ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 72613— 2026

ИССЛЕДОВАНИЯ ЛАБОРАТОРНЫЕ КЛИНИЧЕСКИЕ

Оценка функциональных характеристик качественных лабораторных исследований с бинарным результатом

Издание официальное

Москва
Российский институт стандартизации
2026

gost.e-ecolog.ru

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 9 апреля 2026 г. № 336-ст

4 Настоящий стандарт разработан с учетом основных нормативных положений документа CLSI EP12 «Оценка результатов качественного исследования с использованием бинарных результатов» («Evaluation of Qualitative, Binary Output Examination Performance», NEQ)

5 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 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 (*gost.e-ecolog.ru

© Оформление. ФГБУ «Институт стандартизации», 2026

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 1
2 Нормативные ссылки 1
3 Термины, определения и сокращения 2
4 Основные положения 5
5 Требования безопасности 6
6 Качественные исследования. Общие положения 6
7 Разработка качественного исследования с бинарным ответом 8
8 Валидация качественного исследования с бинарным ответом 22
9 Верификация заявленных характеристик в лаборатории 39
Приложение А (справочное) Примеры исследований 46
Приложение Б (справочное) Связь между распределением внутреннего непрерывного сигнала и пробит- или логит-моделями 48
Приложение В (справочное) Качественное бинарное исследование с визуальной оценкой ответа . . . 51
Приложение Г (справочное) Оценка точности секвенирования следующего поколения 54
Приложение Д (справочное) Определение нижнего предела обнаружения для качественных исследований, основанных на методах полимеразной цепной реакции 62
Приложение Е (справочное) Схемы прецизионных исследований 64
Приложение Ж (справочное) Исследования точности наблюдателей 68
Приложение И (справочное) Метод оценки Уилсона для расчета доверительных интервалов 71
Приложение К (справочное) Примеры анализа клинической эффективности 75
Приложение Л (справочное) Пример полимеразной цепной реакции в реальном времени для ванкомицин-резистентных энтерококков 82
Библиография 85

Введение

Результаты клинико-диагностических лабораторных исследований показывают наличие или характеристики клинического свойства. По типу данных исследования бывают качественные (с получением номинальных или порядковых результатов) и количественные (с получением дискретных или непрерывных результатов). Качественные данные описывают группы или категории, а количественные — числовые измерения. В настоящем стандарте рассматривают качественные исследования, результаты которых представляют собой двоичные (бинарные, например «да»/«нет», «положительный»/«отрицательный») номинальные результаты (примеры приведены в приложении А). Данные исследования охватывают широкий спектр деятельности медицинских лабораторий, медицинских целей, методов измерения и типов представления результатов. Бинарный ответ может быть получен на основе:

- внутреннего непрерывного сигнала устройства и порогового значения для получения двоичных результатов;
- алгоритмических методов принятия решений, определяющих наличие аналита;
- в некоторых случаях результат «да» или «нет» может быть получен при визуальной оценке без использования приборов.

Эффективность клинико-диагностических лабораторных исследований должна быть оценена во время и после их разработки и подтверждена до того, как результаты исследования будут использованы для принятия клинических решений. Настоящий стандарт рассматривает разработку качественных бинарных исследований и способствует единообразию в оценке эффективности:

а) для разработчиков качественных бинарных исследований:
   1) при разработке исследований;
   2) при валидации исследований на основании требований, установленных при разработке;

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

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

Настоящий стандарт предназначен для различных категорий качественных исследований с бинарным ответом. Каждая категория характеризуется тем, определяет ли она наличие целевого состояния, присутствие аналита, или образцы имеют внутренний непрерывный сигнал. Оценка эффективности качественных методов включает исследование неточности, которая сосредоточена на оценке С5 и С95, и классификацию с определением клинической чувствительности и специфичности или положительной и отрицательной процентной согласованности. Среди других параметров эффективности — стабильность по ГОСТ Р ИСО 23640 и интерференция по ГОСТ Р 53079.4. Данный подход распространяется на весь процесс применения медицинского изделия, начиная с определения типа исследования, установления характеристик, валидации утверждений о изделии и, наконец, протоколов верификации в лаборатории.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ИССЛЕДОВАНИЯ ЛАБОРАТОРНЫЕ КЛИНИЧЕСКИЕ

Оценка функциональных характеристик качественных лабораторных исследований с бинарным результатом

Laboratory clinical research.
Evaluation of the functional characteristics of qualitative laboratory tests with a binary result

Дата введения — 2026—09—01

1 Область применения

Настоящий стандарт определяет требования к организации процесса установления, валидации и верификации прецизионности и оценки клинической эффективности качественных методов лабораторных исследований с бинарным результатом на всех этапах жизненного цикла медицинских изделий для диагностики ин витро (МИ ИВД), используемые протоколы для оценки эффективности при проектировании продукции и на этапах создания и внедрения в производство. Настоящий стандарт предназначен для качественных исследований, имеющих только два возможных результата целевого состояния (ЦС), например, «положительный»/«отрицательный», «присутствует»/«отсутствует», «реактивный»/«нереактивный»).

Настоящий стандарт может быть применен как для изготовителей МИ ИВД для проведения качественных бинарных исследований с получением результатов (далее — качественные бинарные исследования), так и для медицинских лабораторий, имеющих разрешение на применение незарегистрированных МИ ИВД для указанных исследований (в обоих случаях именуемых изготовителями). Также настоящий стандарт предназначен пользователям медицинских лабораторий для верификации эффективности используемых в лаборатории качественных бинарных исследований.

Оценка эффективности качественных исследований, результаты которых предоставляются с более чем двумя возможными категориями в неупорядоченном (номинальном) наборе, в порядковых категориях, а также количественных исследований, не входит в сферу применения настоящего стандарта.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 53079.4 Технологии лабораторные клинические. Обеспечение качества клинических лабораторных исследований. Часть 4. Правила ведения преаналитического этапа

ГОСТ Р ЕН 13612 Оценка функциональных характеристик медицинских изделий для диагностики *in vitro*

ГОСТ Р ИСО 15190 Лаборатории медицинские. Требования безопасности

ГОСТ Р ИСО 17511 Изделия медицинские для диагностики *in vitro*. Требования к установлению метрологической прослеживаемости значений, приписанных калибраторам, контрольным материалам правильности и образцам биологического материала человека

ГОСТ Р ИСО 18113-1 Медицинские изделия для диагностики *in vitro*. Информация, предоставленная изготовителем (маркировка). Часть 1. Термины, определения и общие требования

ГОСТ Р ИСО 23640 Изделия медицинские для диагностики *in vitro*. Оценка стабильности реагентов для диагностики *in vitro*

Издание официальное

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины, определения и сокращения

3.1 Термины и определения

В настоящем стандарте применены термины по ГОСТ Р ИСО 18113-1, ГОСТ Р ИСО 17511, ГОСТ Р ЕН 13612, а также следующие термины с соответствующими определениями:

3.1.1 **аналитическая серия:** Совокупность измерений аналита, выполненная в интервале в одних и тех же условиях без перенастройки и калибровки аналитической системы, при которых характеристики аналитической системы остаются стабильными, и он не может быть длиннее периода времени, рекомендованного производителем.

Примечания

1 Между аналитическими сериями 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 Аналитическая серия — интервал времени или условия измерения, для которых неточность измерения является наименьшей. Таким образом, прецизионность измерений в пределах аналитической серии определяет сходимость.

3.1.2 **внутренний непрерывный сигнал;** ICR: Сигнал, получаемый от анализатора при исследовании образца.

Примечания

1 В некоторых качественных исследованиях этот сигнал сравнивают со значением cut-off для получения бинарного результата.

2 Внутренний непрерывный сигнал также можно назвать ответом.

3.1.3 **диагностическая специфичность (специфичность):** Способность исследования давать отрицательный результат для объектов из целевой популяции, не имеющих целевого состояния, на которое они исследуются.

Примечание — Это доля пациентов без ЦС, имеющих отрицательный результат теста, и часто выражается в процентах.

3.1.4 **диагностическая чувствительность (чувствительность):** Способность исследования дать положительный результат для объектов из предполагаемой популяции, имеющих целевое состояние, на которое они исследуются.

Примечание — Это доля объектов с ЦС, имеющих положительный результат исследования, и часто выражается в процентах.

3.1.5 **исследование:** Совокупность процессов, целью которых является определение значений или характеристик объекта.

Примечание — Лабораторные исследования, определяющие значения объекта, являются количественными исследованиями; исследования, определяющие характеристики объекта, являются качественными исследованиями.

3.1.6 **качественное бинарное исследование:** Исследование, дающее бинарный результат.

Примечания

1 Бинарные исследования могут иметь внутренний непрерывный сигнал, который имеет количественные значения. Однако не требуется характеризовать эти значения, за исключением близости к cut-off, используемому 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.

3.1.7 **номинальное исследование:** Исследование, которое характеризует целевую популяцию по нескольким возможным неупорядоченным категориям.

Примечание — Примером исследования номинального является определение групп крови.

3.1.8 **порядковое исследование:** Исследование, которое характеризует целевую популяцию по нескольким возможным упорядоченным категориям.

Примечание — Результаты исследования порядкового представлены как 0, +, ++, +++ и ++++.

3.1.9 **клиническая эффективность:** Способность при бинарном исследовании различать объекты исследования с целевым состоянием и без него.

Пример — Способность разделять беременных и небеременных женщин.

Примечания

1 Клиническую эффективность обычно оценивают с помощью таких параметров, как диагностическая чувствительность и диагностическая специфичность.

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

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

Примечание — Для качественного бинарного исследования неточность может быть охарактеризована диапазоном интервала от C5 до C95.

3.1.11 **нижний предел обнаружения;** LLoD: Минимальная концентрация аналита, которую можно надежно отличить от LoB и при которой возможно обнаружение.

Примечания

1 В количественных процедурах — наименьшая концентрация аналита, которая может быть стабильно обнаружена как превышающая предел бланка (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».

3 При качественном исследовании для обнаружения аналита LLoD определяют с помощью исследования частоты попаданий.

4 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.

3.1.12 **отрицательная процентная согласованность;** NPA: Доля отрицательных результатов сравнительного метода, при которых результат метода исследования является отрицательным, если сравнительный метод не является общепризнанным методом оценки целевого состояния.

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

3.1.14 **ошибка II типа:** Ложное принятие нулевой гипотезы, когда она не верна.

3.1.15 **повторное исследование (повтор):** Значение, полученное в результате повторного исследования того же образца или аналогичных объектов, если это необходимо.

3.1.16 **положительная процентная согласованность;** PPA: Доля положительных результатов сравнительного метода, при которых результат метода исследования является положительным, если сравнительный метод не является общепризнанным методом оценки целевого состояния.

3.1.17 **пороговое значение (cut-off):** Порог внутреннего непрерывного сигнала для бинарного исследования, выше которого результат считают положительным, а ниже — отрицательным.

Примечания

1 Конкурентные исследования ведут себя противоположным образом, результаты при меньших значениях сигнала, чем cut-off, считают положительными.

2 Cut-off устанавливают во время разработки исследования; клинические результаты не могут быть оценены до тех пор, пока не будет установлено cut-off.

3 Некоторые бинарные исследования могут иметь серую зону, для которой результат является сомнительным.

3.1.18 **предел бланка (холостого хода);** LoB: Максимальная наблюдаемая концентрация аналита, которую можно обнаружить при исследовании образцов, не содержащих аналит.

Примечание — Наивысший результат измерения, который может наблюдаться (с установленной вероятностью α) для холостого образца (бланка).

3.1.19 **процент попаданий:** Доля положительных результатов исследования в целевой популяции.

3.1.20 **серая зона:** Интервал внутреннего непрерывного ответа (ICR) вокруг cut-off, в пределах которого окончательный результат не может быть представлен как положительный или отрицательный, а определен как «сомнительный», с указанием или без указания фактического значения результата.

Примечания

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

2 Многие бинарные исследования включают внутренний контроль, который может указать, когда ICR ненадежен или результат не может быть представлен. Неопределенный результат отличается от сомнительного результата.

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

3.1.21 **смещение (исследования):** Оценка систематической ошибки исследования.

Примечание — Для качественного, бинарного исследования смещение обычно определяют на cut-off и выражают в терминах внутреннего непрерывного сигнала инструмента.

3.1.22 **согласованность:** Доля проб, для которых результаты, полученные с помощью разработанного изделия, и результаты, полученные с помощью изделия сравнения, совпадают.

Примечания

1 Согласованность также может быть установлена в пределах исследования с использованием одной и той же популяции пациентов, но с учетом двух факторов, например, таких как разные операторы, приборы или партии реагентов.

2 См. термины «отрицательная процентная согласованность» и «положительная процентная согласованность».

3.1.23 **показатель C5:** Значение по соответствующей шкале (например, внутреннего непрерывного сигнала), при котором при бинарном исследовании образец признает положительным в 5 % случаев.

Примечание — Как и прецизионность, значение C5 связано с установленными условиями прецизионности.

3.1.24 **показатель C95:** Значение по соответствующей шкале (например, внутреннего непрерывного сигнала), при котором при бинарном исследовании образец признается положительным в 95 % случаев.

Примечание — Как и прецизионность, значение C95 связано с установленными условиями прецизионности.

3.1.25 **целевое состояние;** ЦС: Заболевание, стадия заболевания, состояние здоровья или любое другое идентифицируемое состояние или характеристика, представляющая интерес для субъекта, например стадия 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 пациент, имеющий аналит (например, вирус).

3.2 Сокращения

В настоящем стандарте применены следующие сокращения:

ВИЧ — вирус иммунодефицита человека;

ВПЧ — вирус папилломы человека;

ДНК — дезоксирибонуклеиновая кислота;

ДС — диагностическая специфичность;

$ДС_{ПИ}$ — диагностическая специфичность ранее применявшегося изделия;

$ДС_{РИ}$ — диагностическая специфичность разработанного изделия;

ДЧ — диагностическая чувствительность;

$ДЧ_{ПИ}$ — диагностическая чувствительность ранее применявшегося изделия;

$ДЧ_{РИ}$ — диагностическая чувствительность разработанного изделия;

ИФА — иммуноферментный анализ;

КК — контроль качества;

ПЦР — полимеразная цепная реакция;

ЦС — целевое состояние;

ANOVA — дисперсионный анализ;

BDS — двунаправленное секвенирование;

CI — доверительный интервал;

CRISPR — кластеризованные регулярно перемежающиеся короткие палиндромные повторы;

CV — коэффициент вариации;

CX — значение по соответствующей шкале, при котором образец в ходе бинарного исследования признается положительным в X % случаев;

DF — степени свободы;

FN — ложноотрицательный результат;

FP — ложноположительный результат;

HMW-CAP EIA — иммуноферментный анализ высокомолекулярных клеточно-ассоциированных белков;

ICR — внутренний непрерывный сигнал;

IVD — диагностика in vitro;

LLoD — нижний предел обнаружения;

LoB — предел холостого хода;

LR — коэффициент правдоподобия;

NGS — секвенирование нового поколения;

NLR — отрицательное отношение правдоподобия;

NPA — отрицательная процентная согласованность;

NPV — отрицательное прогностическое значение;

OR — отношение шансов;

PLR — положительное отношение правдоподобия;

PPA — положительная процентная согласованность;

PPV — положительное прогностическое значение;

RFU — относительная единица флуоресценции;

РНК — рибонуклеиновая кислота;

SD — стандартное отклонение;

TN — истинно отрицательный;

TP — истинно положительный.

4 Основные положения

В лабораторной практике многие результаты исследования предоставляют в бинарном виде (например, «да» или «нет»). Медицинские работники могут назначить исследование с целью определения наличия заболевания у пациента, присутствия аналита в образце, беременность женщины или положительные результаты теста на наркотики. Для таких исследований используют две основные оценки: клиническая эффективность (чувствительность и специфичность) и прецизионность.

Оценку клинических показателей при создании и валидации процесса бинарной классификации проводят путем исследования образцов пациентов с известным целевым состоянием (ЦС). Это ис-

следование включает в себя сравнение бинарных результатов для предполагаемой популяции, полученных в ходе исследования, с независимой процедурой, которая обеспечивает наилучшую оценку ЦС. Однако во многих случаях такое сравнение бинарных результатов представляет собой процентную согласованность, полученную в процедуре сравнительных исследований, с исследованием, которое не является идеальным.

Определение прецизионности качественного исследования зависит от значений по соответствующей шкале, где при бинарном исследовании образец признается 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).

Стабильность реагентов и интерференция при исследовании — еще два показателя эффективности медицинских изделий, которые рассматриваются в 9.3 (стабильность) и 8.4 (интерференция).

5 Требования безопасности

Предметом клинических лабораторных исследований является биологический материал человека. Поскольку зачастую невозможно определить, какие образцы биологического материала или пробы могут быть инфицированными, то все образцы и пробы должны рассматриваться как потенциально инфицированные и работы должны проводиться с обеспечением биологической безопасности как в отношении сотрудников лаборатории, так и окружающей среды в соответствии с [1] и ГОСТ Р ИСО 15190.

6 Качественные исследования. Общие положения

6.1 Бинарные исследования обеспечивают разделение пациентов на группы с ЦС (ЦС присутствует) и группы без ЦС (ЦС отсутствует). Качественные исследования также используют для широкого спектра клинических целей, включая скрининг, установление диагноза и подтверждение результатов. Даже с учетом различий в методах и целях применения для качественных исследований существуют общие признаки, позволяющие оценить их диагностическую эффективность и прецизионность.

6.2 Классификация качественных исследований

Качественные бинарные исследования можно разделить на группы в зависимости от их целевого назначения и метода определения результата исследования.

Целевое назначение подразделяют на два вида:

- первый вид включает исследования, используемые для определения наличия у пациента, чей образец исследуют, определенного ЦС. Примеры целевых клинических состояний включают, например, беременность, предыдущую вакцинацию или инфицирование микроорганизмами, онкологические заболевания;
- второй вид исследования включает обнаружение в образце целевого аналита. Целью этих исследований является признание образцов, не содержащих целевого аналита, отрицательными, а образцов, содержащих целевой аналит, — положительными.

По методу определения результата исследования можно выделить следующие исследования:

а) в которых для всех образцов используется cut-off по единой шкале внутреннего непрерывного сигнала (ICR);

б) основанные на алгоритме принятия решения, для которого может использоваться несколько ICR и/или категориальные данные для получения бинарного результата, обычно применяют только для предполагаемого использования обнаружения аналитов, такие как:

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

2) исследования с несколькими ICR, для которых алгоритм использует решения «если/то»;

3) исследования, в которых используют многоступенчатые решения о категоризации, исключающие варианты до принятия окончательного решения;

4) исследования, которые визуально оцениваются пользователем без ICR, могут использоваться по назначению в обоих случаях.

6.3 Подкатегории качественных исследований с бинарным результатом

Качественные бинарные исследования можно разделить на следующие категории:

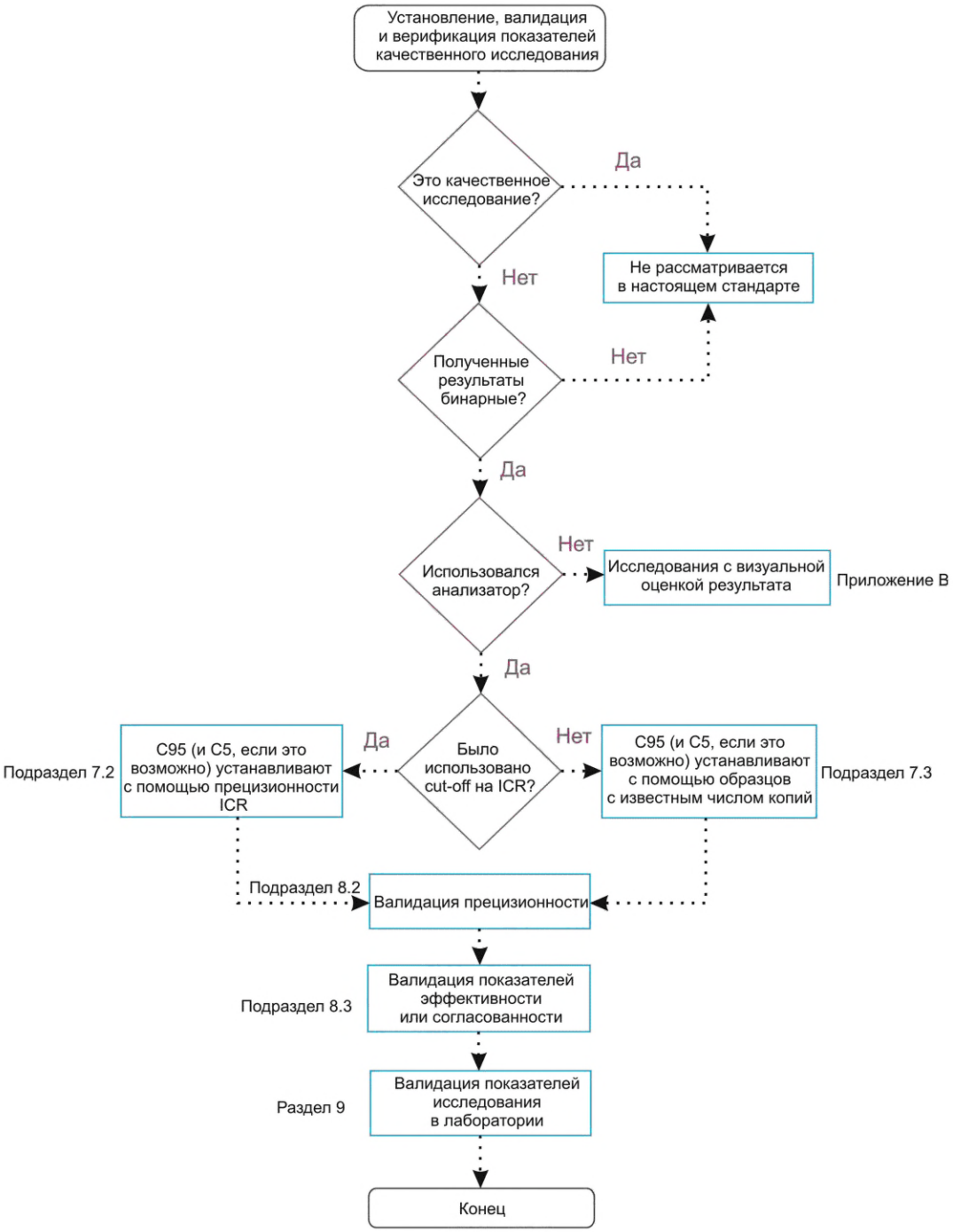
- исследования, в которых определяют наличие и отсутствие ЦС путем сравнения ICR с диагностическим значением cut-off. Для таких исследований cut-off по шкале непрерывных сигналов обычно выбирают для оптимизации диагностической чувствительности и специфичности исследования для целевой популяции с учетом компромисса между пользой и риском или соответствия заранее установленной точке принятия решения. Для таких исследований C5 и C95 могут быть измерены с помощью прецизионности ICR;

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

- исследования, основанные на процессе и алгоритме обнаружения уникальной последовательности нуклеиновых кислот целевого аналита. Методы амплификации нуклеиновых кислот (например, ПЦР, кластеризованные регулярно перемежающиеся короткие палиндромные повторы (CRISPR)) направлены на обнаружение только целевого аналита. Образец с отсутствием целевого аналита признают отрицательным почти в 100 % случаев (без FP), то есть исследование имеет 100 %-ную диагностическую специфичность. Диагностическая чувствительность может меняться в зависимости от процесса исследования. Поскольку ICR не существует, определяют наименьшую концентрацию целевых копий, которая обнаруживается в 95 % случаев в соответствии с пробит-фиттингом для данной частоты попаданий. Определение C5 по этой шкале может быть или не быть осуществимым (см. приложение Б);

- качественные бинарные исследования с визуальной оценкой (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.



C5 — значение по соответствующей шкале, при котором в бинарном исследовании образец признается положительным в 5 % случаев; C95 — значение по соответствующей шкале, при котором образец признается положительным в 95 % случаев; ICR — внутренний непрерывный сигнал

Рисунок 1 — Установление, валидация и верификация показателей качественного исследования

7 Разработка качественного исследования с бинарным ответом

7.1 Общие положения

Все качественные лабораторные исследования, выполняемые на анализаторах, генерируют ICR. В некоторых исследованиях непосредственно используют сигнал и cut-off для ответа «да» или «нет». В качественном исследовании, обеспечивающем ICR для всех образцов, для определения соотношения сигнала и cut-off используют расчетное соотношение сигнала от образца. Для определения значе-

ния cut-off в таких исследованиях используют клинические критерии. В 7.2 приведены этапы разработки качественных исследований.

Качественные исследования для обнаружения аналитов также могут быть разработаны таким образом, чтобы использовать cut-off по шкале значений ICR, доступных для всех образцов. Поскольку шум системы на шкале ICR оказывает влияние на нахождение оптимальной точки разграничения между присутствием и отсутствием аналита, оценку LoB на шкале ICR рекомендуется использовать при выборе значения cut-off. Эти виды исследований рассмотрены в 7.2.4.

