

ГЛАВА 2.6.8

СОПОСТАВИМОСТЬ АНАЛИЗОВ ПОСЛЕ ИЗМЕНЕНИЙ ВАЛИДИРОВАННОГО МЕТОДА ИССЛЕДОВАНИЯ

ВВЕДЕНИЕ

Существует множество причин, по которым валидированные тесты могут подвергаться изменениям с течением времени. Замена отработанных реактивов, вероятно, наиболее типичный пример незначительного изменения валидированного анализа (см. Глава 1.1.6 Принципы и методы валидации диагностических анализов для инфекционных болезней животных, Рисунок 1, Разработка анализа и способы валидации). Небольшие изменения могут быть вызваны, например, наличием менее дорогостоящих или более качественных реактивов для молекулярных исследований; необходимостью повышения стандартизации, например, замена планшетов для иммуноферментного анализа (ИФА), требующих сенсibilизации оператором, на предварительно сенсibilизированные ИФА-планшеты; повышенными требованиями к производительности, например, замена обработки вручную на роботизированную обработку, и т.д. Такие небольшие изменения обычно требуют проведения экспериментального исследования для оценки того, являются ли производственные характеристики валидированного анализа все еще сопоставимыми с характеристиками новой процедуры (Таблица 1, Рисунок 1). Имеются и другие значимые переменные за пределами анализа, которые могут требовать верификации, например, происхождение целевой популяции, вида и образца. Например, опытным лабораторным диагностам следует с осторожностью применять конкурентный ИФА для выявления антител к вирусу бруцеллеза у КРС в Латинской Америке, если анализ был валидирован специально для КРС в Канаде (Gall et al., 1998). Анализы зачастую применяют для видов животных, отличных от тех, для которых они были изначально валидированы, например, домашние птицы против диких птиц или мясной КРС против молочного КРС. Другие изменения включают различные исследуемые образцы, например, трахеальные смывы в сравнении с клоакальными смывами, отобранными от птиц для диагностики гриппа птиц посредством молекулярных анализов. В данных условиях будут необходимы верификационные исследования для валидации рабочих характеристик теста в новых условиях. Существуют разногласия о том, что считать «незначительным» и «значительным» изменением диагностического анализа. Есть некоторые изменения, которые считаются значительными, т.к.

Валидация – процесс, который определяет пригодность анализа, который надлежащим образом разработан, оптимизирован и стандартизирован для предусмотренного назначения.

Верификация представляет собой доказательство того, что рабочие характеристики, например, точность и прецизионность валидированного анализа, сопоставимы при использовании в другой лаборатории.

Сопоставимость – предпочтительный термин, когда рабочие характеристики нового теста, в который внесено незначительное изменение, настолько же хороши, что и характеристики валидированного теста в рамках статистически определенных пределов.

Эквивалентность или равнозначность исторически используется в некоторых диагностических лабораториях для исследования сопоставимости. Однако термин предполагает более жесткое требование, чем пригодность для предусмотренного назначения, а также имеет специальное статистическое значение. По этой причине в данной главе термин не используется.

фундаментально изменяется биологическая основа анализа, например, эволюционные изменения или мутации в структуре нуклеиновых кислот патогена потребуют корректировки праймеров и зондов. Аналогичным образом, замена непрямого ИФА на конкурентный формат с использованием высокоспецифичного моноклонального антитела считается значительным изменением, которое потребует полной повторной валидации анализа. В Таблице 1 приведено несколько примеров незначительных и значительных изменений, которые часто встречаются при использовании тестов с антителами и тестов на выявление нуклеиновых кислот (NAD). Точные и хорошо разработанные исследования сопоставимости обеспечат объективную оценку того, является ли анализ, применяемый с незначительным изменением, сопоставимым с валидированным анализом и настолько же пригодным для предусмотренного применения. Результат экспериментального исследования определит, потребуется ли полная повторная валидация кандидатного анализа и, следовательно, можно ли его использовать с уверенностью или нет.

Таблица 1. Примеры изменения диагностических тестов

Тип изменения	Изменения анализа	Изменения популяции целевых животных или образцов
Незначительные	<p>Замена устаревшего реактива, например положительного контрольного образца, новая партия антигена, планшеты, конъюгат (ИФА)</p> <p>Изменение инструмента/ платформы, например ИФА ридера, инкубатора/ шейкера, амплификатора (ПЦР)</p> <p>Замена планшетов, требующих сенсibilизации вручную, на предварительно сенсibilизированные планшеты</p> <p>Переход от обработки вручную на роботизированную обработку (ИФА, NAD)</p> <p>Изменение процедуры экстрагирования нуклеиновой кислоты (NAD)</p> <p>Использование модифицированного праймера (-ов) (частичная замена аминокислот, например, дегенерация)</p> <p>Модифицированные условия проведения ПЦР при использовании тех же праймеров и зондов</p> <p>Добавление дополнительного зонда в амплифицированную область</p> <p>Изменение химического состава зонда (NAD)</p>	
Значительные	<p>Замена рекомбинантного антигена на антиген, полученный на культуре клеток при проведении ИФА</p> <p>Замена непрямого ИФА на конкурентный ИФА с использованием специфического антитела</p> <p>Замена праймеров и зондов для различных мишеней тех же или отличных генов (NAD)</p>	<p>Разные виды, например, КРС в сравнении с буйволами, домашние куры против диких птиц</p> <p>Различные виды образцов, например, трахеальные смывы в сравнении с клоакальными смывами, кровь против сыворотки, различные ткани или органы</p>

ИФА = иммуноферментный анализ; ПЦР = полимеразная цепная реакция; NAD = выявление нуклеиновых кислот (тесты);

При подготовке эксперимента по определению сопоставимости процедура должна исходить из цели анализа (Рисунок 1, этап 1). Например, скрининговые анализы требуют высокой диагностической чувствительности, и важно сравнивать пределы обнаружения. В данном случае необходимо определить необходимый диапазон разведений и количество репликатов контрольного образца. Необходимо получить, распределить на аликвоты и правильно хранить достаточное количество материала хорошо описанного образца.

