6 = (88,5 – 78,2)2 – 2 ∙ 0,4241(88,5 – 78,2)(97,4 – 93,4) + (97,4 – 93,4)2 = 87,14 %,

(К.11)

(К.12)

95 %-ный ДИ для Δ*Se* = *Secand − Seold* составляет (− 3,6 %, 14,2 %).

Cравнение специфичности выполняют по формуле (К.13)

(К.13)

или

95 %-ные ДИ для исследований-кандидатов и ранее применяемых исследований могут быть рассчитаны для специфичности по формуле (И.1) приложения И. Однако более целесообразно рассчитывать 95 %-ные ДИ для разницы между ними по формулам (И.19) − (И.24) приложения И и формул (К.14) и (К.15), приведенным ниже, где:

- *l*1 = 83,9 % (от ДИ оценки для *Spcand*);

- *u*1 = 98,7 %;

- *l*2 = 68,7 % (от ДИ оценки для *Spold*);

- *u*2 = 91,5 %.

*Q*1 = (2 + 0)(5 + 34)(2 + 5)(0 + 34) = 2 ∙ 39 ∙ 7 ∙ 34 = 18 564,

*Q*2 = 2 ∙ 34 – 0 ∙ 5 = 68.

Потому, что ,

,

*Q*5 = (95,1 – 83,9)2 – 2 ∙ 0,3486(95,1 – 83,9)(91,5 – 82,9) + (91,5 – 82,9)2 = 132,25 %,

*Q*6 = (82,9 – 68,7)2 – 2 ∙ 0,3486(82,9 – 68,7)(98,7 – 95,1) + (98,7 – 95,1)2 = 178,96 %,

(К.14)

(К.15)

95 %-ный ДИ для Δ*Sp* = *Spcand − Spold* составляет (0,7 %, 25,6 %).

Разница в чувствительности составляет 4,9 % с 95 %-ным ДИ (− 3,6 %; 14,2 %). Этот ДИ включает ноль, поэтому пользователь не может сделать вывод о статистической разнице в чувствительности. Разница в специфичности составляет 12,2 % при 95 %-ном   
ДИ (0,7 %; 25,6 %). Этот ДИ не включает нуля; следовательно, есть доказательства того, что специфичность исследования-кандидата статистически значимо выше, чем специфичность сравнительного исследования. Этот результат указывает на то, что клинические показатели исследования-кандидата лучше, чем у ранее применяемого исследования.

Учитывая распространенность заболевания 59,8 % [рассчитанную по формуле (К.1)], *PPV* и *NPV* обоих исследований можно вычислить по формулам (К.16) − (К.19) при данном проценте распространенности. Приведенные ниже уравнения включают использование оценки распространенности. Таким образом, *PPV* и *NPV* допускается рассчитывать для любой целевой популяции, для которой известна распространенность.

(К.16)

,

(К.17)

,

(К.18)

(К.19)

Результаты *PPV* и *NPV* показывают, как используют исследования для ведения пациентов. Результаты, вычисленные по формулам (К.16) − (К.19), указывают на то, что клинические показатели исследования-кандидата лучше, чем у ранее применяемого исследования.

**К.4 Пример − Если оценка состояния цели недоступна**

В типичной ситуации эксперт может не знать диагноза, но хочет сравнить результаты исследования-кандидата с результатами другого исследования. В примере из справочных документов оценивали результаты двух тестов на антитела к *H. pylori*[[26]](#footnote-26)). В исследовании-кандидате использовали иммунохроматографический метод, а в сравнительном исследовании − метод иммуноферментного анализа высокомолекулярных клеточно-ассоциированных белков (*HMW-CAP EIA*) *ELISA*. Таблица случайных чисел 2×2 результатов оценки представлена в таблице К.4.

Таблица К.4 − *HMW-CAP EIA*

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование-кандидат (иммунохроматографи-ческий метод) | Сравнительное исследованиеа | | |
| Положительное | Отрицательное | Всего |
| Положительное | 285 | 15 | 300 |
| Отрицательное | 14 | 222 | 236 |
| Всего | 299 | 237 | 536 |
| aДолжна быть приведена информация о сравнительном исследовании. | | | |

Положительное процентное согласие (*PPA*), обусловленное результатами иммунохроматографического анализа или *HMW-CAP EIA*, равно

100 % − 285/299 = 95,3 %.