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 единую шкалу ICR не используют для всех потенциальных образцов, основным показателем эффективности является C95, измеряемый в количестве копий на единицу объема, а не ICR. Подробные сведения об этапах разработки исследований приведены в 7.3.

7.2 Разработка качественных тестов с использованием сравнения с cut-off внутреннего непрерывного сигнала

Для многих типов качественных бинарных исследований ICR содержит значения значительно ниже и значительно выше cut-off. Отношение сигнала к cut-off, как функция концентрации аналита, должно демонстрировать монотонность вблизи cut-off. При исследовании, например, на антитела такую монотонность можно продемонстрировать с помощью серии взятия крови, которая показывает увеличение количества антител по мере того, как пациент становится иммунологически реактивным к вирусу. Однако нет необходимости демонстрировать линейность между отношением сигнала к cut-off и концентрацией антител в образце. Также данный подход применим для обнаружения аналитов, для которых система исследования обеспечивает значение ICR даже при отсутствии аналита. Если наблюдаемый результат ICR выше значения cut-off, пробу считают положительной, если ниже — отрицательной (см. 7.2.4).

7.2.1 Установление cut-off: ошибки классификации и вариабельность исследования

Качественные бинарные исследования используют для оценки наличия или отсутствия ЦС, и для получения бинарных результатов им требуется cut-off. Например, результат может быть положительным, если ICR больше или равен cut-off, и отрицательным, если ICR меньше cut-off. Обычно cut-off выбирают для оптимизации диагностической чувствительности и диагностической специфичности исследования. Они основаны на сборе клинических данных из популяции предполагаемого использования и учитывают соотношение пользы исследования и риска недостоверных результатов. Например, качественное разработанное изделие для скрининга при исследовании донорской крови на инфекционные заболевания, такие как ВИЧ или вирус гепатита В, cut-off должно быть подобрано таким образом, чтобы обеспечить высокую чувствительность, то есть вероятность того, что инфицированный донор будет исключен из системы обеспечения реципиентов компонентами крови. Та же измерительная система может использоваться для проведения качественного исследования с целью диагностики, при этом cut-off выбирают с учетом высокой специфичности, чтобы не применять инвазивные методы лечения у пациентов без ЦС. Cut-off может быть установлено таким образом, чтобы повторять заранее установленные точки принятия решения.

Для установления cut-off используют образцы из двух популяций (одна с наличием ЦС, другая — с отсутствием ЦС). Если распределение внутренних сигналов для популяций, которые должны различаться, сильно разнятся, cut-off можно установить в широком интервале, не влияя на результаты исследования, как показано на рисунке 2. Кроме того, присущая ICR вариабельность не является критичной, поскольку образцов со значениями сигналов вблизи cut-off мало или они отсутствуют.

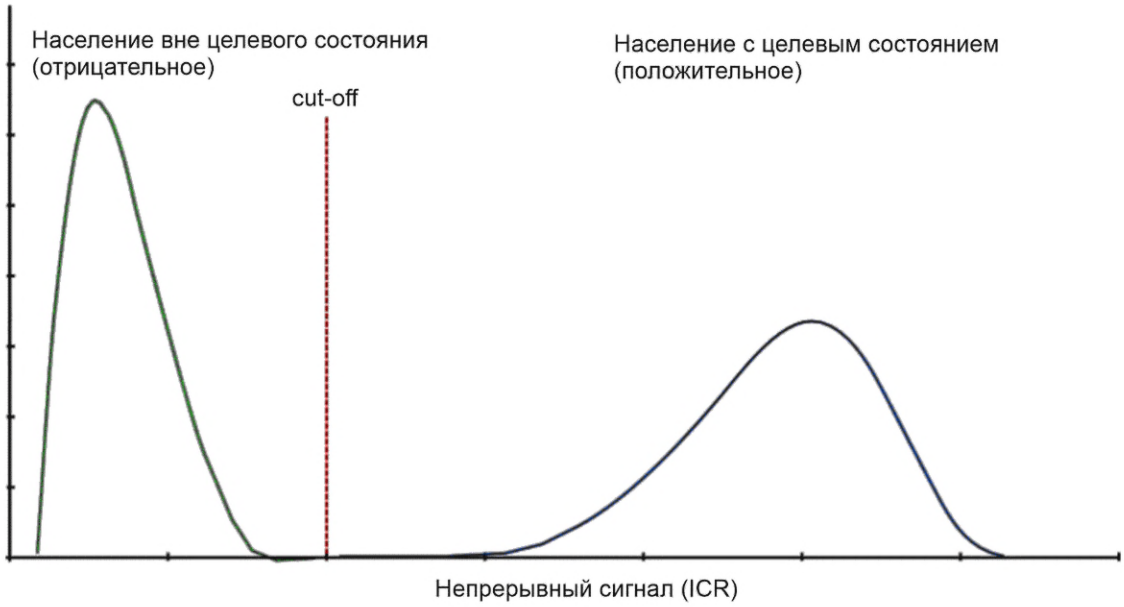


Рисунок 2 — Два клинических состояния с широко разнесенными распределениями

Если распределения ICR из двух популяций близки или перекрываются, как приведено на рисунке 3, выбор cut-off становится вопросом компромисса между частотой FP и FN и способностью исследования стабильно давать одинаковые результаты вблизи cut-off.



Рисунок 3 — Два клинических состояния с перекрывающимися распределениями

По мере сближения этих двух распределений и ограничения частоты ошибочной классификации, допускающей только минимальные показатели FP и FN, необходимо уменьшать прецизионность ICR. Оптимизация включает в себя процессы по изменению состава реагентов, пробоподготовки и производительности оборудования (например, точность объемов дозирования).

Чтобы оценить потребность в оптимизации процесса, необходимо мысленно допустить возможность исследовать всех индивидуумов из популяции предполагаемого использования. На рисунке 4 две сплошные синие линии представляют собой непрерывные распределения сигналов, соответствующие индивидуумам из популяции предполагаемого использования, которые делятся на две субпопуляции (т. е. одна с ЦС, а другая без ЦС).

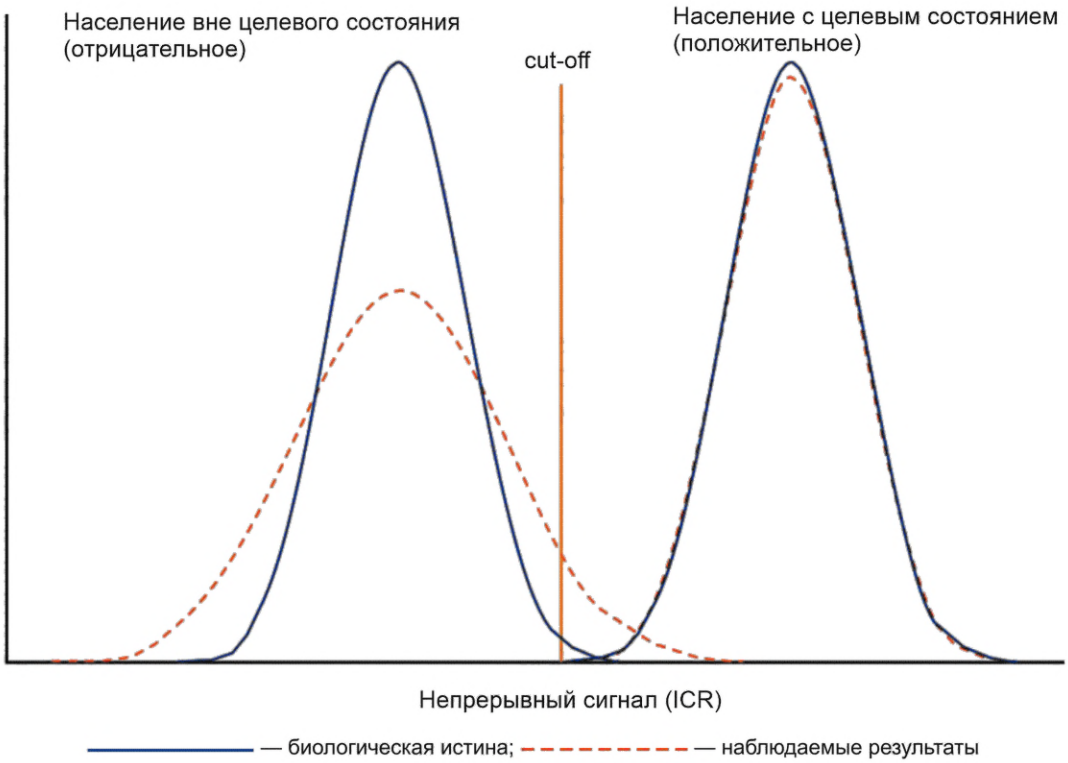


Рисунок 4 — Сравнение частоты ошибочной классификации при качественном исследовании в субпопуляциях с ЦС и без ЦС

В данном примере 5 %-ная ошибка классификации является приемлемой для субпопуляции без ЦС (допускается 5 % FP) и 0,04 % ошибочной классификации допустимо для субпопуляции с ЦС (допускается 0,04 % FN), при этом два непрерывных распределения вероятности сигнала (т. е. сплошные синие линии) пересекаются на приблизительно 1 %. Cut-off (т. е. сплошная желтая вертикальная линия) для исследования была ранее установлена таким образом, чтобы правильно классифицировать как можно больше индивидуумов с ЦС, но при этом допускала неправильную классификацию некоторых индивидуумов без ЦС. Чтобы удовлетворить требованию 0,04 % неправильной классификации для индивидуумов с ЦС (чувствительность около 100 %), учитывая биологическую вариабельность и установленное cut-off, вариабельность, присущая исследованию образцов с ЦС, должна быть очень низкой (пунктирная оранжевая линия представляет дисперсию как от неточности, так и от индивидуальных сигналов 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 того же cut-off, 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 вариабельность, так и смещение и прецизионность исследования (в данном примере смещение равно нулю).

Поскольку невозможно оценить выборку от каждого человека из популяции предполагаемого использования, можно сравнить общую вариабельность процедуры измерения вокруг cut-off с допустимой вариабельностью, которая была рассчитана. Для такого сравнения могут быть применены C5 и C95, которые описаны в 7.2.2. C5 и C95 отражают вариабельность или SD исследования в оговоренных условиях прецизионности (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 C5 — C95 для любого cut-off. 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 cut-off по шкале ICR на выборках, репрезентативных для предполагаемой популяции. Следует изучить диапазон cut-off и выбрать cut-off, удовлетворяющее требованиям диагностической эффективности (см. рисунок 3); этот период изучения часто называют «тренировкой». Затем, после фиксации cut-off, клинические характеристики исследования должны быть оценены в отдельном исследовании с другим, обычно большим набором образцов из популяции предполагаемого использования, как описано в 8.2.

Некоторые качественные исследования дают положительный, отрицательный или сомнительный результат. Такие исследования включают серую зону, в которой ICR действителен, но окончательный результат сообщается как сомнительный или неопределенный, а не как положительный или отрицательный. Например, если серая зона, основанная на ICR, обозначается [C1, C2], результаты могут быть отрицательными, если ICR менее C1, сомнительными, если ICR более или равен C1 и 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 серая зона, производитель, как правило, рекомендует провести повторное исследование. Рекомендация может включать взятие другого образца в более поздние сроки и повторное исследование с использованием той же или другой методики. Для неточности ICR ключевыми факторами при разработке теста являются размер серой зоны и частота сомнительных результатов, а также частота результатов FP и FN в популяции предполагаемого использования. Следует изучить преимущества наличия серой зоны определенного размера для предотвращения сообщения неверного результата, фиксирования сомнительных результатов, которые могут привести к нежелательным последствиям, или необходимости проведения дополнительного тестирования.

Другие качественные исследования дают бинарный результат и позволяют провести повторное исследование в зоне вокруг cut-off. Если первоначальный бинарный результат достоверен, но из-за неточности разработчик может рекомендовать 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 диагностическую эффективность исследования отражают с помощью окончательного бинарного результата, как описано в 8.2, и дополнительно указывают частоту повторных испытаний.

7.2.2 Взаимосвязь между cut-off и C5, C50 и C95

C5 и C95 — это показатели, рассчитываемые при проведении испытания, в котором используют cut-off, основанное на ICR. Образец со значением ICR по соответствующей шкале, где бинарное испытание признает образец положительным в 50 % случаев (C50), дает отрицательный результат в 50 % случаев, если множество повторных исследований этого образца выполняется в оговоренных условиях, как показано на рисунке 5.

Если непрерывные сигналы повторных исследований распределены симметрично (т. е. нормально) для образца на cut-off (уровне), то C50 является оценкой cut-off.

Значения C5 и C95, как показано на рисунке 6, связаны с cut-off на основе вариабельности реакции, присущей исследованию вблизи cut-off.

Если множество повторных исследований образца со значением сигнала C5 будет выполнено в оговоренных условиях, 5 % повторных исследований будут выше cut-off. Аналогично, 95 % повторных исследований образца со значением сигнала C95 будут иметь значение сигнала выше cut-off. Чем больше вариабельность сигнала в исследовании, тем дальше значения C5 и C95 от cut-off. Таким образом, ширина интервала от C5 до C95 является показателем вариабельности сигналов в ходе исследования. Узкий интервал свидетельствует о низкой вариабельности, а широкий — о высокой вариабельности сигналов.

На рисунке 6 приведено распределение результатов непрерывного сигнала для одного образца при C5 1,40 и для другого образца при C95 1,68, оба основаны на cut-off 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 сигнал следует использовать для оценки вариабельности исследования, поскольку это оптимальный способ оценки характеристики в районе порогового значения. Определение C5 и C95 — важные этапы в процессе разработки качественного исследования, поскольку, являясь мерой неточности, они могут определить необходимость дополнительных улучшений в конструкции и как следует проводить последующие исследования эффективности.

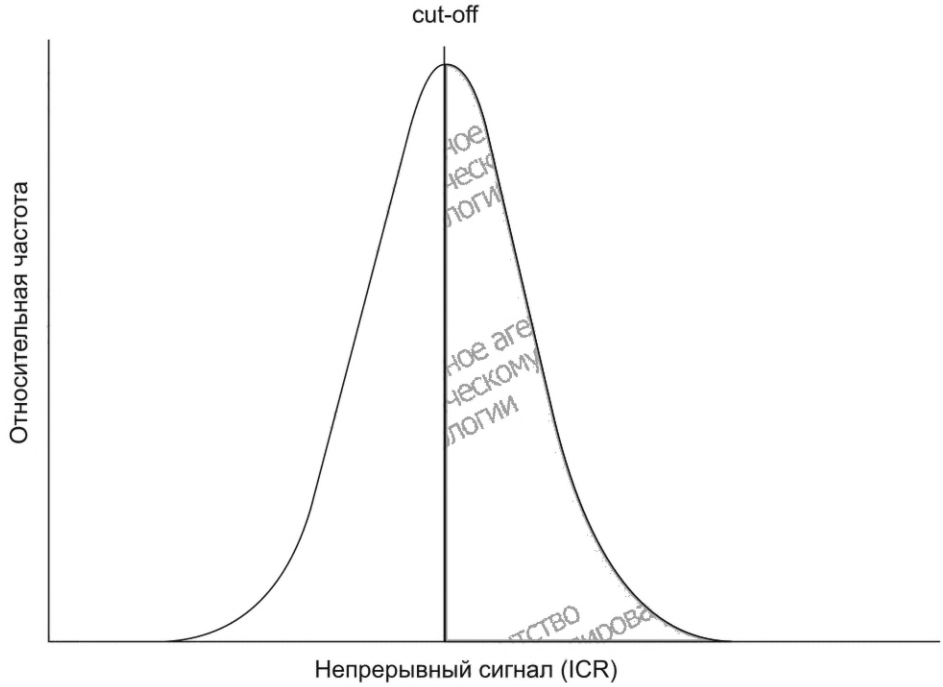
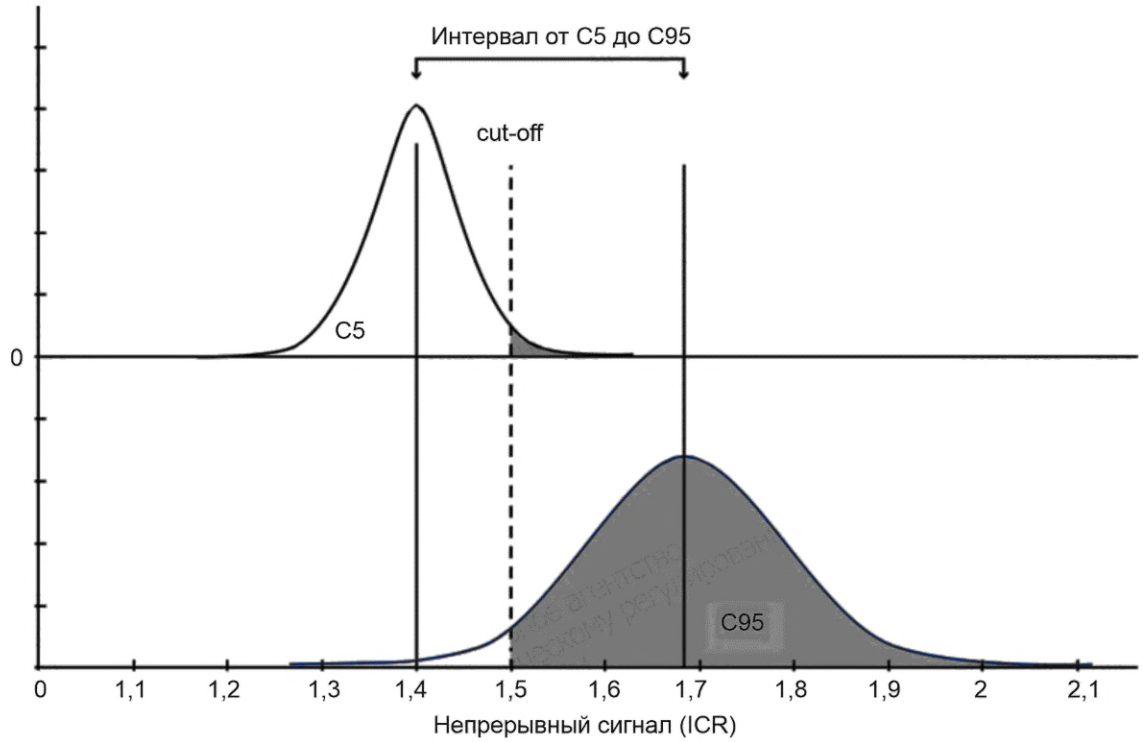


Рисунок 5 — Гипотетическое распределение повторных результатов для выборки с cut-off



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

Рисунок 6 — Гипотетические распределения повторных результатов для двух разных выборок

7.2.3 Установление C5 и C95 для качественных исследований с помощью внутреннего непрерывного сигнала

Если ICR доступен и охватывает широкий интервал вокруг cut-off, C5 и C95 могут быть получены из профиля прецизионности этого сигнала. 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, которое включает исследование сходимости (внутрисерийную воспроизводимость), межсерийную и внутрилабораторную воспроизводимости. В зависимости от их известного влияния на исследование, разработчику рекомендуется включить другие факторы изменчивости, которые могут влиять на оценку воспроизводимости исследования. Общую вариабельность или SD рассчитывают с помощью анализа компонентов дисперсии результатов сигналов от каждого образца.

Предполагаемые уровни C5 и C95 можно оценить, используя оценку SD примерно на уровне cut-off, умножив ее на два, а затем добавив и убавив это значение из значения cut-off. Без первоначальной оценки SD в качестве интервала, используемого для построения профиля прецизионности, можно использовать cut-off ± 20 % от порогового значения.

После получения общих SD и CV из внутрилабораторных данных следует составить профиль прецизионности, 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 сигнала, SD на C5 и C95 следует рассчитать на основе модели профиля прецизионности для данных нескольких образцов (например, четырех-пяти) вокруг cut-off. 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 прецизионности подробно приведено в приложении Г.

Если сходимость результатов или разница между компонентами дисперсии распределены нормально, внутрилабораторные C5 и C95 можно рассчитать непосредственно с помощью cut-off и оцененного общего внутрилабораторного SD на C5 и на C95. При нормальном распределении C95 и SD_{C95} — это среднее и SD распределения внутреннего сигнала, при котором его 5-й процентиль равен cut-off исследования, полученному при повторном тестировании выборки с соответствующими значениями внутреннего сигнала. Аналогично, C5 и SD_{C5} — это среднее и SD распределения внутреннего сигнала, при котором его 95-й процентиль равен cut-off, полученному при повторных исследованиях образца с соответствующим значением внутреннего сигнала. C5 и C95 могут быть вычислены по формулам (1) и (2) соответственно.

$$C5 = \text{cut-off} - 1{,}645 \cdot SD_{C5}, \quad (1)$$

$$C95 = \text{cut-off} + 1{,}645 \cdot SD_{C95}, \quad (2)$$

где 1,645 — 95-й процентиль стандартного нормального распределения ($Z_{0,95}$).

Если вариабельность в области вокруг cut-off имеет приблизительно постоянное SD, то SD_{C5} и SD_{C95} — одно и то же значение. Если образцы около cut-off имеют примерно одинаковый CV, SD_{C5} можно заменить на CV·C5, а SD_{C95} — на CV·C95. Тогда C5 и C95 допускается вычислять по формулам (3) и (4) соответственно:

$$C5 = \frac{\text{cut-off}}{1 + 1{,}645 \cdot CV}, \quad (3)$$

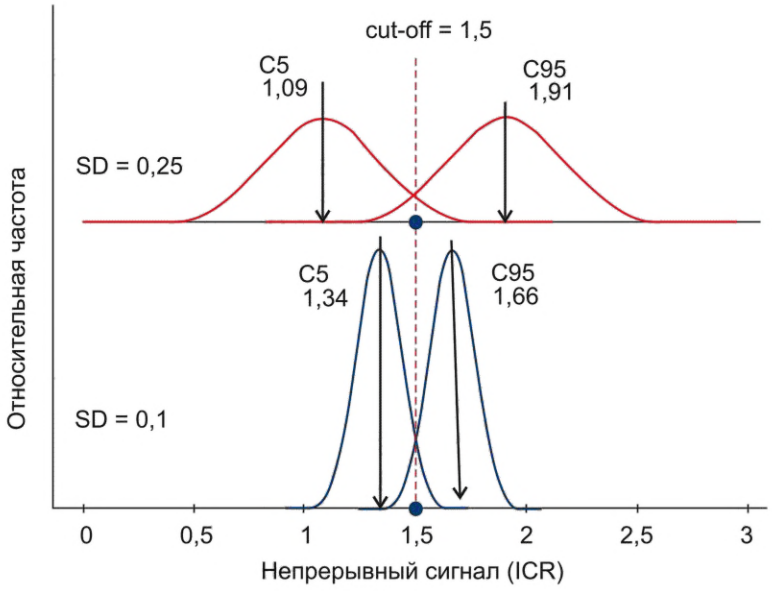
$$C95 = \frac{\text{cut-off}}{1 - 1{,}645 \cdot CV}. \quad (4)$$

Приведенные расчеты предполагают, что распределение результатов для значений ICR является нормальным. Если распределение значений не является нормальным, шкала ICR может быть преобразована. В таких случаях обычно используются различные логарифмические преобразования.

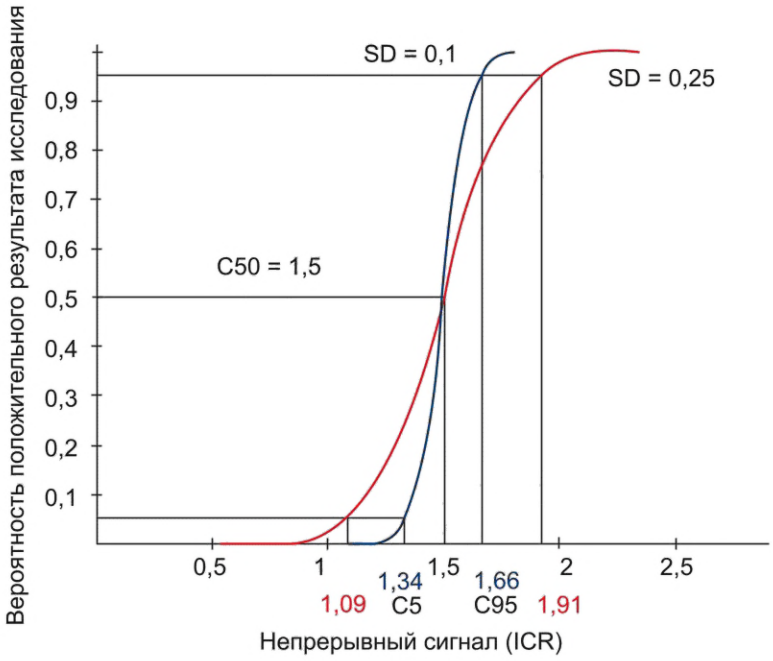
Для определения C5 и C95 допускается использовать другие типы оценок состояния неточности (SD или CV), используя формулы (1)—(4). В некоторых случаях (например, при использовании образцов цельной крови) нестабильность образца не позволяет измерить ежедневную прецизионность. В таких случаях допускается использовать исследования, измеряющие только сходимость, также известную как внутрисерийная прецизионность. Также возможно проведение исследований, охватывающих внутрисерийную и межсерийную прецизионность. C5 и C95 также могут быть определены с помощью

SD или CV, полученных в результате исследований прецизионности, включающих вариабельность между лабораториями и, возможно, другие факторы, такие как партии реагентов. Условия прецизионности, которые были использованы, должны быть указаны, когда сообщается о C5 и C95. Рекомендации по исследованию неточности для качественных исследований приведены в 8.1.