Если целью является оценка и сравнение воспроизводимости, необходимо использовать хорошо описанные репликаты контрольных образцов различных концентраций аналита, охватывая расчетный рабочий диапазон анализа, например, высокая, средняя и низкая концентрация аналита. В практических целях для примера, приведенного в данной главе, используется слабоположительный контрольный образец. Рекомендуется визуально инспектировать корреляцию результатов обоих методов. Например, можно легко построить корреляционную диаграмму и гистограмму и получить оперативную информацию о типе корреляции и распределении данных (Рисунки 2 и 3). Основы статистики помогают устанавливать верхние и нижние пределы и оценивать результаты, например, предел обнаружения и воспроизводимость (Таблицы 2 и 3). Более комплексным подходом к сравнению и анализу результатов является диаграмма Блэнда-Алтмана (Рисунок 4 и Таблица 4).

Рисунок 5 – пример того, как повысить эффективность эксперимента по определению сопоставимости посредством одновременной оценки различных параметров на одном планшете, например, оценка аналитической чувствительности с использованием этапов разведения целевого аналита в трех повторностях с последующей оценкой аналитической специфичности посредством использования ряда отрицательных образцов (в которых отсутствует целевой аналит), и, в конечном счете, оценка аналитической чувствительности с использованием ряда образцов от инфицированных животных при разных концентрациях аналита. Использование репликатов в ходе одного цикла и между циклами позволит определить воспроизводимость (Рисунки 5 и 6).

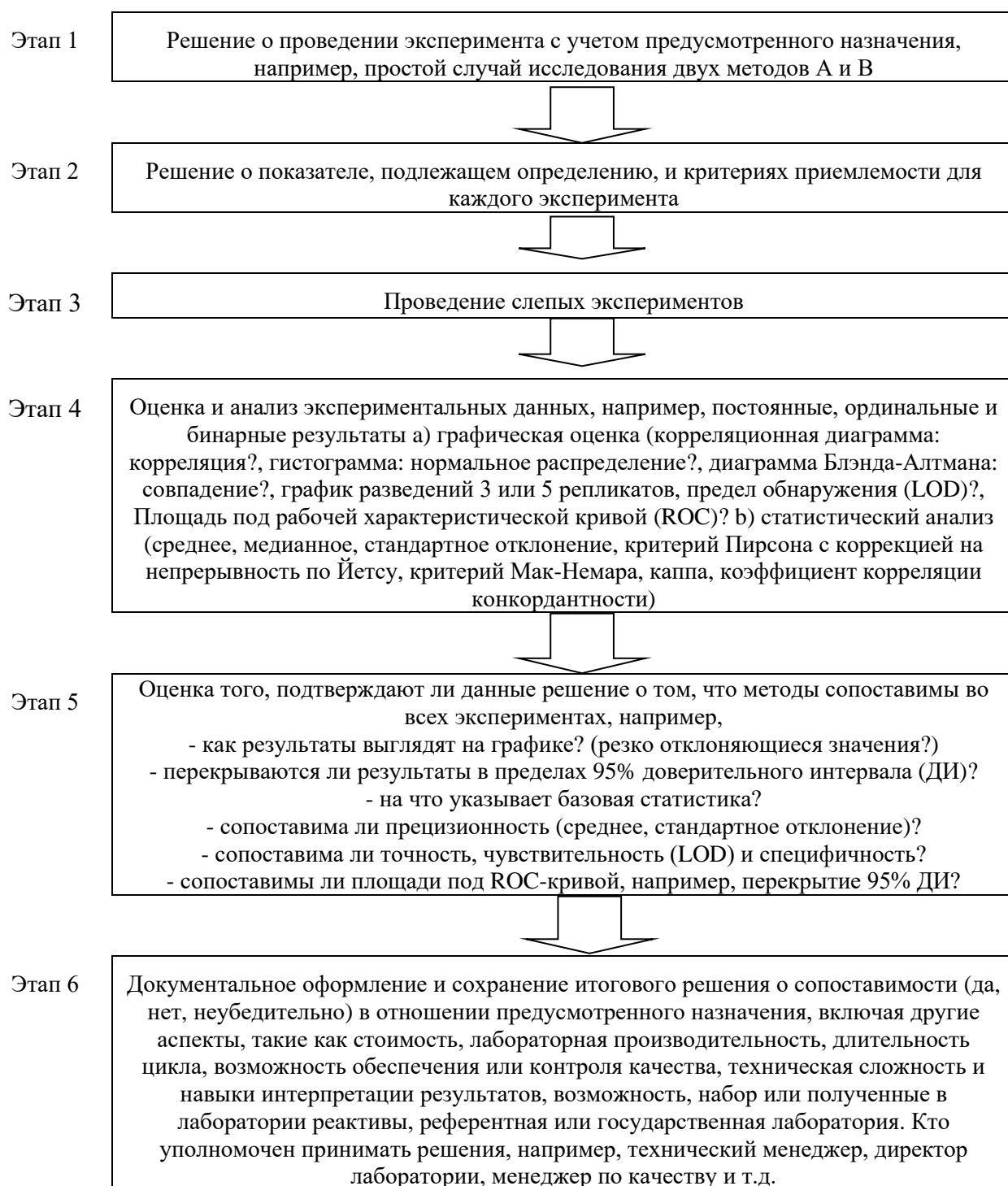
Для оценки результатов эксперимента важно договориться о критериях соответствия (Рисунок 1, этап 2), например, доверительные интервалы могут быть ассиметричными для того, чтобы обеспечить более высокие рабочие характеристики нового анализа. В данной главе консервативный 95% доверительный интервал считается приемлемым для эксперимента по определению предела обнаружения (Рисунки 6 и 7). При сравнении воспроизводимости в качестве критериев соответствия могут быть использованы среднее и стандартное отклонение (СО) или видимые результаты теста плюс и минус заданный диапазон (Таблицы 3 и 4 и Рисунок 4). Результаты исследования панели образцов от инфицированных и неинфицированных особей предоставляют информацию о сравнительной диагностической чувствительности и специфичности (Рисунок 8 и Таблицы 5 и 6).