95 %-ные ДИ *PPA*, которые зависят от результатов *HMW-CAP EIA*, рассчитывают по формуле (И.1) приложения И, как показано в уравнениях ниже

*Q*1 = 2 ∙ 258 + 1,962 = 573,84,

*Q*3 = 2(285 + 14 + 1,962) = 605,68

Нижний и верхний пределы в данном примере вычисляют по формулам (К.5) и (К.6)

Отрицательное процентное согласие (*NPA*), которое обусловлено результатами иммунохроматографического анализа или *HMW-CAP EIA*, равно

100 % − 222/237 = 93,7 %.

95 %-ные ДИ *NPA*, обусловленные результатами иммунохроматографии или *HMW-CAP EIA*, рассчитывают по формуле (И.1) приложения И, как показано в уравнениях ниже

*Q*1 = 2 ∙ 222 + 1,962 = 447,84,

*Q*3 = 2(222 + 15 + 1,962) = 481,68

Нижний и верхний пределы в данном примере вычисляют по формулам (К.5) и (К.6)

#### Приложение Л

#### (справочное)

#### Пример полимеразной цепной реакции в реальном времени для ванкомицин-резистентных энтерококков

Сокращения, применяемые в настоящем приложении:

ДИ − доверительный интервал;

*CT* – порог цикла;

*CV* − коэффициент вариации;

*CX* − значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в *X* % случаев;

ПЦР − полимеразная цепная реакция;

*SD* − стандартное отклонение.

В больничной лаборатории установлен прибор для ПЦР в реальном времени, который выполняет качественные исследования на наличие ванкомицин-устойчивых энтерококков (ген *vanA*). Хотя данные о прецизионности образцов с целевыми условиями при двух различных концентрациях целевого гена приведены во вкладыше к изделию, прецизионность при значении в соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев (C5), и при значении в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев (C95), не приведена. Неточность результатов исследования выражается в *SD* среднего числа циклов ПЦР. Отсечение указано как 39 порогов циклов (*CT*). В этом типе теста *CT* уменьшается по мере увеличения концентрации *vanA*, поскольку требуется меньше циклов ПЦР для достижения порога обнаружения сигнала, так что C95 *CT* меньше, чем значение в соответствующей шкале, где бинарный анализ признает образец положительным в 50 % случаев (C50). *CT* и C5 *CT* больше, чем C50. Заявления о неточности во вкладыше к продукту приведены в таблице Л.1.

Таблица Л.1 − Заявления во вкладыше к препарату о неточности исследования нуклеиновой кислоты ванкомицин-резистентных энтерококков

|  |  |  |  |
| --- | --- | --- | --- |
| Образец с ЦС | *CT* | *SD* | *CV* |
| *VanA* низкая концентрация | 33,1 | 1,5 | 4,5 |
| *VanA* умеренная концентрация | 30,4 | 1,4 | 4,6 |

В таблице Л.1 представлено ограниченное число результатов исследования образцов. Для определения постоянства *SD* или *CV* в течение всего интервала измерений целесообразно использовать большее количество образцов, охватывающих более широкий интервал концентраций. При неверном предположении о постоянстве *CV* или *SD* может потребоваться повторение начальных этапов процедуры, показанной ниже, для проверки заявления производителя с альтернативным предположением.

Поскольку погрешность двух образцов одинакова, предполагается, что *CV* на отсечке, C5 и C95 также будут одинаковыми. Однако количественные расчеты с помощью этого метода невозможны, поскольку, как только *CT* достигает 39, результаты проб представляются как нули, и прибор не сообщает о *CT*. Исходя из предположения о постоянном *CV*, ошибка *CT*, ожидаемая при отсечении, должна быть равна 0,045 ∙ 39 = 1,755 *SD*. Вычитание 1,645 *SD* (односторонний 95 %-ный ДИ) из отсечки должно приблизительно равняться циклам на C95: 39 − 1,645 ∙ 1,755 = 39 − 2,887= 36,1.