На рисунках 7а) и 7б) приведены эффекты изменения прецизионности между двумя различными вариантами разрабатываемых исследований (также см. приложение Б).



а)



б)

C5 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C50 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 50 % случаев; C95 — значение по 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

Версия 1, приведенная на верхнем графике рисунка 7а), имеет более широкое распределение (больше неточности) как на С5, так и на С95, чем версия 2, приведенная на нижнем графике. Версия 2 является более предпочтительной.

Версия 1, приведенная на рисунке 7б) (красная линия) с SD 0,25, имеет более широкий интервал от С5 до С95, чем версия 2 (синяя линия) с SD 0,1. Версия 2 является более предпочтительной.

Аналогичным образом варианты исследований или рабочие процессы пробоподготовки также должны быть оценены на предмет смещения, отражающегося в изменении cut-off. На этапе разработки исследования следует оптимизировать аспекты системы, влияющие на ее работу, чтобы уменьшить прецизионность. На рисунках 7а) и 7б) показано, как интервал от С5 до С95 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 прецизионности исследования. Взаимосвязь между вероятностью положительного результата и неточностью внутреннего сигнала 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 программа, используемая для второй версии, является предпочтительной. Кроме того, область С5—С95 первой версии шире, чем область С5—С95 второй версии, когда эти две версии имеют одинаковые cut-off, как показано на рисунке 7б). Если не принимать во внимание другие соображения, такие как скорость FP или FN, программа, используемая в версии 2, является предпочтительной.

Примечание — Оценка погрешностей различных качественных исследований путем сравнения интервалов С5 и С95 не имеет смысла, если нет возможности проследить ICR к стандартному образцу. Эта несопоставимость справедлива даже при одинаковом значении cut-off (например, соотношение сигнал/cut-off = 1,0).

7.2.4 Использование внутреннего непрерывного сигнала для исследований обнаружения аналитов

Если цель исследования — определить, присутствует ли в образце какой-либо аналит, то при измерении образцов без аналита cut-off по шкале ICR может учитывать шум системы. LoB представляет собой шум системы. При использовании непрерывного сигнала и установлении cut-off на уровне LoB показатель FP составит 5 %. Аналогично можно установить различные cut-off, 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 медицинского изделия, при определении cut-off необходимо учитывать компромиссные решения для баланса показателей FP и FN. После того как cut-off установлено, С95 также может быть определен с помощью методов, описанных в 7.2.3. Часто распределение ниже С50 слишком сжато, чтобы можно было использовать эти методы для определения С5.

7.3 Разработка качественных исследований с использованием методов обнаружения аналитов

Значительное и растущее число качественных исследований, которые можно классифицировать как исследования по обнаружению аналитов, выполняют с помощью ПЦР. Для разработки любого изделия для 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.

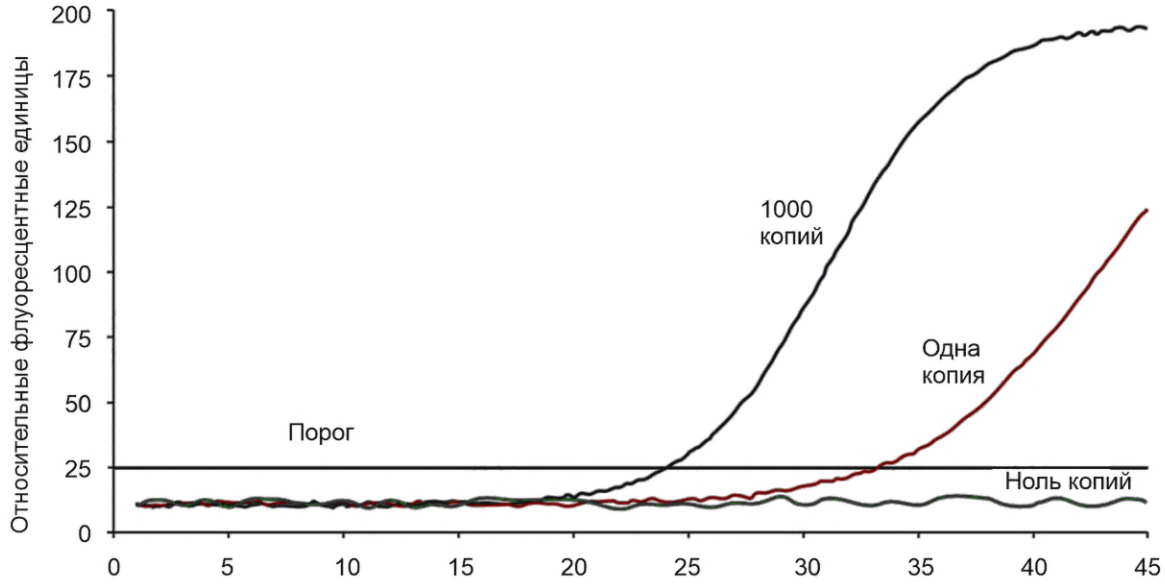


Рисунок 8 — Исследование ПЦР

ПЦР-исследования включают в себя циклы репликации, при которых 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 ICR, например, относительных единиц флуоресценции (RFU) на 45-м цикле, как показано на рисунке 8. В этом примере RFU, определяющие порог цикла, потенциально будут хорошим вариантом для порога. Для таких случаев непосредственно применимы процессы разработки, описанные в 7.2.4. Cut-off, C95 и, возможно, C5 могут быть определены с помощью этой шкалы.

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 прецизионность ICR, а число копий. Несмотря на то, что ПЦР используют 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 установленное cut-off до определенного количества циклов ПЦР, в остальных случаях образец является 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 шума системы 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 % диагностической специфичности. Тестирование популяции отрицательных образцов и оценка интреференции при исследовании помогут продемонстрировать эту надежную диагностическую 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 диагностическую чувствительность, близкую к 100 %.

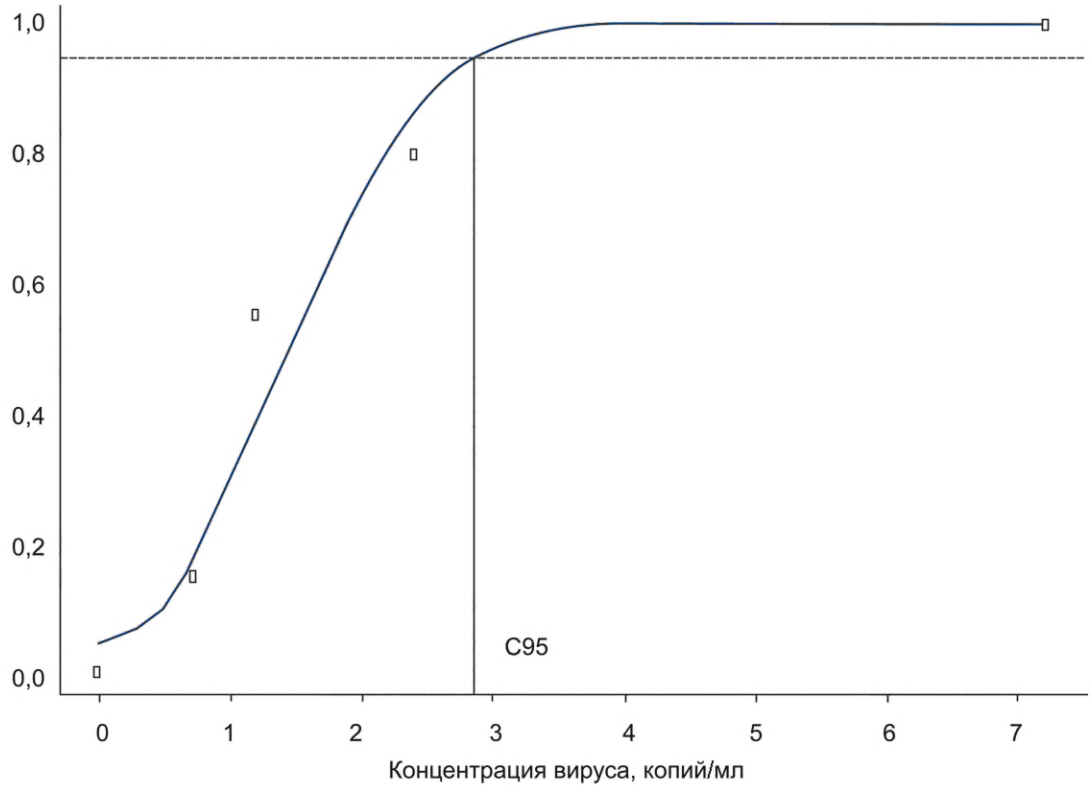
Более сложные алгоритмы могут объединять более одной последовательности целевой ДНК, чтобы дать ответ «да» или «нет» на вопрос «Обнаружен ли целевой аналит?». В других алгоритмах используют несколько этапов процесса, которые поочередно устраняют все возможные нецелевые аналиты, оставляя только образцы с целевым аналитом, которые будут признаны положительными. Про-

изводитель обязан создать алгоритм, обеспечивающий наилучшую диагностическую чувствительность и диагностическую 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 с известной концентрацией аналита. Ключевыми моментами для создания наилучшего процесса измерения C95 являются:

- использование достаточного количества и соответствующего распределения образцов;
- точное определение концентрации аналита (копий в мл) в каждом образце;
- наличие достаточного количества повторных исследований для каждого образца для получения приемлемой оценки частоты попаданий;
- нахождение соответствия данных соответствующей обобщенной линейной модели (например, пробит, логит).

На рисунке 9 приведен пример графика набора данных, полученных в ходе качественного исследования с превосходной диагностической чувствительностью.



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

Рисунок 9 — Использование аппроксимации под кривую пробит для оценки C95

На рисунке представлены образцы в пяти концентрациях, охватывающих интервал от 0 до 7 копий/мл. Самая высокая концентрация включена, чтобы обеспечить 100 %-ное 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 копии/мл.

В отличие от предыдущих описаний измерения C95 в этом разделе, шкала, используемая для определения C95, основана на внешней единице измерения (копии/мл), которая отличается от любой внутренней шкалы. Ее не измеряют с помощью интервала погрешности ICR выше cut-off. Тем не менее, измеряют точку (по шкале копии/мл), при которой образец признается 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 диагно-

стическую 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 измеряют с помощью интервала неточности ICR ниже cut-off. Если его можно достоверно определить, то измеряют точку (по шкале копии/мл), в которой образец признается положительным в 5 % случаев. Как упоминалось ранее, если последовательность целевой нуклеиновой кислоты уникальна для целевого аналита, диагностическая специфичность будет близка к 100 %. C5 коррелирует с диагностической специфичностью, но в процессе разработки 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, часто используемым для определения образцов, которые следует применять 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 используют в настоящем стандарте в качестве показателя эффективности 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 контрольные материалы часто доступны только в высоких концентрациях (например, 10^6 копий/мл). 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 референтной методики измерений, такой как цифровая ПЦР.

В примере, приведенном на рисунке 9, каждый образец исследуют 25 раз, чтобы оценить пропорциональную частоту, при которой этот образец будет признан положительным (т. е. частоту попаданий). В этом случае шаг измерения каждого ответа 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 шага измерений. По этой причине для данного типа исследований рекомендуется использовать не менее 20 повторов.

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 одном повторе 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 большее количество измерений и, соответственно, большее количество повторных исследований образцов. На рисунке 9 показана аппроксимация кривой пробит для нетрансформированных данных о количестве копий; часто такая аппроксимация лучше (т. е. меньше остаточная ошибка) после преобразования подсчетов с помощью функции на основе логарифма, которую измеряют при нулевой концентрации. Такая модель может обеспечить лучшую аппроксимацию 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 cut-off по флуоресценции. Это значение может быть использовано в качестве ICR для повторных исследований, признанных положительными. Однако повторные исследования, считающиеся отрицательными, представляют 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 в 1,5 раза выше C95, что ограничивает полезность такого исследования неточности.

7.4 Контрольные материалы

7.4.1 В данном подразделе приведены правила использования контроля правильности, контрольных материалов КК и внутреннего контроля системы.

7.4.2 Контроль правильности

Качественные исследования сравниваются ICR образца с материалом, имеющим ICR на уровне cut-off. Распространенным типом таких исследований являются исследования соотношения сигнал/порог (S/CO). Материал с пороговым значением метрологически прослеживается либо до внутреннего контрольного материала производителя, установленного при разработке исследования, либо до внешнего контрольного материала, используемого в процессе начальной настройки cut-off. Его приписанное 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 ГОСТ Р ИСО 17511.

7.4.3 Контрольные материалы

Использование соответствующих материалов для контроля качества обеспечивает стабильное качество бинарных исследований. Эти контроли исследуются 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 лаборатории доступен ICR для контрольного материала, то руководство по проведению внутрилабораторного контроля качества количественных методов исследований может быть использовано 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.

Материалы QC должны включать как минимум один образец с ожидаемым отрицательным результатом и как минимум один образец с ожидаемым положительным 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 ряда 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 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 сигналов. Положительные и отрицательные контроли, расположенные вблизи cut-off, более чувствительны к обнаружению ошибок процесса, чем расположенные значительно выше или значительно ниже cut-off; однако контроли, расположенные вблизи cut-off, могут привести к отклонению на основании случайных ошибок (для C5 и C95 такое отклонение происходит в 5 % случаев). Уровни контроля должны быть выбраны таким образом, чтобы обеспечить ожидаемый результат при правильном функционировании изделия, что предполагает, что процесс исследования должен контролироваться на уровнях, более далеких, чем C5 и C95 от cut-off.

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 1.

Таблица 1 — Примеры риска ошибок при тестировании и материалов QC для качественных исследований

Сдвиг в функционировании системы	Направление смещения	Риск для пациента
Ошибка в процессе качественного исследования или отказ реагентов приводят к изменению результатов у пациента	Отрицательное[1)]	Высокий риск: ошибочное медицинское решение не лечить[2)]
	Положительное[3)]	Риск среднего уровня: ненужное или неадекватное решение о лечении
		Низкий риск: подтверждающий тест может дать правильный отрицательный результат

Окончание таблицы 1

[1] Низкоположительный материал QC редко дает отрицательные результаты, когда процедура находится под контролем, но вероятность получения FN-результатов увеличивается при изменении режимов отказа, вызывая отрицательный сдвиг.

[2] Отрицательный сдвиг (FN-результат) может привести к ошибочному решению не лечить пациента или не отбраковывать инфицированный образец крови. Примером положительного сдвига (FP-результат) является серологическое определение 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.

[3] Высокоотрицательный материал QC редко дает положительные результаты, когда процедура находится под контролем, но вероятность получения FP-результатов возрастает при наличии режимов отказа, вызывающих положительный сдвиг.

7.4.4 Контроль валидности МИ ИВД

Для контроля валидности МИ ИВД могут осуществляться процедуры, обеспечивающие контроль процессов, этапов. Контроль выполняют перед исследованием каждого образца 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 этапов процесса (в идеале — через все), чтобы получить результаты для пациента. Сравнение ICR с ожидаемым сигналом 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 исследования, в том числе для выделения исследуемого аналита или для создания сигнала прибора или порогового образца, применяемого для получения непрерывного сигнала по соотношению сигнал/пороговый образец. Во всех случаях, если ICR доступен производителю, сигналы исследуемых образцов могут быть использованы для оценки тенденций с помощью методов регрессии, как описано в ГОСТ Р 53079.4, 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 детекции. Например, результаты визуального исследования мочи с помощью тест-полосок могут быть количественно оценены с помощью колориметрического считывателя полосок. При исследовании на основе ПЦР для обнаружения аналитов может использоваться цикл (cut-off) 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 недоступен, для исследования стабильности можно использовать схему исследования, описанную в 9.3.

Срок годности и стабильность при использовании, при которых реагенты деградировали и, как следствие, изменили наблюдаемое значение тестового образца, могут привести к смещению cut-off и, таким образом, негативно повлиять на все параметры эффективности, включая C5, C95, диагностическую чувствительность и диагностическую 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 разработана программа исследования.

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 аналитов для исследований стабильности.

8 Валидация качественного исследования с бинарным ответом

8.1 Сбор данных для валидации качественного бинарного исследования включает в себя оценку аппаратного и программного обеспечения, а также реагентов (включая «расходные материалы») в конфигурации, находящейся в состоянии аналитической системы (т. е. окончательном рабочем). Все данные собирают с использованием тех же инструкций по применению медицинского изделия, или в окончательных стандартных операционных процедурах (в случае разработанного лабораторией изделия).

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

а) неточность (исследования прецизионности и частоты попаданий):
   1) внутрилабораторное исследование,
   2) межлабораторное исследование;
б) классификация (диагностическая эффективность или процентная согласованность);
в) интерферирующие вещества.

8.2 Исследования прецизионности

Независимо от типа качественного исследования, для оценки неточности обычно проводят два типа испытаний: внутрилабораторные испытания прецизионности (внутри производителя) и испытания на прецизионность, включающие лаборатории вне производителя (в случае распространяемого набора) или референс-лаборатории. Для описания требований к испытаниям прецизионности качественные бинарные исследования рекомендуется группировать согласно 6.2.

Программа прецизионных испытаний включает два основных аспекта: оцениваемые условия прецизионности (источники вариабельности) и панель образцов. Изделие (например, оборудование, программное обеспечение, реагенты, рабочие растворы, принадлежности) определяет программу испытаний и влияет на то, сколько источников изменчивости можно оценить одновременно. Ранее определенные параметры изделия, такие как C95, влияют на выбор образцов, используемых для исследования. Различия в обоих аспектах (логистика и оценки параметров) и то, как они влияют на программу исследований, приведено ниже. Более подробная информация о программах прецизионных исследований содержится в приложении Е.

8.2.1 Испытания для исследований с cut-off при внутреннем непрерывном сигнале

8.2.1.1 В данном подразделе рассмотрены первые два типа исследований, упомянутые выше: наличие или отсутствие ЦС на основании клинически установленного cut-off ICR и обнаружение аналитов (ноль против одной или более копий) на основании одного cut-off ICR шкалы. Поскольку ICR предоставляет числовые значения для всех результатов анализа образцов, для таких исследований допускается внутрилабораторные исследования и исследования прецизионности на разных образцах. Для этих целей допускается определять влияние таких компонентов дисперсии, как повторяемость, аналитическая серия, 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 пробоподготовки), системы исследования (аппаратное и программное обеспечение, реагенты и другие расходные материалы), методы предварительных исследований, таких как выделение аналитов, и предполагаемого использования МИ ИВД.

8.2.1.2 Исследования прецизионности

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

- заведомо отрицательный образец (образец с нулевым содержанием аналитов для исследований на определение аналитов);
- высокоотрицательный образец (около C5 для клинически установленного cut-off исследований ICR);

- образец выше cut-off (около C95);
- умеренно положительный образец (выше C95, но в идеале — ближе к C100);
- высокоположительный образец.

Примечание — Для исследований с клинически установленным cut-off ICR высокоотрицательная проба должна быть направлена на значение C5. Для исследований с cut-off ICR, установленным на пределе шума системы или вблизи него (т. е. LoB), высокоотрицательная проба не требуется. Варианты высокоотрицательной пробы для исследований без cut-off, с обнаружением аналитов, см. в 8.2.2.1.

Рекомендуется все образцы получать из немодифицированных проб пациента. Варианты использования образцов панели прецизионности ниже cut-off, а для исследований с определением аналитов описаны в 8.2.2.1. Поскольку ICR доступен для всех образцов для данного типа исследований, ANOVA следует проводить для прецизионности каждого результата и/или уровня пробы отдельно, если предположения ANOVA уместны. Сильно асимметричные распределения могут потребовать преобразования результатов ICR перед анализом, а бимодальное распределение может исключить проведение ANOVA. 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 прецизионности) должны быть вычислены с использованием формул (1) и (2) и представлены на основе внутрилабораторной прецизионности. Кроме того, необходимо представить описательную сводку, результаты бинарных исследований. Для каждого образца в панели прецизионности следует рассчитать процент положительных, отрицательных и недействительных результатов. Пример такой сводки для внутрилабораторного исследования прецизионности приведен в 8.2.3. Аналогичные результаты должны быть представлены для межлабораторных исследований.

8.2.1.3 Исследования частоты попаданий для определения C95 по количественной шкале

Если существует контрольный материал с известным количественным значением или доступна референтная методика 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 каждого разведения 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 %,
один образец с процентом попадания более 95 %, но в идеале близким к 100 %,
одного образца со 100 %-ными попаданиями.

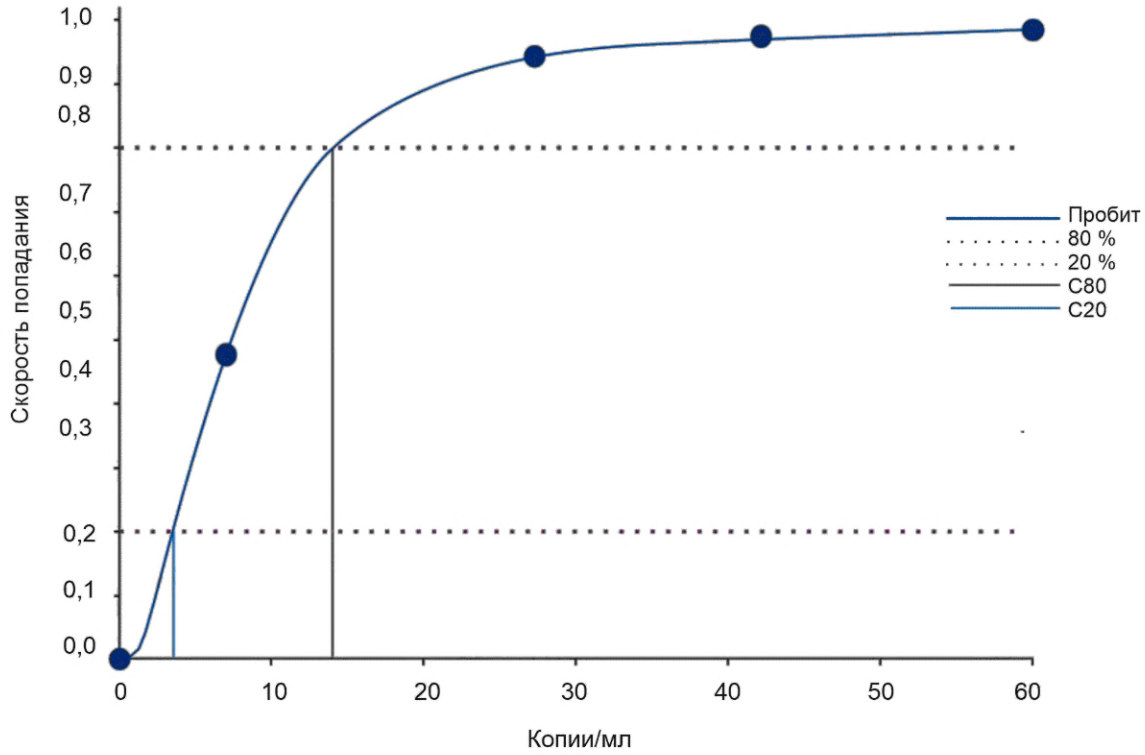
Этот набор образцов обычно создают на уровне, близком к желаемому, но не всегда точно соответствующем ему, путем разбавления. Анализ этих данных для каждого партии реагентов, более подробно описанный в 8.2.2.2, позволяет определить ожидаемую частоту попаданий на всем диапазоне образцов. Наиболее распространенным параметром, определяемым таким образом, является значение C95 (определяется по референтному материалу количественной шкалы). Если cut-off является LoB, то это значение часто называют LLoD.

8.2.2 Исследования по обнаружению аналитов с использованием алгоритма вместо внутренней непрерывной реакции с cut-off

8.2.2.1 Исследования прецизионности

Хотя эти исследования дают только результаты «да» или «нет» для образцов, все равно следует проводить те же внутрилабораторные испытания и межлабораторные испытания прецизионности, которые описаны в 8.2.1. 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 8.2.1.1. Для каждого типа исследования должны быть представлены бинарные отчеты, как показано в 8.2.3. Для панели прецизионности исследования обнаружения аналита уровень образца ниже cut-off может быть определен с помощью результатов разработанного изделия и изделия сравнения (процедуры сравнения).

Если более 10 % всех испытуемых из целевой популяции, признанных изделием сравнения положительными, имеют наблюдаемое значение меньше C95 в разработанном изделии, высокоотрицательный образец не включается в панель прецизионности. Он заменяется образцом в интервале, в котором бинарное исследование признает образец положительным от 20 % до 80 % случаев, как показано на рисунке 10.