Данные, используемые на Рисунках 2, 3 и 4 и в Таблицах 2, 3 и 4 получены при использовании результатов неоднократно протестированного слабоположительного контрольного образца в двух анализах с зондами TaqMan, направленными на гены M (анализ M1) и N1 (анализ N1) вируса Хендра.

Данные экспериментов для определения пределов обнаружения и схема планшета на Рисунках 5, 6 и 7 вымышленные. Данные ROC-кривых на Рисунке 8 и в Таблицах 5 и 6

взяты по результатам экспериментов по определению сопоставимости различных ИФА по выявлению гриппа у свиней.

Рисунок 1. Факторы, которые необходимо учитывать в ходе проведения исследования по определению сопоставимости диагностических тестов



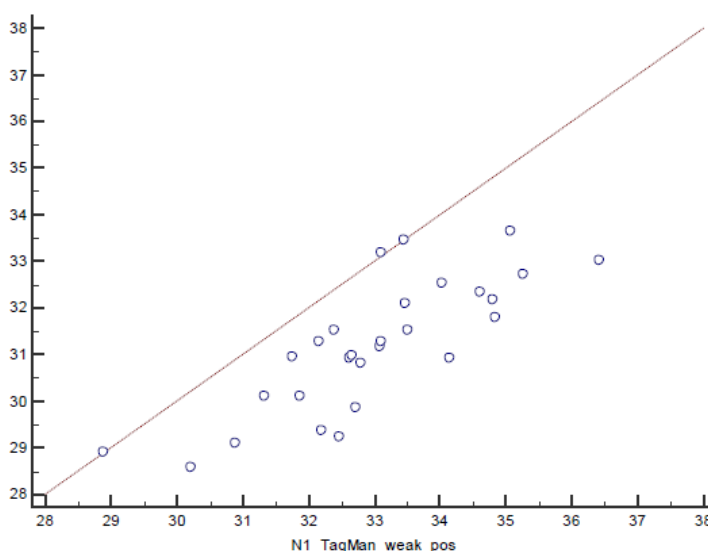
В данной главе представлен обзор различных подходов к схемам эксперимента и интерпретации результатов, полученных в ходе исследований по определению сопоставимости, например, аналитическая (предел обнаружения) и диагностическая чувствительность, аналитическая и диагностическая специфичность, а также воспроизводимость. Существует множество фундаментально разных тестов, но предоставленные примеры обусловлены или относятся к экспериментам с тестами по выявлению нуклеиновых кислот (NAD) и иммуноферментными анализами (ИФА). Можно

предположить, что принципы, представленные в этой главе равно применимы и к другим тестам.

1. Визуальная оценка

Корреляционная диаграмма пригодна для визуальной оценки изначальной корреляции между методами, например, является ли соотношение линейным или логарифмическим? Присутствуют ли резко отклоняющиеся значения или отсутствующие показатели или артефакты? Например, на Рисунке 2 показаны результаты исследования повторного тестирования слабоположительного контрольного образца в анализах с зондами TaqMan, направленными на гены М (анализ М1) и N (анализ N1) вируса Хендра. Результаты показывают, что данные анализа N1 смещены вправо вниз по сравнению с анализом М1, что указывает на наличие устойчивых более высоких показателей анализа N1 в сравнении с анализом М1. Однако корреляционная диаграмма не предоставляет информации о совпадении двух анализов. Из этого правила есть одно исключение: если все результаты обоих тестов расположены вдоль диагонали 45°, совпадение составит 100%. На Рисунке 2 только три результата расположены вдоль диагональной линии. В данной статье мы определяем совпадение в виде набора показателей кандидатного теста, которые находятся в пределах 95% ДИ результатов утвержденного теста после повторных циклов с одним и тем же хорошо описанным контрольным образцом. Корреляция измеряет степень соотношения между измерениями, полученными в ходе этих тестов, и выражается в виде р-значения.

Рисунок 2. Корреляционная диаграмма слабоположительного образца, протестированного 28 раз в двух различных анализах вируса Хендра с зондами TaqMan, M1 и N1 (результаты выражены в виде пороговых значений цикла [ДИ])



Результаты дальнейшего анализа результатов представлены в Таблице 2 ниже, где r (коэффициент корреляции) = 0,8 и указывает на устойчивую положительную корреляцию двух методов. Значение $p < 0,0001$ указывает на то, что вероятность случайности данной ассоциации очень мала, а 95% ДИ (доверительный интервал) указывает на то, что при использовании этих методов в сходных условиях мы на 95% уверены, что истинное неизвестное значение r находится в диапазоне от 0,61 до 0,9.