Чтобы проверить заявление производителя:

- разбавляют аликвоту образца, положительного на микроорганизм, в разбавителе для предварительной обработки в соотношении 1:100;

- исследуют пять реплик за один прогон.

Средние результаты = 22,24 ∙ *CT* обратно рассчитанные аккуратные результаты:   
(100 = 26,64), 22,24 − 6,64= 15,6 *CT*.

Примечание − Каждый цикл удваивает количество нуклеиновой кислоты, поэтому разница в 6,64 цикла − 100-кратная разница в концентрации нуклеиновой кислоты (т.е. кол/см3).

- разбавляют образец, приняв за 1 *CT* = 50 %-ное снижение концентрации. Чтобы сформировать образец C50:

1) вычисляют разность 39 − 15,6 = 23,4.

Примечание – Два, возведенное в степень разности циклов, равно 223,4 = 11,07×106.

2) разбавляют неочищенный высокоположительный препарат 1:2,75×106, 1:11×106 и 1:44×106 в разбавителе для предварительной обработки.

- исследуют пять копий каждого разведения за один прогон;

- исследуют образцы в течение пяти дней с четырьмя повторами в день.

Исходя из ожидаемой неточности исследования при C50, разница в 1,755 цикла примерно эквивалентна 21,76 = 3,41-кратной разнице в концентрации, поэтому четырехкратная разница в разведениях должна давать результаты, отличающиеся чуть более чем на *SD* исследования. В таблице Л.2 приведены средние или медианные результаты каждого разведения. Эти результаты являются средними или медианными в зависимости от того, есть ли данные за пределами предела количественного определения.

Таблица Л.2 − Средние или медианные результаты

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Разбавле-ние | Повтор 1 | Повтор 2 | Повтор 3 | Повтор 4 | Повтор 5 | Среднее или медианное | *SD* |
| 1:2,75×106 | 37,5 | 35,5 | 38,1 | 36,2 | 38.5 | 37,16 | 1,272 |
| 1:11×106 | 37,8 | 38,2 | 38,9 | 0 (> 39) | 0 (> 39) | 38,9 | Неприме-нимо |
| 1:44×106 | 0 (> 39) | 0 (> 39) | 0 (> 39) | 38,7 | 0 (> 39) | 0 (> 39) | Неприме-нимо |

Второй результат, в котором три из пяти реплик являются положительными при разведении 1:11×106, соответствует результатам, которые могут быть получены из истинного образца C50. Из-за ограниченного количества повторов, от одного положительного результата до четырех положительных результатов также соответствуют образцу C50. Для повышения точности и достоверности оценки C50 необходимо дополнительное тестирование. Поэтому образцы C50, C5 и C95 формируют следующим образом:

- разведение C50 = 1:11×106 микробиологических культур в разбавителе для предварительной обработки;

- разведение C5 = 1:22.91×11×106 = 1:83×106 микробиологическая культура в разбавителе для предварительной обработки;

- разведение C95 = 1:11×106/22.91= 1:1.46×106 микробиологическая культура в разбавителе для предварительной обработки

Примечание − Эти результаты, основанные на биномиальном распределении, соответствуют заявлениям производителя о точности. Таким образом, были проверены заявления о точности.

В таблице Л.3 представлены результаты выборочных тестов, представленные в виде значений C (>39 − положительный результат) или нуля (отрицательный результат). Ожидаемый результат для образца C95 − 35,5 *CT*, поскольку это целевой *CT* для 95 % положительных результатов. Для образца C50 указано нулевое значение, поскольку 50 % количества циклов   
≥ 39. Ожидаемое среднее значение C5 также указано как нулевое, поскольку примерно 95 % результатов равны нулю.

В таблице Л.4 представлены доверительные интервалы (ДИ) вокруг «истинных» положительных пропорций.

Таблица Л.3 − Результаты выборочных испытаний

|  |  |  |  |
| --- | --- | --- | --- |
| Образец | C95 | C50 | C5 |
| Ожидаемое среднее | 35,5 | 0 | 0 |
| 1 | 36,2 | 37,5 | 0 (отрицательный) |
| 2 | 31,8 | 0 (отрицательный) | 0 (отрицательный) |
| 3 | 35,7 | 37,0 | 0 (отрицательный) |
| 4 | 34,0 | 37,9 | 0 (отрицательный) |
| 5 | 36,6 | 38,2 | 0 (отрицательный) |
| 6 | 34,9 | 36,9 | 0 (отрицательный) |
| 7 | 0 (отрицательный) | 38,3 | 0 (отрицательный) |
| 8 | 35,1 | 0 (отрицательный) | 0 (отрицательный) |
| 9 | 0 (отрицательный) | 38,8 | 0 (отрицательный) |
| 10 | 32,3 | 0 (отрицательный) | 0 (отрицательный) |
| 11 | 34,2 | 0 (отрицательный) | 0 (отрицательный) |
| 12 | 34,6 | 38,5 | 0 (отрицательный) |
| 13 | 36,4 | 36,2 | 0 (отрицательный) |

*Окончание таблицы Л.3*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Образец | | C95 | | C50 | | C5 | |
| 14 | | 34,9 | | 37,7 | | 0 (отрицательный) | |
| 15 | | 34,6 | | 37,8 | | 0 (отрицательный) | |
| 16 | | 36,8 | | 0 (отрицательный) | | 0 (отрицательный) | |
| 17 | | 34,5 | | 0 (отрицательный) | | 0 (отрицательный) | |
| 18 | | 38,8 | | 0 (отрицательный) | | 0 (отрицательный) | |
| 19 | | 37,8 | | 38,4 | | 0 (отрицательный) | |
| 20 | | 37,9 | | 0 (отрицательный) | | 0 (отрицательный) | |
| Количество положительных результатов | | 18 | | 12 | | 0 | |
| Количество отрицательных результатов | | 2 | | 8 | | 20 | |

В таблице Л.4 приведено ожидаемое количество положительных результатов при истинной исходной концентрации и количестве всех исследований (*N*). Нижние и верхние значения представляют собой 95 % ДИ, или пределы положительных результатов, которые пользователь должен ожидать в 95 % случаев проведения одинакового количества исследований образца. Эти ДИ для числа положительные результаты рассчитывают с использованием оценки Уилсона и округления до ближайшего целого числа для нижней границы и округления до ближайшего целого числа для верхней границы (см. приложение И, где приведены примеры расчета оценки ДИ).

Таблица Л.4 − ДИ вокруг ожидаемого числа положительных результатов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Истинная доля выборки | 95 %-ные ДИ, интервал ожидаемых результатов | *N* | | |
| 20 | 30 | 40 |
| C5 | Наименьшее количество положительных результатов | 0 | 0 | 0 |
| Наибольшее количество положительных результатов | 5 | 6 | 7 |
| C25 | Наименьшее количество положительных результатов | 2 | 3 | 5 |
| Наибольшее количество положительных результатов | 10 | 13 | 17 |
| C50 | Наименьшее количество положительных результатов | 5 | 9 | 14 |
| Наибольшее количество положительных результатов | 15 | 21 | 26 |
| C75 | Наименьшее количество положительных результатов | 10 | 17 | 23 |
| Наибольшее количество положительных результатов | 18 | 27 | 35 |
| C95 | Наименьшее количество положительных результатов | 15 | 24 | 33 |
| Наибольшее количество положительных результатов | 20 | 30 | 40 |
| Примечание – Обозначение: *N* − количество всех экспертиз. | | | | |

**Библиография**

|  |  |  |  |
| --- | --- | --- | --- |
| [1] | CLSI. A Framework for Using CLSI Documents to Evaluate Medical Laboratory Measurement Procedures. 3-е изд. Отчет CLSI EP19. Институт клинических и лабораторных стандартов; 2022. | | |
| [2] | CLSI. Испытания на интерференцию в клинической химии. 3-е изд. CLSI guideline EP07. Институт клинических и лабораторных стандартов; 2018. | | |
| [3] | CLSI. Оценка стабильности реагентов для экстракорпоральной диагностики; утвержденное руководство. Документ CLSI EP25-A. Институт клинических и лабораторных стандартов; 2009. | | |
| [4] | Miller JM, Astles R, Baszler T, et al. Руководство по безопасным методам работы в медицинских диагностических лабораториях для людей и животных. Рекомендации группы экспертов по биобезопасности, созванной CDC. MMWR Suppl. 2012;61(1):1-101 | | |
| [5] | CLSI. Защита работников лабораторий от профессионально приобретенных инфекций; утвержденное руководство - четвертое издание. Документ CLSI M29-A4. Институт клинических и лабораторных стандартов; 2014. | | |
| [6] | ISO 17511:2020 | «In vitro diagnostic medical devices — Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples» («Медицинские изделия для диагностики in vitro. Требования к установлению метрологической прослеживаемости значений, присваиваемых калибраторам, материалам контроля и образцам биологического материала человека») | |
| [7] | ISO 18153:2003 | «In vitro diagnostic medical devices — Measurement of quantities in biological samples — Metrological traceability of values for catalytic concentration of enzymes assigned calibrators and control materials» («Оборудование медицинское для диагностики in vitro. Количественные измерения в биологических образцах. Метрологическая прослеживаемость величин каталитической концентрации ферментов, заданных для калибраторов и контрольных материалов») | |
| [8] | Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: обновленный список основных пунктов для отчетности по исследованиям диагностической точности. Clin Chem. 2015;61(12):1446-1452. doi:10.1373/clinchem.2015.246280 | | |
| [9] | Bossuyt PM, Reitsma JB, Bruns DE, et al. К полной и точной отчетности об исследованиях диагностической точности: инициатива STARD. Ann Intern Med. 2003;138(1):40-44. doi:10.7326/0003-4819-138-1-200301070- 00010 | | |
| [10] | Международное бюро мер и весов (BIPM). Международный словарь по метрологии - Основные и общие понятия и связанные с ними термины (VIM, 3-е издание, JCGM 200:2012). Accessed 18 January 2023. https://www.bipm.org/en/publications/guides/. | | |
| [11] | ISO 15189:2012 | | «Medical laboratories — Requirements for quality and competence» («Медицинские лаборатории. Специальные требования к качеству и компетентности») |
| [12] | ISO 11843-1:1997 | | «Capability of detection − Part 1: Terms and definitions» («Способность обнаружения. Часть 1. Термины и определения») |
| [13] | ISO 15193:2009 | | «In vitro diagnostic medical devices − Measurement of quantities in samples of biological origin − Requirements for content and presentation of reference measurement procedures» («Медицинские изделия для in vitro диагностики. Измерение величин в образцах биологического происхождения. Требования к содержанию и представлению референтных процедур измерения») |
| [14] | ISO 9000:2015 | | «Quality management systems − Fundamentals and vocabulary» («Системы менеджмента качества. Основные положения и словарь») |
| [15] | **CLSI. Оценка диагностической точности лабораторных тестов с помощью кривых операционных характеристик приемника; утвержденное руководство - второе издание. Документ CLSI EP24-A2. Институт клинических и лабораторных стандартов; 2011** | | |
| [16] | CLSI. Оценка способности обнаружения для процедур клинических лабораторных измерений; утвержденное руководство - второе издание. Документ CLSI EP17-A2. Институт клинических и лабораторных стандартов; 2012 | | |
| [17] | CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ  CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014 | | |
| [18] | Box GEP, Cox DR. Анализ преобразований. J Roy Stat Soc B Met. 1964; 26(2):211-243. doi:10.1111/j.2517-6161.1964.tb00553.x | | |
| [19] | Whale AS, Devonshire AS, Karlin-Neumann G, et al. International interlaboratory digital PCR study demonstrating high reproducibility for the measurement of a rare sequence variant. Anal Chem. 2017;89(3):1724-1733.  doi:10.1021/acs.analchem.6b03980 | | |
| [20] | CLSI. Статистический контроль качества для процедур количественных измерений: Принципы и определения. 4-е изд. Руководство CLSI C24. Институт клинических и лабораторных стандартов; 2016 | | |
| [21] | CLSI. Иерархический подход к выбору суррогатных образцов для оценки медицинских лабораторных тестов in vitro; утвержденное руководство. 1-е изд. CLSI guideline EP39. Институт клинических и лабораторных стандартов; 2021 | | |
| [22] | Weusten JJ. Статистический подход к оценке аналитической чувствительности диагностических анализов при наличии ложноположительных результатов. J Virol Methods. 2008;151(2):308-310. doi:10.1016/j.jviromet.2008.05.013 | | |
| [23] | CLSI. Оценка эквивалентности или пригодности типов проб для медицинских лабораторных измерительных процедур. 1-е изд. Руководство CLSI EP35. Институт клинических и лабораторных стандартов; 2019 | | |
| [24] | CLSI. Валидация и верификация мультиплексных анализов нуклеиновых кислот. 2-е изд. Руководство CLSI MM17. Институт клинических и лабораторных стандартов; 2018 | | |
| [25] | CLSI. Микрочипы для диагностики и мониторинга инфекционных заболеваний; утвержденное руководство. Документ CLSI MM22-A. Институт клинических и лабораторных стандартов; 2014 | | |
| [26] | Lijmer JG, Mol BW, Heisterkamp S, et al. Эмпирические доказательства предвзятости, связанной с дизайном, в исследованиях диагностических тестов.  JAMA. 1999;282(11):1061-1066. doi:10.1001/jama.282.11.1061 | | |
| [27] | Ransohoff DF, Feinstein AR. Проблемы спектра и предвзятости при оценке эффективности диагностических тестов. N Engl J Med. 1978;299(17):926-930. doi:10.1056/nejm197810262991705 | | |
| [28] | Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Спектральная предвзятость при оценке диагностических тестов: уроки экспресс-теста на инфекции мочевыводящих путей. Ann Intern Med. 1992;117(2):135-140. doi:10.7326/0003-4819-117-2-135 | | |
| [29] | Medeiros FA, Ng D, Zangwill LM, Sample PA, Bowd C, Weinreb RN. Влияние дизайна исследования и смещения спектра на оценку диагностической точности конфокальной сканирующей лазерной офтальмоскопии при глаукоме. Invest Ophthalmol Vis Sci. 2007;48(1):214-222. doi:10.1167/iovs.06-0618 | | |
| [30] | Уиллис БХ. Предвзятость спектра - почему клиницисты должны быть осторожны при применении диагностических тестов. Fam Pract. 2008; 25(5):390-396. doi:10.1093/fampra/cmn051 | | |
| [31] | Zhou XH, Obuchowski NA, McClish DK. Статистические методы в диагностической медицине. 2nd ed. John Wiley & Sons Inc; 2011 | | |
| [32] | Биггерстафф Б. Дж. Сравнение диагностических тестов: простой график с использованием коэффициентов правдоподобия. Stat Med. 2000;19(5):649-663. doi:10.1002/(sici)1097-0258(20000315)19:5<649::aid-sim371>3.0.co;2-h | | |
| [33] | Альтман Д., Мачин Д., Брайант Т., Гарднер М. Статистика с доверием: Доверительные интервалы и статистические рекомендации. 2-е изд. BMJ Books; 2000 | | |
| [34] | Гарт Дж. Дж., Нам Дж. Приближенное интервальное оценивание отношения биномиальных параметров: обзор и поправки на перекос. Biometrics. 1988;44(2):323-338. doi:10.2307/2531848 | | |
| [35] | Begg CB. Предвзятость в оценке диагностических тестов. Stat Med. 1987;6(4):411-423. doi:10.1002/sim.4780060402 | | |
| [36] | Пепе МС. Статистическая оценка медицинских тестов для классификации и прогнозирования. Oxford University Press; 2004 | | |
| [37] | Meijer BC, Thijs JC, Kleibeuker JH, van Zwet AA, Berrelkamp RJ. Оценка восьми иммуноферментных анализов для выявления иммуноглобулина G против Helicobacter pylori. J Clin Microbiol. 1997;35(1):292-294. doi:10.1128/jcm.35.1.292- 294.1997 | | |
| [38] | Agresti A, Coull BA. Приблизительное лучше точного для интервального оценивания биномиальных пропорций. Am Stat.1998; 52(2):119-126. doi:10.1080/00031305.1998.10480550 | | |
| [39] | Армитидж П., Берри Г., Мэтьюс Дж. Н. С. Статистические методы в медицинских исследованиях. 4th ed. Wiley-Blackwell; 2001 | | |
| [40] | Zwillinger D, Kokoska S. CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall/CRC; 2000 | | |
| [41] | Pradhan V, Evans JC, Banerjee T. Binomial confidence intervals for testing non-inferiority or superiority: a practitioner's dilemma. Stat Methods Med Res. 2016;25(4):1707-1717. doi:10.1177/0962280213498324 | | |
| [42] | Уилсон Э. Б. Вероятное умозаключение, закон наследования и статистическое умозаключение. J Am Stat Assoc. 1927;22(158):209-212. doi:10.1080/01621459.1927.10502953 | | |
| [43] | CLSI. Дополнительные таблицы для тестирования интерференции в клинической химии. 1-е изд. CLSI supplement EP37. Институт клинических и лабораторных стандартов; 2018 | | |
| [44] | Сокал Р., Рольф Ф. Биометрия: Принципы и практика статистики в биологических исследованиях. 3-е изд. WH Freeman and Company; 1995 | | |
| [45] | CLSI. Проверка пользователем точности и оценка погрешности; утвержденное руководство - третье издание. Документ CLSI EP15-A3. Институт клинических и лабораторных стандартов; 2014 | | |
| [46] | CLSI. Оценка пользователем приемлемости смены партии реагентов. 2-е изд. Руководство CLSI EP26. Институт клинических и лабораторных стандартов; 2022 | | |
| [47] | Диксон В.Дж., Масси Ф.Дж. Введениевстатистическийанализ. 4th ed. McGraw-Hill; 1983 | | |
| [48] | Vaks JE. Новый метод оценки предела обнаружения в молекулярной диагностике. Proceedings of Joint Statistical Meetings. 2018:529-543 | | |
| [49] | Vaks JE. Вероятность обнаружения последовательности целевой нуклеиновой кислоты с помощью ПЦР-анализа в молекулярной диагностике: вывод математической модели, проверка и применение. Proceedings of Joint Statistical Meetings. 2017:2460-2476 | | |

УДК 57.065:006.354 ОКС 11.100.10

Ключевые слова: изделия медицинские для диагностики *in vitro*, качественные методы

исследования, бинарные результаты, клиническая чувствительность, клиническая специфичность, отношение правдоподобия, прогностическое значение, помехи, точность

1. ) В рамках настоящего стандарта оценка эффективности количественных исследований должна основываться на рекомендациях, приведенных в CLSI EP19. [↑](#footnote-ref-1)
2. ) В данной технологической карте используются четыре основных символа: овал (обозначает начало или конец процесса), стрелка (соединяет действия процесса), рамка (обозначает действия процесса), ромб (включает вопрос с альтернативными ответами «Да» и «Нет»). [↑](#footnote-ref-2)
3. ) Дополнительная информация приведена в CLSI. A Framework for Using CLSI Documents to Evaluate Medical Laboratory Measurement Procedures. 3-е изд. Отчет CLSI EP19. Институт клинических и лабораторных стандартов; 2022. [↑](#footnote-ref-3)
4. ) Дополнительная информация приведена в McCullagh P, Nelder JA. Обобщенные линейные модели. 2nd ed. Chapman & Hall/CRC; 1989. [↑](#footnote-ref-4)
5. ) Дополнительная информация приведена в Агрести А. Анализ категориальных данных. 3rd ed. John Wiley & Sons, Inc.; 2013. [↑](#footnote-ref-5)
6. ) Создание серии образцов с шипами с использованием стандартного материала осуществляют по CLSI. Установление и проверка увеличенного интервала измерения с помощью разбавления образца и введения шипов. 1-е изд. CLSI guideline EP34. Институт клинических и лабораторных стандартов; 2018. [↑](#footnote-ref-6)
7. ) Набор данных, использованный для этого примера, приведен в EP12-Ed3-DS, доступном по адресу https://clsi.org/ep12ed3-ds. [↑](#footnote-ref-7)
8. ) Метод частоты попаданий описан в CLSI. Оценка способности обнаружения для процедур клинических лабораторных измерений; утвержденное руководство - второе издание. Документ CLSI EP17-A2. Институт клинических и лабораторных стандартов; 2012. [↑](#footnote-ref-8)
9. ) Более подробная информация приведена в Снедекор Г. У., Кокран Р. Г. Статистические методы. 6-е изд. Издательство Университета штата Айова; 1967. [↑](#footnote-ref-9)
10. ) Информация о том, как характеристики исследований могут быть использованы для разработки исследований неточности, приведена в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-10)
11. ) Концепция и важность *DF* как естественного показателя неопределенности в оценках точности описаны в в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-11)
12. ) Дополнительная информация приведена в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-12)
13. ) Дополнительная информация приведена в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-13)
14. ) Рекомендации приведены в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-14)
15. ) Оценка компонентов дисперсии описана в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство - третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов; 2014. [↑](#footnote-ref-15)
16. )Дополнительная информация приведена в Dice LR. Меры величины экологической ассоциации между видами. Ecology. 1945;26(3):297-302. [↑](#footnote-ref-16)
17. ) Дополнительная информация приведена в Zou KH, Warfield SK, Bharatha A, et al. Статистическая проверка качества сегментации изображений на основе индекса пространственного перекрытия: научные доклады. Acad Radiol. 2004;11(2):178-189. [↑](#footnote-ref-17)
18. ) Дополнительная информация приведена в Fleiss JL. Измерение согласия между двумя судьями относительно наличия или отсутствия признака. Биометрия. 1975;31(3):651-659. doi:10.2307/2529549 [↑](#footnote-ref-18)
19. ) Дополнительная информация приведена в Cicchetti DV, Feinstein AR. Высокое согласие, но низкая каппа: II. Разрешение парадоксов. J Clin Epidemiol. 1990;43(6):551-558. doi:10.1016/0895-4356(90)90159-M [↑](#footnote-ref-19)
20. ) Дополнительная информация приведена в Samsa GP. Выборочные распределения p(pos) и p(neg). J Clin Epidemiol. 1996;49(8):917-919. doi:10.1016/0895-4356(96)00042- x [↑](#footnote-ref-20)
21. ) Дополнительная информация приведена в Graham P., Bull B.. Приблизительные стандартные ошибки и доверительные интервалы для индексов положительного и отрицательного согласия. J Clin Epidemiol. 1998;51(9):763-771. doi:10.1016/S0895-4356(98)00048-1. [↑](#footnote-ref-21)
22. ) Дополнительная информация приведена в Kottner J, Audigé L, Brorson S, et al. Предложено руководство по составлению отчетов об исследованиях надежности и согласия (GRRAS). Int J Nurs Stud. 2011;48(6):661-671. doi:10.1016/j.ijnurstu.2011.01.016. [↑](#footnote-ref-22)
23. ) Дополнительная информация приведена в Meijer BC, Thijs JC, Kleibeuker JH, van Zwet AA, Berrelkamp RJ. Оценка восьми иммуноферментных анализов для выявления иммуноглобина *G* против *Helicobacter pylori*. J Clin Microbiol. 1997;35(1):292-294. doi:10.1128/ jcm.35.1.292-294.1997. [↑](#footnote-ref-23)
24. ) *Z*-стандарт должен быть подробно описан как наилучшая доступная оценка ЦС. [↑](#footnote-ref-24)
25. ) Дополнительная информация приведена в Meijer BC, Thijs JC, Kleibeuker JH, van Zwet AA, Berrelkamp RJ. Оценка восьми иммуноферментных анализов для выявления иммуноглобина *G* против *Helicobacter pylori*. J Clin Microbiol. 1997;35(1):292-294. doi:10.1128/ jcm.35.1.292-294.1997. [↑](#footnote-ref-25)
26. ) Дополнительная информация приведена в Schrier WH, Schoengold RJ, Baker JT, et al. Разработка FlexSure HP − иммунохроматографического метода для выявления антител против Helicobacter pylori. Clin Chem. 1998;44(2):293-298. doi:10.1093/clinchem/44.2.293. [↑](#footnote-ref-26)