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

Рисунок 10 — Уровни прецизионной панели

Если менее 10 % всех испытуемых из популяции предполагаемого использования, признанных изделием сравнения положительными, имеют наблюдаемое значение меньше C95 в разработанном изделии, высокоотрицательный образец исключается из панели прецизионности.

8.2.2.2 Исследования частоты попаданий для определения нижнего предела обнаружения

В исследованиях с коэффициентом попадания значения образцов получают путем их измерения с помощью процедуры количественного референтного измерения или создания путем разбавления контрольного материала. Схема исследования коэффициента попадания такая же, как описана в 8.2.1.2 для оценки значения образца, при котором 95 % повторов 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 аналита X (для $X > 0$), вычисляют по формуле

$$P_{pos}(X) = \Phi[a + b \log(X)], \quad (5)$$

где X — концентрация аналита (для $X > 0$);

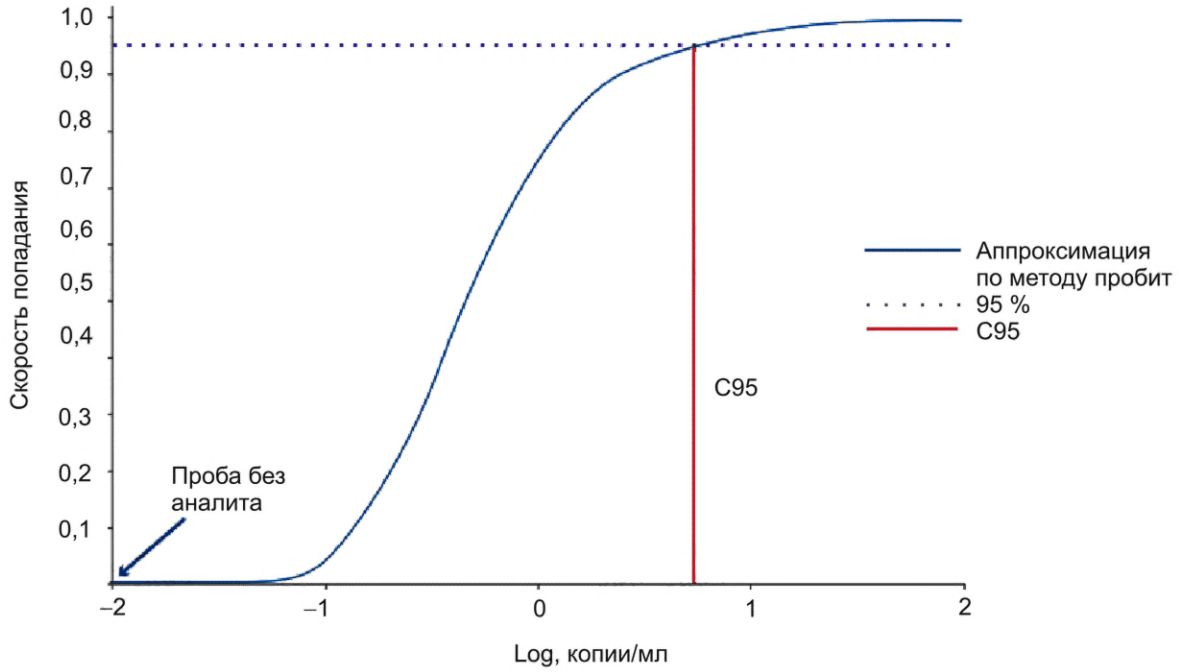
$P_{pos}(X)$ — вероятность получения положительного результата исследований в зависимости от концентрации X;

Φ — кумулятивная стандартная нормальная функция вероятности;

a и b — отрезок (Intercept) и наклон (Slope) 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.

На рисунке 11 показано соответствующее преобразование к $\log_{10}$ набора данных при c = 0,01.



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

Рисунок 11 — Стандартная кривая пробит

Статистическое программное обеспечение выполняет нелинейную аппроксимацию 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 шума системы или несоответствия образцов, образцы без аналитов могут иметь FP-результаты. 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).

Расширенную пробит-модель вычисляют по формуле

$$P_{pos}(X) = \Phi[a + b \log(X)] + \{1 - \Phi[a + b \log(X)]\}, \quad (6)$$

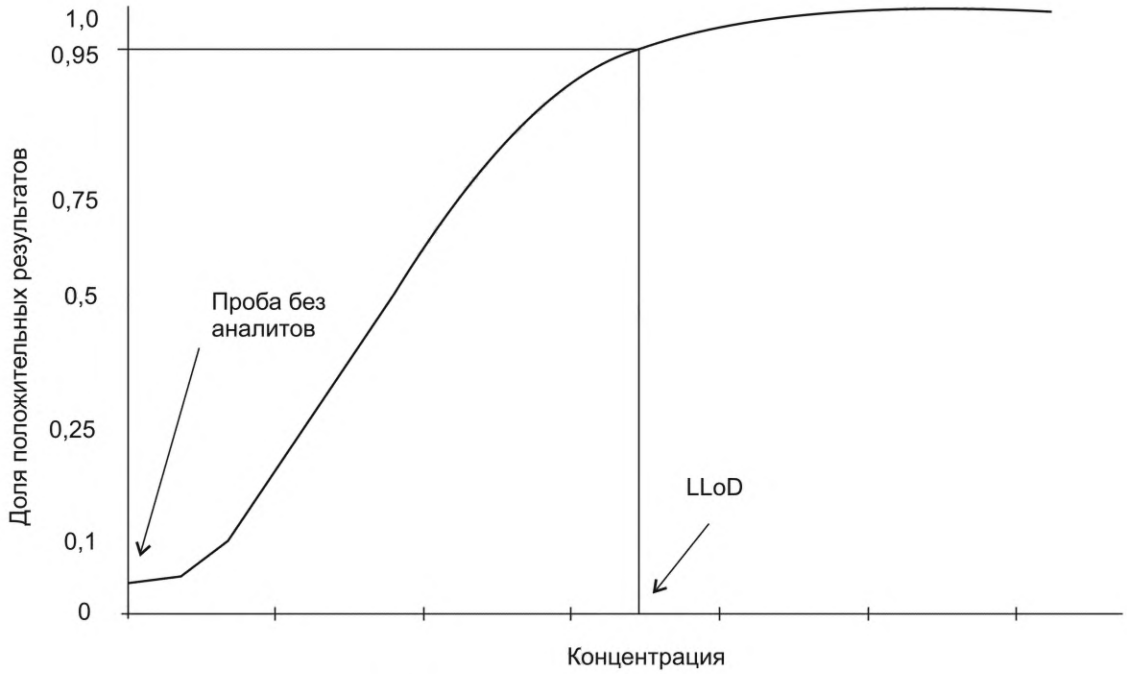
где все общие параметры те же, что и в стандартной пробит-модели, $X > 0$, а P — вероятность FP. Замена X на $X' = X + c$, причем c — небольшое положительное число, означает, что модель больше не ограничена нулем. Статистическое программное обеспечение также может быть использовано в этом случае для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

Из формулы (6) следует, что $P_{pos}(X)$ приближается к P_{FP} по мере приближения X к нулю (или X' к c) (см. рисунок 12).

Расширенная пробит-модель обычно дает меньшее значение LLoD, чем стандартная пробит-модель.

8.2.3 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 2 для ежедневной логистики), бинарные результаты за все дни для образцов на разных уровнях (например, C5, C95 и C100) обобщены в таблице 3.



LLoD — нижний предел обнаружения

Рисунок 12 — Расширенная кривая пробит с P_{FP} 5 %

Таблица 2 — Количество повторов для программы внутрилабораторного прецизионного исследования

Исследования	День 1											
	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

Таблица 3 — Пример подсчета результатов бинарного исследования

Источник	N (кол-во)	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

Окончание таблицы 3

Источник	N (кол-во)	Уровень 1, C5			Уровень 2, C95			Уровень 3, C100		
		Положительный	Отрицательный	Недействительный	Положительный	Отрицательный	Недействительный	Положительный	Отрицательный	Недействительный
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 — C5 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; C100 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 100 % случаев; N — количество общих повторов для каждого уровня.

Цель исследования на обнаружение аналитов для образца без аналитов — никогда не признавать его положительным. Примеры программ внутрилабораторных и межлабораторных исследований приведены в приложении Е.

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, можно прояснить, были ли получены ожидаемые результаты и какие факторы требуют изучения. Например, если у одного оператора доля положительных результатов на C95 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 для исследования методом NGS.

8.2.4 Положения, касающиеся прецизионности исследований

Для исследований с визуальной оценкой ответа в двоичном формате ICR не существует. Как указано в 8.2.2.1 и таблице 2, оба исследования, внутрилабораторной прецизионности и межлабораторной, допускается проводить с использованием бинарного отчета, поскольку это единственный доступный анализ неточности. Если изделие предназначено в первую очередь для использования вне лаборатории, в испытании прецизионности следует задействовать не менее трех внешних мест. Более подробное описание оценки изделия, учитывающее вариативность визуальной оценки, приведено в приложении Ж.

Если разработчик хочет заявить об использовании нескольких типов образцов для исследования, то основной или преобладающий тип образца должен быть тщательно валидирован в соответствии с описанными выше схемами прецизионных испытаний. 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 или для микрочипов.

8.3 Оценка диагностической эффективности

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. Диагностическая эффективность разработанного изделия для проведения 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 изделие.

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 диагностическую эффективность разработанного изделия. Наиболее распространенной парой показателей диагностической эффективности являются диагностическая чувствительность и диагностическая специфичность. Диагностическая чувствительность — это доля испытуемых с ЦС, признанных положительными по результатам разработанных изделий. Диагностическая специфичность — это доля лиц без ЦС, у которых разработанное изделие выявило отрицательный результат.

8.3.3 Сравнение с альтернативным изделием

Если разработанное изделие оценивают путем сравнения со сравнительным медицинским изделием, которое не является общепризнанным лучшим методом оценки ЦС, диагностическая чувствительность и диагностическая специфичность не могут быть оценены.

Вместо этого разработанное изделие 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 положительной процентной согласованностью (PPA) и отрицательной процентной согласованностью (NPA), чтобы отразить, что эти оценки не дают количественной оценки диагностической эффективности, а скорее определяют согласованность разработанного изделия со сравнительным изделием. Кроме того, такие величины, как положительная прогностическая ценность (PPV), отрицательная прогностическая ценность (NPV), а также положительное и отрицательное отношения правдоподобия (PLR и NLR) не могут быть рассчитаны, поскольку состояние испытуемого неизвестно.

8.3.4 Требования к исследованию эффективности классификации

8.3.4.1 Положения данного пункта содержат общие рекомендации по выбору изделия сравнения и отбору образцов для различных ситуаций. Для изделий, предназначенных для оценки диагностической эффективности, необходимо включить следующие два ключевых компонента:

- стандартное или сравнительное изделие 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 разработанного изделия. 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 изделия, поскольку они не отражают вариабельность оценок. Исследование должно быть достаточного объема, чтобы оценить разработанное изделие 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 изделия сравнения

Выбор изделия сравнения должен определяться типом необходимой характеристики. Если целью является характеристика диагностической чувствительности и специфичности, а также соответствующих отношений правдоподобия (LR) 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 изделий, предназначенных для выявления 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 диагностическая 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, приведена в 8.3.6.

В некоторых случаях оценка ЦС основана исключительно на диагностической картине у испытуемых, например на 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 согласованности (исследование, оценивающее согласованность 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 диагностической эффективности. Каждое исследование имеет свою цель, и разработанное изделие должно оцениваться отдельно путем сравнения с подходящими изделиями сравнения.

Для лаборатории или разработчика изделия для качественного бинарного исследования, предназначенного для замены существующего метода, целью может быть демонстрация способности разработанного изделия в достаточной степени соответствовать сравнительному изделию, используемому в настоящее время. Для этой цели сравнительным методом может быть используемое в настоящее время изделие. Однако если новое изделие лучше (например, более чувствительно), чем ранее применяемый метод, эти два метода могут не совпадать. Эта относительная характеристика может также применяться к изделию сравнения, которое использовалось для оценки ЦС. Обсуждение расхождения результатов приведено в 8.3.6.3.

8.3.4.4 Отбор субъектов и образцов исследования

Образцы, включенные в исследование диагностической эффективности, должны представлять популяцию предполагаемого использования. Если участвующие в исследовании субъекты не являются репрезентативными для предполагаемой популяции, оценки диагностической эффективности подвержены эффекту спектра. 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 испытуемым 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.

Для обеспечения легитимности выбора субъектов исследования и образцов биоматериала соблюдают следующие рекомендации:

а) процедура отбора субъектов и образцов исследования

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:

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

2) образцы 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 объем исследования, разработчик может принять во внимание:

1) знание истинной эффективности разработанных медицинских изделий и изделий сравнения и уверенность в этом знании;

2) риски, связанные с невыполнением критериев приемлемости эффективности;

3) рабочий процесс или технологические требования;

4) 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. Дополнительное рассмотрение данной ситуации не входит в область применения настоящего стандарта.

8.3.5 Сравнение бинарного исследования с оценкой целевого состояния

8.3.5.1 Диагностическую эффективность качественного бинарного исследования обычно оценивают с точки зрения правильности классификации (диагностическая чувствительность и специфичность) и прогностической ценности (положительная и отрицательная). Таблица 4 представляет собой таблицу случайностей 2 × 2, в которой сравниваются бинарные результаты качественного исследования и результаты оценки ЦС, используемые для определения наличия или отсутствия ЦС.

Таблица 4 — Качественный результат разработанного изделия в сравнении с оценкой ЦС

Результат исследования для разработанного изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	Количество TP = A	Количество FP = B	$A + B$
Отрицательный	Количество FN = C	Количество TN = D	$C + D$
Всего	$N_1 = A + C$	$N_0 = B + D$	$N = N_1 + N_0 = A + B + C + D$

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — FN — ложноотрицательный; FP — ложноположительный; N — количество образцов; ЦС — целевое состояние; TN — истинно отрицательный; TP — истинно положительный.

Запись в каждой ячейке таблицы представляет собой количество образцов (или субъектов, предполагая один образец на субъект), соответствующих заголовкам в таблице 4.

Формулу (7) используют для расчета диагностической чувствительности ДЧ, а формулу (8) — для расчета диагностической специфичности ДС.

$$\text{ДЧ} = \frac{A}{A + C}, \tag{7}$$

$$\text{ДС} = \frac{D}{B + D}. \tag{8}$$

Примечание — Все коэффициенты, относящиеся к клиническим показателям, включая диагностическую 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, которое может быть вычислено 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.

$$\text{PLR} = \frac{\text{ДЧ}}{1 - \text{ДС}} = \left(\frac{A}{N_1}\right)\left(\frac{N_0}{B}\right), \tag{9}$$

$$\text{NLR} = \frac{1 - \text{ДЧ}}{\text{ДС}} = \left(\frac{C}{N_1}\right)\left(\frac{N_0}{D}\right). \quad (10)$$

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. Исследования 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).

$$\text{OR} = \frac{\text{PLR}}{\text{NLR}}. \quad (11)$$

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

$$\text{Оценка распространенности} = \frac{N_1}{N}. \quad (12)$$

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 диагностической эффективности метода, как диагностическая чувствительность и диагностическая специфичность, и не должны рассчитываться на основе исследования, предназначенного для автоматического определения диагностической 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 низкой 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 диагностической 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:

$$\text{PPV} = 100\frac{A}{A + B}, \quad (13)$$

$$\text{NPV} = 100\frac{D}{C + D}. \quad (14)$$

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

$$\text{PPV при фиксированной распространенности} = \frac{\text{ДЧ} \cdot \text{Распространенность}}{\text{ДЧ} \cdot \text{Распространенность} + (1 - \text{ДС})(1 - \text{Распространенность})}, \quad (15)$$

$$\text{NPV при фиксированной распространенности} = \frac{\text{ДС}(1 - \text{Распространенность})}{\text{ДС}(1 - \text{Распространенность}) + (1 - \text{ДЧ})(\text{Распространенность})}. \quad (16)$$

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 высокую PPV. Более точно, PLR — относительное увеличение шансов на ЦС, связанное с положительным результатом исследования. Например, если PLR равно 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.

$$\frac{\text{PPV}}{1 - \text{PPV}} = \text{PLR}\left(\frac{\text{Распространенность}}{1 - \text{Распространенность}}\right). \quad (17)$$

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.

$$\frac{1 - \text{NPV}}{\text{NPV}} = \text{NLR}\left(\frac{\text{Распространенность}}{1 - \text{Распространенность}}\right). \quad (18)$$

Результатом является монотонное возрастание PPV при PLR и монотонное убывание NPV 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. 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. Показатели оценки эффективности могут быть подвержены смещению. К аспектам исследования, которые могут внести неточность, относятся набор испытуемых и процесс отбора, не соответствующий предполагаемому использованию, отсутствие слепого контроля за клиническими результатами со стороны производящих отбор, 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. Информация по отбору субъектов для исследования приведена в 8.3.4.4.

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 %; поэтому методы расчета ДИ, основанные на нормальном приближении, не рекомендуются. Рекомендуется использовать метод баллов Уилсона, поскольку он обладает хорошими статистическими свойствами и может быть рассчитан напрямую. Информация о получении необходимых расчетных значений приведена в приложении И [см. формулу (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 о серой зоне

Серая 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 разработанного изделия в клиническом исследовании часто нецелесообразно. Включение результатов последующего исследования в оценку результативности разработанного изделия может быть не обобщающим и не полезным, если последующее исследование и интерпретация результатов не указаны. Поэтому при такой оценке результаты разработанного изделия являются положительными, двусмысленными или отрицательными, а результаты исследования могут быть представлены в виде таблицы случайностей 3 × 2, как показано в таблице 5.

Таблица 5 — Результат качественного исследованиея разработанного изделия с неоднозначными результатами в сравнении с оценкой ЦС

Результат разработанного изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	A_1	B_1	$A_1 + B_1$
Сомнительный	A_2	B_2	$A_2 + B_2$
Отрицательный	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

1) Наилучшая имеющаяся оценка ЦС.

Примечание — 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 сомнительные результаты (B_2) считаются положительными. Таким образом, доля сомнительных результатов вычисляется по формуле (21).

$$\text{Диагностическая чувствительность } (\text{ДС}) = \frac{A_1}{N_1}. \quad (19)$$

$$\text{Диагностическая специфичность } (\text{ДС}) = \frac{B_1}{N_0}. \quad (20)$$

$$\text{Доля равновероятных результатов в популяции испытуемых} = \frac{A_2 + B_2}{N}. \quad (21)$$

Сомнительный результат можно рассматривать как третью категорию при оценке таблицы 2 × 2 случайностей. Тогда LR сомнительного результата может быть рассчитан и интерпретирован как изменение шансов на ЦС, которые дают сомнительный результат по отношению к случайно выбранному (не тестированному) субъекту. Если первый результат оказался неоднозначным, повторное исследование разработанного изделия 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 2 × 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 Согласованность качественного исследования с бинарным выходом

8.3.6.1 В распространенной ситуации, когда компаратор не является общепринятой оценкой ЦС, диагностическая чувствительность и специфичность не могут быть легко оценены. Вместо этого проводятся те же численные расчеты, но оценки называются PPA и NPA, а не диагностической чувствительностью и специфичностью, поскольку речь идет не о клинических показателях, а о согласованности разработанного изделия с изделием сравнения. 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 согласованности между разработанным изделием и изделием сравнения

Если метод разработанного изделия сравнивают с чем-то другим, а не с наилучшей доступной оценкой ЦС, полезным может оказаться использование таблицы 2 × 2 результатов и отчета о том, насколько часто совпадают результаты разработанного изделия и изделия сравнения. Пример представления результатов приведен в таблице 6.

Таблица 6 — Таблица 2 × 2 случайностей для сравнительного исследования

Результат разработанного изделия	Результат изделия сравнения		
	Положительный	Отрицательный	Всего
Положительный	*A*	*B*	*A* + *B*
Отрицательный	*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, графы 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, и, соответственно, не могут быть использованы для оценки диагностической чувствительности или диагностической специфичности; они могут быть использованы только для оценки PPA и NPA.

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

$$\text{PPA} = 100\frac{A}{A + C}, \tag{22}$$

$$\text{NPA} = 100\frac{D}{B + D}. \tag{23}$$

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

8.3.6.3 Разрешение несоответствий

Расхождения между разработанным изделием и изделием сравнения могут возникнуть из-за ошибок в одном из них. Кроме того, с развитием новых технологий возможно, чтобы разработанное изделие имело лучшие клинические характеристики, чем лучшая из доступных в настоящее время оценок ЦС, 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 разработанного изделия, так и изделия сравнения, используя контрольные материалы (в идеале — сертифицированные стандартные образцы) для оценки каждого C95. Образцы вблизи cut-off и C95 по-прежнему следует оценивать с помощью образцов пациентов и их разведений соответствующим разбавителем (обычно образцы пациентов с нулевым содержанием аналитов), но критерии приемлемости должны учитывать различия в C95. В целом для тех исследований, в которых используется ICR, список числовых результатов разработанного изделия и изделия сравнения изучают, близки ли расхождения в результатах к cut-off для обоих исследований или расхождения могут быть обусловлены различиями в C95. Если C95 ближе к cut-off для разработанного изделия, 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, согласуются ли характеристики разработанного изделия и изделия сравнения в большей степени с экспертным мнением, но сама по себе эта информация не может быть использована для прямой оценки эффективности исследования 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 разработанного изделия, результат изделия сравнения и результат реферирования для каждой дискордантной выборки. Чтобы рассчитать оценку эффективности разработанного изделия 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 диагностической чувствительности и специфичности двух изделий

8.3.7.1 Пользователь может захотеть оценить различия в чувствительности и специфичности между двумя обследованиями. Допускается использовать формулу (И.1) приложения И для оценки ДИ диагностической 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 репрезентативные субъекты исследования.

При сравнении двух изделий диагностическая 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).

8.3.7.2 Сравнение парных диагностических чувствительностей и специфичностей

Схема исследования трехстороннего сравнения между разработанным изделием, ранее применяемым изделием и оценкой ЦС представлена в таблице 7.

Таблица 7 — Трехстороннее сравнение между разработанным изделием, ранее применявшимся изделием и оценкой ЦС

Оценка ЦС[1)]				Оценка ЦС[1)]			
ЦС присутствует				ЦС отсутствует			
Результат разработанного изделия	Результат изделия сравнения		Всего	Результат разработанного изделия	Результат изделия сравнения		Всего
	Положительный	Отрицательный			Положительный	Отрицательный	
Положительный	A_1	B_1	$A_1 + B_1$	Положительный	A_0	B_0	$A_0 + B_0$
Отрицательный	C_1	D_1	$C_1 + D_1$	Отрицательный	C_0	D_0	$C_0 + D_0$
Всего	$A_1 + C_1$	$B_1 + D_1$	N_1	Всего	$A_0 + C_0$	$B_0 + D_0$	N_0

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — N — количество образцов; ЦС — целевое состояние.

Данные считаются парными, так как одни и те же образцы изучают путем сравнения разработанного изделия и ранее применявшегося изделия. Результаты разработанного изделия сравнивают с результатами ранее применявшегося изделия при наличии оценки ЦС (для сравнения чувствительности) и отдельно, если оценка ЦС отсутствует (для сравнения специфичности). Когда сравнивают чувствительность и специфичность двух изделий, план исследования должен гарантировать, что порядок тестирования образцов не повлияет на результат. Формулы (25)—(28) используют для расчета разницы в чувствительности между разработанным изделием и ранее применявшимся изделием. Детали того, как эти расчеты могут быть модифицированы для аналогичных расчетов специфичности, описаны в формулах (29)—(31).

$$N_1 = A_1 + B_1 + C_1 + D_1, \quad (24)$$

$$N_0 = A_0 + B_0 + C_0 + D_0. \quad (25)$$

По данным таблицы 7 можно создать две таблицы по форме таблицы 4 (TP A, FP B, FN C и TN D в сравнении с оценкой ЦС), одну для разработанного изделия (см. таблицу 8) и одну для ранее применявшегося изделия (см. таблицу 9).

Таблица 8 — Оценка ЦС для сравнения разработанного изделия с ранее применявшимся изделием

Результат разработанного изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	$A_1 + B_1$	$A_0 + B_0$	$A_1 + B_1 + A_0 + B_0$
Отрицательный	$C_1 + D_1$	$C_0 + D_0$	$C_1 + D_1 + C_0 + D_0$
Всего	N_1	N_0	N

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — N — количество образцов; ЦС — целевое состояние.

Таблица 9 — Ранее применявшееся изделие в сравнении с оценкой ЦС

Результат ранее применявшегося изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	$A_1 + C_1$	$A_0 + C_0$	$A_1 + C_1 + A_0 + C_0$
Отрицательный	$B_1 + D_1$	$B_0 + D_0$	$B_1 + D_1 + B_0 + D_0$
Всего	N_1	N_0	N

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — N — количество образцов; ЦС — целевое состояние.

Однако обратное утверждение не верно. То есть пользователь не может построить таблицу 7, если известны только данные, приведенные в таблице 4, для каждого изделия. Чтобы построить таблицу 7, необходим трехсторонний вывод данных по разработанному изделию, ранее применявшемуся изделию и оценке ЦС для каждой отдельной выборки.