Таблица 2. Статистический анализ слабоположительного контрольного образца после 28 исследований в двух анализах вируса Хендра с зондами TaqMan, M1 и N1

Переменная Y	Слабоположительный TaqMan M1
Переменная X	Слабоположительный TaqMan N1
Размер образца	28
Коэффициент корреляции r	0,8015
Уровень достоверности	p<0,0001
95% доверительный интервал для r	от 0,6112 до 0,9042

Гистограмма, где отражено смещение влево или вправо или другие значимые характеристики, такие как выявленное бимодальное распределение, показана на Рисунке 3.

Рисунок 3. Гистограмма слабоположительного контрольного образца после 28 исследований в двух анализах вируса Хендра с зондами TaqMan, M1 и N1 (результаты выражены в виде пороговых значений цикла [ДИ])

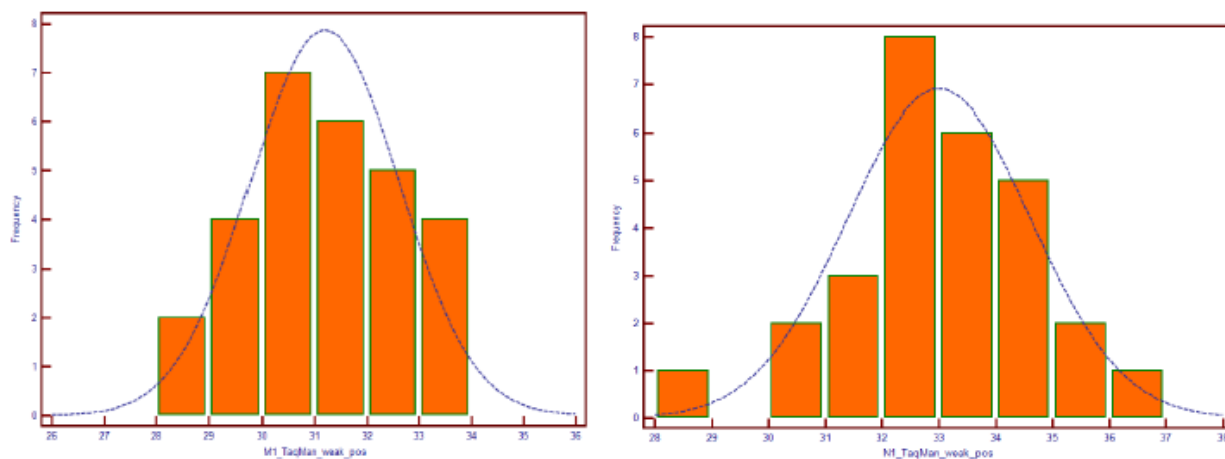


Таблица 3. Статистический анализ слабоположительного контрольного образца после 28 исследований в двух анализах вируса Хендра с зондами TaqMan, M1 и N1 (результаты выражены в виде пороговых значений цикла [ДИ])

	Слабоположительный, Taqman M1	Слабоположительный, Taqman N1
Размер образца	28	28
Нижнее значение	28,58	28,88
Верхнее значение	33,63	36,42
Среднее	31,19	33,00
Медианное	31,22	32,94
Стандартное отклонение	1,42	1,62

2. Воспроизводимость

Погрешность анализа можно определить при использовании репликатов внутреннего контрольного образца, исследуемого в последовательности циклов в течение определенного времени. В данном примере воспроизводимость сравнивали для двух различных анализов на выявление вируса Хендра с зондами TaqMan, направленными на гены N и M, с использованием слабоположительного внутреннего контрольного образца после проведения 28 циклов одним и тем же оператором на 14 день и в течение 18-дневного периода. Результаты порогового значения цикла (Ct) обобщали в виде среднего и стандартного отклонения (CO) в Таблице 3.

Так как оценки основываются только на одном контрольном образце, отсутствует необходимость проведения официального сравнения. Напротив, при наличии заранее определенных критериев приемлемости, например, CO составляет от 2 до 3 или значения Ct находится в пределах от ± 2 до ± 3 , оба анализа будут считаться сопоставимыми.

3. Диаграмма Блэнда-Алтмана

Очень эффективным способом одновременной демонстрации и анализа результатов сравнительного исследования, в котором проводятся 2 различных измерения по каждому образцу (так называемый дизайн связанных пар), является график разности средних показателей по критерию Тьюки (Bland & Altman, 1999, 2007; Kozk & Wnuk, 2014). Данный график полезен при определении взаимодействия между различающимися и средними значениями, оценивая любое систематическое отклонение или возможный выброс. Средние показатели Ct, полученные при использовании методов А (анализ M1 TaqMan) и В (анализ N1 TaqMan) по всем результатам, располагают по оси x, а разницу средних показателей, например А минус В, наносят по оси y (Рисунок 4, Таблица 4). В данном примере анализ M1 TaqMan сравнивают с анализом N1 TaqMan, используя результаты исследования слабоположительного контрольного образца, полученные в ходе 28 циклов. Все различающиеся значения отрицательные, т.к. средние показатели для анализа N1 TaqMan (Ct 33) выше, чем для анализа M1 TaqMan (31,99) (Таблица 3). Вычитание показателей N1 из показателей M1 показывает, что почти все значения отрицательные, что указывает на систематическую ошибку, например, средняя разница между показателями N1 и M1 составила $-1,81$ (систематическая ошибка). Горизонтальные линии отражают среднюю разницу ($-1,81$) и пределы согласия, которые определяют как средняя разница плюс и минус стандартное отклонение разницы, умноженное на 1,96, которое, выраженное как средние разницы Ct, составляет от 0,1 до $-3,71$. Результаты в Таблице 4 показывают, что 95% ДИ средней разницы (от $-2,18$ до $-1,43$) исключает ноль и, следовательно, нельзя сделать заключение о сопоставимости двух методов. Корректировка точки раздела могла бы содействовать уравниванию данной вероятной систематической ошибки при сравнении двух методов после подтверждения данных результатов при использовании большего объема образца и за пределами ожидаемых результатов.