Оцененная диагностическая чувствительность разработанного изделия ($ДЧ_{РИ}$) и ранее применявшегося изделия ($ДЧ_{ПИ}$), а также их разница могут быть вычислены по формулам:

$$ДЧ_{РИ} = \frac{A_1 + B_1}{N_1}, \quad (26)$$

$$ДЧ_{ПИ} = \frac{A_1 + C_1}{N_1}, \quad (27)$$

$$\Delta_{ДЧ} = \frac{A_1 + B_1}{N_1} - \frac{A_1 + C_1}{N_1} = \frac{B_1 - C_1}{N_1}. \quad (28)$$

Оценки диагностической специфичности двух сравниваемых изделий могут быть вычислены по формулам:

$$ДС_{РИ} = \frac{C_0 + D_0}{N_0}, \quad (29)$$

$$ДС_{ПИ} = \frac{B_0 + D_0}{N_0}, \quad (30)$$

$$\Delta_{ДС} = \frac{C_0 + D_0}{N_0} - \frac{B_0 + D_0}{N_0} = \frac{C_0 - B_0}{N_0}. \quad (31)$$

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) диагностических чувствительностей и специфичностей

Чувствительность и специфичность новых разработанных изделий и ранее применявшихся изделий также можно сравнить, если для каждого изделия использовать независимые наборы образцов (непарные). Каждый набор образцов состоит из одного изделия и оценки ЦС с помощью таблицы 2 × 2, подобной таблице 4. Для любого такого сравнения образцов двух наборов необходимо использовать одинаковые программы (например, спектр популяции пациентов). Трехстороннее сравнение результатов, как, например, результаты, представленные в таблице 8, неприменимо. Рекомендуется метод, основанный на оценке ДИ для двух отдельных пропорций, который описан ниже.

Пусть Δ обозначает разницу между двумя оцененными чувствительностями, $\Delta_{ДЧ} = ДЧ_{РИ} - ДЧ_{ПИ}$. Пусть l_1, u_1, l_2 и u_2 обозначают нижнюю и верхнюю границы двухсторонних 95 %-ных ДИ для $ДЧ_{РИ}$ и $ДЧ_{ПИ}$ соответственно [получены с помощью формулы (И.8)]. Двусторонний 95 %-ный ДИ для разницы между непарными чувствительностями, Δ, может быть вычислен по формуле

$$CI = \Delta_{ДЧ} - \sqrt{\left(ДЧ_{РИ} - l_1\right)^2 + \left(u_2 - ДЧ_{ПИ}\right)^2}\ \Delta_{ДЧ} + \sqrt{\left(ДЧ_{ПИ} - l_2\right)^2 + \left(u_1 - ДЧ_{РИ}\right)^2}. \quad (32)$$

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 интерферирующих веществ при проведении качественных исследований

Разработчик должен изучить влияние интерферирующих 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 ГОСТ Р 53079.4. Если ICR оценить невозможно, разработчик должен провести испытание на интерференцию, описанное в 8.6.

Интерферирующие вещества классифицируют на эндогенные и экзогенные. Эндогенные интерферирующие 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 интерферирующих 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 интерферирующих веществ

При отборе проб учитывают как уровень аналита, выявляемого в ходе исследования, так и концентрацию исследуемого интерферента. Как указано в ГОСТ Р 53079.4, два различных уровня аналитов должны проходить испытание, соответствующее лицам с ЦС и без ЦС.

Для качественных исследований с обнаружением аналитов информация о С95 должна использоваться при подготовке образцов для интерференционного исследования. Образец для интерференционного исследования должен содержать наименьшее количество аналита, при котором разработанное изделие с очень высокой вероятностью получит положительный результат (С100). Обычно рекомендуется рассматривать образец с 1,5—2-кратным превышением С95. 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 разработанных изделий с cut-off С50 по шкале ICR следует оценить отрицательные и положительные погрешности, вызванные потенциальными помехами. Для этой оценки необходимо использовать информацию о С5 и С95. Если С5 установлен, расстояние от cut-off до С5, умноженное на 1,5—2, должно использоваться в качестве целевого значения для образца, который, как ожидается, будет признан отрицательным [т. е. С50 – 2(С50 – С5)]. Целевым уровнем ответа для образцов, которые будут признаны положительными, должно быть расстояние от cut-off до С95, умноженное на 1,5—2 [т. е. С50 + 2(С95 – С50)]. Этот результат является разумной оценкой С100.

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 проба любого из образцов называется контрольным образцом. Чтобы проверить интерференцию, во вторую «контрольную пробу» следует добавить тестируемый интерферант. В общем случае, если микроорганизмы проверяются на интерференцию в исследовании, бактерии добавляются на уровне примерно 1×10^6 колониеобразующих единиц/мл, а вирусы — на уровне 1×10^5 бляшкообразующих единиц/мл. Информация о предпочтительных концентрациях многих целевых аналитов-интерферентов приведена в ГОСТ Р 53079.4.

Оба образца сравнивают с соответствующим контрольным образцом, чтобы определить, оказывает ли интерференция вредное влияние на исследование. Одного повтора может быть достаточно, так как контрольная концентрация известна, но можно рассмотреть и большее количество повторов, 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 повторов для каждого исследования или контрольного сравнения. Если любой из этих повторов 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 повторов). Если имеется ICR, интерференцию следует оценивать с помощью числовых значений ICR в соответствии с ГОСТ Р 53079.4.

8.6 Отчеты о результатах исследований

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

8.6.2 Прецизионность

В инструкции к качественному исследованию должно быть описано, какие источники вариабельности оценивают и как каждый из них используют в исследовании прецизионности. Следует также провести различие между внутрилабораторными и межлабораторными исследованиями прецизионности. Если возможно, следует указать значения ICR и, возможно, концентрации, установленные для значений

по соответствующей шкале, при которых бинарное исследование признает образец положительным в 0 % случаев (C0), C5, C95 и C100, чтобы облегчить пользователям проверку заявлений производителя. Должно быть представлено количество операторов, повторов, постановок в день и общее количество дней. 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 контрольный материал или референтную процедуру измерения, использованную для образцов исследования), а также рассчитанный LLoD.

8.6.3 Клинические показатели, связанные с классификацией

Клиническая чувствительность и специфичность (или положительная и отрицательная согласованность) 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 компаратор 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 программ исследований клинической эффективности приведены в приложении К.

8.6.4 Другие требования к эффективности медицинских изделий

Отчеты по другим показателям должны соответствовать форматам, предложенным в ГОСТ Р 53079.4 для интерферирующих веществ и ГОСТ Р ИСО 23640 для стабильности реагентов.

9 Верификация заявленных характеристик в лаборатории

9.1 Когда лаборатория решает предложить какое-либо IVD-исследование, заявления о прецизионности и клинической эффективности или относительной согласованности, содержащиеся в упаковке, должны быть проверены до получения результатов пациентами. Верификация 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 и прогнозирования. Кроме того, при поступлении новых партий реагентов следует проверить их соответствие используемым в настоящее время партиям реагентов. Эти рекомендации предназначены для медицинских изделий, уже зарегистрированных 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 исследования 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 МИ ИВД. Степень согласованности между результатами клинического лабораторного исследования и заявлениями производителя на упаковке можно оценить более глубоко, используя методы, описанные в разделе 8. Примеры лабораторной проверки утверждений о МИ ИВД приведены в приложении Л.

Если качественное исследование имеет ICR, доступный 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 согласованности разница в рекомендуемых процессах невелика, за исключением того, что исследование с ICR может помочь сотрудникам лаборатории 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 к упаковке, 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 приемлемости. По завершении анализа допускается подготовить протокол верификации, рабочую документацию, средства контроля и образцы для верификации и приступить к верификации.

9.2 Верификация прецизионности измерений

9.2.1 Прецизионность измерений 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 МИ ИВД. 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 прецизионности

9.2.2.1 Хотя те же факторы, которые влияют на количественные исследования, могут влиять на прецизионность 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.

Прецизионность качественных исследований, не имеющих зарегистрированных ICR, лучше всего оценивать с использованием образцов с ответами, достаточно близкими к cut-off, чтобы поставить под сомнение способность исследования правильно определять 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 прецизионность 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 воспроизводимости следует включить межлабораторный компонент.

9.2.2.2 Сходимость

Сходимость — близость 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 прецизионности оценку сходимости обычно проводят в течение многих серий или дней. Такой подход необходим, поскольку

однократные измерения сходимости могут не отражать тонких, неконтролируемых, временных изменений в методе испытания и/или работе прибора. В рамках каждой аналитической серии условия (например, оператор, партия реагентов, прибор) остаются постоянными, тестируется несколько повторов и оценивается прецизионность при различных концентрациях измеряемого вещества.

9.2.2.3 Воспроизводимость

Воспроизводимость — близость 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 ежедневную неточность и обеспечить адекватную оценку этого компонента погрешности. Если единственными неповторяющимися факторами, учитываемыми при исследовании, являются день за днем или постановка за постановкой, то можно сказать, что исследование воспроизводимости измеряет внутрилабораторную неточность. Другие типы исследований воспроизводимости могут измерять другие факторы, такие как оператор, прибор и/или место проведения исследования.

9.2.3 Верификация прецизионности без внутреннего непрерывного сигнала

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 C50 с использованием образцов, заявленных как положительные, более практичным способом установления разведения образца C50 для проверки может быть поэтапный подход с серийными разведениями, исследуемыми с меньшим количеством повторов на разведение. Контрольные материалы с установленными 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 повторов 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.

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 10).

Таблица 10 — ДИ вокруг ожидаемого числа положительных результатов

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

Окончание таблицы 10

Истинная доля выборки	95 % ДИ	*N*		
		20	30	40
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 — C5 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C25 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 25 % случаев; C50 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 50 % случаев; C75 — значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 75 % случаев; C95 — значение в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; ДИ — доверительный интервал; *N* — количество всех исследований.

Если исследуют 20 копий обоих образцов, а количество положительных результатов уменьшается с восьми положительных результатов в образце C50 до шести положительных результатов в 20 %-ном разведении, лаборатория может сделать вывод, что качественное исследование неприемлемо, поскольку оно не обладает необходимой прецизионностью. 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 %.

9.2.3.2 Анализ данных

Хотя исследование верификации по умолчанию должно проводиться как минимум в течение пяти дней, количество повторов 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 10 приведено ожидаемое количество положительных результатов при истинной базовой концентрации и количестве всех исследований (*N*). Нижнее и верхнее значения представляют собой 95 %-ные 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 Верификация прецизионности при наличии внутреннего непрерывного сигнала

Аналогично процедурам количественных исследований при качественном исследовании необходимо проверить неточность. Порог для качественного исследования — cut-off, установленный производителем для ICR, при превышении которого результат считается положительным, а при снижении — отрицательным или наоборот. Как уже было рассмотрено в 8.1, исследования прецизионности для качественных исследований должны давать оценку неточности исследования при концентрациях аналита вблизи cut-off.

9.3 Клинические показатели эффективности или согласованность

9.3.1 Клинические показатели эффективности

Для оценки клинических показателей лаборатории должны иметь практически равное количество образцов от субъектов, у которых, как известно, присутствуют ЦС, и от субъектов, у которых ЦС отсутствуют. Такой сценарий является наиболее желательным типом сравнения. Подходящие показатели для описания согласованности включают оценки диагностической чувствительности, диагностической 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(и) была(и) репрезентативной(ыми) для предполагаемой популяции. Однако некоторые образцы должны быть несколько близки к отсечению (например, более C5 и менее C95), чтобы эффективность исследования 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 референтные панели могут быть выгодны для эффективного и рационального использования ресурсов и времени эксперта, такие панели могут быть недоступны. Кроме того, контрольные 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 8.2 рассматривается анализ данных для тех видов сравнений, которые лаборатория, скорее всего, будет проводить не только для проверки клинических показателей или согласованности, но и для оценки величины очевидных различий в чувствительности и специфичности.

9.3.2 Согласованность

В менее идеальном случае сравнение может характеризоваться только согласованностью разработанного изделия 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 согласованности со сравнительным исследованием. Критерии согласованности зависят от сравниваемых изделий. Например, хотя одно или меньшее количество дискордантных отрицательных или дискордантных положительных результатов из 20 пар результатов может свидетельствовать о достаточной согласованности при сравнении исследований с аналогичными возможностями, для исследования, которое, как ожидается, будет более чувствительным, может потребоваться большее количество положительных результатов.

Расхождения между разработанным изделием и изделием сравнения могут возникнуть из-за ошибок в одном из них. Если сравнительное исследование не дает стопроцентного результата, образцы с расхождениями между результатами разработанного изделия и изделия сравнения 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 cut-off 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 согласованности

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

прецизионность 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 разработанного изделия вблизи cut-off.

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, соответствуют ли результаты разработанного изделия заявлениям производителя. Даже при 100 %-ной согласованности 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 11.

Таблица 11 — Примеры ДИ по шкале Уилсона

Результаты		Количество 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 %-ный ДИ	Средний	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 %-ный ДИ	Средний	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 — ДИ — доверительный интервал.						

9.3.4 Анализ данных

Во многих случаях лаборатории оценивают разработанное изделие только по изделию сравнения (см. таблицу 6); 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 4). В этих случаях диагностическая чувствительность и специфичность могут быть оценены по формулам (7) и (8) (см. 8.3.6) 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 с использованием разработанного изделия.

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 разработанным изделием и компаратором.

9.4 Исследование стабильности реагентов

Если лаборатории должны проверять стабильность реагентов, указанную во вкладыше производителя, следует обратиться к ГОСТ Р ИСО 23640. Если имеется ICR, его следует использовать в

контексте протоколов исследования, 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 С5 (или образец, не содержащий ЦС, если С5 не представляется возможным) и С95.

Если ICR недоступен, для отбора образцов для исследования стабильности реагента можно использовать положения согласно 9.3.4. В каждой выбранной временной точке можно использовать размер выборки и количество положительных результатов из таблицы 10, 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 временно́й 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 рекомендуется проверить ожидаемые характеристики медицинского изделия. Если ICR недоступен, следует провести 20 повторов образца с реакциями, близкими к cut-off (С50), используя новую и старую партии реагентов (общее количество тестов равно 40), и получить сопоставимую долю положительных результатов (см. таблицу 10 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 ($\alpha = 0,05$). Партия с меньшим числом положительных результатов должна считаться менее чувствительной. В этом случае для получения истинного С50 и применения критерия отказа более 7 количество положительных результатов в образцах С50 из ранней партии реагентов должно составлять от 8 до 12.

Приложение А
(справочное)

Примеры исследований

А.1 В таблицах А.1—А.3 приведены три различных типа исследований с примерами. Первый тип — качественные бинарные, отчитывающиеся о результатах исследования (качественные бинарные исследования). Второй тип — небинарные исследования с порядковыми и номинальными результатами с более чем двумя категоризациями. Третий тип — процедуры количественных измерений.

Таблица А.1 — Примеры качественных бинарных исследований

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 менее или более 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
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 анализа 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 ΔС между тестовыми образцами и контрольным образцом
Примечание — βhCG — хорионический гонадотропин человека; MRSA — метициллин (оксациллин) — устойчивый золотистый стафилококк; ПЦР — полимеразная цепная реакция; NAAT — тест амплификации нуклеиновых кислот; ΔС — изменение порога цикла.		

Таблица А.2 — Примеры небинарных исследований категоризации

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
Примечание — ANA — антинуклеарные антитела.		

Таблица А.3 — Примеры количественных процедур измерения

Процедуры количественных измерений		
Отчетный результат (значение)	Основание результата исследования	Пример технологии
Концентрация, пг/мл	Концентрация в пг/мл, но с cut-off клинического решения	Автоматизированный иммуноферментный анализ тропонина в сыворотке крови
Концентрация, мкг/мл	Концентрация в мкг/мл, но с клиническим отсечением	Автоматизированное иммунотурбидиметрическое определение димера D в плазме крови
Концентрация, нг/мл	Концентрация в нг/мл, но порог обнаружения 150 нг/мл	Иммуноферментный анализ для выявления наличия метаболитов кокаина в моче
Количество окрашенных клеток, на $мм^2$	Количество окрашенных клеток на $мм^2$ 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 α_1-, α_2-, β- и γ-глобулин)
Примечание — IHC — иммуногистохимия.		

Приложение Б
(справочное)

Связь между распределением внутреннего непрерывного сигнала и пробит- или логит-моделями

Б.1 В настоящем приложении применены следующие сокращения:

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

ICR — внутренний непрерывный сигнал.

Б.2 В настоящем приложении применены следующие обозначения:

P_{pos} — вероятность быть признанным положительным;

Pr — вероятность;

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 сигналов;

X — независимый (по оси *x*) масштаб;

Y — бинарный результат (1 = положительный, 0 = отрицательный);

Φ — кумулятивная нормальная функция распределения;

μ — среднее;

σ — стандартное отклонение.

Б.3 Значение внутреннего непрерывного сигнала (ICR) в соответствующей шкале, где бинарное исследование признает образец положительным в 5 % случаев (C5), и значение ICR по соответствующей шкале, где бинарное исследование признает образец положительным в 95 % случаев (C95), можно оценить с помощью представления измерений непрерывного сигнала или путем моделирования (построение кривой) доли положительных результатов исследования (процент попаданий) относительно истинных значений сигналов. Для определения этой зависимости обычно используют модели пробит и логит с использованием среднего значения нескольких определений в оговоренных условиях как неизвестного истинного значения сигнала. 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 cut-off.

Долю 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.

Пробит-модель для одного количественного прогностического фактора вычисляют по формуле

$$P_{pos}(X) = \Phi(a + bX), \quad (Б.1)$$

где $P_{pos}(X)$ — вероятность положительного результата обследования на уровне *X*;

Φ — кумулятивная функция нормального распределения;

a и *b* — неизвестные параметры, подлежащие оценке.

Преобразованная вероятность положительного результата линейна согласно формуле

$$\Phi^{-1}[P_{pos}(X)] = a + bX, \quad (Б.2)$$

где Φ^{-1} — обратное кумулятивное нормальное распределение.

Пробит-модель может быть описана нормальной моделью скрытых переменных или пороговой моделью. Нормальная модель латентных переменных предполагает наличие истинного ненаблюдаемого непрерывного отклика *R*, как показано в выражениях

$$Y = 0 \text{ (отрицательный), если } R \le \text{Cut-off}, \quad (Б.3)$$

$$Y = 1 \text{ (положительный), если } R \ge \text{Cut-off}. \quad (Б.4)$$

Если $R = a + b(X + e)$, а ошибки *e* независимы и нормально распределены со средним μ и дисперсией, равной σ^2, то вероятность того, что $Y = 1$, учитывая *X*, равна $Pr(Y = 1|X) = P_{pos}(X) = \Phi(a + bX)$, как показано в формуле (Б.1), при этом $\mu = -a/b$ = C50 и $\sigma = 1/|b|$. Пробит-модель — обобщенная линейная модель, которая имеет вид $\Phi^{-1}[Pr(Y = 1|X)] = a + bX$ с обратной нормальной функцией связи Φ^{-1}. Если вместо нормального распределения ошибок используют стандартное логистическое, то вместо пробит-модели используют логистическую модель.

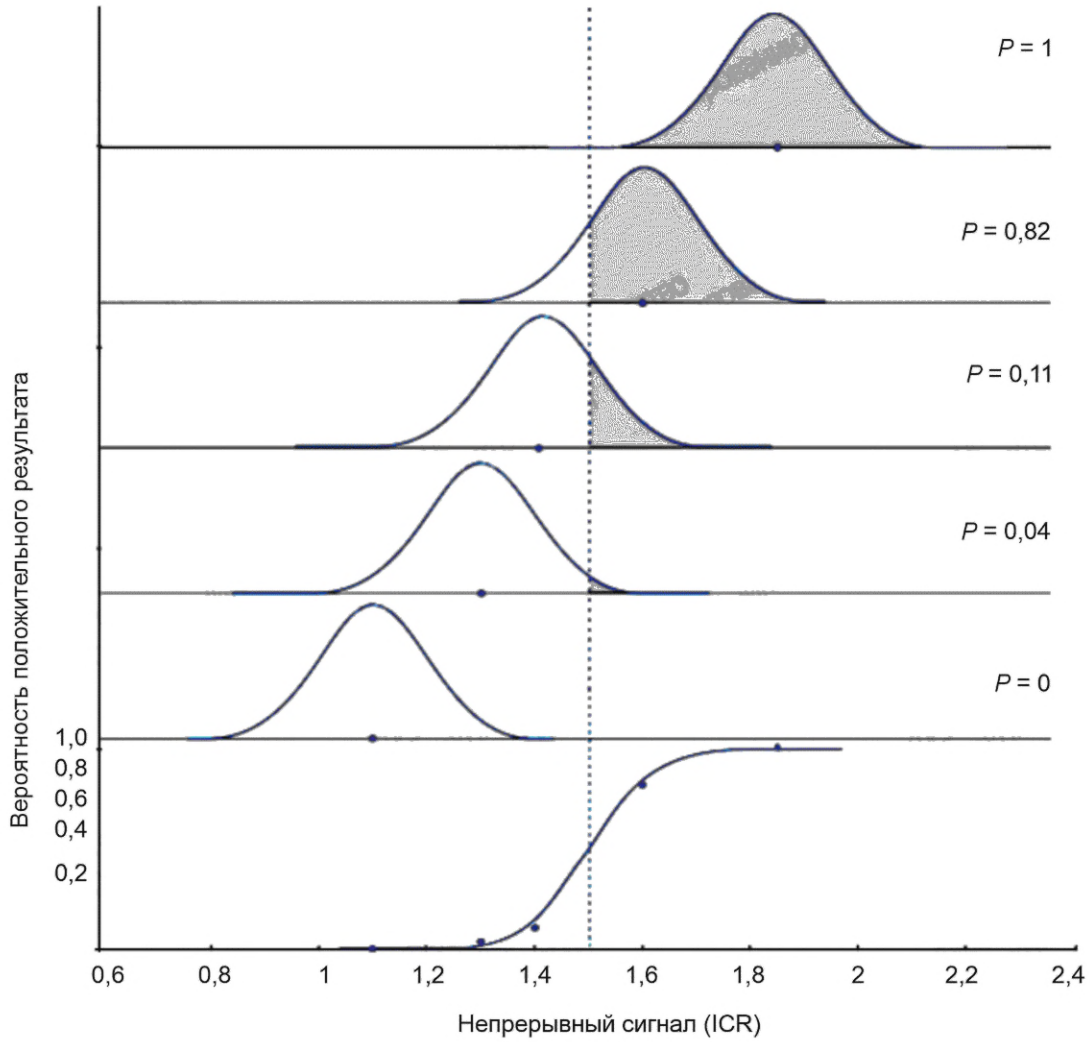


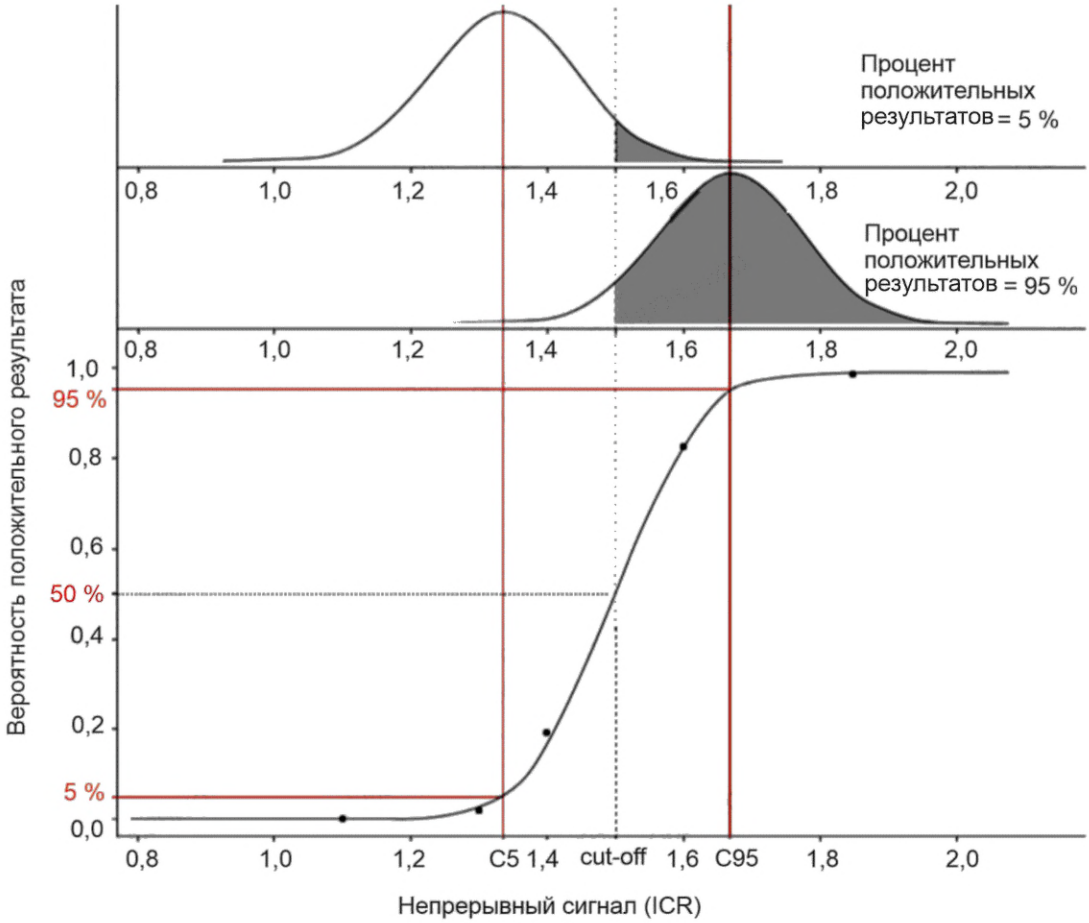
Рисунок Б.1 — Взаимосвязь между долей положительных результатов исследования и средним значением множественных гипотетических определений

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)

$$C5 = \frac{\Phi^{-1}(0{,}05) - a}{b} \text{ или } C5 = \frac{-1{,}645 - a}{b}, \quad (Б.5)$$

$$C50 = \frac{\Phi^{-1}(0{,}5) - a}{b} \text{ или } C50 = \frac{0 - a}{b}, \quad (Б.6)$$

$$C95 = \frac{\Phi^{-1}(0{,}95) - a}{b} \text{ или } C95 = \frac{1{,}645 - a}{b}. \quad (Б.7)$$



C5 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 — значение по соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; ICR — внутренний непрерывный сигнал

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

Если ошибки распределены нормально, но дисперсии не равны по уровням, стандартная регрессионная пробит-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 сигнал доступен (не скрытый), 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 внутренний сигнал 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 визуальной оценкой ответа

В.1 В настоящем приложении применены следующие сокращения:

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

ICR — внутренний непрерывный сигнал.

В.2 Качественное исследование с визуальной оценкой ответа — исследование, при котором категоризация проводится на основе наблюдения пользователя, а не с помощью приборов. Этот тип бинарного исследования встречается довольно часто и не очень легко вписывается в схему категоризации, как другие качественные бинарные исследования, описанные в настоящем стандарте. Такие исследования не имеют внутреннего непрерывного сигнала (ICR), поскольку их бинарные результаты не зависят от инструментов. Однако ответ «да» или «нет» может имитировать порог 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 8.4 и 9.3.

Примерами исследований с визуальной оценкой являются, например, картриджи для теста на беременность, контейнеры для скрининга мочи на наркотики, 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 большей по сравнению с оценкой прибором. 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 визуальной оценкой ответа

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, проведение 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. Должна быть построена зависимость частоты попаданий, или доля случаев, когда результат пробы признается положительным (ось *y*), от значений измеряемого вещества для каждой пробы (ось *x*), и должна быть добавлена кривая, которая 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 неточностью, что может быть достигнуто путем:

- выполнения измерений с помощью процедуры количественных референтных измерений;
- усреднения нескольких определений, которые охватывают соответствующие компоненты дисперсии из выбранной процедуры количественного измерения;
- создания серии образцов с добавлением референсного материала (если есть возможность, следует использовать сертифицированный референсный материал);
- внесения референсного материала в один образец с высокой концентрацией и проведения 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 аппроксимации 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.

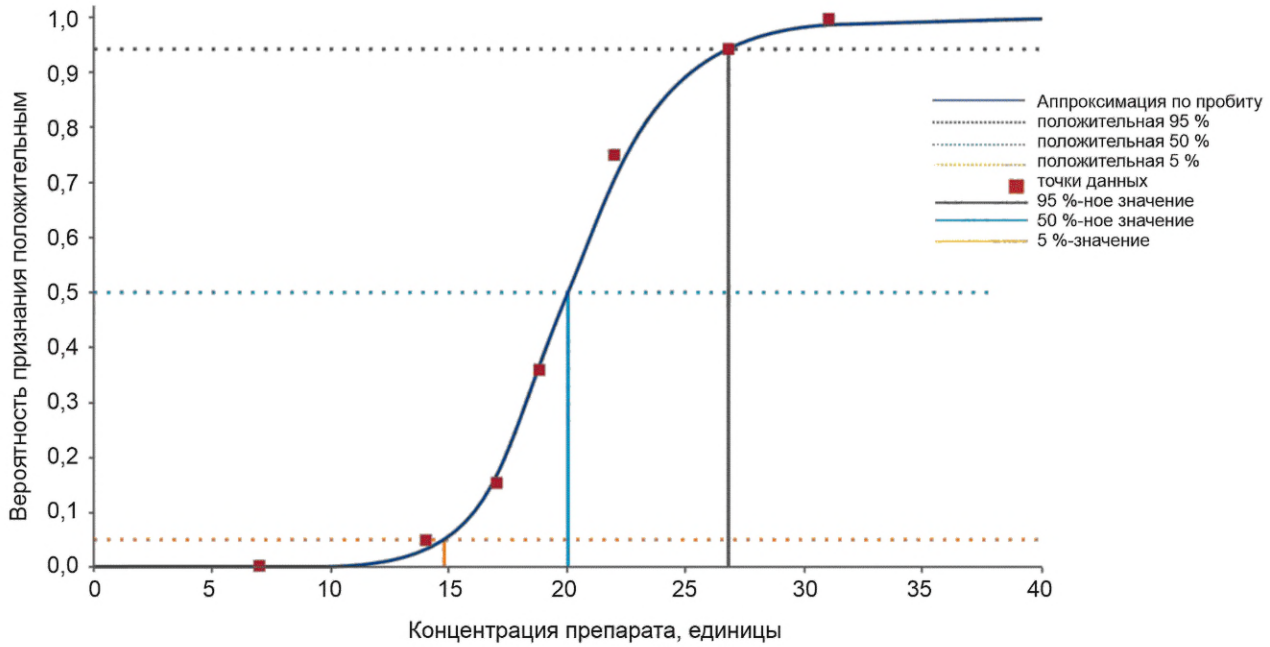


Рисунок В.1 — Пример качественного исследования с визуальной оценкой

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 контрольные материалы. В процессе разработки такие образцы могут использоваться для определения характеристик тестовой полоски. После определения свойств исследования 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 по 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 тест-систем для исследования (например, тест-полосок). Метод измерения вариабельности между наблюдателями приведен в приложении Ж. Несколько партий тест-системы должны использоваться несколькими разными операторами, которые представляют «типичных пользователей». Количество необходимых операторов 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 визуальной оценкой — отсутствие ICR. Эта особенность влияет на некоторые оценки эффективности. Исследования точности для таких исследований рассматривают в 8.3.7. На оценку клинических характеристик, описанную в 8.4, отсутствие ICR не влияет. Измерение стабильности реагентов для исследований с визуальной оценкой описано в 7.5. Наконец, измерение влияния интерферирующих веществ или перекрестных реактивов, а также определение влияния других интерферирующих факторов для

таких исследований рассматривают в 8.6. Определение влияния интерферирующих 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 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 полоски (например, полоска для анализа мочи на наркотики). В этих случаях исследование больше не зависит от пользователя, поскольку теперь имеется ICR. Разработку данного типа инструментальных исследований рассматривают в 7.2.

Приложение Г
(справочное)

Оценка точности секвенирования следующего поколения

Г.1 В настоящем приложении применены следующие сокращения:

ДИ — доверительный интервал;

СХ — внутрилабораторное значение по 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 cut-off 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 по соответствующей шкале, где бинарное исследование признает образец положительным в 5 % случаев (C5), и внутрилабораторного значения по 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.

Таблица Г.1 — Программа исследования: перекрестная и вложенная модель с указанием дня исследования, оператора и прибора

Исследования	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

Этот план исследования охватывает 24 дня и считается перекрестной и вложенной моделью, поскольку инструменты и операторы перекрещиваются, а день полностью вложен в каждую комбинацию инструментов и операторов.

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

Исследования	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 пробы каждого уровня образца в исследовании. Таблица Г.3 описывает логистику плана постановки исследования на основе таблицы Г.2.

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

Исследования	День начала																							
	1				2				3				4				5				6			
	Неделя 1 — понедельник				Неделя 1 — среда				Неделя 1 — пятница				Неделя 2 — понедельник				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 одной и той же программы 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 такой программе для каждой выборки используется модель со случайными эффектами согласно формуле

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_{k(ij)} + \varepsilon_{l(ijk)}, \quad (Г.1)$$

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;

$\delta_{k(ij)}$ — ошибки, связанные с днем, включая оператора и прибор;

$\varepsilon_{l(ijk)}$ — ошибки, связанные с любой остаточной вариацией, включая повторы, каждая из которых нормально распределена со средним нулем и дисперсией σ^2, описанной ниже.

$$\alpha_i \sim N(0, \sigma^2_{\text{прибора}}),$$

$$\beta_j \sim N(0, \sigma^2_{\text{оператора}}),$$

$$\gamma_{ij} \sim N(0, \sigma^2_{\text{прибора} \cdot \text{оператора}}),$$

$$\delta_{k(ij)} \sim N(0, \sigma^2_{\text{день (прибора} \cdot \text{оператора}}),$$

$$\varepsilon_{l(ijk)} \sim N(0, \sigma^2).$$

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

Окончание таблицы Г.4

Фактор	Расчет DF	Уравнение DF
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 сигнала

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.

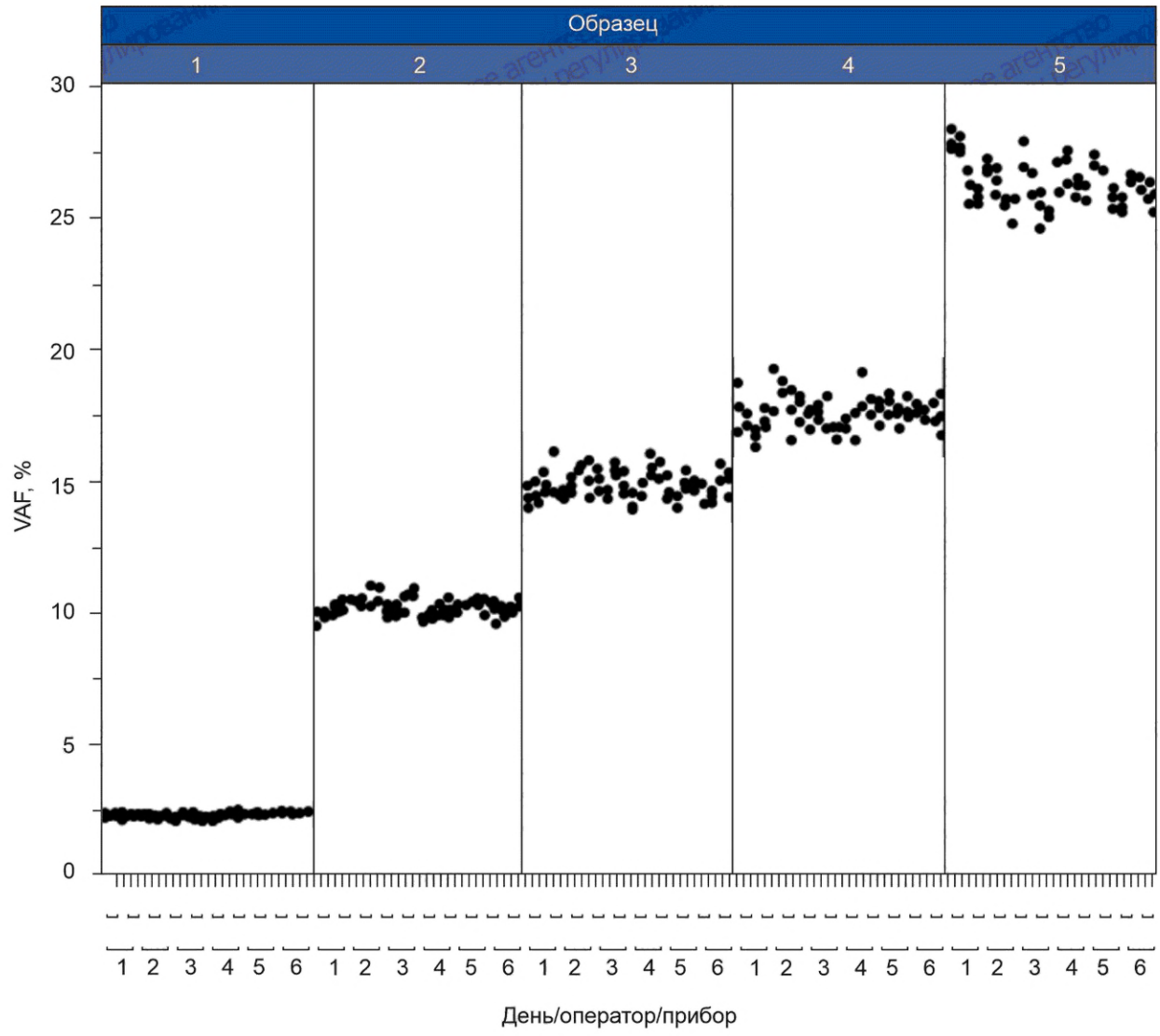


Рисунок Г.1 — Точечные графики наблюдаемых VAF для каждого образца

Разброс (изменчивость) между наблюдениями меньше для образцов с более низким уровнем VAF и увеличивается для образцов с более высоким уровнем VAF.

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 cut-off. 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

Идентификатор образца	*N*	Среднее	Прибор		Оператор		Прибор-оператор		День, прибор, оператор		Повтор		Итого[1)]	
			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

[1)] Включает прибор, оператора, прибор-оператор, день (прибор, оператор) и повтор.

Примечание — CV — коэффициент вариации, %; *N* — количество образцов; SD — стандартное отклонение.

Внутрилабораторные значения C5 и C95 оценивают по формулам:

$$C5 = \text{cut-off исследования} - 1{,}645 \cdot SD_{C5}, \quad (Г.2)$$

$$C95 = \text{cut-off исследования} + 1{,}645 \cdot SD_{C95}, \quad (Г.3)$$

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 общим SD, показанным в таблице Г.5. SD отдельной партии составляет 0,005 (постоянная для всех уровней VAF), а внутрилабораторную изменчивость рассчитывают как квадратный корень из ($0{,}005^2$ + общим 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	Итого[1)]		Внутрилабораторные[2)]	
			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
*S*5	72	25,66	1,04	4,0	1,04	4,0

[1)] Включает прибор, оператора, прибор-оператор, день (прибор, оператор) и повтор.

[2)] Включает прибор, оператора, прибор-оператор, день (прибор, оператор), партию реагентов и повтор.

Примечание — CV — коэффициент вариации, %; *N* — количество образцов; SD — стандартное отклонение.

Взаимосвязь между внутрилабораторным SD и средней VAF изучают путем построения графика зависимости соответствующего внутрилабораторного SD от средней VAF для каждого образца. Внутрилабораторное SD (и SD повторяемости) не является постоянным для всех оцениваемых уровней, а линейно увеличивается по мере увеличения уровня образца (процент мутационных VAF). Параметры линейной зависимости точности представлены в таблице Г.7, а график — на рисунке Г.2.

Таблица Г.7 — Оценки профиля точности

Показатель	Значение	Стандартная ошибка
Отрезок (intersept)	–0,02633	0,057
Наклон (slope)	0,04066	0,004

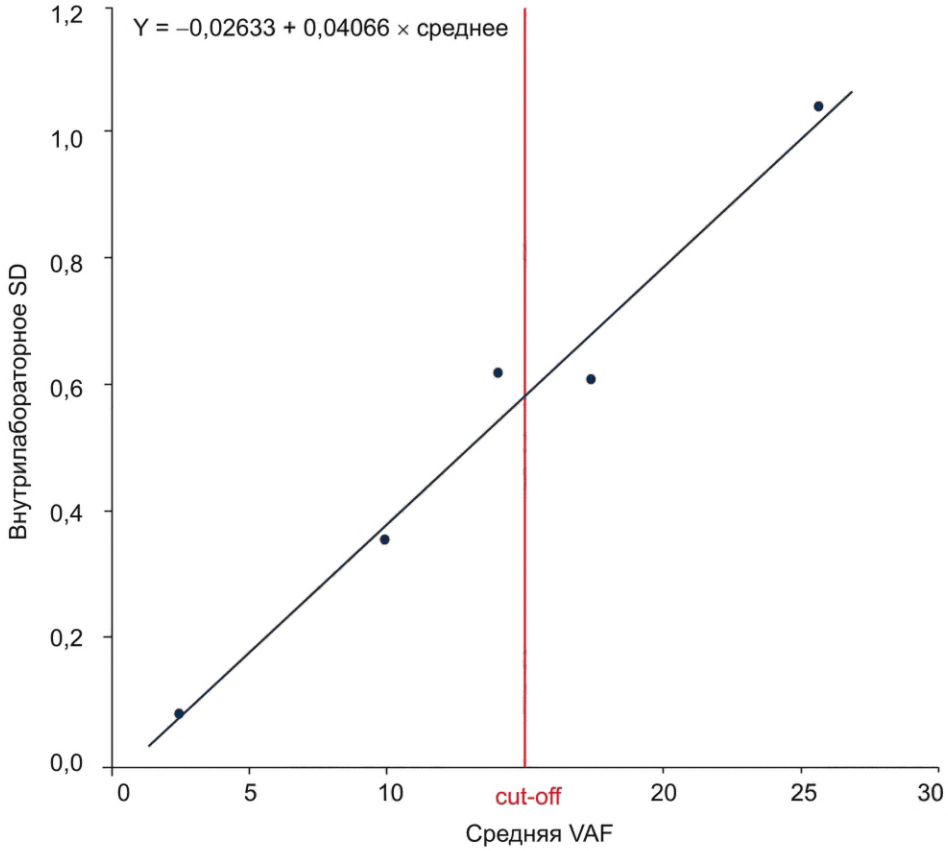


Рисунок Г.2 — Профиль прецизионности внутрилабораторного SD (оператор, прибор, день начала секвенирования, партия и повтор) в зависимости от среднего процента VAF с наложением линейной модели

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

$$SD_{\text{средняя VAF}} = \text{Средняя VAF} \cdot \text{Наклон} + \text{Отрезок}. \quad (Г.4)$$

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

$$C95 = \frac{\text{Cut-off исследования} + 1{,}645 \cdot \text{Отрезок}}{1 - 1{,}645 \cdot \text{Наклон}}. \quad (Г.5)$$

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.

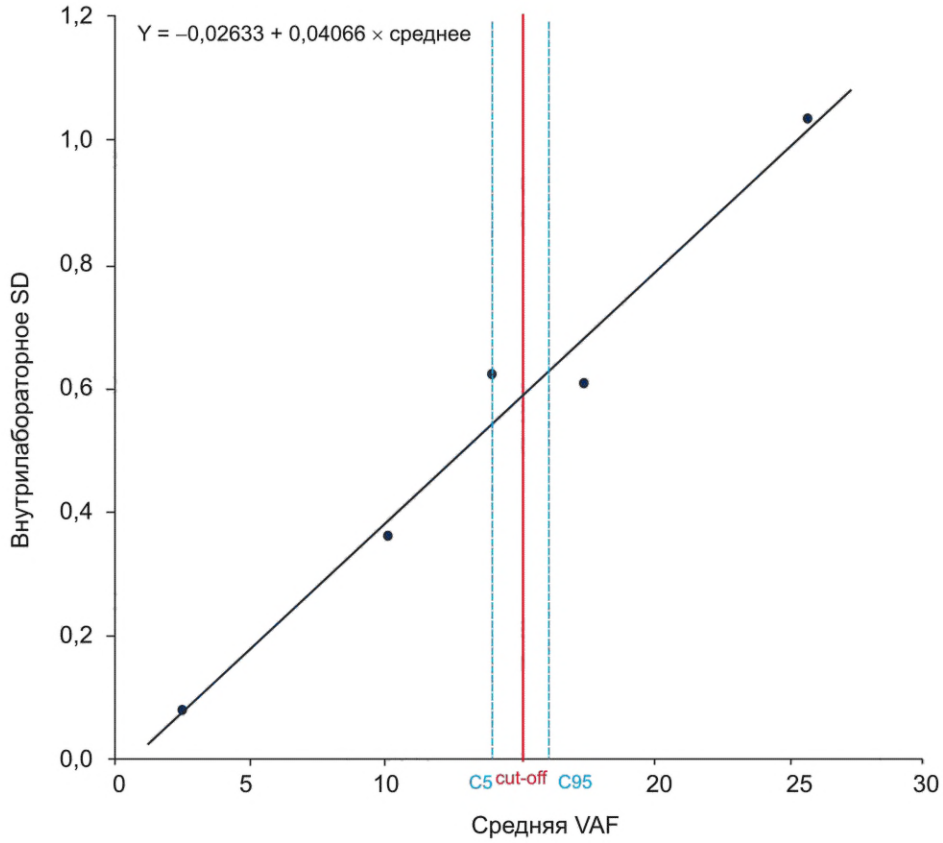
$$C5 = \frac{\text{Cut-off исследования} - 1{,}645 \cdot \text{Отрезок}}{1 + 1{,}645 \cdot \text{Наклон}}. \quad (Г.6)$$

При наклоне 0,04066 и отрезке — 0,02633 значения для C5 и C95 рассчитывают следующим образом:

$$C5 = \frac{15 - 1{,}645 \cdot (-0{,}02633)}{1 + 1{,}645 \cdot 0{,}04066} = \frac{15{,}04331}{1{,}066886} = 14{,}1,$$

$$C95 = \frac{15 + 1{,}645 \cdot (-0{,}02633)}{1 - 1{,}645 \cdot 0{,}04066} = \frac{14{,}95669}{0{,}933114} = 16{,}0.$$

Для этого исследования порог составляет 15 % VAF, поэтому интервал от C5 до C95 составляет 14,1 % — 16,0 %. Расчетное SD при C5 равно 0,547 VAF, а SD при C95 равно 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 были отмечены точки относительно средних значений уровня. Были определены порог 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 линейная модель, соответствующая данным.



C5 — значение внутрилабораторного SD по соответствующей шкале, при котором бинарное исследование признает образец положительным в 5 % случаев; C95 — значение внутрилабораторного SD по соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев

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 наблюдаемая VAF ≥ 15 % (отсечение). Для образцов со средним значением VAF значительно ниже 15 результаты ожидают отрицательными; для образцов со средним значением VAF значительно выше 15 результаты ожидают положительными; а для образцов со средним значением 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 в выборке

Идентификатор образца	Средняя VAF	Ожидаемые результаты	Количество попыток	QC неудачи	Количество положительных результатов из достоверных проб[1)]	Процент положительных результатов (95 % ДИ[2)])	Процент отрицательных результатов (95 % ДИ[2)])
*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 отрицательных; 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)

[1)] В исследовании используют только образцы, прошедшие контроль качества анализа («валидные»).
[2)] Двусторонние 95 %-ные ДИ рассчитаны по методу оценки Уилсона.

Примечание — ДИ — доверительный интервал; QC — контроль качества; VAF — частота вариантов аллелей.

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	Оператор	День	Количество попыток	Количество отказов QC	Количество положительных результатов из достоверных проб[1)]	Количество отрицательных результатов из достоверных проб[1)]	Процент положительных результатов	Процент отрицательных 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

Окончание таблицы Г.9

Прибор	Оператор	День	Количество попыток	Количество отказов QC	Количество положительных результатов из достоверных проб[1)]	Количество отрицательных результатов из достоверных проб[1)]	Процент положительных результатов	Процент отрицательных результатов
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

[1)] Для включения в исследование образцы должны пройти контроль процесса исследования (если применимо) и контроль качества.

Примечание — 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

Д.1 В настоящем приложении применены следующие сокращения:

ДНК — дезоксирибонуклеиновая кислота;

ПЦР — полимеразная цепная реакция;

LLoD — нижний предел обнаружения (согласно методу частоты попаданий*).

Д.2 В настоящем приложении применены следующие обозначения:

Pr — вероятность;

b — параметр логита;

c — параметр логита;

e — уравнение Пуассона;

μ — ожидаемое количество копий.

Д.3 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. Диагностическая 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

$$Pr(\text{Обнаружение}) = Pr(\text{Обнаружение/Аналит присутствует}) \cdot Pr(\text{Аналит присутствует}). \quad (Д.1)$$

Pr(Аналит присутствует) из формулы (Д.1) определяют как 1 — уравнение Пуассона для вероятности отсутствия аналита в образце в соответствии с формулой**

$$Pr(\text{Аналит присутствует}) = 1 - e^{-\mu}, \quad (Д.2)$$

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/мл) на объем образца (например, мл) исследования. Например, μ = 4 копии вируса/мл — 0,5 мл/образец = 2 нити целевой ДНК/образец.

Pr(Обнаружение/Аналит присутствует) из формулы (Д.1) определяют с помощью логит-уравнения, представленного в виде формулы (Д.3).

$$Pr\left(\text{Обнаружение/Аналит присутствует}\right) = \frac{1}{1 + \left(\frac{\mu}{c}\right)^{-b}}, \quad (Д.3)$$

где *Pr*(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 %-ный 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.