Для сравнения воспроизводимости двух методов в диапазоне измерений по каждому методу проводили одно измерение каждого образца (дизайн связанных пар), например, в случае конкурентного ИФА известно, что разница повышается при понижении концентрации аналита в образце. Отрицательный или слабоположительный контрольный образец может демонстрировать значительно более высокую разницу, чем высокоположительные контрольные образцы. На Рисунке 4 показаны три результата по

линии значения 0, например, линия, когда результаты обоих тестов идентичны, т.к. разница равна нулю. Это те же результаты, которые находятся на диагональной линии корреляционной диаграммы на Рисунке 2 при значении 28,9; 33,10 и 33,45 обоих анализов.

Рисунок 4. Диаграмма Блэнда-Алтмана, отражающая разницу значений пороговых значений цикла (Ct), полученную в ходе 28 циклов исследования слабоположительного образца в двух анализах на определение вируса Хендра с зондами TaqMan, M1 и N1

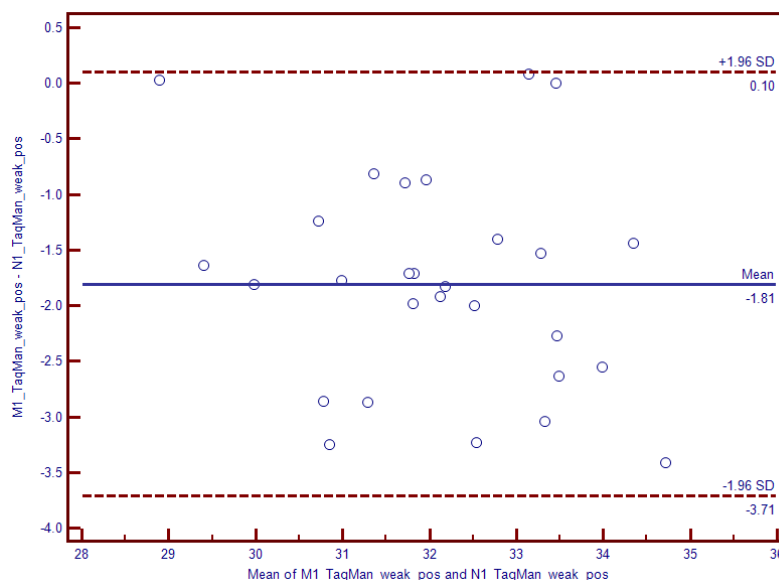


Таблица 4. Статистический анализ по диаграмме Блэнда-Алтмана

Метод А	Слабоположительный, Taqman M1
Метод В	Слабоположительный, Taqman N1
Разница	
Размер образца	28
Арифметическое среднее	-1,8054
95% ДИ	от -2,1831 до -1,4276
Стандартное отклонение	0,9741

4. Предел обнаружения (LOD) эксперимента

Пример схемы планшета для использования в ходе эксперимента по определению сопоставимости представлен на Рисунке 5. Проводят оценку предела обнаружения (три репликата положительных образцов в серии разведений от 10^{-1} до 10^{-8} [аналитическая чувствительность]), диагностической специфичности (отрицательные контрольные образцы от неинфицированных животных или животных, инфицированных нецелевым патогеном [Neg], исследуемые в двух повторностях), диагностической чувствительности (образцы с противоположной активностью от животных, инфицированных полевым вирусом, например, крайне высокоположительные образцы [C+++], очень высокоположительные образцы [C++], высокоположительные образцы [C+] и положительные образцы [C], протестированные в двух повторностях), и

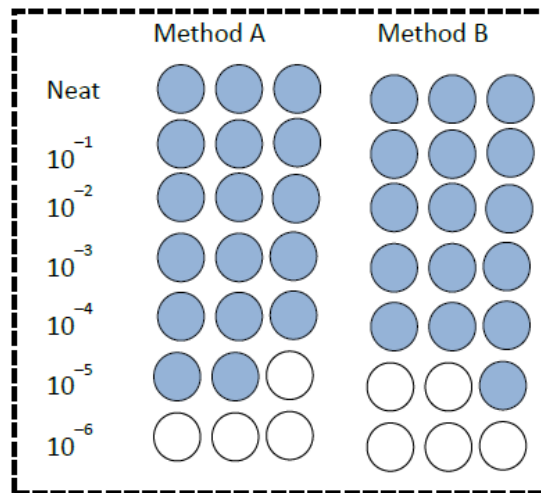
воспроизводимости. Можно также провести оценку перекрестной реактивности, т.к. высокоположительные образцы помещают рядом с отрицательными образцами.

Рисунок 5. Схема 96-луночного планшета для определения аналитической Se, диагностической Se и Sp, а также воспроизводимости

	1	2	3	4	5	6	7	8	9	10	11	12
A	10^{-1}	10^{-1}	10^{-1}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
B	10^{-2}	10^{-2}	10^{-2}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
C	10^{-3}	10^{-3}	10^{-3}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
D	10^{-4}	10^{-4}	10^{-4}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
E	10^{-5}	10^{-5}	10^{-5}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
F	10^{-6}	10^{-6}	10^{-6}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
G	10^{-7}	10^{-7}	10^{-7}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
H	10^{-8}	10^{-8}	10^{-8}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C

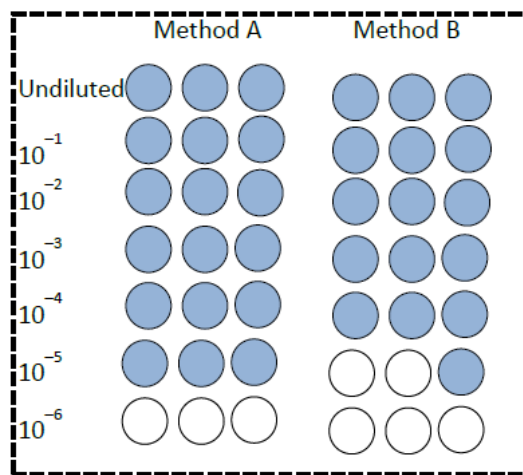
Предел обнаружения (LOD) – это степень аналитической чувствительности (ASe) анализа. LOD – это расчетное количество аналита в специфической матрице, которое обеспечит положительный результат, как минимум в определенном проценте случаев. На Рисунках 6 и 7 представлены гипотетические результаты эксперимента по определению LOD. Например, при титровании с использованием десятикратных разведений все репликаты во всех разведениях могут показать реактивность либо 100%, либо 0%. Это два варианта в данной точке. Последнее разведение, проявляющее 100% реактивность, может быть принято как осторожная оценка нижнего предела обнаружения. Более точная оценка может быть получена на втором этапе эксперимента с использованием более узких интервалов в схеме разведения, уделяя особое внимание области между 100% и 0%. Первый этап заключается в получении, разделении на аликвоты и ослеплении достаточного количества образцов для проведения эксперимента. Второй этап – получение набора разведений аналита, желательно используя вместо буфера ту же матрицу образца в виде разведения, которое отражает диапазон измерений метода, например, в серии десятикратных разведений, 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} . В практических примерах зачастую используют от 3 до 5 репликатов на каждый этап разведения. Используя осторожную оценку 100% (все 3 лунки с самым высоким разведением темные = положительный), на Рисунках 6 и 7 отражены сопоставимые и несопоставимые результаты, соответственно. Метод А представляет валидированный метод, а метод В – новый метод.

Рисунок 6. Пример эксперимента по определению предела обнаружения (LOD) с приемлемым результатом



На Рисунке 6 все репликаты в разведении 10^{-4} теста А и теста В положительные (голубые). В разведении 10^{-5} только две из трех лунок положительные в тесте А, и одна из трех в тесте В. Так как предел обнаружения определяется как разведение, при котором все лунки положительные, результаты, начиная с разведения 10^{-5} и по нисходящей, не могут рассматриваться при оценке сопоставимости. При использовании этих критериев предел обнаружения на Рисунке 6 одинаковый для теста А и теста В, и, следовательно, эти два теста могут считаться сопоставимыми.

Рисунок 7. Пример эксперимента по определению предела обнаружения (LOD) с неприемлемым результатом



На Рисунке 7 наивысшее разведение всех трех репликатов составляет 10^{-5} для теста А. Напротив, при использовании теста В наивысшее разведение, при котором все три репликата все еще положительные, составляет 10^{-4} . Следовательно, пределы обнаружения тестов А и В не могут считаться сопоставимыми, если логарифм разведения неприемлем. Прежде чем принять окончательное решение о сопоставимости, рекомендуется повторить эксперимент несколько раз.

5. Сравнение ROC кривых

Анализ рабочих характеристик приемника (ROC) является полнофункциональным методом для оценки сравнения общей точности диагностического теста, например,

диагностической чувствительности (DSe) и диагностической специфичности (DSp) при разных точках разделения одного и более отличных или измененных тестов (Greiner & Smith, 2000). Основным показателем – площадь под кривой (AUC), например, значение 1 указывает на то, что тест имеет 100% DSe и 100% DSp. В данном случае присутствует отличное разделение между показателями двух групп, т.е. отсутствует наложение обобщенных функций, и ROC-кривая достигает левого угла графика. Напротив, значение 0,5 указывает на однозначное отсутствие разницы между инфицированными и неинфицированными особями; ROC-кривая совпадает с диагональю, указывая на неинформативность теста. Значения в диапазоне от 0,5 до $\leq 0,7$ могут считаться менее точными, от 0,7 до $\leq 0,9$ умеренно точными и от 0,9 до < 1 высокоточными (Greiner & Smith, 2000).

На Рисунке 8 показаны результаты четырех ИФА на выявление антител к вирусу гриппа у свиней. Панель сывороток включала 168 положительных и 77 отрицательных сывороток, в качестве референтного теста использовали реакцию торможения гемагглютинации ($n=245$). При сравнении данных диагностических образцов установили следующую классификацию: тест 1 (AUC = 0,988), тест 4 (AUC = 0,988), тест 3 (AUC = 0,929) и тест 2 (AUC = 0,814) (Таблица 5). В ходе данного эксперимента, проведенного в соответствии с дизайном связанных пар, в качестве показателя статистической значимости для разницы площади под кривой (AUC) может быть использован 95% ДИ (Таблица 6). Таким образом, тест 1 является тестом, имеющим наилучшие DSe и DSp и самую большую площадь под кривой (0,988). Результат 0,988 означает, что случайным образом отобранная особь из положительной группы в 99% случаев продемонстрировала показатели теста выше, чем случайным образом отобранная особь из отрицательной группы (Zweig & Campbell, 1993).

Еще одним способом сравнить результаты является оценка разницы AUC. Например, в случае тестов 1 и 2 разница между AUC составила 0,0173, в случае теста 1 и теста 3 результат составил 0,058, для тестов 1 и 4 – 0,00019, для тестов 2 и 3 – 0,1150, для тестов 2 и 4 – 0,174 и для тестов 3 и 4 – 0,0586 (Таблица 6). Тестами с самыми высокими AUC и самой низкой разницей AUC были тесты 1 и 4, например, оба теста показали AUC 0,988, наложение на 95% ДИ; разница между AUC составила только 0,00019, а p -значение 0,98 указало на отсутствие статистически значимой разницы при уровне значимости равном 5%. Низкие показатели AUC, отсутствие наложения на 95% ДИ, значительно повышенные AUC и p -значения $< 0,05$ указали на отсутствие согласованности в других комбинациях тестов, например, 1 и 3, 1 и 2, 2 и 3, 2 и 4, а также 3 и 4. Тест 1 0,964 к 0,997, тест 2 0,760 к 0,861, тест 3 0,889 к 0,958, тест 4 0,965 к 0,997.

Опубликованы описания более сложных сравнительных исследований с использованием тестов, основанных на аналогичных диагностических и биологических принципах и использующих частотные и классические статистические подходы (Broschi *et al.*, 2006; Engel *et al.*, 2008). См. также Главу 3.6.5. *Статистические подходы к валидации.*

Рисунок 8. Сравнение площади под кривой (AUC) рабочих характеристик приемника (ROC) четырех различных ИФА для выявления антител к вирусу гриппа у свиней

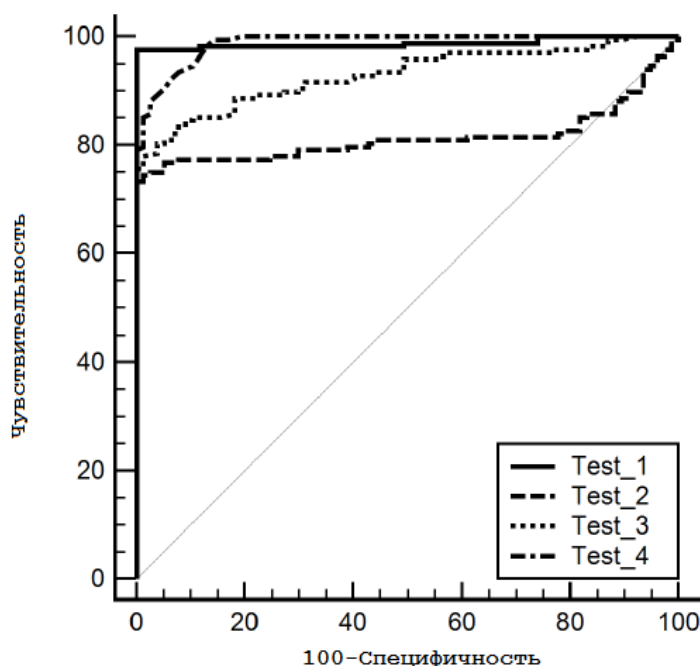


Таблица 5. Сравнение площади под кривой (AUC) рабочих характеристик приемника (ROC) и р-значений четырех различных ИФА для выявления антител к вирусу гриппа у свиней

Параметр	Тест 1	Тест 2	Тест 3	Тест 4
Площадь под ROC-кривой	0,988	0,814	0,929	0,988

Таблица 6. Парные сравнения ROC-кривых

	Тесты 1 и 4	Тесты 1 и 3	Тесты 3 и 4	Тесты 2 и 3	Тесты 1 и 2	Тесты 2 и 4
Разница между AUC	0,0002	0,055	0,0589	0,115	0,173	0,174
95% ДИ	от -0,016 до 0,016	от -0,025 до 0,092	от -0,027 до 0,090	от -0,069 до 0,161	от -0,117 до 0,230	от -0,118 до 0,229
Уровень значимости	$p=0,9808$	$p=0,0007$	$p=0,0003$	$p<0,0001$	$p<0,0001$	$p<0,0001$

6. Обсуждение и заключения

Для заключения о сопоставимости обоих методов необходимо провести оценку результатов сравнительных экспериментов, используя статистические анализы и объективные оценки, который помогут принять окончательное решение (NATA, 2013). Однако при принятии решения зачастую следует также учесть другие критерии, такие как затраты/оборудование, производительность, длительность цикла, возможность обеспечения качества, техническая сложность, приемлемость с нормативной и научной точки зрения, а также навыки интерпретации (Рисунок 1, Этап 6). Важно, чтобы процесс уже функционировал, что позволяет определить, кто имеет полномочия принимать

окончательное решение о сопоставимости методов, например, технический менеджер, руководитель лаборатории, менеджер по качеству.

Например, сравнение двух анализов с зондами TaqMan показало сильную и положительную корреляцию. Воспроизводимость обоих методов была сопоставима при SD 2–3 или Ct $\pm 2-3$. Значения Ct слабopоложительного контрольного образца были стабильно ниже для M1 TaqMan, чем для N1 TaqMan, что указывало на несколько лучшую Se M1 анализа. При использовании в качестве скринингового теста M1 анализ будет более пригоден вследствие его более высокой Se. Результаты были подтверждены при тестировании диагностических образцов, полученных в местах вспышек вируса Хендра в период с 2011 по 2013 гг. (данные не приводятся). С другой стороны, наличие двух тестов, направленных на разные гены, повышает шанс не пропустить новый вариант вируса. В случае смертельных зоонозов, таких как вирус Хендра, это является важным фактором.

Результаты анализа диаграммы Блэнда-Алтмана на Рисунке 4 и в Таблице 4 показывают, что иногда статистические результаты сложно интерпретировать, например 95% ДИ средней разницы исключает ноль. Данный факт может быть интерпретирован как отсутствие сопоставимости, но для компенсации данного результата может быть применено изменение значения точки деления.

ROC-анализ 4 различных ИФА для выявления антител к вирусу гриппа (Рисунок 8 и Таблицы 5 и 6) показал, что два теста имеют почти идентичные DSe и DSr (Тесты 1 и 4). В тоже время, это помогает классифицировать производительность других тестов. В данных условиях стоимость, доступность и другие критерии будут предопределять окончательное решение о том, какой тест является наиболее подходящим для указанной цели.

С другой стороны, не все параметры следует рассматривать при проведении каждого эксперимента по сравнению методов. Например, оценка перекрестной контаминации будет необходима при смене оборудования, например, при переходе с процедуры экстрагирования нуклеиновых кислот вручную на процедуру роботизированного экстрагирования, но не в случае смены основного реактива, например, праймера или зонда. Перед проведением эксперимента рекомендуется принять решение обо всех важных параметрах и критериях допустимости/недопустимости. Самый важный вопрос заключается в том, соответствует ли новый тест целевому назначению.

Для ветеринарных диагностических исследований в природе вещей предусматривать гибкость, когда речь заходит о спецификации пределов допустимости, например, при сравнении двух различных молекулярных тестов может быть принято решение, что различие не должно превышать 1, 2 или 3 Ct для 95% (99%) исследуемых образцов, не более 10% среднего значения тестирования 2 образцов для не менее 95% (99%) исследуемых образцов, не более 1, 2, или 3 SD образцов. Рекомендуется заранее учесть значимые параметры и пределы, хотя качественные факторы (например, стоимость, простота проведения теста, оперативность получения результатов), возможно, появятся в конце. Какой бы параметр не был выбран в качестве основного правила, кандидатный тест по своим рабочим характеристикам, в среднем, не должен значительно уступать валидированному тесту.

Полученные данные и процесс принятия решения о приемлемости изменения должны быть подробно задокументированы и заархивированы для отслеживания при проверке.

7. Анализ данных

Данные хранили и группировали в Microsoft Excel. Анализ и построение графиков для корреляционных диаграмм, диаграмм Блэнда-Алтмана, сравнительных графиков и ROC-анализа проводили с использованием программного обеспечения MedCalc (MedCalc®, Version 12.4.0.0, 64 bit, Window XP/Vista 7/8, www.medcalc.org, Copyright 1993–2013, MedCalc software bvba).

СПИСОК ЛИТЕРАТУРЫ

BLAND J.M. & ALTMAN D.G. (1999). Measuring agreement in methods comparison studies. *Stat. Methods Med. Res.*, **8**, 135–160.

BLAND J.M. & ALTMAN D.G. (2007). Agreement between methods of measurement with multiple observations per individual. *J. Biopharm. Stat.*, **17**, 571–582.

BROCCHI E., BERGMANN I.E., DEKKER A. PATON D.J., SAMMIN D.J., GREINER M., GRAZIOLI S., DE SIMONE F., YADIN H., HAAS B., BULUT N., MALIRAT V., NEITZERT E., GORIS N., PARIDA S., SØRENSEN K. & DE CLERCQ K. (2006). Comparative evaluation of six ELISAs for the detection of antibodies to the non-structural proteins of foot-and-mouth disease virus. *Vaccine*, **24**, 6966–6979.

ENGEL B., BUIST W., ORSEL K., DEKKER A., DE CLERCQ C., GRAZIOLI S. & VAN ROERMUND H. (2008). A Bayesian evaluation of six diagnostic tests for food-and-mouth disease for vaccinated and non-vaccinated cattle. *Prev. Vet. Med.*, **86**, 124–138.

GALL D., COLLING A., MARINO O., MORENO E., NIELSEN K., PEREZ B. & SAMARTINO L. (1998). Enzyme immunoassays for serological diagnosis of bovine brucellosis: a trial in Latin America. *Clin. Diagn. Lab. Immunol.*, **5**, 654–651.

GREINER M., PFEIFFER D., SMITH R.D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.*, **45**, 23–41.

KOZAK M. & WNUK A. (2014). Including the Tukey mean-difference (Bland–Altman) plot in a statistics course. *Teaching Statistics*, **36**, 83–87.

NATIONAL ASSOCIATION OF TESTING AUTHORITIES (OF AUSTRALIA) (NATA) (2013). NATA Technical Note 17, October 2013. Guidelines for the validation and verification of quantitative and qualitative test methods. http://www.nata.com.au/nata/phocadownload/publications/Guidance_information/tech-notes-information-papers/technical_note_17.pdf (accessed 8 April 2015)

ZWEIG M.H. & CAMPBELL G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–577.

*

* *