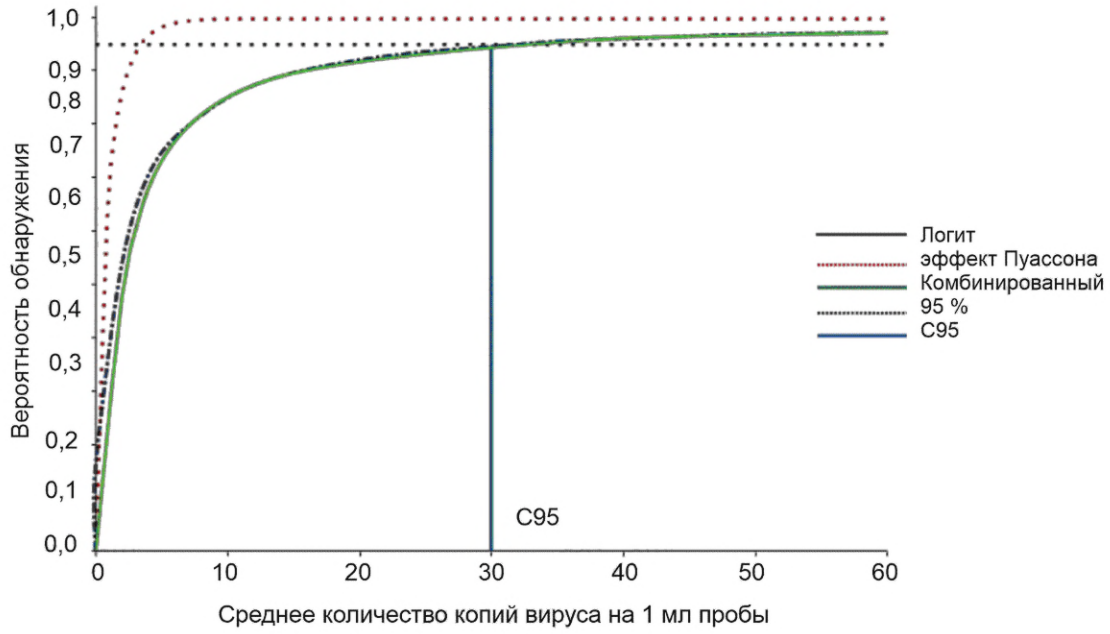
На рисунке Д.1 LLoD определен на уровне 30. Модель Пуассона предполагает, что для исследования используется образец объемом 1 мл; следовательно, измеряемой величиной является концентрация в копиях/мл

---

* 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.

** Более подробная информация приведена в работе Снедекора Г.У., Кокрана Р.Г. «Статистические методы». 6-е изд. Издательство Университета штата Айова; 1967.

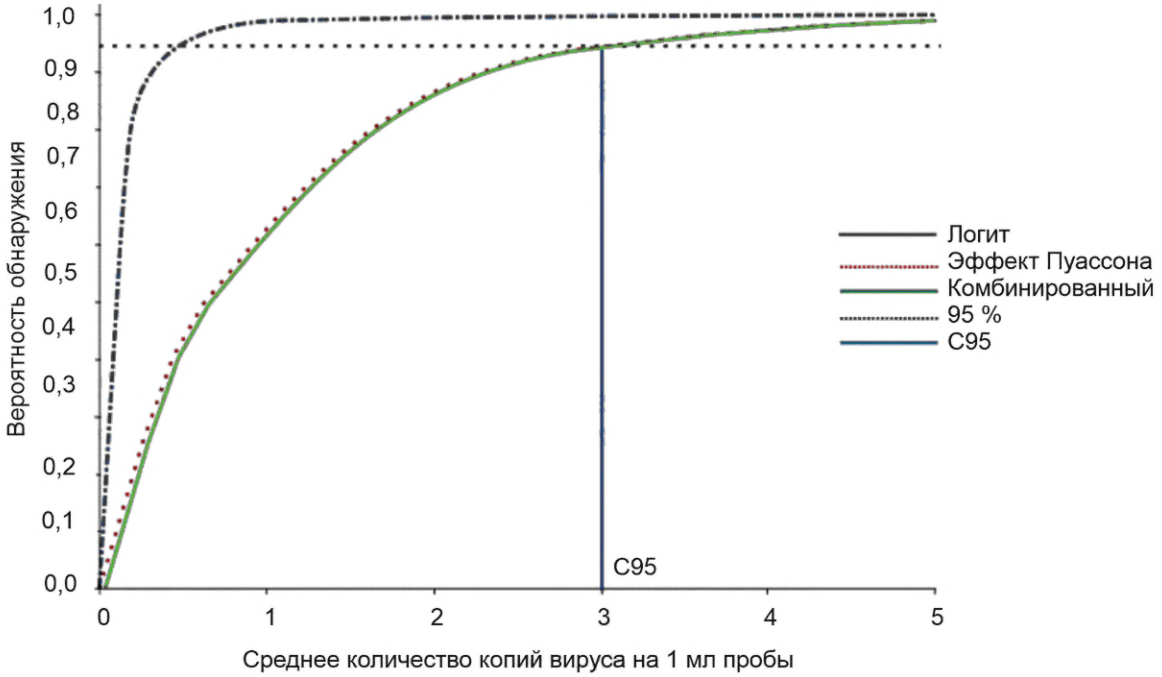
целевого вируса в образце. LLoD измеряют при гораздо более высоком значении, чем то, которое может дать пуассоновский расчет. Поэтому предполагается, что комбинированное соответствие обусловлено почти исключительно диагностической чувствительностью исследования.



C95 — число копий аналита в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; логит — *Pr*(Обнаружение/Аналит присутствует); эффект Пуассона — *Pr*(Аналит присутствует)

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

На рисунке Д.2 приведено изделие, которое является высокочувствительным при обнаружении аналита. При тех же допущениях, что и на рисунке Д.1, измерение LLoD будет зависеть почти исключительно от того, присутствует ли в образце нить целевой нуклеиновой кислоты (т. е. от распределения Пуассона). Таким образом, диагностическая чувствительность изделия не может быть измерена напрямую. Пример такого исследования приведен в 7.3.



C95 — число копий аналита в соответствующей шкале, при котором бинарное исследование признает образец положительным в 95 % случаев; логит — *Pr*(Обнаружение/Аналит присутствует); эффект Пуассона — *Pr*(Аналит присутствует)

Рисунок Д.2 — Высокочувствительное обнаружение NAAT (тест амплификации нуклеиновых кислот)

Приложение Е
(справочное)

Схемы прецизионных исследований

Е.1 В настоящем приложении применено следующее сокращение:

DF — степени свободы.

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

Е.3 Внутрилабораторные исследования

Программу можно рассматривать как полностью перекрестную, 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 два повтора для каждой из 12 комбинаций из трех партий реагентов, двух приборов и двух операторов, оцениваемых в один день. Вся схема исследования повторяется в течение пяти разных дней для каждого оцениваемого образца и уровня.

Таблица Е.1 — Программа примера ежедневного внутрилабораторного прецизионного исследования

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 этой программы, можно определить DF, как показано в таблице Е.2.

Таблица Е.2 — Влияние DF на пример программы 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 · оператор[1)]	(Количество приборов – 1) · (количество операторов – 1)	1 · 1 = 1
Прибор · партия реагентов[1)]	(Количество приборов – 1) · (количество партий реагентов – 1)	1 · 2 = 2
Прибор · день[1)]	(Количество приборов – 1) · (количество дней – 1)	1 · 4 = 4
Оператор · партия реагентов[1)]	(Количество операторов – 1) · (количество партий реагентов – 1)	1 · 2 = 2
Оператор · день[1)]	(Количество операторов – 1) · (количество дней – 1)	1 · 4 = 4
Партия реагентов · день[1)]	(Количество партий реагентов – 1) · (количество дней – 1)	2 · 4 = 8

Окончание таблицы Е.2

Фактор	Расчет DF	Уравнение DF
Прибор · оператор · партия реагентов[1)]	(Количество приборов – 1) · (количество операторов – 1) · (количество партий реагентов – 1)	1 · 1 · 2 = 2
Прибор · оператор · день[1)]	(Количество приборов – 1) · (количество операторов – 1) · (количество дней – 1)	1 · 1 · 4 = 4
Прибор · партия реагентов · день[1)]	(Количество приборов – 1) · (количество партий реагентов – 1) · (количество дней – 1)	1 · 2 · 4 = 8
Оператор · партия реагентов · день[1)]	(Количество операторов – 1) · (количество партий реагентов – 1) · (количество дней – 1)	1 · 2 · 4 = 8
Прибор · оператор · партия реагентов · день[1)]	(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

[1)] Условия взаимодействия должны быть изучены на стадии разработки и, как ожидается, будут близки к нулю. На стадии валидации условия взаимодействия не включают*.

Примечание — DF — степени свободы.

Если невозможно, чтобы каждый оператор запускал один и тот же прибор или лот 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 лоты реагентов в разные дни (с двумя повторами 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 — Пример альтернативной программы 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

* Дополнительная информация приведена в CLSI. Оценка точности процедур количественных измерений; утвержденное руководство — третье издание. Документ CLSI EP05-A3. Институт клинических и лабораторных стандартов, 2014.

Таблица Е.4 — Пример программы внутрилабораторного подтверждения точности исследования по дням

Очередность исследований	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-й и 7-й	4-й и 10-й	4-й и 10-й	1-й и 7-й	2-й и 8-й	5-й и 11-й	2-й и 8-й	5-й и 11-й	3-й и 9-й	6-й и 12-й	3-й и 9-й	6-й и 12-й

Для каждого образца имеется 48 результатов (два прибора · два оператора · три лота 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 программы внутрилабораторного подтверждения точности 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 · оператор[1)]	(Количество инструментов – 1) · (количество операторов – 1)	1 · 1 = 1
Прибор · партия реагентов[1)]	(Количество приборов – 1) · (количество партий реагентов – 1)	1 · 2 = 2
Оператор · партия реагентов[1)]	(Количество операторов – 1) · (количество партий реагентов – 1)	1 · 2 = 2
Прибор · оператор · партия реагентов[1)]	(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 лота реагентов – 1)	2 · 2 · 3 · (2 – 1) = 12
Повторы	Количество приборов · количество операторов · количество партий реагентов · количество дней внутри прибора и операторов · (количество повторов – 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 повторов – 1	2 · 2 · 3 · 2 · 2 – 1 = 47

[1)] Условия взаимодействия должны быть изучены на стадии разработки и, как ожидается, будут близки к нулю. На стадии валидации условия взаимодействия не включают.

Примечание — DF — степени свободы.

Е.4 Воспроизводимость

При реализации программы воспроизводимости исследований место проведения становится источником вариабельности и должно быть частью программы, а также модели исследования. В таблицах Е.6 и Е.7 приведены примеры программы с учетом двух операторов на участок, одного прибора на участок и, в данном случае, шести повторов на уровень за аналитическую серию.

Таблица Е.6 — Пример программы исследования воспроизводимости

Исследования	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-й	День 3-й	День 5-й	День 2-й	День 4-й	День 6-й	День 1-й	День 3-й	День 5-й	День 2-й	День 4-й	День 6-й	День 1-й	День 3-й	День 5-й	День 2-й	День 4-й	День 6-й
Повторы	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 программу 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 повторов – 1)	3 · 2 · 3 · (6 – 1) = 90
Всего	Количество участков · количество операторов · количество дней · количество повторов – 1	3 · 2 · 3 · 6 – 1 = 107
Примечание — 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, чтобы план исследования воспроизводимости обеспечивал не менее 20 DF для повторяемости. Схема, представленная в таблице Е.7, обеспечивает 90 DF для изменчивости повторов (повторяемости).

Приложение Ж
(справочное)

Исследования точности наблюдателей

Ж.1 В настоящем приложении применены следующие сокращения:

ДИ — доверительный интервал;

ANA — средняя отрицательная согласованность;

APA — средняя положительная согласованность;

IHC — иммуногистохимический;

NPA — отрицательная процентная согласованность;

PPA — положительная процентная согласованность.

Ж.2 В настоящем приложении применены следующие обозначения:

k — каппа-статистика;

i — наблюдатель 1;

j — наблюдатель 2;

L — нижний;

p — особая пропорция;

U — верхний;

w — вес;

y — количество образцов.

Ж.3 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 HER2 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 согласованности между двумя наблюдателями допускается рассматривать положительную процентную согласованность (PPA) и отрицательную процентную согласованность (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 8.3.7. Желательно использовать меры согласованности между двумя наблюдателями, которые не зависят от того, кто из наблюдателей выступает в качестве компаратора для оценки другого. Одной из таких пар мер является средняя отрицательная согласованность (ANA) и средняя положительная согласованность (APA), определяемые по формулам (Ж.1) и (Ж.2), соответственно

$$\mathrm{ANA} = \frac{2y_{11}}{y_{1\cdot} + y_{\cdot 1}}, \qquad (Ж.1)$$

$$\mathrm{APA} = \frac{2y_{22}}{y_{2\cdot} + y_{\cdot 2}}, \qquad (Ж.2)$$

где y_{11} — выборка, где оба наблюдателя дали отрицательный результат;

$y_{1\cdot}$ — количество отрицательных результатов у первого наблюдателя;

$y_{\cdot 1}$ — количество отрицательных результатов у второго наблюдателя;

y_{22} — выборка, где оба наблюдателя дали положительный результат;

$y_{2\cdot}$ — количество положительных результатов у первого наблюдателя;

$y_{\cdot 2}$ — количество положительных результатов у второго наблюдателя (см. таблицу Ж.1).

Таблица Ж.1 — Условные обозначения для подсчета парных бинарных результатов от наблюдателей 1 и 2

Наблюдатель 1	Наблюдатель 2		Итого[1), 2)]
	–	+	
–	y_{11}	y_{12}[4)]	$y_{1\cdot}$
+	y_{21}[5)]	y_{22}	$y_{2\cdot}$
Итого[2), 3)]	$y_{\cdot 1}$	$y_{\cdot 2}$	$y_{\cdot\cdot}$

Окончание таблицы Ж.1

1) $y_{1\cdot} = y_{11} + y_{12}$, $y_{2\cdot} = y_{21} + y_{22}$.
2) $y_{\cdot\cdot} = y_{1\cdot} + y_{2\cdot} = y_{\cdot 1} + y_{\cdot 2}$.
3) $y_{\cdot 1} = y_{11} + y_{21}$, $y_{\cdot 2} = y_{12} + y_{22}$.
4) y_{12} — выборка, где первый наблюдатель дал отрицательный результат, а второй наблюдатель — положительный результат.
5) y_{21} — выборка, где первый наблюдатель дал положительный результат, а второй наблюдатель — отрицательный результат.

ANA допускается пересчитывать как средневзвешенное значение согласно формуле

$$ANA = w_{-} NPA_{12} + (1 - w_{-}) NPA_{21}, \quad (Ж.3)$$

где $w_{-} = y_{\cdot 1}/(y_{1\cdot} + y_{\cdot 1})$ — вес, предельное общее количество отрицательных образцов $y_{\cdot 1}$ для наблюдателя 2 над суммой предельных общих количеств отрицательных образцов для наблюдателей 1 и 2;

$NPA_{12} = y_{11}/y_{\cdot 1}$ — NPA наблюдателя 1 с наблюдателем 2 (выражают как долю отрицательных образцов наблюдателя 2, которые наблюдатель 1 также считает отрицательными);

$NPA_{21} = y_{11}/y_{1\cdot}$ — NPA наблюдателя 2 с наблюдателем 1.

Аналогичным образом APA допускается пересчитывать как средневзвешенное значение по формуле

$$APA = w_{+}PPA_{12} + (1 - w_{+}) PPA_{21}, \quad (Ж.4)$$

где $w_{+} = y_{2\cdot}/(y_{2\cdot} + y_{\cdot 2})$ — вес, основанный на предельных суммах положительных выборок для двух наблюдателей;

$PPA_{12} = y_{22}/y_{\cdot 2}$ — PPA наблюдателя 1 с наблюдателем 2;

$PPA_{21} = y_{22}/y_{2\cdot}$ — PPA второго наблюдателя 2 с наблюдателем 1.

В качестве альтернативы, ANA и APA допускается интерпретировать как NPA и PPA для оценки случайно выбранного наблюдателя против другого (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

$$ANA = \frac{2 \cdot 70}{75 + 72} = 0{,}952 \text{ или } 95{,}2\ \%,$$

$$APA = \frac{2 \cdot 13}{15 + 18} = 0{,}788 \text{ или } 78{,}8\ \%.$$

Доверительный интервал (ДИ) для ANA может быть получен, если учесть, что $ANA = 2p_{-}/(1 + p_{-})$, где $p_{-} = y_{11}/(y_{12} + y_{\cdot 1})$ — 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 $[2p_{-}^{L}/(1 + p_{-}^{L}), 2p_{-}^{U}/(1 + p_{-}^{U})]$.

Аналогично, ДИ на APA можно получить, отметив, что $APA = 2p_{+}/(1 + p_{+})$, где $p_{+} = y_{22}/(y_{21} + y_{\cdot 2})$ — специальная пропорция, на которой можно получить ДИ $[p_{+}^{L}, p_{+}^{U}]$. Поэтому ДИ на APA — это $[2p_{+}^{L}/(1 + p_{+}^{L}), 2p_{+}^{U}/(1 + p_{+}^{U})]$.

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 значения 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 была первоначально предложена Дайсом, который назвал его индексом совпадения. В радиологии его обычно называют коэффициентом сходства Дайса. В обзоре индексов согласованности между двумя наблюдателями Флейсс отметил, что коррекция ANA или APA на случайную согласованность дает каппа-статистику (k). Чтобы разрешить кажущиеся парадоксы с общей согласованностью и k, Чиккетти и Файнштейн предложили называть ANA и APA p_- и p_+, соответственно. Исследователи спорят о клиническом значении p_- и p_+ и следует ли учитывать их выборочные распределения. Были предложены приблизительные ДИ для p_- и p_+, но предпочтительным является метод ДИ, описанный выше. Были предложены общие рекомендации по представлению результатов исследований надежности и согласованности.

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 В настоящем приложении применены следующие сокращения:

ДИ — доверительный интервал;

$ДС_{РИ}$ — диагностическая специфичность разработанного изделия;

$ДС_{ПИ}$ — диагностическая специфичность ранее применяемого изделия;

$ДЧ_{РИ}$ — диагностическая чувствительность разработанного изделия;

$ДЧ_{ПИ}$ — диагностическая чувствительность ранее применяемого изделия.

И.2 В настоящем приложении применены следующие обозначения:

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;

$\Delta_{ДЧ}$ — разница между чувствительностями;

$\Delta_{ДС}$ — разница между специфичностями.

И.3 Доверительный интервал пропорции

Примером использования метода баллов является расчет доверительного интервала (ДИ) для оценки диагностической чувствительности с использованием данных из таблицы И.1.

Таблица И.1 — Качественное разработанное изделие в сравнении с оценкой ЦС

Результат разработанного изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	A	B	$A + B$
Отрицательный	C	D	$C + D$
Всего	$A + C$	$B + D$	$A + B + C + D$
[1)] Наилучшая имеющаяся оценка ЦС.			

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

$$ДИ = \left(\frac{Q_1 - Q_2}{Q_3}, \frac{Q_1 + Q_2}{Q_3} \right), \quad (И.1)$$

где Q_1, Q_2 и Q_3 вычисляют по данным таблицы И.1 по формулам (И.2)—(И.4):

$$Q_1 = 2A + 1{,}96^2 = 2A + 3{,}84, \quad (И.2)$$

$$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\frac{AC}{A+C}} = 1{,}96\sqrt{3{,}84 + 4\frac{AC}{A+C}}, \quad (И.3)$$

$$Q_3 = 2(A + C + 1{,}96^2) = 2(A + C + 3{,}84). \quad (И.4)$$

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 разработанное изделие и ранее применяемое изделие с оценкой ЦС, можно получить следующие значения, как показано в таблице И.2.

Таблица И.2 — Трехстороннее сравнение между разработанным изделием, ранее применяемым изделием и оценкой ЦС

Оценка ЦС[1)]				Оценка ЦС[1)]			
ЦС присутствует				ЦС отсутствует			
Результат разработанного изделия	Результат ранее применяемого изделия		Всего	Результат разработанного изделия	Результат ранее применяемого изделия		Всего
	Положительный	Отрицательный			Положительный	Отрицательный	
Положительный	A_1	B_1	$A_1 + B_1$	Положительный	A_0	B_0	$A_0 + B_0$
Отрицательный	C_1	D_1	$C_1 + D_1$	Отрицательный	C_0	D_0	$C_0 + D_0$
Всего	$A_1 + C_1$	$B_1 + D_1$	N_1	Всего	$A_0 + C_0$	$B_0 + D_0$	N_0

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — N — количество образцов; ЦС — целевое состояние.

Таблица И.3 — Разработанное изделие в сравнении с оценкой ЦС

Результат разработанного изделия	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	$A_1 + B_1$	$A_0 + B_0$	$A_1 + B_1 + A_0 + B_0$
Отрицательный	$C_1 + D_1$	$C_0 + D_0$	$C_1 + D_1 + C_0 + D_0$
Всего	N_1	N_0	N

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — ЦС — целевое состояние.

Таблица И.4 — Ранее применяемое изделие в сравнении с оценкой ЦС

Ранее применяемое изделие/ результат	Оценка ЦС[1)]		Всего
	ЦС присутствует	ЦС отсутствует	
Положительный	$A_1 + C_1$	$A_0 + C_0$	$A_1 + C_1 + A_0 + C_0$
Отрицательный	$B_1 + D_1$	$B_0 + D_0$	$B_1 + D_1 + B_0 + D_0$
Всего	N_1	N_0	N

[1)] Наилучшая имеющаяся оценка ЦС.

Примечание — ЦС — целевое состояние.

Предполагаемая диагностическая чувствительность разработанного изделия ($ДЧ_{РИ}$) и ранее применяемого изделия ($ДЧ_{ПИ}$), их разность и ДИ этой разности могут быть рассчитаны с помощью формул (И.5)—(И.8). Такие же расчеты для диагностической специфичности выполняют по формулам (И.15)—(И.24).

$$ДЧ_{РИ} = \frac{A_1 + B_1}{N_1}, \quad (И.5)$$

$$ДЧ_{ПИ} = \frac{A_1 + C_1}{N_1}, \quad (И.6)$$

$$\Delta_{ДЧ} = \frac{A_1 + B_1}{N_1} - \frac{A_1 + C_1}{N_1} = \frac{B_1 - C_1}{N_1}, \quad (И.7)$$

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

$$ДИ = (\Delta_{ДЧ} - \sqrt{Q_5},\ \Delta_{ДЧ} + \sqrt{Q_6}), \quad (И.8)$$

где Q_5 и Q_6 вычисляют из данных с помощью формул (И.9)—(И.14). Сначала вычисляют отдельные балльные ДИ для $ДЧ_{РИ}$ и $ДЧ_{ПИ}$ 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 $ДЧ_{РИ}$;
- u_1 — верхняя граница двухстороннего 95 %-ного ДИ для $ДЧ_{РИ}$;
- l_2 — нижний предел двухстороннего 95 %-ного ДИ для $ДЧ_{ПИ}$;
- u_2 — верхняя граница двухстороннего 95 %-ного ДИ для $ДЧ_{ПИ}$.

$$Q_1 = (A_1 + B_1)(C_1 + D_1)(A_1 + C_1)(B_1 + D_1), \quad (И.9)$$

$$Q_2 = (A_1 D_1) - (B_1 C_1), \quad (И.10)$$

$$Q_3 = \begin{pmatrix} Q_2 - \frac{N_1}{2}, & \text{если } Q_2 > \frac{N_1}{2} \\ 0, & \text{если } 0 \le Q_2 \le \frac{N_1}{2} \\ Q_2, & \text{если } Q_2 < 0 \end{pmatrix}, \quad (И.11)$$

$$Q_4 = \begin{pmatrix} 0, & \text{если } Q_1 = 0 \\ \frac{Q_3}{\sqrt{Q_1}}, & \text{если } Q_1 \ne 0 \end{pmatrix}, \quad (И.12)$$

$$Q_5 = (ДЧ_{РИ} - l_1)^2 - 2Q_4(ДЧ_{РИ} - l_1)(u_2 - ДЧ_{ПИ}) + (u_2 - ДЧ_{ПИ})^2, \quad (И.13)$$

$$Q_6 = (ДЧ_{ПИ} - l_2)^2 - 2Q_4(ДЧ_{ПИ} - l_2)(u_1 - ДЧ_{РИ}) + (u_1 - ДЧ_{РИ})^2. \quad (И.14)$$

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

$$ДС_{РИ} = \frac{C_0 + D_0}{N_0}, \quad (И.15)$$

$$ДС_{ПИ} = \frac{B_0 + D_0}{N_0}, \quad (И.16)$$

$$\Delta_{ДС} = \frac{C_0 + D_0}{N_0} - \frac{B_0 + D_0}{N_0} = \frac{C_0 - B_0}{N_0}, \quad (И.17)$$

$$ДИ = \left(\Delta_{ДС} - \sqrt{Q_5},\ \Delta_{ДС} + \sqrt{Q_6}\right), \quad (И.18)$$

где l_1 — нижняя граница двустороннего 95 %-ного ДИ для $ДС_{РИ}$;
u_1 — верхняя граница двустороннего 95 %-ного ДИ для $ДС_{РИ}$;
l_2 — нижняя граница двухстороннего 95 %-ного ДИ для $ДС_{ПИ}$;
u_2 — верхняя граница двустороннего 95 %-ного ДИ для $ДС_{ПИ}$.

$$Q_1 = (A_0 + B_0)(C_0 + D_0)(A_0 + C_0)(B_0 + D_0), \quad (И.19)$$

$$Q_2 = (A_0 D_0) - (B_0 C_0), \quad (И.20)$$

$$Q_3 = \begin{pmatrix} Q_2 - \frac{N_0}{2}, & \text{если } Q_2 > \frac{N_0}{2} \\ 0, & \text{если } 0 \le Q_2 \le \frac{N_0}{2} \\ Q_2, & \text{если } Q_2 < 0 \end{pmatrix}, \quad (И.21)$$

$$Q_4 = \begin{pmatrix} 0, & \text{если } Q_1 = 0 \\ \dfrac{Q_3}{\sqrt{Q_1}}, & \text{если } Q_1 \neq 0 \end{pmatrix}, \quad (И.22)$$

$$Q_5 = (ДС_{РИ} - I_1)^2 - 2Q_4(ДС_{РИ} - I_1)(u_2 - ДС_{ПИ}) + (u_2 - ДС_{ПИ})^2, \quad (И.23)$$

$$Q_6 = (ДС_{ПИ} - I_2)^2 - 2Q_4(ДС_{ПИ} - I_2)(u_1 - ДС_{РИ}) + (u_1 - ДС_{РИ})^2. \quad (И.24)$$

Приложение К
(справочное)

Примеры анализа клинической эффективности

К.1 В настоящем приложении применены следующие сокращения:
ДИ — доверительный интервал;
$ДЧ_{РИ}$ — диагностическая чувствительность разработанного изделия;
$ДЧ_{ПИ}$ — диагностическая чувствительность ранее применяемого изделия;
$ДС_{РИ}$ — диагностическая специфичность разработанного изедия;
$ДС_{ПИ}$ — диагностическая специфичность ранее применяемого изделия;
ЦС — целевое состояние;
ИФА — иммуноферментный анализ;
FN — ложноотрицательный результат;
FP — ложноположительный результат;
HMW-CAP EIA — иммуноферментный анализ высокомолекулярных клеточно-ассоциированных белков;
NPA — отрицательная процентная согласованность;
NPV — отрицательная прогностическая ценность;
PPA — положительная процентная согласованность;
PPV — положительная прогностическая ценность;
TP — истинно положительный;
TN — истинно отрицательный.
К.2 В настоящем приложении применены следующие обозначения:
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;
$\Delta_{ДЧ}$ — разница между чувствительностями;
$\Delta_{ДС}$ — разница между специфичностями.

К.3 Пример — Если имеется оценка состояния цели

Представлены данные, которые позволят сравнить разработанное изделие и ранее применяемое изделие 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*. В этом примере все коэффициенты, относящиеся к клиническим показателям, выражены в процентах.

Показатели разработанного изделия и ранее применяемого изделия определяют в сравнении с наилучшей доступной оценкой результата целевого состояния (ЦС) для каждого образца пациента (т. е. *Z*-стандарт). Таблица 2 × 2 случайных чисел результатов разработанного изделия и ранее применяемого изделия приведены в таблицах К.1 и К.2. Для оценки диагностической чувствительности и специфичности используются формулами (7) и (8), приведенными в 8.3.6. Обе эти таблицы получены из объединенной таблицы К.3.

Таблица К.1 — Таблица 2 × 2 зависимостей между результатами разработанного изделия и *Z*-стандартом

Результат разработанного изделия	*Z*-стандарт для *Helicobacter pylori*/результат		Всего
	Положительный	Отрицательный	
Положительный	TP = 57	FP = 2	59
Отрицательный	FN = 4	TN = 39	43
Всего	61	41	102
Примечание — FN — ложноотрицательный результат; FP — ложноположительный результат; TN — истинно отрицательный; TP — истинно положительный.			

Распространенность заболевания, %, вычисляют по формуле

$$\text{Распространенность заболевания} = 100\frac{TP + FN}{N}. \quad (К.1)$$

В данном примере $\text{Распространенность заболевания} = 100\frac{61}{102} = 59{,}8$ %.

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

$$\text{PPV} = 100\,\frac{\text{TP}}{\text{TP} + \text{FP}}. \quad \text{(К.2)}$$

В данном примере $\text{PPV} = 100\frac{57}{59} = 96{,}6$ %.

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

$$\text{NPV} = 100\,\frac{\text{TN}}{\text{FN} + \text{TN}}. \quad \text{(К.3)}$$

В данном примере $\text{NPV} = 100\frac{39}{43} = 90{,}7$ %.

Оценочную чувствительность для разработанного изделия $\text{ДЧ}_{\text{РИ}}$, %, вычисляют по формуле

$$\text{ДЧ}_{\text{РИ}} = 100\,\frac{\text{TP}}{\text{TP} + \text{FN}}. \quad \text{(К.4)}$$

В данном примере $\text{ДЧ}_{\text{РИ}} = 100\frac{57}{61} = 93{,}4$ %.

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

$Q_1 = 2 \cdot 57 + 1{,}96^2 = 117{,}84,$

$$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\,\frac{57 \cdot 4}{57 + 4}} = 8{,}496,$$

$Q_3 = 2(57 + 4 + 1{,}96^2) = 129{,}68.$

Нижний предел, %, вычисляют по формуле

$$\text{Нижний предел} = 100\,\frac{Q_1 - Q_2}{Q_3}. \quad \text{(К.5)}$$

В данном примере $\text{Нижний предел} = 100\frac{117{,}84 - 8{,}496}{129{,}68} = 84{,}3$ %.

Верхний предел, %, вычисляют по формуле

$$\text{Верхний предел} = 100\,\frac{Q_1 + Q_2}{Q_3}. \quad \text{(К.6)}$$

В данном примере $\text{Верхний предел} = 100\frac{117{,}84 + 8{,}496}{129{,}68} = 97{,}4$ %.

Расчетную специфичность для разработанного изделия $\text{ДС}_{\text{РИ}}$, %, вычисляют по формуле

$$\text{ДС}_{\text{РИ}} = 100\,\frac{\text{TN}}{\text{FP} + \text{TN}}. \quad \text{(К.7)}$$

В данном примере $\text{ДС}_{\text{РИ}} = 100\frac{39}{41} = 95{,}1$ %.

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

$Q_1 = 2 \cdot 39 + 1{,}96^2 = 81{,}84,$

$$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\,\frac{39 \cdot 2}{39 + 2}} = 6{,}632,$$

$Q_3 = 2(39 + 2 + 1{,}96^2) = 89{,}68.$

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

$$\text{Нижний предел} = 100\,\frac{81{,}84 - 6{,}632}{89{,}68} = 83{,}9\ \%;$$

$$\text{Верхний предел} = 100\,\frac{81{,}84 + 6{,}632}{89{,}68} = 98{,}7\ \%.$$

Таблица К.2 — Таблица 2 × 2 зависимостей между ранее применяемым изделием и оценкой ЦС

Результат ранее применяемого изделия	*Z*-стандарт для *Helicobacter pylori*/результат[1)]		Всего
	Положительный	Отрицательный	
Положительный	TP = 54	FP = 7	61
Отрицательный	FN = 7	TN = 34	41
Всего	61	41	102

[1)] *Z*-стандарт должен быть подробно описан как наилучшая доступная оценка ЦС.

Примечание — FN — ложноотрицательный результат; FP — ложноположительный результат; ЦС — целевое состояние; TN — истинно отрицательный; TP — истинно положительный.

Распространенность заболевания, PPV и NPV вычисляют по формулам (К.1)—(К.3).

В данном примере распространенность заболевания $= 100\frac{61}{102} = 59{,}8$ %, $\text{PPV} = 100\frac{54}{61} = 88{,}5$ %; $\text{NPV} = 100\frac{34}{41} = 82{,}9$ %.

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

$$\text{ДЧ}_{\text{ПИ}} = 100\frac{\text{TP}}{\text{TP} + \text{FN}}. \quad \text{(К.8)}$$

В данном примере $\text{ДЧ}_{\text{ПИ}} = 100\frac{54}{61} = 88{,}5$ %.

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

$Q_1 = 2 \cdot 54 + 1{,}96^2 = 111{,}84$,

$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\frac{54 \cdot 7}{54 + 7}} = 10{,}487$,

$Q_3 = 2(54 + 7 + 1{,}96^2) = 129{,}68$.

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

$$\text{Нижний предел} = 100\frac{111{,}84 - 10{,}487}{129{,}68} = 78{,}2\ \%;$$

$$\text{Верхний предел} = 100\frac{111{,}84 + 10{,}487}{129{,}68} = 94{,}3\ \%.$$

Расчетную специфичность для ранее применяемого изделия $\text{ДС}_{\text{ПИ}}$, %, вычисляют по формуле

$$\text{ДС}_{\text{ПИ}} = 100\frac{\text{TN}}{\text{FP} + \text{TN}}. \quad \text{(К.9)}$$

В данном примере $\text{ДС}_{\text{ПИ}} = 100\frac{34}{41} = 82{,}9$ %.

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

$Q_1 = 2 \cdot 34 + 1{,}96^2 = 71{,}84$,

$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\frac{34 \cdot 7}{34 + 7}} = 10{,}196$,

$Q_3 = 2(34 + 7 + 1{,}96^2) = 89{,}68$.

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

$$\text{Нижний предел} = 100\frac{71{,}84 - 10{,}196}{89{,}68} = 68{,}7\ \%;$$

$$\text{Верхний предел} = 100\frac{71{,}84 + 10{,}196}{89{,}68} = 91{,}5\ \%.$$

В данном случае имеет смысл проведение совместного статистического сравнения пар чувствительность/специфичность, поскольку предполагаемая чувствительность и специфичность для разработанного изделия больше, чем предполагаемая чувствительность и специфичность для ранее применяемого изделия. Чтобы сравнить чувствительность и специфичность двух экзаменов, необходимо провести трехстороннее сравнение между ранее применяемым изделием, разработанным изделием и результатами оценки ЦС.

Таблица К.3 — Трехстороннее сравнение между разработанным изделием, ранее применяемым изделием и оценкой ЦС

Z-стандарт: ЦС присутствует[1)]				*Z*-стандарт: ЦС отсутствует[1)]			
Результат разработанного изделия	Результат ранее применяемого изделия		Всего	Результат разработанного изделия	Результат ранее применяемого изделия		Всего
	Положительный	Отрицательный			Положительный	Отрицательный	
Положительный	53	4	57	Положительный	2	0	2
Отрицательный	1	3	4	Отрицательный	5	34	39
Всего	54	7	61	Всего	7	34	41
[1)] *Z*-стандарт должен быть подробно описан как наилучшая доступная оценка ЦС.							

Сравнение чувствительности $\Delta_{ДЧ}$, %, вычисляют по формуле

$$\Delta_{ДЧ} = ДЧ_{РИ} - ДЧ_{ПИ}. \quad (К.10)$$

В данном примере $\Delta_{ДЧ} = 100\left(\frac{57}{61} - \frac{54}{61}\right) = 100(0,934 - 0,885) = 100 \cdot 0,049 = 4,9$ % или $\Delta_{ДЧ} = 100\frac{4-1}{61} = 4,9$ %.

95 %-ные ДИ для разработанных изделий и ранее применяемых изделий 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 $ДЧ_{РИ}$);
- $u_1 = 97,4$ %;
- $l_2 = 78,2$ % (от ДИ оценки для $ДЧ_{ПИ}$);
- $u_2 = 94,3$ %.

$Q_1 = (53 + 4)(1 + 3)(53 + 1)(4 + 3) = 57 \cdot 4 \cdot 54 \cdot 7 = 86\,184$,

$Q_2 = 53 \cdot 3 - 4 \cdot 1 = 155$.

В приведенных ниже примерах n_1 — общее количество образцов от субъектов с ЦС, а n_0 — общее количество субъектов без ЦС.

Потому, что $\frac{n_1}{2} = \frac{61}{2} = 30,5 < 155 = Q_2$,

$$Q_3 = Q_2 - \frac{n_1}{2} = 155 - 30,5 = 124,5,$$

$$Q_4 = \frac{Q_3}{\sqrt{Q_1}} = \frac{124,5}{\sqrt{86184}} = \frac{124,5}{293,57} = 0,4241,$$

$$Q_5 = (93,4 - 84,3)^2 - 2 \cdot 0,4241(93,4 - 84,3)(94,3 - 88,5) + (94,3 - 88,5)^2 = 71,68\ \%,$$

$$Q_6 = (88,5 - 78,2)^2 - 2 \cdot 0,4241(88,5 - 78,2)(97,4 - 93,4) + (97,4 - 93,4)^2 = 87,14\ \%,$$

$$\Delta_{ДЧ} - \sqrt{Q_5}, \quad (К.11)$$

$$4,9 - \sqrt{71,68} = -3,6\ \%,$$

$$\Delta_{ДЧ} + \sqrt{Q_6}, \quad (К.12)$$

$$4,9 + \sqrt{87,14} = 14,2\ \%.$$

95 %-ный ДИ для $\Delta_{ДЧ} = ДЧ_{РИ} - ДЧ_{ПИ}$ составляет (–3,6 %, 14,2 %).

Сравнение специфичности выполняют по формуле (К.13)

$$\Delta_{ДС} = ДС_{РИ} - ДС_{ПИ}. \quad (К.13)$$

$$\Delta_{ДС} = 100\left(\frac{39}{41} - \frac{34}{41}\right) = 100\left(0{,}951 - 0{,}829\right) = 100 \cdot 0{,}122 = 12{,}2\ \%$$

или $\Delta_{ДС} = 100\dfrac{5-0}{41} = 12{,}2\ \%$.

95 %-ные ДИ для разработанных изделий и ранее применяемых изделий могут быть рассчитаны для специфичности по формуле (И.1) приложения И. Однако целесообразно рассчитывать 95 %-ные ДИ для разницы между ними по формулам (И.19)—(И.24) и (К.14) и (К.15):

- $l_1 = 83{,}9\ \%$ (от ДИ оценки для $ДС_{РИ}$);
- $u_1 = 98{,}7\ \%$;
- $l_2 = 68{,}7\ \%$ (от ДИ оценки для $ДС_{ПИ}$);
- $u_2 = 91{,}5\ \%$.

$Q_1 = (2 + 0)(5 + 34)(2 + 5)(0 + 34) = 2 \cdot 39 \cdot 7 \cdot 34 = 18\,564$,

$Q_2 = 2 \cdot 34 - 0 \cdot 5 = 68$.

Потому, что $\dfrac{n_0}{2} = \dfrac{41}{2} = 20{,}5 < 68 = Q_2$,

$$Q_3 = Q_2 - \frac{n_0}{2} = 68 - 20{,}5 = 47{,}5,$$

$$Q_4 = \frac{Q_3}{\sqrt{Q_1}} = \frac{47{,}5}{\sqrt{18\,564}} = \frac{47{,}5}{136{,}25} = 0{,}3486,$$

$$Q_5 = (95{,}1 - 83{,}9)^2 - 2 \cdot 0{,}3486(95{,}1 - 83{,}9)(91{,}5 - 82{,}9) + (91{,}5 - 82{,}9)^2 = 132{,}25\ \%,$$

$$Q_6 = (82{,}9 - 68{,}7)^2 - 2 \cdot 0{,}3486(82{,}9 - 68{,}7)(98{,}7 - 95{,}1) + (98{,}7 - 95{,}1)^2 = 178{,}96\ \%,$$

$$\Delta_{ДС} - \sqrt{Q_5}, \qquad (К.14)$$

$$\Delta_{ДС} = 12{,}2 - \sqrt{132{,}25} = 0{,}7\ \%,$$

$$\Delta_{ДС} + \sqrt{Q_6}, \qquad (К.15)$$

$$\Delta_{ДС} = 12{,}2 + \sqrt{178{,}96} = 25{,}6\ \%.$$

95 %-ный ДИ для $\Delta_{ДС} = ДС_{РИ} - ДС_{ПИ}$ 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 разработанного изделия статистически значимо выше, чем специфичность изделия сравнения. Этот результат указывает на то, что клинические показатели разработанного изделия лучше, чем у ранее применяемого изделия.

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.

$$PPV_{РИ} = \frac{ДЧ_{РИ} \cdot Распространенность}{ДЧ_{РИ} \cdot Распространенность + (1 - ДС_{РИ})(1 - Распространенность)}, \qquad (К.16)$$

$$PPV_{РИ} = \frac{\dfrac{57}{61} \cdot 59{,}8}{\dfrac{57}{61} \cdot 59{,}8 + \left(1 - \dfrac{39}{41}\right)(1 - 59{,}8)} = 96{,}6\ \%,$$

$$PPV_{ПИ} = \frac{ДЧ_{ПИ} \cdot Распространенность}{ДЧ_{ПИ} \cdot Распространенность + (1 - ДС_{ПИ})(1 - Распространенность)}, \qquad (К.17)$$

$$PPV_{ПИ} = \frac{\dfrac{54}{61} \cdot 59{,}8}{\dfrac{54}{61} \cdot 59{,}8 + \left(1 - \dfrac{34}{41}\right)(1 - 59{,}8)} = 88{,}5\ \%,$$

$$NPV_{РИ} = \frac{ДС_{РИ}(1 - Распространенность)}{ДС_{РИ}(1 - Распространенность) + (1 - ДЧ_{РИ})Распространенность}, \quad (К.18)$$

$$NPV_{РИ} = \frac{\frac{39}{41}(1 - 59{,}8)}{\frac{39}{41}(1 - 59{,}8) + \left(1 - \frac{57}{61}\right)59{,}8} = 90{,}7\ \%,$$

$$NPV_{ПИ} = \frac{ДС_{ПИ}(1 - Распространенность)}{ДС_{ПИ}(1 - Распространенность) + (1 - ДЧ_{ПИ})Распространенность}, \quad (К.19)$$

$$NPV_{ПИ} = \frac{\frac{34}{41}(1 - 59{,}8)}{\frac{34}{41}(1 - 59{,}8) + \left(1 - \frac{54}{61}\right)59{,}8} = 82{,}9\ \%.$$

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 разработанного изделия лучше, чем у ранее применяемого изделия.

К.4 Пример — Если оценка состояния цели недоступна

В типичной ситуации эксперт может не знать диагноза, но хочет сравнить результаты разработанного изделия с результатами другого изделия. Например, в разработанном изделии использовали иммунохроматографический метод, а в изделии сравнения — метод иммуноферментного анализа высокомолекулярных клеточно-ассоциированных белков (HMW-CAP EIA) ELISA. Таблица 2 × 2 случайных чисел результатов оценки представлена в таблице К.4.

Таблица К.4 — HMW-CAP EIA

Результат разработанного изделия (иммунохроматографический метод)	Изделие сравнения[1)]		
	с положительным результатом	с отрицательным результатом	Всего
Положительный	285	15	300
Отрицательный	14	222	236
Всего	299	237	536
[1)] Должна быть приведена информация о сравнительном исследовании.			

Положительная процентная согласованность (PPA), обусловленная результатами иммунохроматографического анализа или HMW-CAP EIA, равна

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.

$$Q_1 = 2 \cdot 258 + 1{,}96^2 = 573{,}84,$$

$$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\frac{285 \cdot 14}{285 + 14}} = 14{,}826,$$

$$Q_3 = 2(285 + 14 + 1{,}96^2) = 605{,}68.$$

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

$$\text{Нижний предел} = 100\frac{573{,}84 - 14{,}826}{605{,}68} = 92{,}3\ \%,$$

$$\text{Верхний предел} = 100\frac{573{,}84 + 14{,}826}{605{,}68} = 97{,}2\ \%.$$

Отрицательная процентная согласованность (NPA), которая обусловлена результатами иммунохроматографического анализа или HMW-CAP EIA, равна

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.

$Q_1 = 2 \cdot 222 + 1{,}96^2 = 447{,}84,$

$$Q_2 = 1{,}96\sqrt{1{,}96^2 + 4\frac{222 \cdot 15}{222 + 15}} = 15{,}187,$$

$Q_3 = 2(222 + 15 + 1{,}96^2) = 481{,}68.$

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

$$\text{Нижний предел} = 100\frac{447{,}84 - 15{,}187}{481{,}68} = 89{,}8\ \%,$$

$$\text{Верхний предел} = 100\frac{447{,}84 + 15{,}187}{481{,}68} = 96{,}1\ \%.$$

Приложение Л
(справочное)

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

Л.1 В настоящем приложении применены следующие сокращения:

ДИ — доверительный интервал;

ПЦР — полимеразная цепная реакция;

CV — коэффициент вариации;

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

SD — стандартное отклонение.

Л.2 В медицинской 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 (C_T). В этом типе теста C_T уменьшается по мере увеличения концентрации *vanA*, поскольку требуется меньше циклов ПЦР для достижения cut-off обнаружения сигнала, так что C95 C_T меньше, чем значение в соответствующей шкале, где бинарный анализ признает образец положительным в 50 % случаев (C50). C_T и C5 C_T больше, чем C50. Заявления о неточности во вкладыше к МИ ИВД приведены в таблице Л.1.

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

Образец с ЦС	C_T	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 неточность двух образцов одинакова, предполагается, что CV на cut-off, C5 и C95 также будут одинаковыми. Однако количественные расчеты с помощью этого метода невозможны, поскольку, как только C_T достигает 39, результаты проб представляются как нули, и прибор не сообщает о C_T. Исходя из предположения о постоянном CV, ошибка C_T, ожидаемая при отсечении, должна быть равна $0{,}045 \cdot 39 = 1{,}755$ SD. Вычитание 1,645 SD (односторонний 95 %-ный ДИ) из cut-off должно приблизительно равняться циклам на C95: $39 - 1{,}645 \cdot 1{,}755 = 39 - 2{,}887 = 36{,}1$.

Чтобы проверить заявление производителя:

- разбавляют аликвоту образца, положительного на микроорганизм, в разбавителе для предварительной обработки в соотношении 1:100;
- исследуют пять повторов за один прогон.

Средние результаты = $22{,}24 \cdot C_T$, обратно рассчитанные корректные результаты: ($100 = 2^{6{,}64}$), $22{,}24 - 6{,}64 = 15{,}6\ C_T$.

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

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

1) вычисляют разность $39 - 15{,}6 = 23{,}4$.

Примечание — Два, возведенное в степень разности циклов, равно $2^{23{,}4} = 11{,}07 \times 10^6$.

2) разбавляют неочищенный высокоположительный препарат 1:$2{,}75 \times 10^6$, 1:11×10^6 и 1:44×10^6 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 С50 разница в 1,755 цикла примерно эквивалентна $2^{1,76}$ = 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

Разбавление	Повтор 1	Повтор 2	Повтор 3	Повтор 4	Повтор 5	Среднее или медианное	SD
$1:2,75 \times 10^6$	37,5	35,5	38,1	36,2	38,5	37,16	1,272
$1:11 \times 10^6$	37,8	38,2	38,9	0 (> 39)	0 (> 39)	38,9	Неприменимо
$1:44 \times 10^6$	0 (> 39)	0 (> 39)	0 (> 39)	38,7	0 (> 39)	0 (> 39)	Неприменимо

Второй результат, в котором три из пяти повторов являются положительными при разведении $1:11 \times 10^6$, соответствует результатам, которые могут быть получены из истинного образца С50. Из-за ограниченного количества повторов, от одного положительного результата до четырех положительных результатов также соответствуют образцу С50. Для повышения точности и достоверности оценки С50 необходимо дополнительное тестирование. Поэтому образцы С50, С5 и С95 формируют следующим образом:

- разведение С50 = $1:11 \times 10^6$ микробиологических культур в разбавителе для предварительной обработки;
- разведение С5 = $1:2^{2,91} \times 11 \times 10^6 = 1:83 \times 10^6$ микробиологических культур в разбавителе для предварительной обработки;
- разведение С95 = $1:11 \times 10^6/2^{2.91} = 1:1.46 \times 10^6$ микробиологических культур 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 С (>39 — положительный результат) или нуля (отрицательный результат). Ожидаемый результат для образца С95 — 35,5 C_T, поскольку это целевой C_T для 95 % положительных результатов. Для образца С50 указано нулевое значение, поскольку 50 % количества циклов ≥ 39. Ожидаемое среднее значение С5 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	С95	С50	С5
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 — *N* — количество всех исследований.				

Библиография

[1] СанПиН 3.3686-2021 Санитарно-эпидемиологические требования по профилактике инфекционных болезней

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, интерференция, прецизионность

Редактор *Л.В. Коретникова*
Технический редактор *В.Н. Прусакова*
Корректоры *Р.А. Ментова, Л.С. Лысенко, И.А. Королева*
Компьютерная верстка *И.А. Налейкиной*

Сдано в набор 14.04.2026. Подписано в печать 28.04.2026. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 10,23. Уч.-изд. л. 9,27.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.

gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru