Приложение № 2 к протоколу
РГ МСИ НТКМетр № 3-2012

ГОСУДАРСТВЕННЫЙ КОМИТЕТ ПО СТАНДАРТИЗАЦИИ РЕСПУБЛИКИ БЕЛАРУСЬ

Республиканское унитарное предприятие «Белорусский государственный институт метрологии» (БелГИМ)

Провайдер проверки квалификации поверочных, калибровочных и испытательных лабораторий

THE INTERNATIONAL HARMONIZED PROTOCOL FOR THE PROFICIENCY TESTING OF ANALYTICAL CHEMISTRY LABORATORIES

(IUPAC Technical Report, 2006)

Michael Thompson, Stephen L. R. Ellison, Roger Wood

Международный гармонизированный протокол для проверки квалификации аналитических химических лабораторий

(Технический отчет IUPAC)

Переводчик:
Ефремова Н. Ю

Минск 2012 год

Аннотация: Международные организации по стандартизации – AOAC международная, ISO, IUPAC– объединились, чтобы создать Международный гармонизированный протокол для проверки квалификации (химических) аналитических лабораторий. Рабочая группа, которая создала протокол, согласилась пересмотреть этот протокол в свете последних разработок и опыта, достигнутых с тех пор, когда он был первый раз опубликован. Эта пересмотренная версия была подготовлена и согласована в свете комментариев, полученных в ходе последующих открытых консультаций.

***Ключевые слова*:** гармонизированный, аналитическое химическое отделение IUPAC, неопределенность, анализ, проверка квалификации, протокол.

Оглавление

ЧАСТЬ 1: Предисловие и введение 5
1.0 Предисловие 5
1.1 Обоснование для проверки квалификации 6
1.2 Проверка квалификации по отношению к другим методам обеспечения качества 7
Часть 2: Гармонизированный протокол; Организация программ проверки квалификации 8
2.1 Сфера действия и область применения 8
2.2. Терминология 8
2.3 Схема проверки квалификации 9
2.4 Организация 9
2.5 Обязанности консультативной группы 9
2.6 Анализ программы 10
2.7 Контрольные материалы 10
2.8 Частота распределения 11
2.9 Приписанное значение 11
2.10 Выбор участниками аналитического метода 12
2.11 Оценивание функционирования 12
2.12 Критерии функционирования 12
2.13 Представление результатов участниками 12
2.14 Отчеты предоставляемые провайдером программы 12
2.15 Взаимодействие с участниками 13
2.16 Сговор и фальсификация результатов 14
2.17 Повторяемость 14
2.18 Конфиденциальность 14
ЧАСТЬ 3: Практическое внедрение 15
3.1 Преобразование результатов участников в показатели 15
3.2 Методы определения приписанного значения 16
3.3 Оценивание приписанного значения как согласованного на основании результатов участников 20
3.4 Неопределенность приписанного значения 22
3.5 Определение стандартного отклонения для оценки квалификации 23
3.6 Данные участников, представленные с неопределенностью 25
3.7 Определение количественных показателей возле предела обнаружения 28
3.8 Осторожность при использовании z-показателей 29
3.9 Классификация, ранжирование и другие способы оценивания данных квалификации 29
3.10 Частота проведения туров 30
3.11 Испытания для определения достаточной однородности и стабильности 30

Итоговые рекомендации 36
Библиография 38
Приложение 1: Рекомендуемая процедура для проверки материала на достаточную однородность 40
Приложение 2: Пример проведения проверки на стабильность 44
Приложение 3: Примеры из практики по определению согласованного значения от участников для использования его в качестве приписанного значения 45
Приложение 4: Оценивание z-показателей в долгосрочной перспективе: суммарные показатели и графические методы 49
Приложение 5: Метод валидации на основании результатов программ проверок квалификации 52
Приложение 6: Как участники должны реагировать на результаты проверки квалификации 53
Приложение 7: Руководство по проверкам квалификации для конечных пользователей данных 58

ЧАСТЬ 1: Предисловие и введение

1.0 Предисловие

В течение 10 лет с тех пор как была опубликована первая версия протокола [1], проверка квалификации очень быстро развивалась. Метод стал широко использоваться во многих секторах химического анализа, и много новых программ проверок квалификации получили распространение во всем мире [2]. Было предпринято детальное изучение проверок квалификации для аналитических химических лабораторий [3]. Международная организация по стандартизации (ISO) опубликовала руководство по проверке квалификации [4] и стандарт по статистическим методам для использования при проверке квалификации [5]. Международная ассоциация по аккредитации (ILAC) опубликовала документ о требованиях к качеству проверки квалификации [6], и многие программы проверки квалификации сейчас аккредитованы. Внесение ясности в течение последнего десятилетия в применение концепции неопределенности в химических измерениях повлияло на направление, в котором мы видим проверку квалификации. Такое необычайное развитие проверки квалификации обусловлено как признанием ее исключительной возможности обнаруживать неожиданные проблемы при проведении анализов, так и современным требованием принимать участие в программе проверки квалификации как части аккредитации аналитических лабораторий.

Как результат всей этой деятельности во многих различных секторах, вместе со значительным количеством проведенных исследований, аналитическое сообщество нарастило большой объем нового опыта в области проверки квалификации. Приятно отметить, что каких-либо глобальных изменений фундаментальных идей и принципов Гармонизированного протокола редакции 1993 года [1] не потребовалось для того, чтобы включить этот новый опыт. Однако, дополнительный опыт показал необходимость и обеспечил основу для усовершенствования нашего подхода ко многим аспектам проверки квалификации и для более специфических и конкретных рекомендаций в некоторых областях. Более того, первоначальный Протокол в большей степени касался организации программ проверки квалификации, и поэтому был предназначен в основном провайдерам программ. Однако возрастающая значимость данных программ проверки квалификации привело к необходимости для дополнительного руководства по интерпретации результатов программ как участниками программ, так и конечными пользователями аналитических данных (таких как заказчики лаборатории, регулятивные органы, и другие стороны, заинтересованные в качественной работе лаборатории). Все эти факторы потребовали обновления Гармонизированного Протокола редакции 1993 года.

Новая версия также предоставляет возможность обратить внимание на то, что некоторые важные аспекты проверки квалификации остаются не полностью задокументированными в настоящее время. В дополнение, мы должны признать, что ряд возможных подходов, приведенных в документах ИСО, является комплексным и охватывает все области измерений. Практический опыт в химическом анализе приводит к тому, что ограниченный набор из этого широкого диапазона подходов обеспечивает оптимальный подход для повседневной аналитической работы. Поэтому этот обновленный Гармонизированный протокол является не только сопоставлением выжимок из других документов, но и оптимальным набором методов, основанных на детальном практическом опыте в управлении программами проверки квалификации, интерпретированном специально для аналитических лабораторий, а также объединения новейших идей.

Создание обновленного Протокола позволяет нам также акцентировать внимание на важности профессионального суждения и опыта как для тех, кто обеспечивает выполнение программ проверки квалификации, так и для участников, соответственно влияющих на результаты. Соблюдение протокола ясно предполагает, что определенные действия должны выполняться. Но способы, которыми они выполняются, должны зависеть в некоторой

степени от конкретной области применения, а диапазон возможных применений не только большой, но и может изменяться со временем. Более того, любой обладающий опытом практикующий аналитик должен быть готов признать, что реально существующие материалы для испытания и методы в интенсивно изменяющейся области аналитической химии будут постоянно демонстрировать случайное необычное поведение, которое требует обсуждения и бдительности экспертов. Соответственно, мы рассматриваем его как ненадежное, чтобы исключить все возможности для экспертной оценки и заменить ее жесткими правилами. Структура этого документа отражает такую философию. Собственно сам Протокол идет первым и включает серию относительно коротких разделов, излагающих основные действия, необходимые для программ, в которых декларируются правила соблюдения протокола. Далее следует ряд более длинных разделов и приложений, которые обсуждают наличие вариантов для выполнения протокола и причины, почему даются конкретные рекомендации. Приложения включают независимые разделы конкретно для участников и для конечных пользователей данных, чтобы помочь им с интерпретацией данных программ проверки квалификации.

Наконец отметим, что хотя этот документ оставляет название "Протокол", мы не пропагандируем философию о том, что существует только один хороший способ проведения проверки квалификации. Этот документ в простой форме излагает только то, что оказалось хорошей и эффективной практикой для аналитических химиков в большинстве случаев. Мы признаем, что обстоятельства могут диктовать, что могут потребоваться альтернативные процедуры, и что такой выбор является ответственностью провайдера, прибегает к помощи консультативного комитета программы. Отметим также, что этот протокол в большей степени ограничивается научными и техническими аспектами проверки квалификации, используемой как средство для улучшения выполнения измерений. Поэтому в нем не затрагиваются такие вопросы, как квалификация или дисквалификация лабораторий или персонала для конкретных целей, или аккредитация провайдеров проверки квалификации или конкретных программ.

1.1 Обоснование для проверки квалификации

Чтобы постоянно производить надежные данные, лаборатория должна внедрить соответствующую систему обеспечения качества и процедуры мониторинга выполнения работ (функционирования). Проверка квалификации является одной из таких процедур. Обычно программы проверки квалификации в аналитической химии основаны на распределении образцов испытываемого материала участникам. Участвующие лаборатории ("участники"), как правило, знают материал для испытаний, который высылается провайдером проверки квалификации, но иногда материал может быть получен "в слепую" (т. е. его получают от обычного заказчика лаборатории). Участники анализируют материал без имеющихся сведений о корректном результате и возвращают результат измерений провайдеру программы. Провайдер преобразует результаты в индексы (показатели), которые отражают характеристику функционирования участвующей лаборатории. Это предупреждает участника о неожиданных проблемах, которые могут иметь место, и стимулирует руководство к принятию корректирующих действий.

Преобладающая черта этого Гармонизированного Протокола состоит в том, что проверка квалификации должна обеспечить информацией о соответствии назначению аналитических результатов, предоставляемых участниками, с тем, чтобы помочь им удовлетворить (выполнить) требования. Это может быть достигнуто если

- критерии для оценивания результатов принимают во внимание аспект соответствия назначению с помощью индексов, которые информируют участников о необходимости улучшения своего функционирования для удовлетворения потребностей потребителя или заинтересованной стороны;

- условия проверки квалификации являются настолько близкими к превалирующим в повседневной практике анализам, так что выходные данные представляют "реальную жизнь";
- метод количественной оценки должен быть прост, и, где, это возможно, быть согласующимся со всей сферой аналитических измерений, чтобы гарантировать быструю (легкую) интерпретацию участниками и заказчиками.

Не смотря на то, что первая (главная) возможность проверки квалификации - это обеспечение основы для самопомощи для каждого участника, было бы неразумно игнорировать тот факт, что существуют и другие применения результатов проверки квалификации. Участники обычно используют свои показатели для демонстрации компетенции потенциальным заказчикам и экспертам (оценщикам) по аккредитации, и это приводит к неблагоприятному влиянию на оцениваемых аналитиков, которые стремятся стать лучшими в проверках квалификации вместо того, чтобы просто оценивать свои повседневные процедуры. Участники должны направить все усилия, чтобы избежать такой тенденции, так как в большинстве случаев провайдеры программы не могут определить или устранить ее. Участники должны также стремиться избегать любой неверной интерпретации накопленных показателей.

1.2 Проверка квалификации по отношению к другим методам обеспечения качества

Всесторонняя программа обеспечения качества (QA) в лабораториях аналитической химии в дополнение к проверке квалификации должна включать следующие элементы: валидация аналитических методов [7]; использование сертифицированных стандартных образцов (CRMs) (когда возможно) [8]; применение повседневного внутреннего контроля качества (IQC) [9]. Традиционно, валидация аналитического метода предполагает, что его характеристики (показатели) - правильность, прецизионность при различных условиях, и линейность калибровки (калибровочная зависимость) и так далее – являются полностью известными. В других современных терминах это означает, что мы имеем оцененную для метода неопределенность измерений в единичной операции при приемлемых условиях использования, и находим ее потенциально соответствующей цели. Идеально метод валидации включает среди прочего использование соответствующих по матрице CRMs, если они существуют, для калибровки или для проверки существующих калибровок, если имеет место эффект влияния матрицы. Когда CRMs не доступны, должны использоваться другие способы.

Внутрилабораторный контроль качества должен проводиться повседневно и включать в себя один или более "контрольных" материалов внутри каждой серии анализа. Эта процедура, комбинированная с использованием контрольных карт, гарантирует, что факторы, определяющие величину неопределенности, существенно не изменяются с тех пор, как было продемонстрировано соответствие цели при валидации процесса. Другими словами, демонстрируется (внутри пределов статистической изменчивости), что оцененная при валидации неопределенность может применяться к каждой отдельной индивидуальной серии, выполненной позднее. Подготовка контрольного материала также в идеале включает использование CRMs, чтобы установить прослеживаемость измеряемых значений, приписываемых ему.

В принципе, метод валидации и внутренний контроль качества сами по себе достаточны для обеспечения точности. Но на практике, они часто менее совершенны. Поэтому проверка квалификации является средством гарантии того, что эти две внутрилабораторные процедуры работают успешно. При валидации метода, неизвестные факторы влияния могут вносить помехи в процесс измерения, и во многих секторах CRMs являются не доступными. При таких условиях прослеживаемость трудно установить и неизвестные источники ошибки могут присутствовать в процессе измерения. Без внешнего эталона лаборатории могут работать в течение длительного периода со смещением или

случайными вариациями значительной величины. Проверка квалификации является средством определения и инициирования исправления таких проблем (смотри Приложение 6). Ее основное достоинство в том, что она предоставляет средство, с помощью которого участники могут получать внешнюю и независимую оценку точности своих результатов.

Часть 2: Гармонизированный протокол; Организация программ проверки квалификации

2.1 Сфера действия и область применения

Этот протокол применим, когда:

- основная цель состоит в оценивании функционирования лаборатории по отношению к установленным критериям, основанным на соответствии общей (единой) цели;
- о соответствии этим критериям можно судить на основании отклонения результатов измерений от приписанного значения; и
- результаты участников представляются с помощью шкалы интервалов или отношений.

Примечание: Эти условия широко применяются при оценивании химических аналитических лабораторий, выполняющих повседневные испытания, но также и в других областях измерений и испытаний.

Этот протокол не предназначен для оценивания калибровочных служб и поэтому не предусматривает использования провайдером программы информации о неопределенности, предоставляемой с результатами участников. Он также не предоставляет критерии для оценивания, сертификации или аккредитации провайдеров программ квалификации.

2.2. Терминология

Технические слова в этом документе соответствуют своим определениям ИСО, когда такие существуют. Аббревиатуры (сокращения) соответствуют Сборнику аналитических терминов IUPAC(1997). Следующие дополнительные термины часто встречаются в этом документе:

- Провайдер проверки квалификации ("провайдер программы" или "провайдер") *Proficiency testing provider*(*“the scheme provider” or “provider”*): Организация ответственная за координацию конкретной программы квалификации.
- Контрольный материал (для проверки) *(Proficiency) test material*: Материал, который при проверке квалификации распределяется для анализа участникам.
- Распределяемый образец (*Distribution unit*): Упакованная порция контрольного материала, которая отправляется или готова для отправки в участвующую лабораторию;
- Порция для испытаний (*Test portion*): Часть распределяемого образца, которая используется для анализа.
  Примечание: Порция для испытаний может включать весь распределяемый образец или его часть;
- Серия (цикл) (*Series*): Часть программы квалификации, определяемая конкретным диапазоном контрольных материалов, аналитами, аналитическими методами или другими общими свойствами.
- Тур (*Round*): Единичный эпизод распределения образцов в серии.

2.3 Схема проверки квалификации

2.3.1 Выполнение программы

- Контрольные материалы будут распределяться на постоянной основе участникам, от которых требуется представлять результаты к указанной дате.
- Значение приписывается для каждой измеряемой величины, или до или после распределения; значение не разглашается участникам до тех пор, пока не наступит окончательный срок для предоставления отчета.
- Результаты подвергаются статистическому анализу и/или преобразуются в показатели провайдером программы, и участники своевременно уведомляются о своих показателях.
- Для плохих исполнителей будут доступны консультации, и все участники будут иметь полную информацию о выполнении программы.

2.3.2 Структура единичного тура

Структура программы для любого одного аналита или тура в серии должна включать следующее:

- провайдер программы организует подготовку и валидацию контрольного материала;
- провайдер программы распределяет образцы для испытаний в соответствии с принятым календарным планом;
- участники анализируют контрольные материалы и предоставляют результаты провайдеру;
- результаты подвергаются статистическому анализу и/или присвоению количественных показателей;
- участники информируются о своем функционировании;
- провайдер программы предоставляет при возможности консультации плохим исполнителям по запросу; и
- провайдер программы анализирует выполнение программы для конкретного тура, и осуществляет необходимые корректировки.

  Примечание: Подготовка для тура программы должна будет часто организовываться тогда, когда проходит предыдущий тур.

2.4 Организация

- За повседневное выполнение программы несет ответственность провайдер программы.
- Провайдер программы должен документировать все инструкции и процедуры в своем руководстве по качеству, и краткое изложение соответствующих процедур должно быть предоставлено всем участникам.
- Провайдер программы должен также информировать участников об эффективности программы в целом, о любых изменениях, которые вносятся и о том, как решаются любые возникающие проблемы.
- Выполнение программы должно периодически анализироваться (смотри ниже).
- Общее руководство программы должно находиться под наблюдением консультативной группы, имеющей представителей (которые должны включать практикующих аналитиков-химиков в соответствующей области) от, например, провайдера программы, субподрядных лабораторий (если они имеются), соответствующих профессиональных органов, участников, и конечных пользователей аналитических данных. Консультативная группа должна включать также эксперта по статистике.

2.5 Обязанности консультативной группы

Консультативная группа будет рассматривать и консультировать по следующим вопросам:

- выбор типов контрольных материалов, аналитов, и диапазонов концентраций аналитов (веществ, определяемых при анализе);

- частота туров;
- система количественных показателей и статистические процедуры (включая те, которые используются при проверке однородности);
- консультации, которые могут быть предложены участникам;
- специфические и общие проблемы, возникающие при выполнении программы;
- инструкции, посылаемые участникам;
- форма протокола участника для представления результатов;
- содержание отчета, посылаемого участникам;
- другие средства связи с участниками;
- комментарии от участников или конечных пользователей, касающиеся выполнения программы;
- уровень конфиденциальности, соответствующий программе.

2.6 Анализ программы

- Выполнение программы должно постоянно анализироваться.
- Провайдер программы должен анализировать результаты по каждому туру программы, отмечая, например, любые достоинства, недостатки, специфические проблемы и возможности для улучшения.
- Провайдер и консультативная группа должны рассматривать обычно не менее раза в год каждый аспект выполнения программы, включая вопросы, выявленные при анализе каждого тура программы провайдера.
- Краткое изложение такого анализа должно быть доступно участникам и при необходимости другим заинтересованным сторонам и согласовано консультативной группой.

2.7 Контрольные материалы

- Провайдер программы должен принять меры по подготовке контрольных материалов. Подготовка контрольных материалов и другие аспекты программы могут выполняться по договору подряда, но провайдер остается ответственным и должен осуществлять соответствующий контроль.
- Организация, осуществляющая подготовку контрольных материалов должна иметь очевидный (демонстрируемый) опыт в области анализа, в которой выполняется проверка квалификации.
- Контрольные материалы, распределяемые в программе, должны быть в целом похожи по типу на материалы, которые обычно анализируются в повседневной практике (в отношении состава матрицы и диапазона концентраций, количества или уровня аналита).
- Находящиеся в большом объеме материалы (навалом), подготовленные для проверки квалификации должны быть достаточно однородными и стабильными в отношении каждого аналита, чтобы гарантировать, что все лаборатории получили распределяемые образцы, которые не отличаются в значительной степени от средней концентрации аналита (см. раздел 3.11). Провайдер программы должен четко установить процедуру, используемую для установления однородности контрольного материала.
  Примечание: Так как между образцами требуется достаточная однородность, участник не должен предполагать, что распределенный образец сам по себе является в полной мере однородным для его конкретной аналитической процедуры. Участники сами несут ответственность, чтобы гарантировать, что порция для испытаний, используемая для анализа, является представительной для всего контрольного материала в распределяемом образце.
- Количество материала в распределяемом образце должно быть достаточным для требуемого анализа, включая любой повторный анализ в рамках протокола программы.

- Когда оценивается нестабильный аналит, может быть необходимым для провайдера программы установить дату выполнения анализа или дату, к которой анализ должен быть завершен.
- Провайдеры программ должны рассматривать любую опасность, которую могут представлять контрольные материалы, и принимать соответствующие действия, чтобы проконсультировать любые стороны (например, распространители контрольного материала, испытательные лаборатории и т. п.), которые могут подвергнуться риску из-за возможных потенциальных вредных факторов.
  Примечание: "Соответствующие действия" включают, но не ограничиваются соответствием специальному законодательству. Многие страны также вводят дополнительные "обязанности соблюдать предосторожность", которые могут выходить за рамки минимальных требований законодательства.
- Одновременно с контрольными материалами участники должны обеспечиваться достаточной информацией о материалах и любых критериях соответствия цели, которые будут использоваться, чтобы позволить им выбрать соответствующие методы анализа. Такая информация не должна включать приписанное значение.

2.8 Частота распределения

Выбор соответствующей частоты для распределения контрольных материалов должен осуществляться провайдером программы с учетом рекомендации, полученной от консультативной группы (см. раздел 3.10). Частота будет нормальной между 2 и 10 турами в год.

2.9 Приписанное значение

Приписанное значение является оценкой значения измеряемой величины и используется с целью расчета показателей.

- Приписанное значение может быть определено одним из следующих методов:
  - измерением в эталонной лаборатории[1];
  - сертифицированным значением(ями) для CRM, используемого в качестве контрольного материала;
  - прямым сличением контрольного материала для проверки квалификации с CRMs;
  - на основании соглашения (консенсуса) между экспертными лабораториями;
  - приготовлением (т. е. приписывание значения на основании пропорций, используемых в растворе или другой смеси ингредиентов с известным содержанием аналита);
  - согласованным значением (которым является значение, полученное непосредственно из заявленных результатов).

Приписанное значение не должно разглашаться участникам до окончания отчетного срока для результатов.

- Провайдер программы должен сообщить приписанное значение и оценку его неопределенности участникам вместе с отчетными результатами и показателями и должен предоставить достаточные пояснения о том, как приписанное значение и неопределенность были получены. Методы для определения приписанного значения обсуждаются ниже (см. раздел 3.2).
- В секторах, где используются эмпирические методы анализа, приписанное значение обычно рассчитывается из результатов, полученных с помощью четко определенной аналитической процедуры. Альтернативно приписанное значение может быть рассчитано из результатов двух или более эмпирических методов, которые являются практически эквивалентными.

[1]"Эталонная лаборатория" в данном контексте является лабораторией, установленной провайдером программы и консультативной группой, в качестве обеспечивающей эталонные значения с соответствующей достоверностью для целей программы.

- Иногда может быть необходимым для программы использовать разные приписанные значения для разных методов, но такую рекомендацию следует использовать в случае четкой необходимости.
- Когда приписанное значение основано на эмпирическом методе, участникам следует сообщить заранее, какая эмпирическая процедура будет использоваться для определения приписанного значения.

2.10 Выбор участниками аналитического метода

- Участники, как правило, должны использовать аналитический метод по своему выбору. В некоторых случаях, однако, например, когда законодательство так требует, участники могут быть проинструктированы использовать определенный задокументированный метод.
- Методы должны быть теми, которые используют участники в своей повседневной работе в соответствующем секторе, а не варианты метода, специально приспособленные (адаптированные) для проверки квалификации.

2.11 Оценивание функционирования

Лаборатории будут оцениваться на основании разницы между их результатом и приписанным значением. Показатель функционирования будет рассчитываться для каждой лаборатории, используя статистическую программу, детально описанную в разделе 3.1.

Примечание: Индекс z, основанный на критерии соответствия назначению, является единственным типом показателей рекомендованных в этом протоколе.

2.12 Критерии функционирования

Для каждого аналита в туре должен быть установлен критерий функционирования, по отношению к которому можно судить о функционировании лаборатории. Критерий функционирования будет устанавливаться так, чтобы гарантировать, что аналитические данные, производимые в повседневной работе лаборатории, будут такого качества, которое соответствует их использованию по назначению. Как правило, он не будет устанавливаться так, чтобы представить лучшее функционирование, чем способны обеспечить типичные методы (смотри раздел 3.5).

2.13 Представление результатов участниками

- Участники должны представить результаты способом и в формате, требуемым программой.
- Провайдер программы должен установить дату, к которой должны быть представлены результаты. Результаты, представляемые после предельного срока должны отклоняться.
- Представленные результаты не могут быть откорректированы или исправлены.
  Примечание: Причина для такого строго подхода заключается в том, что проверка квалификации подразумевает каждый аспект получения и производства аналитического результата, включая расчет, проверку и представление результата.

2.14 Отчеты, предоставляемые провайдером программы

- Провайдер программы должен предоставить отчет о функционировании каждому участнику для каждого тура.
- Отчеты, изданные для участников, должны быть четкими и полными и показывать распределение результатов от всех лабораторий вместе с показателем функционирования каждого участника.
- Результаты испытаний, используемые провайдером программы, должны быть также доступны, чтобы участники имели возможность проверить, что их данные были корректно введены.

- Отчеты должны быть сделаны как можно быстрее после представления результатов в координирующую лабораторию и, если это вообще возможно, до следующего распределения образцов.
- Участники должны получить по крайней мере:
  a) отчеты в четком и простом виде и
  b) результаты всех лабораторий в графической форме (т. е. в виде гистограммы, столбиковой диаграммы или другого графика распределения) с соответствующими итоговыми статистиками.
  Примечание: Хотя в идеале все результаты должны быть представлены участникам, этого невозможно достичь в некоторых очень протяженных программах (например, где участвуют сотни участников, каждый из которых определяет 20 аналитов в каждом туре).

2.15 Взаимодействие с участниками

- Присоединяясь к программе, участники должны обеспечиваться детальной информацией, которая должна описывать
  - диапазон возможных испытаний и критерии участия для участников;
  - метод установки критериев функционирования;
  - критерии функционирования, действующие на момент присоединения к программе, если только критерии не устанавливаются отдельно для каждого контрольного материала;
  - метод(ы) определения приписанных значений, включая методы измерения в соответствующих случаях;
  - краткое содержание статистических процедур, используемых для получения показателей участников;
  - информация об интерпретации показателей;
  - условия в отношении участия (например, последний срок предоставления результатов, избежание сговора с другими участниками и т. п.);
  - состав и метод выбора консультативной группы; и
  - контактная информация провайдера и других имеющих отношение к делу организаций.
    Примечание: Связь с участниками может осуществляться через соответствующие средства связи, включая, например, периодические информационные бюллетени, постоянные отчеты по анализу программы, периодические открытые семинары, или связь по электронной почте.
- Следует информировать участников о любых непредвиденных изменениях в проекте или в ходе выполнения программы.
- Должны быть доступны консультации для плохих исполнителей (участников с плохим функционированием), хотя это может быть достигнуто в форме списка консультантов, которые являются экспертами в данной области.
- Участники, которые считают, что их оценка функционирования ошибочна, должны иметь возможность сообщить об этом провайдеру программы.
- Для того чтобы участники вносили вклад в разработку программы и чтобы позволить участникам предупредить провайдера программы о любых непредвиденных трудностях с контрольными материалами должен существовать механизм, в соответствии с которым участники могут комментировать аспекты выполнения программы и проблемы с отдельными контрольными материалами,.
  Примечание: Отзывы участников должны приветствоваться.

2.16 Сговор и фальсификация результатов

- Избегать сговора или фальсификации результатов является ответственностью участвующих лабораторий. Это должно быть задокументированным условием участия в программе, которое включается в инструкции для участников программы.
- Провайдер программы должен принимать меры, чтобы препятствовать сговору с помощью соответствующего проекта программы. (Например, может быть объявлено, что внутри одного тура может быть случайным образом распределено более одного контрольного материала, так что лаборатории не могут сравнить результаты напрямую, и не должно быть никакого опознаваемого (идентифицируемого) повторного использования материалов в последующих турах)

  Примечание: Сговоры, или между участниками или между отдельными участниками и провайдером программы, противоречат профессиональному научному поведению и служат только для сведения к нулю выгод от проверок квалификации для потребителей, органов по аккредитации и аналитиков. Поэтому сговоры крайне не одобряются.

2.17 Повторяемость

Представлять среднее значение повторных определений по образцам для проверки квалификации следует, если только это является нормой для повседневной работы. (Процедуры, используемые лабораториями, участвующими в программах проверки квалификации, должны быть такими же, которые используются в повседневном анализе образцов.)

Примечание: Отдельное представление повторных результатов, выполненных внутри лабораторий, разрешается в проверках квалификации, но не рекомендуется. Если такая практика применяется, то провайдеры программы и участники должны остерегаться неверной интерпретации стандартных отклонений, рассчитанных по многим участникам. Например, внутригрупповая сумма квадратов, полученная с помощью дисперсионного анализа, не должна интерпретироваться как "средняя" дисперсия повторяемости, если используются различные аналитические методы.

2.18 Конфиденциальность

Степень конфиденциальности, определяемая провайдером программы и участниками в отношении информации, касающейся программы и данных, представляемых участниками, должна устанавливаться в условиях участия и доводиться до сведения участников до начала присоединения к программе.

Примечание: При установлении условий конфиденциальности организаторы должны использовать преимущества открытой доступности для аналитического сообщества основных данных о функционировании, и рекомендуется предоставлять для открытой публикации такую информацию при условии надлежащей защиты информации отдельных участников.

За исключением установленных в условиях участия других случаев:

- Провайдер программы не должен открывать индивидуальности (личности) участников любой третей стороне, а также другим участникам, без высказанного разрешения касающихся этого участников.
- Участники должны идентифицироваться в отчетах только с помощью кодов (шифров).

  Примечание: Случайное приписывание лабораторных кодов (шифров) для каждого тура предотвращает идентификацию на основании истории участия и рекомендуется в случае возможности выполнения на практике.
- Участники могут сообщать свои личные результаты, включая постоянные отчеты по программе, частным образом органу по аккредитации лабораторий и другим оценочным органам, когда это требуется для цели оценивания или клиентам (включая вышестоящую

организацию для лаборатории, если применимо) с целью демонстрации способности выполнять аналитические измерения.

- Участники могут публиковать информацию о своем личном функционировании, но не должны публиковать сравнительную информацию о других участниках, включая ранжирование (упорядочение) показателей.

ЧАСТЬ 3: Практическое внедрение

3.1 Преобразование результатов участников в показатели

3.1.1 Особенности определения количественных показателей

В 1993 году Гармонизированный протокол рекомендовал преобразование результатов участников в z-показатели, и опыт в промежуточные годы продемонстрировал широкую применимость и признание z-показателя в проверках квалификации. Результат участника x преобразуется в z-показатель в соответствии с выражением:

$$z = (x - x_a)/\sigma_p \qquad (1)$$

где x_a – "приписанное значение", наилучшая оценка провайдера программы для значения измеряемой величины (истинное значение концентрации аналита в материале для проверки квалификации)

σ_p – "стандартное отклонение для оценки квалификации", основанное на соответствии назначению. Рекомендации по вычислению x_aи σ_p даются ниже (см. разделы 3.2 – 3.5).

Примечание 1: В Гармонизированном Протоколе 1993 года [1] σ_p было обозначено "заданным значением". Такое использование сейчас представляется обманчивым.

Примечание 2: В Руководстве ИСО 43 [4] и Статистическом Руководстве ИСО 13528 [5] символ $\hat{\sigma}$ используется для стандартного отклонения для оценки квалификации. Здесь σ_p используется для подчеркивания важности приписывания диапазона, соответствующего конкретной цели.

Основная идея z-показателя состоит в том, чтобы сделать все показатели проверки квалификации сравнимыми, так чтобы смысл показателя стал сразу же очевиден вне зависимости от концентрации или подлинности аналита, природы контрольного материала, физического принципа, лежащего в основе аналитического измерения или организации, предлагающей программу. В идеале показатель, скажем z = -3,5, вне зависимости от его происхождения должен сразу же иметь одинаковый смысл для любого, провайдера, участника, или конечного пользователя, связанного с проверкой квалификации. Это требование тесно связано с идеей соответствия назначению. В формуле, определяющей z, разность ($x - x_a$) является погрешностью измерения. Параметр σ_p описывает стандартную неопределенность, которая наилучшим образом подходит для области применения результатов анализа, или другими словами "соответствует назначению". Она будет необязательно близка к неопределенности, связанной с представленными результатами. Поэтому, несмотря на то, что мы можем интерпретировать z-показатели на основании стандартного нормального распределения, мы не ожидаем от них соответствия этому распределению.

Неопределенность, которая соответствует назначению результата измерения, зависит от использования. Например, если относительная стандартная неопределенность [т. е. $u(x)/x$] величиной 10 % вероятно подходит для многих измерений в окружающей среде, более малая относительная неопределенность требуется для определения содержания металла в партии металлолома, содержащего золото, для определения ее коммерческого значения. Но это еще не все. Решение о неопределенности для соответствия назначению - это компромисс между затратами на анализ и затратами на принятие некорректного решения. Получение меньшей неопределенности требует непропорционально больших

расходов на анализ. Но использование методов с большей неопределенностью означает большую вероятность принятия дорогого некорректного решения, основанного на данных. Соответствие назначению определяется неопределенностью, которая балансирует между этими факторами, т. е. которая минимизирует ожидаемые общие потери [10]. Аналитики и их заказчики обычно не проводят формальный математический анализ ситуации, но они должны, по крайней мере, прийти к соглашению о том, что подразумевается под соответствием назначению для каждого определенного применения.

3.1.2 Как должен интерпретироваться z-показатель?

Важно подчеркнуть, что интерпретация *z*-показателей, как правило, не основана на итоговых статистиках, которые описывают наблюдаемые результаты участника. Наоборот, она использует предполагаемую модель, основанную на предложенном провайдером программы критерии соответствия назначению, который выражается в виде стандартного отклонения для оценки квалификации σ_p. В частности, интерпретация основана на нормальном распределении $x \sim N(x_{true}, \sigma_p^2)$, где x_{true} является истинным значением для измеряемой величины. Для такой модели и в предположении близости приписанного к x_{true} z-показатели подчиняются стандартному нормальному распределению:

- Показатель равный нулю предполагает отличный результат. Это будет случаться редко даже в большинстве компетентных лабораториях.
- Примерно 95 % z-показателей будут попадать между -2 и +2. Знак (т. е. "–" или "+") показателя обозначает отрицательную или положительную погрешность соответственно. Показатели в этом диапазоне, как правило, объявляются "приемлемыми" или "удовлетворительными".
- Показатель вне диапазона от – 3 до 3 будет встречаться крайне редко и используется, чтобы показать, что причина такого попадания должна быть исследована и исправлена. Показатели в этой группе обычно объявляются "неприемлемыми" или "неудовлетворительными", хотя менее уничтожительная фраза такая как "требующие принятия мер" является предпочтительнее.
- Показатели в диапазонах от -2 до -3 и от 2 до 3 ожидаются примерно 1 случай из 20, так что отдельное событие даже такого рода не является важным. Показатели в этой группе иногда объявляются "сомнительными".

Почти никто из лабораторий точно не соответствует вышесказанному. Многие участники будут оперировать со смещенным средним и скользящим стандартным отклонением, которое отличается от σ_p. Некоторые будут генерировать экстремальные результаты вследствие больших ошибок. Однако, модель служит подходящим руководством для принятия действий на основании z-показателей, полученных всеми участниками по следующим причинам. Смещенное среднее или стандартное отклонение большее, чем σ_p будет, в конечном счете всегда производить большую долю результатов, приводящих к значениям $|z| > 2$ и $|z| > 3$, чем стандартная нормальная модель (это примерно 0,05 или 0,003 соответственно). Это будет правильно предупреждать участника о проблеме. И наоборот, участник с несмещенным средним значением и стандартным отклонением равным или меньшим, чем σ_p, будет производить малую долю таких результатов, и будет получать только несколько неблагоприятных отчетов.

3.2 Методы определения приписанного значения

Существует несколько возможных подходов, которые провайдер проверки квалификации может использовать для определения приписанного значения и его неопределенности. Все имеют достоинства и недостатки. Поэтому выбор соответствующего метода для приписывания в различных программах и даже турах в рамках программы или серий будет зависеть от целей программы. При выборе методов приписывания значений организаторы программы и консультативные группы должны рассмотреть следующее:

- стоимость для организаторов и участников – высокая стоимость может отпугивать участников и посредством этого снижать эффективность программы;
- любые требования законодательства для согласованности с эталонной лабораторией или другими организациями;
- потребность в независимых приписанных значениях для обеспечения проверки смещения для совокупности в целом;
- любые специфические требования для прослеживаемости к конкретным эталонным значениям.

Примечание: Выполнение метрологической прослеживаемости участниками рассматривается как важный элемент хорошего аналитического обеспечения качества (QA). Когда метрологическая прослеживаемость и соответствующее методы QA/QC– в частности валидация с использованием соответствующих матриц CRMs – соответствующим образом осуществляются, в итоге ожидается хорошая согласованность и небольшое рассеяние результатов. Обычное наблюдение за рассеянием результатов (и в частности за частью лабораторий, получающих приемлемые показатели) является непосредственной проверкой эффективной прослеживаемости, вне зависимости от приписанного значения. Однако, контроль относительно независимого приписанного прослеживаемого значения может предоставить полезную дополнительную проверку эффективной прослеживаемости.

3.2.1 Измерение в эталонной (референтной) лаборатории

В принципе, приписанное значение и неопределенность могут быть получены в подходящей квалифицированной измерительной лаборатории с помощью метода с достаточно малой неопределенностью. Для большинства практических целей это вполне эквивалентно использованию CRM (ниже). Выгоды состоят в том, что материал надлежащим образом подобран специально для участников программы. Принципиальный недостаток состоит в том, что может потребоваться несоразмерное усиление расходов, например, значительные исследования требуются для валидации методологии для спорного материала или для устранения возможности значительных влияний (помех).

3.2.2 Использование сертифицированного эталонного материала

Если существует CRM в достаточном количестве для использования в проверке квалификации, сертифицированное значение (я) и связанная с ним неопределенность может использоваться непосредственно. Это быстро и просто в применении, и (обычно) обеспечивает независимость значения от результатов участников. Соответствующая прослеживаемость для эталонного значения также автоматически обеспечивается (по определению). Однако, у этого метода имеются свои недостатки. CRM с натуральной (природной) матрицей обычно не имеются в наличие в достаточном количестве и/или по приемлемой цене для постоянного использования в программах проверки квалификации. Они могут быть легко опознаны участниками, которые впоследствии способны догадаться о сертифицированном значении. И наконец, хотя ценность (значимость) большинства проверок квалификации в основном велика, наиболее ценными (значимыми) они являются в аналитическом секторе, где стандартные образцы находятся в дефиците или не существуют совсем.

3.2.3 Непосредственное сравнение материала для проверки квалификации с сертифицированным стандартным образцом

В этом методе материал для испытаний анализируется несколько раз параллельно с соответствующим CRM случайным образом в условиях повторяемости (т. е. в единичной серии) методом с приемлемо малой неопределенностью. При условии, что CRM близко сопоставимы по матрице, концентрации, видообразованию и способу выделению аналита с предполагаемым материалом для проверки квалификации, результаты для материала проверки квалификации, определяемые с помощью калибровочной зависимости, основанной на сертифицированных значениях CRM, будут прослеживаемы к значениям CRM, а затем через CRM и к более высоким эталонам. Неопределенность будет включать

только члены из-за неопределенностей CRM и ошибки (погрешности) повторяемости анализа.

Примечание: Эта практика описывается в ИСО 13528 [5]. Она идентична выполнению измерений с помощью согласованных CRM в качестве основы для калибровки, и поэтому может быть описана как «измерение, использующее согласованные калибровочные материалы».

На практике бывает трудно определить является ли CRM достаточно подобным во всех отношениях материалу для проверки квалификации. Если они не подобны, то должен быть включен соответствующий дополнительный вклад в расчет неопределенности для приписанного значения. Трудно определить величину этого дополнительного вклада. Как и ранее (в предыдущем пункте), проверки квалификации являются наиболее значимыми в аналитических секторах, где наблюдается отсутствие стандартных образцов.

3.2.4 Согласованное значение от экспертных лабораторий

Приписанное значение берется как согласованное от группы экспертных лабораторий, которые достигли взаимного согласия по материалу для проверки квалификации с помощью тщательного выполнения известных (признанных) эталонных методов. Такой метод особенно ценен, когда измеряются определяемые с помощью операций («эмпирические») параметры или, например, когда ожидается, что повседневные результаты лаборатории будут согласованы с результатами из небольшой совокупности лабораторий, определенных законом для арбитража или законодательного регулирования. Он имеет также преимущество в том, что предоставляет возможность перекрестного контроля среди экспертных лабораторий, который помогает предотвратить большую ошибку.

На практике, однако, усилия, требуемые для достижения согласия и пригодно малой неопределенности являются примерно такими же как те, которые требуются для сертификации стандартного образца. Если эталонные (референтные) лаборатории использовали повседневную процедуру для анализа материала для проверки квалификации, их результаты будут, в среднем, не лучше, чем результаты большинства участников самой проверки квалификации. Кроме того, так как количество имеющихся эталонных (референтных) лабораторий в силу сложившихся обстоятельств мало, неопределенность и/или изменчивость согласованного значения такой небольшой группы может быть достаточно большая и повредит проверке квалификации.

Если используется согласованное значение от экспертных лабораторий, приписанное значение и связанная с ним неопределенность оценивается с помощью соответствующей оценки среднего значения (обычно, среднее арифметическое или его робастная оценка). Неопределённость приписанного значения затем базируется или на комбинированных (суммарных) заявленных неопределенностях (если это подходит) или на соответствующих статистических неопределенностях, объединяемых (суммируемых) с любыми дополнительными членами, требуемыми для учета (принятия во внимание) неопределенностей калибровочной цепочки, влияний матрицы и любых других влияний.

3.2.5 Процедура приготовления

Процедура приготовления включает в себя добавление известного количества или концентрации аналита (или материала, содержащего аналит) в основной материал, который его не содержит. Должны быть рассмотрены следующие обстоятельства:

- Основной материал должен действительно не содержать аналит или его концентрация должна быть точно известна;
- Может быть сложным получить достаточную однородность (см. раздел 3.11), если микроаналит добавляется к твердому основному материалу;
- Даже если видообразование является адекватным (подходящим), добавляемый аналит может быть более свободно связан с матрицей, чем аналит, являющийся родным для

типичных материалов для испытаний, и поэтому извлечение добавленного аналита может быть нереалистично высоким.

При условии, что эти проблемы будут преодолены, приписанное значение определяется просто из пропорций используемых материалов и известных концентраций (или чистого вещества, если добавляется чистый аналит). Его неопределенность обычно оценивается из неопределенностей чистого вещества или концентраций аналита используемых материалов и весовых и объемных неопределенностей, хотя такие вопросы как содержание влаги и другие изменения, происходящие во время смешивания должны также приниматься во внимание, если они значимы. Метод является относительно простым в выполнении, если материал для проверки квалификации является однородной жидкостью и аналит является действительно раствором. Однако, он может быть не пригоден для твердых натуральных материалов, где аналит уже присутствует ("родной" или "приобретенный").

3.2.6 Согласованное значение от участников

Согласованное значение от участников в настоящее время является наиболее широко используемым методом для определения приписанного значения: Действительно, трудно найти сопоставимую по затратам альтернативу. Идея согласованного значения состоит не в том, что все участники согласованно работают внутри пределов, определенных уровнем повторяемости, а в том, что результаты, производимые большинством, являются несмещенными, и их рассеяние имеет определенную форму (вид). Чтобы получить наиболее вероятное значение для измеряемой величины (т. е. приписанное значение) мы используем соответствующую меру среднего значения распределения результатов и мы (как правило) используем ее среднеквадратическую ошибку в качестве оценки его неопределенности (см. раздел 3.3).

Преимущество согласованного значения от участников включает низкую стоимость, так как приписанное значение не требует дополнительных аналитических работ. Согласованность является часто хорошей среди участников, так как ни один из членов или группа не имеет более высокого статуса. Расчет значения обычно является непосредственным. В заключение, опыт показывает, что согласованные значения обычно на практике очень близки к надежным эталонным значениям, обеспечиваемыми процедурой приготовления, к согласованным значением от экспертных лабораторий и эталонным значениям (сертифицированных стандартных образцов (CRM) или выданных эталонными лабораториями).

Принципиальный недостаток согласованного значения от участников заключается в том что, во-первых, они являются зависимыми от результатов участников и, во-вторых, их неопределенности могут быть слишком большими, когда количество лабораторий мало. Недостаток независимости имеет два возможных последствия: (i) смещение для совокупности в целом может быть сразу не обнаружено, так как приписанное значение будет следовать за ним; (ii) если большинство результатов смещено, то участники, результаты которых не смещены, могут несправедливо получить экстремальные z-показатели. На практике последний случай встречается редко за исключением малых совокупностей, использующих одинаковый метод; существование нескольких отдельных подсовокупностей является более общей проблемой. Как провайдеры проверки квалификации, так и участники должны быть внимательны к таким возможным ситуациям (хотя точно также они должны быть внимательны к возможности допущения ошибки в любом другом методе получения приписанного значения). Ситуация обычно быстро исправляется, если о ней становится известно. Одно из преимуществ проверки квалификации заключается в том, что участники могут узнать о неизвестных основных проблемах, а также о проблемах, касающихся конкретных лабораторий.

Ограничения, вызванные малым размером группы, являются часто более серьезными. Если количество участников меньше, чем 15, даже статистическая неопределенность

согласованного значения (определяемая как среднеквадратическая ошибка) будет слишком высока и информационная ценность z-показателей будет соответственно уменьшена.

Несмотря на очевидные недостатки, однако, имеется большой опыт, демонстрирующий, что проверки квалификации выполняются очень хорошо с использованием согласованного значения, при условии, что организаторы проявляют бдительность к возможным случайным трудностям и применяют соответствующие методы расчета. Точные методы оценивания согласованного значения на основании результатов участников подробно обсуждаются ниже.

3.3 Оценивание приписанного значения как согласованного на основании результатов участников

3.3.1 Оценки среднего значения

Если результаты участников в туре являются унимодальными (одновершинными) и, после исключения выбросов, достаточно близки к симметричным, то различные меры среднего значения практически совпадают. Соответственно, мы чувствуем уверенность, принимая одно из них, такое как мода, медиана, или робастное среднее, в качестве приписанного значения. Нам нужно использовать метод оценки, который является нечувствительным к присутствию выбросов и тяжелых хвостов, чтобы избегать чрезмерного влияния от плохих результатов, и именно поэтому медиана или робастное среднее является ценным (полезным).

Робастные статистики основаны на предположении, что данные являются выборкой из, по существу, нормального распределения, засоренного тяжелыми хвостами и небольшой долей выбросов. Статистики рассчитываются с помощью понижающего взвешивания точек данных, которые удалены от среднего значения, и последующего компенсирования для понижающего взвешивания. Существует много версий робастных статистик [5, 11]. Медиана является простым типом робастного среднего. Робастное среднее Хубера, полученное с использованием алгоритма, рекомендованного Комитетом аналитических методов (Analytical Methods Committee, AMC) [11] и ИСО 5725 и ИСО 13528 в качестве "алгоритма А", предоставляет большую возможность для использования информации в данных, чем это делает медиана, и поэтому в большинстве случаев имеет в некоторой степени малую среднеквадратическую ошибку. Медиана, однако, является более робастной (устойчивой), если распределение частот является сильно асимметричным. Поэтому робастное среднее предпочитается в случаях, когда распределение является близким к симметричному. Мода не определяется точно для выборок из непрерывных распределений, и для ее оценки требуются специальные методы [12]. Несмотря на это обстоятельство, мода может быть особенно полезна, когда получены результаты, имеющие бимодальные или мультимодальные распределения (см. Приложение 3).

Рекомендованная схема для оценивания согласованного значения и его неопределенности приводится ниже. В этой схеме прописан элемент анализа, основанного на опыте в аналитической химии и статистике, и это скорее его сильная, чем слабая сторона и такой анализ следует рассматривать как неотъемлемую часть этой схемы. Это нужно делать потому, что очень сложно или невозможно порекомендовать набор правил, которые могут быть выполнены механически для обеспечения соответствующего согласованного значения для набора произвольных данных.

3.3.2 Рекомендованная схема для получения согласованного значения и его неопределенности

Рекомендованная схема для получения приписанного значения x_a и его неопределенности на основании соглашения представлена в наборе процедур в рамке ниже. Логическое обоснование для основных деталей обсуждается в разделе 3.3.3. Примеры использования этой схемы представлены в Приложении 1.

Рекомендация 1

a. Исключите из данных любые результаты, которые идентифицируются как недействительные (ошибочные) (например, если они выражены в ошибочных единицах или получены с использованием несоответствующего метода) или как экстремальные выбросы (например, они находятся вне диапазона ± 50 % от медианы).
b. Исследуйте визуальное представление оставшихся результатов с помощью точечного графика (для малых (n < 50) набора данных), столбиковой диаграммы или гистограммы (для больших наборов данных). Если выбросы являются причиной чрезмерного сжатия основной доли результатов, постройте новый график с удаленными выбросами. Если распределение, без учета выбросов, очевидно унимодальное и приблизительно симметричное, то переходите к выполнению пункта (c), в других случаях – к пункту (d).
c. Рассчитайте робастное среднее μ_{rob} и стандартное отклонение σ_{rob} для n результатов. Если σ_{rob} примерно меньше, чем $1{,}2\sigma_p$, то используйте μ_{rob} в качестве приписанного значения x_a и $\sigma_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. Если $\sigma_{rob} > 1{,}2\sigma_p$, то переходите к выполнению пункта (d).
d. Сделайте оценку ядерной плотности (kernel density) распределения по результатам, используя нормальные ядра с шириной h равной $0{,}75\sigma_p$. Если эти результаты имеют унимодальную или приблизительно симметричную ядерную плотность, и мода и медиана практически совпадают, то используйте μ_{rob} в качестве приписанного значения x_a и $\sigma_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. В других случаях переходите к выполнению пункта (e).
e. Если второстепенные моды могут быть с уверенностью приписаны выбросовым результатам, и их вклад составляет менее примерно 5 % в общую площадь, то еще используйте μ_{rob} в качестве приписанного значения x_a и $\sigma_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. В других случаях переходите к выполнению пункта (f).
f. Если второстепенные моды вносят значительные вклады в площадь ядра, рассмотрите возможность того, что две или более несходных совокупности представлены в результатах участников. Если возможно, то сделайте вывод на основании независимой информации (например, детали аналитических методов участников) о том, что одна из этих мод является корректной, а другие – некорректны, используйте выбранную моду в качестве приписанного значения x_a и ее среднеквадратическую ошибку в качестве ее стандартной неопределенности. В других случаях переходите к выполнению пункта (g).
g. Если методы, приведенные выше, оказались неудачными, то оставьте попытку определить согласованное значение и не представляйте в отчетах по туру индивидуальные показатели функционирования для лабораторий. Однако может быть полезным предоставить участникам итоговые статистики, полученные на основании набора данных в целом.

3.3.3 Примечания относительно логического обоснования схемы для определения приписанного значения

Логическое обоснование для приведенной выше схемы является следующее:

Использование $\sigma_{rob}/\sqrt{n}$ в качестве стандартной неопределенности приписанного значения открыто для возражения с теоретической точки зрения, так как влияние некоторых из n результатов взвешивается с понижающим коэффициентом при расчете σ_{rob} и их выборочное распределение является сложным. Тем не менее, это один из методов,

рекомендованных ИСО 13528. На практике $u(x_a) = \mathcal{E}_{rob} / \sqrt{n}$ используется только в качестве грубого ориентира для определения пригодности приписанного значения и теоретическое возражение не имеет большого значения.

В пункте (b) выше мы ожидаем, что $\mathcal{E}_{rob} = \sigma_p$, так как участники будут стараться достичь соответствия назначению. Если мы находим, что $\mathcal{E}_{rob} < 1{,}2\sigma_p$, то разумным предположением будет то, что либо лаборатории имеют трудности при достижении требуемого уровня воспроизводимости в результатах из одной совокупности, либо в результатах могут быть представлены две или более различных совокупности. Ядерная плотность может помочь сделать выбор между двумя этими возможностями. Так или иначе последняя ситуация приводит к двум (или более) модам, зависящим от разделения средних значений совокупностей и количества результатов в каждой выборке.

Использование для построения ядерной плотности ширины интервала *h,* равной $0{,}75\sigma_p$, является компромиссом, который препятствует возникновению ложных мод без неоправданного увеличения дисперсии плотности ядра по отношению к $\mathcal{E}_{rob}$.

3.4 Неопределенность приписанного значения

Если имеется неопределенность $u(x_a)$ приписанного значения x_a и участник функционирует в соответствии с стандартным отклонением σ_p, определенным на основании соответствия назначения, то неопределенность отклонения участника от x_a будет $\sqrt{u^2(x_a) + \sigma_p^2}$, так что мы могли бы ожидать значения *z*-показателей с рассеянием в интервале отличном от N (0; 1). Поэтому уместно сравнить $u^2(x_a)$ с σ_p^2 для проверки того, что $u^2(x_a)$ не имеет негативного влияния на z-показатели. Например, если $u^2(x_a) = \sigma_p^2$, то *z*-показатели должны быть увеличены на коэффициент примерно равный 1,4, что приведет к неблагоприятному последствию. С другой стороны, если $u^2(x_a) > \sigma_p^2$, увеличивающий коэффициент будет примерно 1,05 и его влияние будет незначительно для практических целей. Следовательно, рекомендуется, чтобы *z*-показатели не представлялись участникам в безоговорочной форме, если обнаружено, что $u^2(x_a) > 0{,}1\sigma_p^2$. (Значение коэффициента, равное 0,1, было присвоено, но точное его значение по сути выбирается произвольно, и должно назначаться провайдером программы.) Если неравенство было превышено незначительно (но не сильно), программа может выдать *z*-показатели с привязанной к ним условной квалификацией, например, с помощью пометки "предварительный (временный)" с соответствующим пояснением. Поэтому, от провайдеров проверки квалификации требуется назначать подходящее значение коэффициенту *l* в выражении $u^2(x_a) > l\sigma_p^2$, выше которого не должны рассчитываться *z*-показатели.

ИСО 13528 [5] предлагает модифицированный *z*-показатель, обозначаемый как *z*' и определяемый по формуле $z' = \dfrac{x - x_a}{\sqrt{u^2(x) + \sigma_p^2}}$, который может использоваться, если неопределенность приписанного значения не является пренебрежимо малой. Однако, *z*' не рекомендуется использовать в этом протоколе. Не смотря на то, что он, как правило, приводит к значениям похожим по рассеянию на сами *z*-показатели, использование *z*' будет маскировать факт, что неопределенность приписанного значения слишком высока. Поэтому будет уместна следующая рекомендация.

Рекомендация 2

Провайдер проверки квалификации должен назначить соответствующий программе множитель в интервале $0{,}1 < l < 0{,}5$, и оценив $u^2(x_a)/\sigma_p^2$ для тура, действовать следующим образом:

- если $u^2(x_a)/\sigma_p^2 \leq 0{,}1$, выдать абсолютные z-показатели;
- если $0{,}1 < u^2(x_a)/\sigma_p^2 \leq 1$, выдать условные z-показатели (такие как «предварительные (условные)» z-показатели); и
- если $u^2(x_a)/\sigma_p^2 > 1$ не выдавать z-показатели.

Примечание: В неравенстве $0{,}1 < l < 0{,}5$ границы могут быть несколько изменены с целью соответствия точным требованиям конкретных программ.

3.5 Определение стандартного отклонения для оценки квалификации

Стандартное отклонение для оценки квалификации σ_p – это параметр, который используется для обеспечения для масштабирования лабораторных отклонений $(x - x_a)$ от приписанного значения и тем самым определяет z-показатель (раздел 3.1). Существует несколько способов определения значения этого параметра, и их относительные преимущества рассматриваются ниже.

3.5.1 Значение, определяемое соответствием назначению

В этом методе провайдер проверки квалификации определяет уровень неопределенности, который является общепринятым в качестве подходящего для участников и конечных пользователей данных для сектора применения результатов, и определяет его в терминах σ_p. Под "подходящим", мы понимаем, что неопределенность является достаточно малой, чтобы решения основанные на данных оказывались очень редко некорректными, но не такой малой, чтобы расходы на анализ оказались чрезмерно высоки. Предлагаемое определение этого "соответствия назначению" заключается в том, что оно должно включать в себя неопределенность, которая минимизирует суммарные расходы на анализ и финансовых санкций, связанных с неправильными решениями, умноженными на их вероятности возникновения [10]. Следует подчеркнуть, что σ_p здесь не представляет основную идею о том, как лаборатории функционируют, а то как они должны функционировать, чтобы выполнить свои обязательства перед своими клиентами. Числовые значения параметра должны быть такими, чтобы результирующие z-показатели могли бы интерпретироваться, ссылаясь на стандартное нормальное распределение вероятностей. Вероятно они (значения параметра) будут определяться на основании профессионального суждения, осуществляемого консультативным комитетом программы проверки квалификации. В некоторых аналитических областях уже существуют готовые общепризнанные рабочие требования (стандарты) для соответствия назначению. Например, в области пищевых продуктов, функция Хорвитца часто рассматривается для определения соответствие назначению, она также является простой в описании [13].

Каким бы не было установлено значение параметра, оно должно быть оговорено и опубликовано перед началом распределения материалов для проверки квалификации, с тем чтобы участники могли проверить соответствует ли ему их аналитические процедуры. В некоторых программах возможный диапазон концентрации аналита является малым и один (общий) уровень неопределенности может быть определен, чтобы охватить все возможные случаи. Осложнение возникает в случаях, когда концентрация аналита может варьироваться в широких пределах. Так как приписанное значение неизвестно участникам заранее, критерий соответствия назначению должен быть определен с помощью функции зависимости от концентрации. Наиболее общие подходы являются следующими:

- Установите критерий как относительное стандартное отклонение (relative standard deviation, RSD). Конкретные значения σ_p затем получаются умножением этого RSD на приписанное значение.
- Там где существует интерес к нижнему пределу аналитических результатов, установите RSD применимо ко всему установленному диапазону в сочетании с (нижним) ограничивающим значением для σ_p. Например, при определении концентрации свинца в вине, было бы разумно стремиться к RSD, равному 20 %, в

отношении широкого диапазона концентрации аналита, но при концентрациях значительно ниже максимально допустимой концентрации x_{max} такой уровень прецизионности будет ни необходимым, ни экономически эффективным. Об этом факте можно узнать с помощью формулирования критерия в виде функции в форме $\sigma_p = x_{max}/m + 0{,}2x_a$, где m – это подходящая константа. Если m было установлено, например, равным 4, σ_p никогда не будет ниже, чем $x_{max}/4$.

- Установите основное выражение для соответствия назначению, такое как функция Хорвитца [13], а именно, (используя современное представление):

$$\sigma_p = 0{,}02\mathrm{x}_{\mathrm{a}}^{0{,}8496}$$

где x_a и σ_p выражаются массовыми долями. Обратите внимание, что исходное (оригинальное) отношение Хорвитца теряет применимость при концентрациях ниже, чем примерно 10 ppb (ppb = 10^9 массовых долей), и измененная форма функции была рекомендована [14].

3.5.2 Значение, определяемое законодательством

В некоторых случаях максимальное стандартное отклонение воспроизводимости для аналитических результатов для определенной цели устанавливается законодательством или международным соглашением. Это значение может быть использовано в качестве значения для σ_p. Аналогичным образом, если установлен предел допускаемой погрешности, это может быть использовано для установления σ_p посредством, например, деления на соответствующее значение коэффициента Стьюдента t, если уровень доверия также известен. Однако, может быть еще предпочтительным использовать значение σ_p ниже, чем законодательный предел. Это дело провайдера и консультативной группы программы проверки квалификации.

3.5.3 Другие подходы

Определение количественных показателей в некоторых программах проверки квалификации не основано на идее соответствия назначению, которая значительно преуменьшает значимость определения количественных показателей. Хотя некоторые методы определения количественных показателей охвачены ISO GUIDE 43 [4] (и обсуждаются в предыдущих версиях Гармонизированного протокола [1]), они не рекомендуются здесь для проверок квалификации в химии. Существует по сути две версии таких систем определения количественных показателей. В одной из них значение σ_p определяется экспертным восприятием (пониманием) (perception) функционирования лаборатории для рассматриваемого типа анализа. Очевидно, лаборатории могут функционировать лучше или хуже, чем соответствие назначению, так что система определения количественных показателей говорит нам только о том, какие лаборатории отличаются от других участников, а не о том, какие из них являются достаточно хорошими. Другая версия этого метода, кажущееся более авторитетной, так как она опирается на стандартные статистические идеи, заключается в использовании в качестве σ_p робастного стандартного отклонения результатов участников в туре. Последствия этой стратегии в том, что в каждом случае примерно 95 % участников получают очевидно приемлемые значения z-показателей. Это удобный результат как для участников, так и для провайдеров проверки квалификации, но, опять же, он предназначен только для определения результатов, которые отличаются от других. Кроме того существует дополнительная трудность, заключающаяся в том, что значение, используемое для σ_p, будет изменяться от тура к туру, так что не будет стабильной основы для сравнения показателей из нескольких туров. Хотя метод может быть улучшен с помощью фиксированного значения, полученного объединением результатов из нескольких туров, он все же не обеспечивает стимулом по улучшению своего функционирования лаборатории, производящие не соответствующие назначению результаты.

Могут быть обстоятельства, когда является оправданным для программы проверки квалификации не предоставлять руководства по соответствию назначению. Это может быть

случай, когда участники выполняют свою рутинную работу для множества различных целей, так что не может быть универсального применимого для соответствия назначения критерия. В таких условиях, может быть лучше для провайдера проверки квалификации не предоставлять определения количественных показателей в целом, а только представить приписанное значение (с его неопределенностью) и возможно погрешность лаборатории. (Иногда это обеспечивается терминами относительной погрешности, называемой «Q-показатель»). Некоторые предоставляемые количественные показатели должны быть понятно обозначены как «для неформального (неофициального) использования», чтобы уменьшить случаи, когда эксперты при аккредитации или предполагаемые клиенты составляют неверные суждения, основанные на показателях. Отдельные участники в таких программах затем должны предоставить свои личные критерии для соответствия назначению, и программа для того, как это выполнить, приводится ниже (смотри раздел 3.6 и Приложение 6).

Рекомендация 3

Везде, где возможно, программа проверки квалификации должна использовать для стандартного отклонения оценки квалификации σ_p значение, которое отражает соответствие назначению для сектора. Если не существует единого уровня, который подходит в большинстве случаев, провайдер должен воздержаться от расчета показателей или должен ясно показать в отчетах, что показатели предназначены только для неофициального описательного (наглядного) использования и не должны рассматриваться в качестве показателей функционирования участников.

3.5.4 Модифицированный z-показатель для индивидуальных требований

Некоторые программы проверки квалификации не оперируют положением «соответствия назначению». Провайдер проверки квалификации рассчитывает показатель исключительно из результатов участников (т. е. без внешних ссылок на действительные требования). И наоборот, участник может найти, что критерий соответствия назначению, используемый провайдером программы, является неподходящим для основных видов работ, выполняемых лабораторией. На самом деле это не может быть необычным для лаборатории, имеющей нескольких клиентов, желающих определить один и тот же аналит в одном и том же материале, но имеющих различные требования к неопределенности.

В программах проверки квалификации, выполняемых с учетом любого из этих положений, участники могут рассчитывать показатели, основанные на своих личных требованиях соответствия назначению. Это может быть достигнуто в непосредственном виде. Участник должен договориться с клиентом о критерии специфического соответствия назначения σ_{ffp} для каждого специфического применения, и использовать его для расчета соответствующего модифицированного z-показателя, определяемого как

$$z_L = (x - x_a)/\sigma_{ffp} \qquad (4)$$

для замены установленного z-показателя [15]. Критерий σ_{ffp} должен выражаться как функция концентрации, если необходимо. Он должен использоваться как значение для сигмы в z-показателе, то есть он должен быть в форме стандартной неопределенности, которая представляет согласованное соответствие назначению. Если имеется несколько клиентов с различными требованиями по точности, то должно быть несколько адекватных показателей, полученных для каждого единичного результата. Модифицированный z-показатель может быть интерпретирован точно так же, как это рекомендуется для z-показателей (смотри Приложение 6).

3.6 Данные участников, представленные с неопределенностью

Этот Протокол не рекомендует представление вместе с результатом неопределенности измерения участников. Такая рекомендация соответствует ISO GUIDE 43. В самом деле, относительно немногие программы проверки квалификации для аналитической химии в

настоящее время требуют снабжать результаты участников оценкой неопределенности. Это происходит главным образом из-за того, что предполагается, что после тщательного рассмотрения экспертом, программа, как правило, устанавливает значение σ_p, которое представляет соответствие назначению для всего сектора применения. Поэтому это оптимальное требование неопределенности подразумевается в программе. От участников ожидается, что они будут функционировать способом, согласованным с этой спецификацией и поэтому (в данном контексте) от них не требуется сообщать неопределенности в явном виде. Такое функционирование в соответствии с требованиями программы будет обычно приводить к значениям z-показателей в диапазоне ± 2. Те участники, которые значительно недооценили неопределенность, имеют гораздо больше шансов получить «неприемлемые» z-показатели. Другими словами, корректно оцененные неопределенности будут ожидаться главным образом похожими на значение σ_p и недооценки будут приводить к плохим z-показателям. В таких обстоятельствах представление неопределенности не добавляет ценности программе. Кроме того, программы проверки квалификации до сегодняшнего дня были достаточно эффективны без данных о неопределенности от участников; это означает, что данные по неопределенности от участников не требуются для улучшения повседневного аналитического функционирования.

Однако, обстоятельства, изложенные выше, не могут быть универсальными. Лаборатории, работающие со своими частными критериями соответствия назначению, следовательно, должны оцениваться с помощью индивидуальных критериев, а не по общему для программы значению σ_p. Кроме того, данные по неопределенности все чаще требуются клиентами лабораторий, и лаборатории должны соответственно проверять свои процедуры для выполнения этого требования (оценки неопределенности). Таким образом, следующие разделы обсуждают три важных момента, касающихся использования неопределенностей участников: определение согласованных значений, использование количественных показателей для проверки заявляемой неопределенности, и использование неопределенности участника при оценивании индивидуального соответствия назначению.

3.6.1 Согласованные значения

Когда имеются оценки неопределенности, провайдерам программ возможно придется подумать о том, как наилучшим образом определить согласованное значение, когда участники заявили данные о неопределенности, и о том, как наилучшим образом оценить неопределенность этого согласованного значения. Простой версией проблемы является установление консенсуса и его неопределенности из набора несмещенных оценок измеряемой величины с различными неопределенностями. Реальность такова, что: (a) часто имеются несогласующиеся результаты среди тех, которые заявляются (то есть данные включают выборки из распределений с различными средними); и (b) оценки неопределенности часто являются некорректными и, в частности, привязанные к выбросовым результам вероятно будут слишком малы.

В настоящее время, не существует хорошо разработанных методов для обеспечения робастных оценок среднего или дисперсии для межлабораторных данных с переменными неопределенностями. Эта область, однако, находится в активной разработке и несколько интересных предложений уже обсуждалось [16, 17]. Например, методы, основанные на оценке ядерной плотности [12], в настоящее время выглядят весьма продуктивно.

К счастью, большинство участников в определенной программе работают с похожими требованиями, и от них ожидается, что они будут обеспечивать похожие неопределенности. При таких обстоятельствах взвешенные и невзвешенные оценки среднего значения распределения будут очень схожи. Робастные невзвешенные оценки могут, следовательно, быть применены с преимуществом меньшей чувствительности к существенным недооценкам неопределенности или удаленным выбросам.

Учитывая, что эти вопросы находятся в стадии активной разработки, и что оценки неопределенности являются весьма ненадежными, рекомендуется использовать невзвешенные робастные методы для расчета приписанного значения.

Рекомендация 4

Даже если имеются оценки неопределенности, невзвешенные робастные методы (т. е. методы, не принимающие во внимание индивидуальные неопределенности) должны использоваться для получения согласованного значения и его неопределённости в соответствии с методами описанными в разделах 3.3. и 3.4.

3.6.2 Показатели дзета

Для результатов и заявленных неопределенностей, на основании которых вычисляются количественные показатели, ИСО 13528 определяет показатель «дзета» (ζ) как:

$$\zeta = (x - x_a) / \sqrt{u^2(x) + u^2(x_a)} \qquad (5)$$

где $u(x)$ – заявленная стандартная неопределенность представленного результата x и
$u(x_a)$ – стандартная неопределенность приписанного значения.

Показатель «дзета» позволяет судить о том, согласуется ли оценка неопределённости участника с наблюдаемым отклонением от приписанного значения. Интерпретация похожа на интерпретацию z-показателей; абсолютные значения больше 3 должны рассматриваться в качестве причины для дальнейшего исследования. Причиной может быть недооценка неопределенности $u(x)$, но может быть также большая погрешность, вызванная слишком большим отклонением $(x - x_a)$. Последнее условие обычно приводит к высоким z-показателям, поэтому важно рассмотреть z-показатели и показатели «дзета» вместе. Отметим также, что устойчиво малые показатели «дзета» в течение определенного периода времени могут указывать на завышенную оценку неопределенности.

Примечание: ИСО 13528 определяет дополнительные методы расчета количественных показателей, которые используют расширенные неопределённости; рекомендуется ссылка на ИСО 13528, если это признано целесообразным консультативной группой программы.

3.6.3 Определение количественных показателей для результатов, которые представлены с неопределенностью

Достаточно легко для участников использовать показатели «дзета» для проверки своих личных оценок неопределенности. Тем не менее, сейчас мы рассмотрим действия, предпринимаемые организаторами программ проверки квалификации.

Спорным вопросом является то, должен ли стремиться провайдер программы (а не отдельные участники) принимать во внимание неопределенность при расчете показателей по необработанным результатам. Не существует никакой трудности сделать это – единственным вопросом является лишь то, будет ли результат полезным. Баланс выгод противопоставляется расчету показателей «дзета» в программе. Все программы рекомендуется снабжать диаграммами, показывающими результаты участников и количественные показатели. Такие диаграммы, основанные на показателях будут неоднозначны, так как результаты могут быть не приемлемы для представления на двухмерных графиках. Конкретное значение показателя «дзета» (скажем, -3,7) может быть результатом как слишком большой погрешности и большой неопределенности, так и малой погрешности и соразмерно малой неопределенности. Кроме того, организатор программы не имеет возможности судить о том, является ли представленное участником значение неопределенности соответствующим его потребностям, так что показатели «дзета», полученные таким способом, не будут иметь смысла при оценивании результатов участников.

Рекомендация 5

Программы не предоставляют показатели «дзета» до тех пор, пока не будет специальных причин для этого. Если участник имеет требования, не совпадающие с теми, которые представлены в программе, он может рассчитать показатели «дзета» или эквивалентные им показатели.

3.7 Определение количественных показателей в области предела обнаружения

Многие аналитические задачи включают измерение концентраций аналита, которые близки к пределу обнаружения метода, или даже почти к нулю. Проверка квалификации для таких методов в идеале должна имитировать истинную жизненную картину; материалы для испытаний должны содержать типично низкие концентрации аналита. Однако существуют трудности в использовании привычного метода определения z-показателей к результатам таких испытаний. Эти трудности частично вызваны следующими способами записи данных:

- Многие практикующие аналитики, получившие низкий результат будут записывать результат «по умолчанию», как «не определен» или «ниже чем c_L», где c_L является произвольно определенным пределом. Такие результаты, хотя возможно и пригодные для соответствия назначению, не могут быть преобразованы в z-показатели. z-Показатели требуют, чтобы аналитический результат был представлен с помощью шкалы интервалов или отношений. Не рекомендуется замена результата «по умолчанию» произвольно выбранным способом (т. е. нулем, или половиной интервала обнаружения).
- Некоторые программы проверки квалификации позволяют избежать трудностей, но не с помощью обработки результатов по умолчанию. Если многие участники работают близко к своим пределам обнаружения, вне зависимости от того, представляют ли они результат по умолчанию или нет, это создает трудность в оценивании достоверного согласованного значения для приписанного значения. Распределение предположительно достоверных результатов, как правило, имеет строго положительную асимметрию, и большинство видов средних значений, как правило, имеют большое смещение.

Эти трудности можно обойти, работая на чуть более высоких концентрациях, чем обычно можно найти в представляющих интерес материалах. Эта практика не является полностью удовлетворительной, так как образцы тогда являются нереалистичными. Если участники записали фактический (действительный) полученный результат плюс (корректная) неопределенность результата, то в принципе оказывается возможным оценить достоверное согласованное значение с неопределенностью. Хотя это рекомендуется там, где это возможно, другие факторы, такие как установившаяся практика требований, по которым представляется отчет для заказчика, превращает это в маловероятный сценарий для рутинного (повседневного) анализа. Таким образом, представляется, что z-показатели могут быть использованы при низких концентрациях, только когда применимо следующее:

- Участники записывают фактические полученные результаты.
- Приписанное значение является независимым от результатов. Этого можно достичь, если известно, что приписанное значение равно нулю или очень маленькому значению или его можно определить с помощью процедуры приготовления или в эталонной (референтной) лаборатории (см. раздел 3.2).
- Стандартное отклонение для оценки квалификации является независимым соответствующим назначению критерием; его значение заранее устанавливается, то есть независимо от результатов участников. Это было бы относительно просто для достижения эффекта.

В настоящее время в проверках квалификации не существует альтернативной хорошо разработанной системы определения количественных показателей для результатов

измерения низких концентраций, и тема по-прежнему очень много обсуждается. Если от результатов требуется, что они по существу должны быть биномиальными ($\leq x$ или $> x$), то система определения количественных показателей может быть разработана на основании доли правильных результатов (ответов), но она будет ограничена по мощности (т. е. менее информативной), чем система оценки с помощью z-показателей. Нельзя предусмотреть смешанную систему (способную иметь дело со смесью биноминальных, порядковых и количественных результатов).

3.8 Осторожность при использовании z-показателей

Соответствующее использование z-показателей участниками и конечными пользователями обсуждается более подробно в Приложениях 6 и 7. Следующие слова предостережения адресованы провайдерам.

Распространенной практикой является требование по выполнению нескольких различных анализов в рамках каждого тура проверки квалификации. Хотя каждое отдельное испытание предоставляет полезную информацию, заманчиво определить единственный показатель качества, который будет подводить итог для общего функционирования лаборатории в рамках тура. Это опасно, так как такой комбинированный показатель будет неправильно интерпретироваться или употребляться не экспертами, особенно вне контекста отдельных показателей. Таким образом, общее предоставление комбинированных показателей в отчетах для участников не рекомендуется, но следует признать, что такие показатели могут иметь специфические применения, если они основаны на известных статистических принципах и публикуются с надлежащим предостерегающим уведомлением. Используемые для этого процедуры описаны в Приложении 4.

Следует особо подчеркнуть, что существуют ограничения и недостатки в любой программе, которая комбинирует z-показатели из разнородных анализов. Если единичный показатель из нескольких полученных лабораторией является выбросовым, комбинированный показатель вполне может быть не выбросовым. В некотором отношении, это полезное свойство, заключающаяся в том, что ошибка в одном анализе взвешивается с понижением в комбинированном показателе. Однако, существует опасность в том, что лаборатория может последовательно допускать ошибку только в определенном анализе и часто представлять неприемлемое значение для этого анализа в последовательных турах проверки. Этот фактор может быть скрыт с помощью комбинирования показателей.

3.9 Классификация, ранжирование и другие способы оценивания данных квалификации

Классификация лабораторий не является целью проверки квалификации, и провайдерам лучше ее избегать, так как она скорее приведет к путанице, чем к разъяснению. Замена непрерывной меры, такой как z-показатель несколькими классами с присвоенными названиями имеет мало преимуществ с научной точки зрения: информация выбрасывается. Поэтому классификация не рекомендуется в проверках квалификации. Пределы для принятия решений, основанные на z-показателях, могут быть использованы при необходимости в качестве рекомендаций. Например, z-показатель вне диапазона ± 3 можно рассматривать в качестве требования для выполнения исследований, если необходимо изменить процедуры. Тем не менее, такие пределы являются произвольными. Показатель со значением 2,9 должна всегда внушать беспокойство, как и показатель со значением 3,1. Кроме того, эти вопросы скорее касаются отдельных участников, чем провайдеров программ.

Ранжирование лабораторий по их абсолютным значениям z-показателей, полученным в туре проверки, с целью формирования «таблицы лиги» является более оскорбительным, чем классификация. Ранги участников гораздо сильнее изменяются от тура к туру, чем показатели, на основании которых они были выведены, и лаборатория с малым абсолютным значением показателя в туре вряд ли окажется «наилучшей».

Рекомендация 6

Провайдеры проверки квалификации, участники и конечные пользователи должны избегать классификации и ранжирования лабораторий на основании их z-показателей.

3.10 Частота проведения туров

Соответствующая частота распределения является балансом между факторами, наиболее важными из которых являются следующие:

- трудность выполнения эффективного аналитического контроля качества;
- лабораторная пропускная способность (производительность) образцов для испытаний;
- согласованность (совместимость) результатов в конкретной области деятельности, охваченной программой;
- затраты/выгоды программы;
- наличие сертифицированных стандартных образцов (CRM) в секторе аналитики; и
- скорость изменения аналитических требований, методологии, оборудования и персонала в интересующем секторе.

Редко обнаруживается объективное свидетельство влияния частоты тура на эффективность проверки квалификации. Только одно надежное исследование частоты было представлено в [18], и оно показало, что (в конкретной программе) изменение частоты туров от трех до шести раз в год не имеет значительного влияния (полоижительного или другого) на характеристики функционирования участников.

На практике частота будет попадать где-то в интервал от одного раза в каждые две недели до одного раза в каждые четыре месяца. Частота, больше чем один раз в каждые две недели, может привести к проблемам во времени обращения образцов и результатов. Это может также способствовать уверенности в том, что программа проверки квалификации может быть использована в качестве замены для внутреннего контроля качества (IQC), идея, которая настоятельно не одобряется. Если период между распределениями (образцов) составляет гораздо больше четырех месяцев, то будут возникать неприемлемые задержки в идентификации и корректировке аналитических проблем, и влияние программы на участников может быть незначительным. Для повседневной (рутинной) аналитической работы принятие участия в проверках квалификации мене двух раз в год будет иметь очень малую практическую ценность.

3.11 Испытания для определения достаточной однородности и стабильности

3.11.1 Испытания для «достаточной однородности»

Несмотря на прилагаемые усилия для обеспечения однородности, материалы, подготовленные для проверок квалификации и других межлабораторных исследований, обычно являются в некоторой степени неоднородными. Если такой материал в большом объеме (целиком, в партии) разделяется для распределения различным лабораториям, то получаемые образцы будут незначительно отличаться по составу друг от друга. Этот протокол требует, чтобы эта вариация была достаточно малой для данной цели. Рекомендуемая процедура проверки однородности представлена в Приложении 1. Обоснование для этой процедуры обсуждается в следующих абзацах.

Если мы проверяем так называемую «достаточную однородность» в таких материалах, то мы стремимся показать, что вариация в составе среди распределяемых образцов (характеризуемая стандартным отклонением выборки σ_{sam}) является пренебрежимо малой по сравнению с вариацией, представленной измерениями, выполняемыми участниками в проверке квалификации. Как мы ожидаем, стандартное отклонение межлабораторной вариации в проверках квалификации будет приближаться к,

стандартному отклонению для оценки квалификации σ_p, и естественно использовать этот критерий в качестве эталонного значения. Гармонизированный протокол версии 1993 года [1] требовал, чтобы оцененное стандартное отклонение выборки s_{sam} было менее 30 % от целевого стандартного отклонения σ_p, то есть

$$s_{sam} < \sigma_{all}$$

где допускаемое стандартное отклонение выборки $\sigma_{all} = 0{,}3\sigma_p$.

Это условие, если оно выполняется, было названо «достаточная однородность». При таком пределе стандартное отклонение получающегося в результате z-показателя может быть завышено за счет неоднородности не более, чем на 5 % в относительной форме, например, от 2,0 до 2,1, что посчиталось приемлемым. Если условие не выполняется, z-показатели отражали бы, до неприемлемого уровня, вариацию в материале, а также вариацию в показателях функционирования лабораторий. Участников программ проверок квалификации следует убедить, что распределяемые образцы материала для испытаний являются достаточно подобными, и эти требования обычно нуждаются в проверке.

Установленная проверка на однородность требует выполнить случайный отбор десяти или более образцов после того, как предположительно однородный материал был разделен и упакован в качестве отдельных образцов для распределения. Материал от каждого образца затем анализируется с двумя повторами в условиях случайной (рандомизированной) повторяемости (т.е. все в одной серии), используя метод с достаточной аналитической прецизионностью. Значение σ_{sam} затем оценивается из средних квадратов после однофакторного дисперсионного анализа (ANOVA).

Испытания для достаточной однородности на практике никогда полностью не являются удовлетворительными (достаточными). Основная проблема состоит в высокой стоимости анализа и малом количестве образцов, взятых для испытания. Это делает мощность статистических критериев (т.е. вероятность отклонения материала, когда он на самом деле является неоднородным) очень низкой. Дальнейшая проблема состоит в том, что неоднородность по существу является скорее всего фрагментарной, и неоднородные распределенные образцы могут быть не достаточно представлены среди тех, которые выбраны для проверки на однородность. Испытания на однородность следует рассматривать как важную, но не надежную меру предосторожности.

Однако, учитывая, что достаточная однородность является предварительным обоснованным предположением (потому что провайдеры проверки квалификации делают все возможное для его обеспечения), и стоимость испытаний на однородность часто является высокой, целесообразно сделать основной упор на то, чтобы избежать «ошибки 1 рода» (т. е., ошибочный отказ от удовлетворительного материала). Такое решение будет иметь эффект при создании модифицированного критерия, который будет менее чувствителен к отклонению хороших образцов.

Чтобы проверить достаточную однородность, мы должны оценить σ_{sam} по результатам случайно повторяемого эксперимента используя методы дисперсионного анализа (ANOVA). В эксперименте, каждый выбираемый для распределения образец отдельно гомогенизируется и анализируется дважды. Многое зависит от качества аналитических результатов. Если аналитический метод является достаточно прецизионным, то σ_{sam} может быть оценена достоверно, и любое отсутствие в данный момент достаточной однородности обнаруживается с достаточно высокой вероятностью. В действительности, процедура проверки может быть слишком чувствительной. Материал может быть достаточно неоднородным статистически, но дисперсия выборки будет незначительной по отношению к σ_p. Тем не менее, если аналитическое стандартное отклонение σ_{an} мало, важная вариация выборки может быть скрыта аналитической вариацией. Мы можем получить незначимый результат при проверке на неоднородность, но не потому, что это так в данный момент, а потому, что процедура испытаний не имеет мощности для ее обнаружения.

Гармонизированный протокол 1993 года не смотря на то, что он определял необходимость для достаточно хорошей аналитической прецизионности, он не определял никакие числовые пределы для σ_{an}, но приведенное выше обсуждение настоятельно рекомендует это делать. При определении этого значения, должен быть компромисс между стоимостью задания очень точных аналитических методов и риском невозможности определения важной вариации выборки. Таким образом мы рекомендуем, чтобы аналитическая (повторяемость) прецизионность метода, используемого при испытаниях на однородность должна удовлетворять условию

$\sigma_{an}/\ \sigma_p < 0{,}5$

Однако, мы должны признать, что иногда оказывается невозможным выполнить это требование, поэтому нужна статистическая процедура, которая приводит к разумному результату вне зависимости от значения σ_{an}.

Рекомендация 7

Аналитическая (повторяемость) прецизионность метода, используемого для испытаний на однородность, должна удовлетворять условию $\sigma_{an}/\sigma_p < 0{,}5$, где σ_{an} является стандартным отклонением повторяемости, соответствующим методу испытаний, используемому при проверке на однородность.

3.11.2 Новая статистическая процедура

Вместо того, чтобы выразить критерий для достаточной однородности в терминах оцененной дисперсии выборки s^2_{sam}, как в Гармонизированном Протоколе 1993 года, более логичным представляется наложить ограничения на истинную дисперсию выборки σ^2_{sam} [19]. Именно она является истинной дисперсией выборки, которая в большей степени относится к изменчивости в (неиспытанных) образцах, посылаемых в лаборатории. Поэтому, наш новый критерий для достаточной однородности заключается в том, что дисперсия выборки σ^2_{sam} не должна превышать допустимую величину $\sigma^2_{all} = 0{,}09\sigma^2_p$ (то есть $\sigma_{all} = 0{,}3\sigma_p$). Затем при проверке на однородность имеет смысл проверить гипотезу $\sigma^2_{sam} \le \sigma^2_{all}$ против альтернативной гипотезы. (Обычный *F*-критерий в однофакторном дисперсионном анализе (ANOVA) проверяет довольно строгую гипотезу $\sigma^2_{sam} = 0$ против альтернативной гипотезы $\sigma^2_{sam} < 0$, которая будет обеспечивать доказательство наличия дисперсии выборки, но при этом она не обязательно будет иметь недопустимо большое значение). Новая процедура разработана для обеспечения этого требования и урегулирования других упомянутых выше трудностей. Полная процедура и рабочий пример приведены в Приложении 1.

Рекомендация 8

Примените определенную проверку гипотезы H: $\sigma^2_{sam} \le \sigma^2_{all}$, с помощью нахождения одностороннего 95 %-ого доверительного интервала для σ^2_{sam} и отклоните гипотезу H, если этот интервал не включает σ^2_{sam}. Это эквивалентно отклонению H, если

$$s^2_{sam} > F_1\sigma^2_{all} + F_2 s^2_{an}$$

где s^2_{sam} и s^2_{an} - это обычные оценки дисперсии выборки и аналитической дисперсии, полученные из дисперсионного анализа (ANOVA) и

F_1 и F_2 – это константы, которые могут быть получены из стандартных статистических таблиц.

3.11.3 Обращение с выбросовыми (сильно отклоняющимися) результатами при проверке однородность

Единичные аналитические выбросы довольно часто влияют на наборы данных, полученных при проведении проверки на однородность, так как по меньшей мере 20 аналитических результатов производится при каждом испытании. Аналитические выбросы проявляются как непредвиденно большое отклонение между двумя повторными результатами, полученными на одном из образцов. Так как процедура требует, чтобы каждый образец должным образом гомогенизировался перед тем как будут отобраны из него две порции для испытаний, то вне зависимости от неоднородности или других причин, связанных с исходным не расфасованным материалом, любые выбросовые разницы между повторными парами должны возникать только из-за процедуры анализа, а не материала.

Влияние единичного (то есть аналитического) выбросового результата будет возможно неожиданным: Хотя он раздувает оценку дисперсии между образцами, выброс помогает материалу пройти проверку с помощью *F*-критерия, так как еще в большей степени он раздувает оценку аналитической дисперсии. Чем более экстремальным является аналитический выброс, тем ближе *F*-значение становится к единице (другие результаты остаются равными). Таким образом, хотя Гармонизированный Протокол 1993 года требовал оставлять все результаты, является очевидным исключение экстремальных аналитических выбросов, когда они могут быть однозначно идентифицированы. Если, однако, набор данных очевидно содержит более, чем одну пару несогласованных аналитических результатов, достоверность всего мероприятия попадает под сомнение и данные, полученные при проверке на однородность, должны быть забракованы.

Следует отметить, что рекомендация исключить единичную выбросовую пару относится только к образцам с отдельными выбросовыми результатами, но не к образцам с взаимно согласующимися результатами, но выбросовыми средними значениями. Если результаты, полученные на образце, согласуются друг с другом, но средний результат не согласуется с другими данными, результаты должны быть оставлены – они содержат доказательство неоднородности между образцами. Это отличие проиллюстрировано на рисунке 1. Образец 9 обеспечивает несогласующийся результат, который должен быть исключен. Образец 12 обеспечивает согласующиеся результаты с выбросовым средним, и не должен быть исключен. Проверка дисперсий с помощью критерия Кохрена является подходящей для выявления экстремальных разностей между наблюдениями (Приложение 1).

Примечание 1: Автоматическое исключение выбросов дисперсии при 95 %-ом уровне доверия будет существенно увеличивать долю неверных (неправильных) неблагоприятных исходов (т. е. на самом деле они благоприятные, но мы их считаем неблагоприятными и поэтому они неверные, некорректные - примечание переводчика) проверки на однородность и не рекомендуется.

Примечание 2: В некоторых редких случаях рекомендация отказаться от единичной пары аналитических выбросов может быть неуместной. Это исключение может возникнуть, когда аналит представлен повсеместно в контрольном материале при низких концентрациях, но почти полностью ограничен фазой следов (остаточных количеств), которая не поддается измельчению, но содержит аналит в высоких концентрациях. Примером является порода, содержащая золото. Решение о том, следует ли исключать выбросы, принимается консультативным комитетом провайдера, принимая во внимание вышеизложенное обсуждение.

Рекомендация 9

При проверке на достаточную однородность, повторные результаты для единичного распределяемого образца должны быть удалены до начала выполнения дисперсионного анализа, если они показали значительную разницу между собой с помощью проверки по критерию Кохрена при уровне доверия 99 % или эквивалентной проверке для экстремальной внутригрупповой дисперсии. Наборы данных, содержащие несовместимости

в двух таких распределяемых образцах должны быть полностью отброшены. Пары результатов с выбросовыми средними значениями, но не содержащими доказательств экстремальной дисперсии, не должны исключаться.

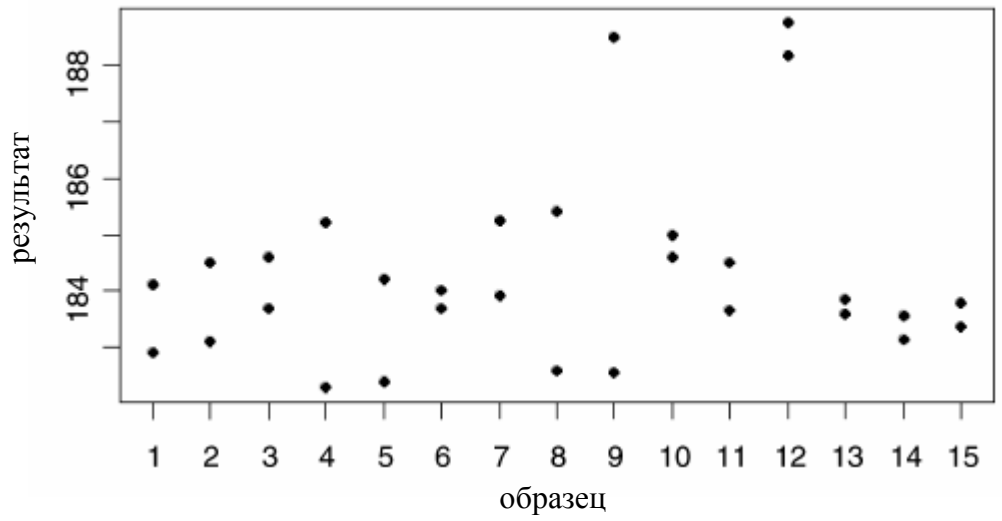


Рисунок 1 – Данные проверки на однородность

3.11.4 Другие отклонения в наборах данных при проверке на однородность

Все аспекты проверки на однородность зависят от лаборатории, выполняющей испытание корректно и, в частности, отбирающей случайным образом образцы для испытания, гомогенизирующей их перед началом анализа, анализирующей повторные пробы в строго рандомизированных условиях и регистрируя результаты с достаточным количеством разрядов для предоставления возможности выполнения дисперсионного анализа. Любые нарушения могут сделать результат проверки недостоверным. Если жесткий контроль не поддерживается, то, как правило, встречаются (возникают) наборы данных, где, по крайней мере, некоторые из этих требований не были выполнены. Поэтому мы рекомендуем, чтобы (a) выдавались детальные инструкции для лаборатории, выполняющей испытания на однородность, и (b) данные регулярно (постоянно) проверялись на подозрительные свойства. Предложения по этим инструкциям и испытаниям приводятся в Приложении 1.

Рекомендация 10

a. Лаборатории, выполняющей проверку на однородность, должны выдаваться детальные инструкции.

b. Результирующие данные должны проверяться на подозрительные свойства.

3.11.5 Стабильность материалов для испытаний

Материалы, распределяемые в проверках квалификации, должны быть достаточно стабильны во время всего периода времени, на протяжении которого приписанное значение является достоверным. Термин «достаточно стабилен» включает в себя то, что любые изменения, которые возникают во время соответствующего периода времени должны быть несущественными по величине по отношению к интерпретации результатов тура. Поэтому, если было признано, что изменение в z-показателе в пределах ±1 будет несущественным, то нестабильность, равная изменению в концентрации аналита $0{,}1\sigma_p$ может быть признана допускаемой. Обычно, рассматриваемый период является интервалом между подготовкой материала и крайним сроком для возвращения результатов, хотя период может быть дольше, если остаточный материал должен быть повторно использован в последующих турах или для других целей. При испытании на стабильность материал должен

подвергаться воздействию наиболее экстремальных условий, похожих на те, с которыми сталкивается материал при распределении и хранении или должен быть проверен с помощью ускорения состояния неисправности. Испытываемый материал должен находиться в упаковке, в которой он должен распределяться.

Полное испытание на достаточную стабильность может потребовать чрезвычайно много ресурсов (смотри ниже). Поэтому обычно не практикуется испытывать каждую партию материала для каждого тура в серии. Однако, благоразумной предварительной предосторожностью будет являться проведение испытания каждой новой комбинации материала и аналита перед первым использованием в проверке квалификации, а иногда и после него, и в следующих абзацах это обсуждается. Дополнительно может быть полезно наблюдать (выполнять мониторинг) за стабильностью с помощью, например, организации анализа образцов в отдельной лаборатории до и после распределения, предусматривая возврат нескольких распределенных образцов для прямого сравнения с хранящимися образцами, или сравнивая результаты анализа после распределения с предварительной информацией, такой как данные по испытаниям на однородность.

Основные испытания на стабильность включают сличение очевидных уровней аналита между материалом, подвергаемым возможным условиям разложения, и материалом, который не был подвержен таким условиям. Обычно это требует наличия выборки распределяемых образцов, которые случайным образом разделены на (как минимум) два равных набора. «Экспериментальный» набор подвергается соответствующей обработке, в то время как «контрольный» набор остается в условиях максимальной стабильности, например, при низких температурах и низком кислородном потенциале (содержание кислорода вокруг образцов сводится к минимуму для предотвращения разложения – примечание переводчика). Альтернативно, и особенно если интересует стабильность в течение длительного периода времени, контрольный набор может содержаться в условиях окружающей среды, в то время как экспериментальный набор содержится в условиях ускоренного разложения (т. е. при высоких температурах). Затем материалы анализируются одновременно или, если это невозможно, как план с рандомизированными блоками.

Такие эксперименты должны быть тщательно продуманы во избежание смешения эффектов из-за изменения в материале с вариацией, вызванной эффективностью используемого аналитического метода. Анализ контрольного материала в начале периода испытания и экспериментального материала в конце автоматически включает любые возникающие между сериями аналитические различия в результатах, и могут повлиять на неверное заключение о существовании значительной нестабильности. Рекомендуемый подход состоит в том, чтобы, насколько это возможно, анализировать экспериментальный и контрольный наборы вместе, в случайном порядке внутри единичной серии анализа, то есть в условиях повторяемости. Любая значимая разница между средними значениями результатов двух наборов может затем смело рассматриваться как доказательство нестабильности.

Как при испытаниях на однородность должно быть сделано концептуальное различие между статистически значимой нестабильностью и логической нестабильностью. Например, в аналитических результатах может быть обнаружено значительное изменение, но это изменение может оказаться слишком малым для оказания значительного влияния на z-показатели участников. На практике, критерии значимости не являются достаточно мощными для подтверждения такой малой нестабильности, если только не используется исключительно точный аналитический метод и/или не анализируется чрезмерное количество распределенных образцов. Следовательно, испытание на стабильность будет обнаруживать только серьезную (очевидную, большую) нестабильность.

Итоговые рекомендации

Рекомендация 1: Схема для получения согласованного значения и его неопределенности (смотри Раздел 3.3.2)

a. Исключите из данных любые результаты, которые идентифицируются как недействительные (ошибочные) (например, если они выражены в ошибочных единицах или получены с использованием несоответствующего метода) или как экстремальные выбросы (например, они находятся вне диапазона ± 50 % от медианы).

b. Исследуйте визуальное представление оставшихся результатов с помощью точечного графика (для малых (n < 50) набора данных), столбиковой диаграммы или гистограммы (для больших наборов данных). Если выбросы являются причиной чрезмерного сжатия основной доли результатов, постройте новый график с удаленными выбросами. Если распределение, без учета выбросов, очевидно унимодальное и приблизительно симметричное, то переходите к выполнению пункта (c), в других случаях – к пункту (d).

c. Рассчитайте робастное среднее x^*_{rob} и стандартное отклонение s^*_{rob} для n результатов. Если s^*_{rob} примерно меньше, чем $1{,}2\sigma_p$, то используйте x^*_{rob} в качестве приписанного значения x_a и $s^*_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. Если $s^*_{rob} > 1{,}2\sigma_p$, то переходите к выполнению пункта (d).

d. Сделайте оценку ядерной плотности (kernel density) распределения по результатам, используя нормальные ядра с шириной h равной $0{,}75\sigma_p$. Если эти результаты имеют унимодальную или приблизительно симметричную ядерную плотность, и мода и медиана практически совпадают, то используйте x^*_{rob} в качестве приписанного значения x_a и $s^*_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. В других случаях переходите к выполнению пункта (e).

e. Если второстепенные моды могут быть с уверенностью приписаны выбросовым результатам, и их вклад составляет менее примерно 5 % в общую площадь, то еще используйте x^*_{rob} в качестве приписанного значения x_a и $s^*_{rob}/\sqrt{n}$ в качестве его стандартной неопределенности. В других случаях переходите к выполнению пункта (f).

f. Если второстепенные моды вносят значительные вклады в площадь ядра, рассмотрите возможность того, что две или более несходных совокупности представлены в результатах участников. Если возможно, то сделайте вывод на основании независимой информации (например, детали аналитических методов участников) о том, что одна из этих мод является корректной, а другие – некорректны, используйте выбранную моду в качестве приписанного значения x_a и ее среднеквадратическую ошибку в качестве ее стандартной неопределенности. В других случаях переходите к выполнению пункта (g).

g. Если методы, приведенные выше, оказались неудачными, то оставьте попытку определить согласованное значение и не представляйте в отчетах по туру индивидуальные показатели функционирования для лабораторий. Однако может быть полезным предоставить участникам итоговые статистики, полученные на основании набора данных в целом.

Рекомендация 2: Использование неопределенности приписанного значения (смотри Раздел 3.4)

Провайдер проверки квалификации должен назначить соответствующий программе множитель в интервале $0{,}1 < l < 0{,}5$, и оценив $u^2(x_a)/\sigma_p^2$ для тура, действовать следующим образом:

- если $u^2(x_a)/\sigma_p^2 \le 0{,}1$, выдать абсолютные z-показатели;
- если $0{,}1 < u^2(x_a)/\sigma_p^2 \le 1$, выдать условные z-показатели (такие как «предварительные (условные)» z-показатели); и

• если $u^2(x_a)/\sigma_p^2 > l$ не выдавать z-показатели.

Примечание: В неравенстве $0{,}1 < l < 0{,}5$ границы могут быть несколько изменены с целью соответствия точным требованиям конкретных программ.

Рекомендация 3: Определение стандартного отклонения для оценки квалификации (смотри Раздел 3.5)

Везде, где возможно, программа проверки квалификации должна использовать для стандартного отклонения оценки квалификации σ_p значение, которое отражает соответствие назначению для сектора. Если не существует единого уровня, который подходит в большинстве случаев, провайдер должен воздержаться от расчета показателей или должен ясно показать в отчетах, что показатели предназначены только для неофициального описательного (наглядного) использования и не должны рассматриваться в качестве показателей функционирования участников.

Рекомендация 4: Использование взвешивания при расчете согласованных значений (смотри Раздел 3.6)

Если даже имеются оценки неопределенности, то невзвешенные методы (т.е. принимающие в расчет индивидуальные неопределенности) должны быть использованы для получения согласованного значения и его неопределенности, в соответствии с методами, представленными в разделах 3.3 и 3.4.

Рекомендация 5: Определение количественных показателей для результатов, которые представлены с неопределенностью (смотри Раздел 3.6.2)

Программы не предоставляют показатели «дзета» до тех пор, пока не будет специальных причин для этого. Если участник имеет требования, не совпадающие с теми, которые представлены в программе, он может рассчитать показатели «дзета» или эквивалентные им показатели.

Рекомендация 6: Классификация и ранжирование лабораторий (смотри Раздел 3.9)

Провайдеры проверки квалификации, участники и конечные пользователи должны избегать классификации и ранжирования лабораторий на основании их z-показателей.

Рекомендация 7: Требования к повторяемости при испытаниях на однородность (смотри раздел 3.11.1)

Аналитическая (повторяемость) прецизионность метода, используемого для испытаний на однородность, должна удовлетворять условию $\sigma_{an}/\sigma_p < 0{,}5$, где σ_{an} является стандартным отклонением повторяемости, соответствующим методу испытаний, используемому при проверке на однородность.

Рекомендация 8: Статистическая проверка при испытаниях на однородность (смотри Раздел 3.11.2)

Примените определенную проверку гипотезы H: $\sigma_{sam}^2 \le \sigma_{all}^2$, с помощью нахождения одностороннего 95 %-ого доверительного интервала для σ_{sam}^2 и отклоните гипотезу H, если этот интервал не включает σ_{sam}^2. Это эквивалентно отклонению H, если

$$s_{sam}^2 > F_1\sigma_{all}^2 + F_2 s_{an}^2$$

где s_{sam}^2 и s_{an}^2 - это обычные оценки дисперсии выборки и аналитической дисперсии, полученные из дисперсионного анализа (ANOVA) и

F_1 и F_2 – это константы, которые могут быть получены из стандартных статистических таблиц.

Рекомендация 9: Обращение с выбросами в проверках на однородность (смотри раздел 3.11.3)

При проверке на достаточную однородность, повторные результаты для единичного распределяемого образца должны быть удалены до начала выполнения дисперсионного анализа, если они показали значительную разницу между собой с помощью проверки по критерию Кохрена при уровне доверия 99 % или эквивалентной проверке для экстремальной внутригрупповой дисперсии. Наборы данных, содержащие несовместимости в двух таких распределяемых образцах должны быть полностью отброшены. Пары результатов с выбросовыми средними значениями, но не содержащими доказательств экстремальной дисперсии, не должны исключаться.

Рекомендация 10: Управление испытаниями на однородность (смотри Раздел 3.11.4)

a. Лаборатории, выполняющей проверку на однородность, должны выдаваться детальные инструкции.

b. Результирующие данные должны проверяться на подозрительные свойства.

Библиография

1. M. Thompson and R. Wood. “The International Harmonised Protocol for the proficiency testing of (chemical) analytical laboratories”, *Pure Appl. Chem*. **65**, 2123–2144 (1993). [Also published in *J. AOAC Int*. **76**, 926–940 (1993)].
2. See: (a) M. Golze. “Information system and qualifying criteria for proficiency testing schemes”, *Accred. Qual. Assur.* **6**, 199–202 (2001); (b) J. M. F. Nogueira, C. A. Nieto-de-Castro, L. Cortez. “EPTIS: The new European database of proficiency testing schemes for analytical laboratories”, *J. Trends Anal. Chem*. **20**, 457–61 (2001); (c) <http://www.eptis.bam.de>.
3. R. E. Lawn, M. Thompson, R. F. Walker. *Proficiency Testing in Analytical Chemistry*, The Royal Society of Chemistry, Cambridge (1997).
4. International Organization for Standardization. *ISO Guide 43: Proficiency testing by interlaboratory comparisons—Part 1: Development and operation of proficiency testing schemes*, Geneva, Switzerland (1994).
5. International Organization for Standardization. *ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons*, Geneva, Switzerland (2005).
6. ILAC-G13:2000. *Guidelines for the requirements for the competence of providers of proficiency testing schemes*. Available online at <http://www.ilac.org/>.
7. M. Thompson, S. L. R. Ellison, R. Wood. “Harmonized guidelines for single laboratory validation of methods of analysis”, *Pure Appl. Chem.* **74**, 835–855 (2002).
8. International Organization for Standardization. *ISO Guide 33:2000, Uses of Certified Reference Materials*, Geneva, Switzerland (2000).
9. M. Thompson and R. Wood. “Harmonised guidelines for internal quality control in analytical chemistry laboratories”, *Pure Appl. Chem.* **67**, 649–666 (1995).
10. T. Fearn, S. Fisher, M. Thompson, S. L. R. Ellison. “A decision theory approach to fitness-forpurpose in analytical measurement”, *Analyst* **127**, 818–824 (2002).
11. Analytical Methods Committee. “Robust statistics—how not to reject outliers: Part 1 Basic concepts”, *Analyst* **114**, 1693 (1989).
12. M. Thompson. “Bump-hunting for the proficiency tester: Searching for multimodality”, *Analyst* **127**,1359–1364 (2002).
13. M. Thompson. “A natural history of analytical methods”, *Analyst* **124**, 991 (1999).
14. M. Thompson. “Recent trends in interlaboratory precision at ppb and sub-ppb concentrations in relation to fitness-for-purpose criteria in proficiency testing”, *Analyst* **125**, 385–386 (2000).
15. Analytical Methods Committee. “Uncertainty of measurement—implications for its use in analytical science”, *Analyst* **120**, 2303–2308 (1995).
16. W. P. Cofino, D. E. Wells, F. Ariese, J.-W. M. Wegener, R. I. H. M. Stokkum, A. L. Peerboom. “A new model for the inference of population characteristics from experimental

data using uncertainties. Application to interlaboratory studies", *Chemom. Intell. Lab Systems* **53**, 37–55 (2000).
17. T. Fearn. "Comments on 'Cofino Statistics'", *Accred. Qual. Assur*. **9**, 441–444 (2004).
18. M. Thompson and P. J. Lowthian. "The frequency of rounds in a proficiency test: does it affect the performance of participants?", *Analyst* **123**, 2809–2812 (1998).
19. T. Fearn and M. Thompson. "A new test for 'sufficient homogeneity'", *Analyst* **126**, 1414–1417 (2001).

Приложение 1: Рекомендуемая процедура для проверки материала на достаточную однородность

А.1.1 Процедура

1. Подготовьте весь (не расфасованный) материал в виде, который предполагает его однородность, с помощью соответствующего метода.

2. Разделите материал на контейнеры, которые будут использоваться для отправки участникам.

3. Выберите минимум 10 контейнеров строго в случайном порядке.

4. Отдельно гомогенизируйте содержимое каждого из m выбранных контейнеров и выберите по две порции для испытаний из каждого контейнера.

5. Выполните анализ $2m$ порций для испытаний в случайном порядке в условиях повторяемости с помощью соответствующего метода. Используемый аналитический метод должен быть достаточно прецизионен для выполнения удовлетворительной оценки значения s_{sam}. Если возможно, то $\sigma_{an} < 0{,}5\sigma_p$.

Первый шаг заключается в проверке данных на явные аномалии. Такая проверка должна быть сделана визуально на простом графике, показывающем результаты напротив номера образца, посредством поиска таких диагностических признаков как: (a) трендов (тенденций) или скачков; (b) неслучайных распределений разностей между первым и вторым результатами испытания; (c) избыточного округления; и (d) выбросовых результатов внутри образцов.

Если все хорошо, мы используем данные для оценки аналитической и выборочной дисперсий. Если имеется программа для выполнения однофакторного дисперсионного анализа, то она может быть использована. В противном случае (если программа отсутствует), схема полного расчета приводится ниже.

А1.2 Статистические методы

А1.2.1 Процедура проверки повторных результатов с помощью критерия Кохрена

Рассчитайте сумму S_{i} и разность D_{i} для каждой пары повторных результатов для $i = 1, \ldots, m$.

Рассчитайте сумму квадратов S_{DD} разностей m по формуле

$$S_{DD} = \sum_{m} D_i^2$$

Статистика для критерия Кохрена является отношением наибольшего квадрата разности $D_{\max}^2$ к сумме квадратов разностей

$$C = \frac{D_{\max}^2}{S_{DD}}$$

Рассчитайте отношение и сравните его с соответствующим критическим значением из таблиц. Таблица 1 дает критические значения для 95 % и 99 % доверия для m от 7 до 20 пар.

Результаты для выбросовых по критерию Кохрена пар, обнаруженных при уровне доверия 95 % или более высоком уровне, следует всегда тщательно проверять на наличие ошибок в записях (обозначениях) или других ошибок в анализе и принимать соответствующие действия, если любые ошибки были обнаружены. Выбросовая пара не должна отклоняться, пока не будут обнаружены значимость при уровне доверия 99 % или неисправимые ошибки в аналитической процедуре. Единичный выброс по критерию Кохрена при уровне доверия 99 % должен быть исключен из дисперсионного анализа (ANOVA), если нет причин для обратного (смотри раздел 3.11).

А1.2.2 Проверка на значительную неоднородность

Используйте ту же сумму квадратов разностей для расчета

$$s_{an}^2 = \sum D_i^2 / 2m$$

Рассчитайте дисперсию V_S для суммы S_i

$$V_S = \sum (S_i - \overline{S})^2 / (m-1)$$

где $\overline{S} = (\frac{1}{m}) \sum S_i$ – это среднее значение S_i.

Рассчитайте дисперсию выборки s_{sam}^2 как

$$s_{sam}^2 = (\frac{V_s}{2} - s_{an}) / 2$$

или как $s_{sam}^2 = 0$, если предыдущая оценка имеет отрицательное значение.

Если имеется в распоряжении программа для однофакторного эксперимента, то приведенные выше величины $V_S/2$ и s_{an} могут быть получены из таблиц дисперсионного анализа, как «между(групповые)» и «внутри(групповые)» средние квадраты, соответственно.

Рассчитайте допустимую дисперсию выборки σ_{all}^2 как

$$\sigma_{all}^2 = (0{,}3\sigma_p)^2$$

где σ_p – это стандартное отклонение для оценки квалификации.

Взяв значения F_1 и F_2 из таблицы 2, рассчитайте критическое значение для проверки как $c = F_1\sigma_{all}^2 + F_2 s_{an}^2$

Если $s_{sam}^2 > c$, то это свидетельствует (значимо при 95 %-ом уровне доверия) о том, что стандартное отклонение выборки в генеральной совокупности образцов превышает допустимую долю целевого стандартного отклонения, и проверка на однородность привела к неудаче.

Если $s_{sam}^2 < c$, то нет таких доказательств, и проверка однородности принимается (считается удачной).

А1.2.3 Таблицы критических значений для проверки на однородность

Таблица 1 – Критические значения для статистик критерия Кохрена для повторных результатов

m	**95 %**	**99 %**
7	0,727	0,838
8	0,680	0,794
9	0,638	0,754
10	0,602	0,718
11	0,570	0,684
12	0,541	0,653
13	0,515	0,624
14	0,492	0,599
15	0,471	0,575
16	0,452	0,553
17	0,434	0,532
18	0,418	0,514
19	0,403	0,496
20	0,389	0,480

Таблица 2 – Коэффициенты F_1 и F_2 для использования в проверках на достаточную однородность

m*	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**
F_1	1,59	1,60	1,62	1,64	1,67	1,69	1,72	1,75	1,79	1,83	1,88	1,94	2,01	2,10
F_2	0,57	0,59	0,62	0,64	0,68	0,71	0,75	0,80	0,86	0,93	1,01	1,11	1,25	1,43

*m** - это количество образцов, которые были измерены повторно

Две константы в таблице 2 были получены из стандартных статистических таблиц как $F_1 = \chi^2_{m-1;0,95}/(\mathrm{m}-1)$, где $\chi^2_{m-1;0,95}$ – это значение, которое с вероятностью 0,05 превысит случайная величина, распределенная по закону хи-квадрат, с *m*-1 степенями свободы, и $F_2 = (\mathrm{F}_{m-1;0,95} - 1)/2$, где $\mathrm{F}_{m-1;0,95}$ – это значение, которое с вероятностью 0,05 превысит случайная величина, имеющая *F*-распределение, с *m*-1 степенями свободы.

А1.3 Рекомендуемые инструкции для лабораторий, выполняющих проверку на достаточную однородность

Лаборатория должна иметь опыт в используемых аналитических методах.

- Выберите строго наугад $m \leq 10$ распределяемых образцов из полного комплекта. Это должно быть сделано надлежащим образом посредством приписывания порядковых номеров образцам либо в явном виде (с помощью этикетирования) или в неявном виде (например, с помощью местопложения в массиве). Сделайте выбор, используя случайные числа из таблицы или сгенерированные с помощью компьютерной программы (например, Microsoft Excel) случайные числа (каждый раз с новым начальным числом). Выбор образцов другим способом является неприемлемым (например, перетасовкой их). Не применяйте случайную последовательность, использованную ранее.
- Перемешайте содержимое каждой распределяемой единицы соответствующим образом (например, в блендере) и из каждого образца отвесьте по две порции для испытаний. Пометьте порции для испытаний как показано.

Последовательный код распределяемого образца	**Пометка первой порции для испытаний**	**Пометка второй порции для испытаний**
1	1.1	1.2
2	2.1	2.2
3	3.1	3.2
.	.	.
.	.	.
m	*m*.1	*m*.2

- Отсортируйте 20 порций для испытаний в случайном порядке и выполните с ними в этом порядке все аналитические операции. Снова следует использовать таблицы случайных чисел или компьютерную программу для генерации новой последовательности случайных чисел. Пример последовательности случайных чисел (не для использования): 7,1; 3,1; 5,2; 5,1; 10,2; 1,1; 2,1; 9,2; 8,2; 1,2; 4,1; 2,2; 9,1; 10,1; 7,2; 3,2; 8,1; 6,1; 4,2; 6,2.
- Проведите анализы, если это возможно, в условиях повторяемости (например, в одной серии) или, если это невозможно, в последовательных сериях с минимально возможным для аналитической системы изменением, используя метод, который имеет стандартное отклонение менее $0,5\sigma_p$. Запишите результаты, если это возможно, с таким количеством значащих цифр, как то, которое требуется для $0,01\sigma_p$.
- Верните используемый порядок для 20 аналитических результатов, включая метки.

А1.4 Проверка однородности: Пример

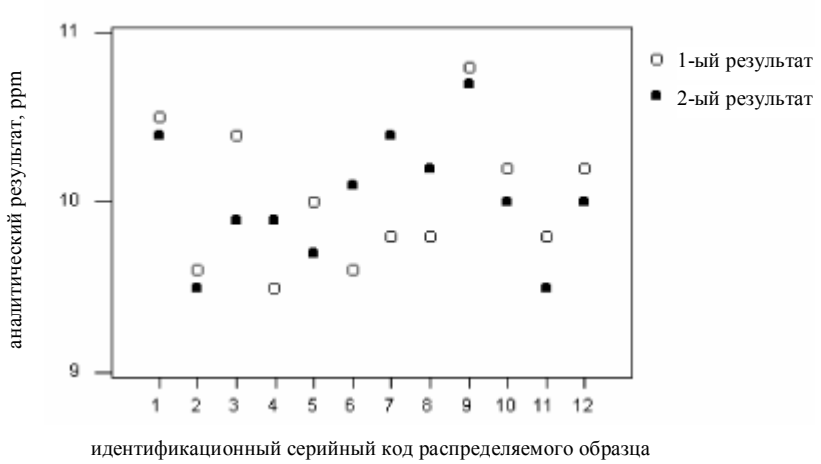
А1.4.1 Данные

Данные, показанные в таблице 3, взяты из Гармонизированного Протокола 1993 года [ссылка 1 в 3 части].

Таблица 3 – Повторные результаты для 12 распределяемых образцов соевой муки, анализируемых на содержание меди (ppm), вместе с некоторыми промежуточными стадиями расчетов для дисперсионного анализа (ANOVA)

Образец	**Результат *a***	**Результат *b***	***D = a - b***	***S = a + b***	$D^2 = (a - b)^2$
1	10,5	10,4	0,1	20,9	0,01
2	9,6	9,5	0,1	19,1	0,01
3	10,4	9,9	0,5	20,3	0,25
4	9,5	9,9	-0,4	19,4	0,16
5	10,0	9,7	0,3	19,7	0,09
6	9,6	10,1	-0,5	19,7	0,25
7	9,8	10,4	-0,6	20,2	0,36
8	9,8	10,2	-0,4	20,0	0,16
9	10,8	10,7	0,1	21,5	0,01
10	10,2	10,0	0,2	20,2	0,04
11	9,8	9,5	0,3	19,3	0,09
12	10,2	10,0	0,2	20,2	0,04

А1.4.2 Визуальный анализ



Данные представлены в визуальном виде, и не показывают никаких подозрительных особенностей, таких как несогласующиеся повторные результаты, выбросовые образцы, тенденции, скачки или любые другие систематические эффекты (можно использовать диаграмму Юдена, первый результат напротив второго повторного результата).

А1.4.3 Проверка с помощью критерия Кохрена

Наибольшее значение D^2 равно 0,36 и сумма D^2 равна 1,47, так что статистика Кохрена равна 0,36/1,47 = 0,24. Это меньше, чем 5 % -ое критическое значение, равное 0,54, так что не существует доказательства существования аналитических выбросов, и мы продолжаем обрабатывать полный набор данных.

А1.4.4 Проверка однородности

Аналитическая дисперсия: s_{an}^2 = 1,47/24 = 0,061

Дисперсия между образцами: Дисперсия суммы $S = a + b$ равна 0,463, так что

s^2_{sam} = (0,463/2 – 0,061)/2 = (0,231 – 0,061)/2 = 0,085

Приемлемая дисперсия между образцами: Целевое стандартное отклонение равно 1,14 ppm, так что допустимая дисперсия между образцами равна σ^2_{all} = $(0{,}3\times1{,}14)^2$ = 0,116.

Критическое значение: Критическое значение для испытания равно 1,79

$\sigma^2_{all} + 0{,}86s^2_{an}$ = 1,79×0,116 + 0,86×0,061 = 0,26.

Так как s^2_{sam} = 0,085 < 0,26, то материал является достаточно однородным.

Приложение 2: Пример проведения проверки на стабильность

Была выполнена процедура, описанная в Разделе 3.11.5. Стандартное отклонение для оценки квалификации (σ_p) было установлено в размере 0,1*c* (т. е., RSD составляет 10 %), а результаты анализа в условиях повторяемости в случайном порядке, в котором проводились анализы, представлены в таблице ниже.

Материал	Результат, ppm
Экспериментальный	11,5
Контрольный	13,4
Контрольный	12,2
Экспериментальный	12,3
Контрольный	12,7
Экспериментальный	10,9
Контрольный	12,5
Экспериментальный	11,4
Экспериментальный	12,4
Контрольный	12,5

Основанный на двойной выборке *t*-критерий с объединенным стандартным отклонением приводит к следующим статистикам:

	n	$\bar{x}$
Экспериментальный	5	12,66
Контрольный	5	11,70
Разность		0,96

Объединенное стандартное отклонение: 0,551.

95 %-ый доверительный интервал для (μ_{cont} - μ_{expt}): (0,16; 1,76)

t-критерий для H_0: $\mu_{cont} = \mu_{expt}$ против H_A: $\mu_{cont} \neq \mu_{expt}$ дает значение t = 2,75 с 8 степенями свободы, соответствующее вероятности (значение p) 0,025. Поэтому разница нестабильности, равная 0,96 ppm, является значимой при 95 %-ом уровне доверия. (Это может быть также выведено из 95 %-ого интервала доверия, который не включает ноль.)

Используя среднее, равное примерно 12, в качестве концентрации аналита, мы находим σ_p = 0,1×12 = 1,2. Разность вследствие нестабильности является намного большей, чем желаемый предел $0{,}1\sigma_p$, поэтому следует признать наличие значительной нестабильности, а материал - непригодным для использования.

Приложение 3: Примеры из практики по определению согласованного значения от участников для использования его в качестве приписанного значения

А3.1 Пример 1

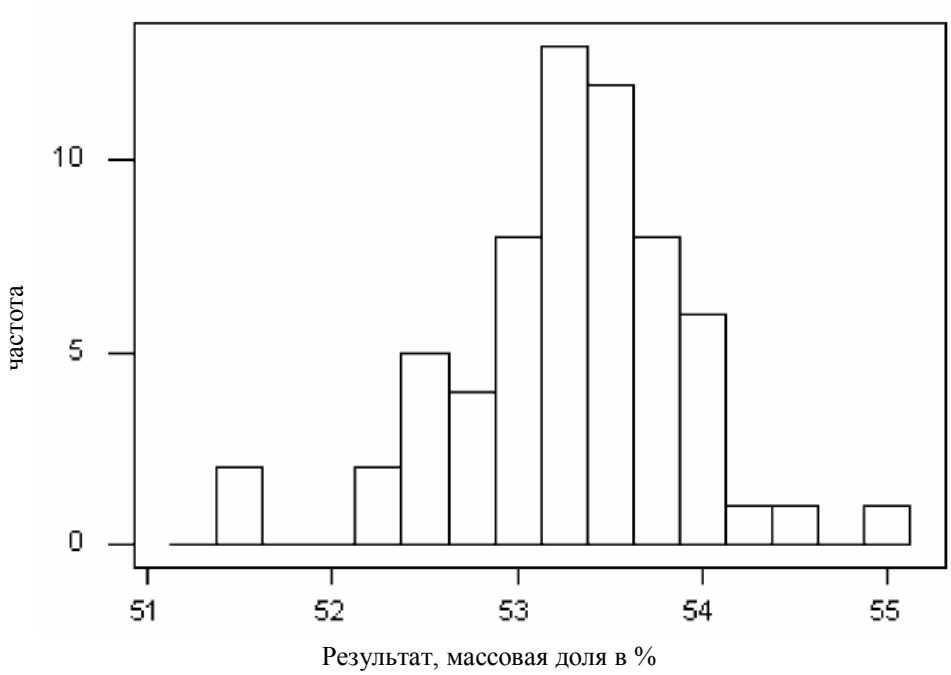
Результаты участников приведены в следующей таблице, которая также показывает соответствующие суммарные статистики. Единицы являются массовыми долями, выраженными в процентах (%); числовая прецизионность (количество разрядов) такая, какая была заявлена участниками.

Представленные результаты						
54,09	53,15	53,702	52,9	53,65	52,815	53,5
52,95	52,35	53,49	55,02	53,32	54,04	53,15
53,41	53,4	53,3	54,33	52,83	53,4	53,38
53,19	52,4	52,9	53,44	53,75	53,39	53,661
54,09	53,09	53,21	53,12	53,18	53,3	53,62
53,7	53,51	53,294	53,57	52,44	53,04	53,23
63,54	46,1	53,18	54,54	53,76	54,04	53,64
53	54,1	52,2	52,54	53,42	53,952	50,09
53,06	48,07	52,51	51,44	52,72	53,7	
53,16	53,54	53,37	51,52	46,85	52,68	

Итоговые статистики	
n	68
Среднее	53,10
Стандартное отклонение	1,96
Медиана	53,30
оценка среднего H15*	53,24
оценка стандартного отклонения H15*	0,64

*Смотри ссылки [1] и [2] для этого Приложения

Стандартное отклонение для оценки квалификации σ_p составляет 0,6 %.



Гистограмма (без выбросов) наводит на мысль об унимодальном и приблизительно симметричном распределении.

Итоговые статистики показывают почти полное совпадение робастного среднего и медианы. Робастное стандартное отклонение меньше чем $1{,}2\sigma_p$, так что не существует проблем, имеющих место в широких распределениях. Значение $\sigma_{rob} / \sqrt{n}$ = 0,079, которое значительно ниже рекомендуемой границы $0{,}3\sigma_p$ = 0,17.

Согласованное значение и его стандартная неопределенность принимаются как массовые доли равные 53,24 % и 0,08 % соответственно.

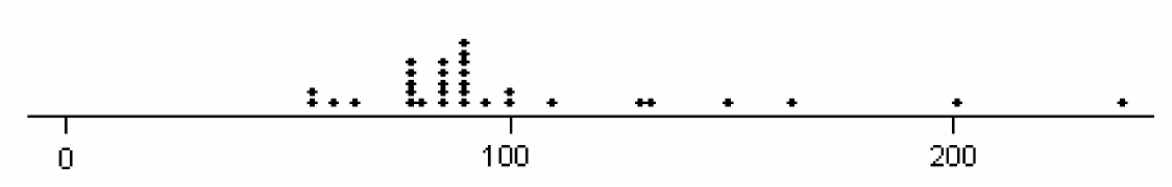
А3.2 Пример 2

Участники представили следующие результаты (единицами измерения являются ppb, миллиардная, 10^9, массовая доля):

Представленные результаты, ppb						
133	89	55	84,48	84,4	90,4	66,6
77	80	60,3	84	78	85	130
90	79	99,7	149	91	164	
78	84	110	77	91	89	
95	55	90	100	200,56	237	

Суммарные статистики	
n	32
Среднее	99,26
Стандартное отклонение	39,76
Медиана	89,0
оценка среднего H15*	91,45
оценка стандартного отклонения H15*	23,64

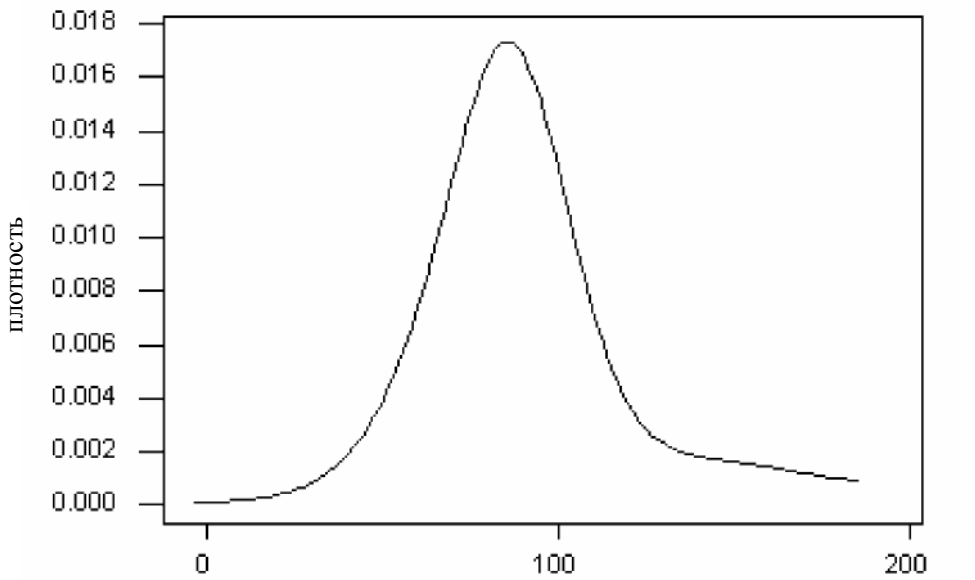
Точечная диаграмма (ниже) представленных результатов показывает набор данных с сильной положительной асимметрией, которая может поставить под сомнение достоверность робастных статистик.



Аналитический результат, ppb

Предварительное стандартное отклонение для оценки квалификации, полученное на основании робастного среднего, было задано функцией Хорвитца: $\sigma_p = 0{,}452 \times 91{,}4^{0{,}8495} = 20{,}8$ ppb.

В связи с наличием асимметрии и большого робастного стандартного отклонения, робастное среднее предполагается сомнительным, так что распределение ядерной плотности было построено с шириной интервала h, равной $0{,}075\sigma_p$:



Аналитический результат, ppm

Ядерная плотность показывает единственную моду, равную 85,2 ppb, и применение процедуры бутстрапа (bootstrapping) к данным предоставляет стандартную ошибку для моды равную 2,0 ppb. Измененное σ_p, основанное на концентрации 85,2, составило 19,7 ppb. Предполагаемая неопределенность моды (2,0) находиться ниже рекомендуемой границы $0{,}3\sigma_p = 5{,}9$ ppb.

Согласованное значение и его стандартная неопределенность принимаются как 85 и 2 *ppb* соответственно.

А3.3 Пример 3

Это первый тур серии после того как представленный метод был изменен с целью количественно квалифицировать различные «формы взвешивания». Отношение относительных молекулярных масс по новой и по старой форме взвешивания составляет 1,35.

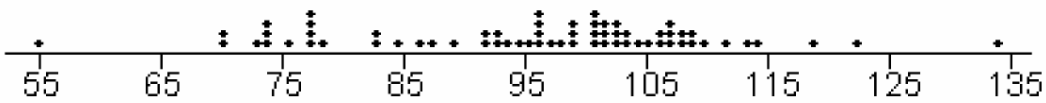
Стандартное отклонение для оценки квалификации для серии определяется функцией Хорвитца, а именно $\sigma_p = 0{,}16 \times c^{0{,}8495}$, где σ_p и c выражаются в ppm (миллионная (10^6) массовая доля).

Результаты участников (в ppm):

Представленные результаты, ppm						
102,5	97,9	102	101	99	75,9	101
74	94	93	70	82,9	106	113
122	97	114	101	70	103,88	93
96	107	103	96	119	99	83
107	101	134	109	103,8	106	77
95	108	96	104	101,33	92,2	
94,5	102	77	98,91	107	109	
89	110	103	112	55	87	
108	105,4	86	74	73	77	
96	77,37	73,5	78	92	84,6	

Суммарные статистики	
n	65
Среднее	95,69
Стандартное отклонение	14,52
Медиана	98,91
оценка среднего H15*	95,78
оценка стандартного отклонения H15*	14,63

Точечная диаграмма (ниже) показывает на возможное бимодальное распределение совокупности.



Аналитический результат, ppm

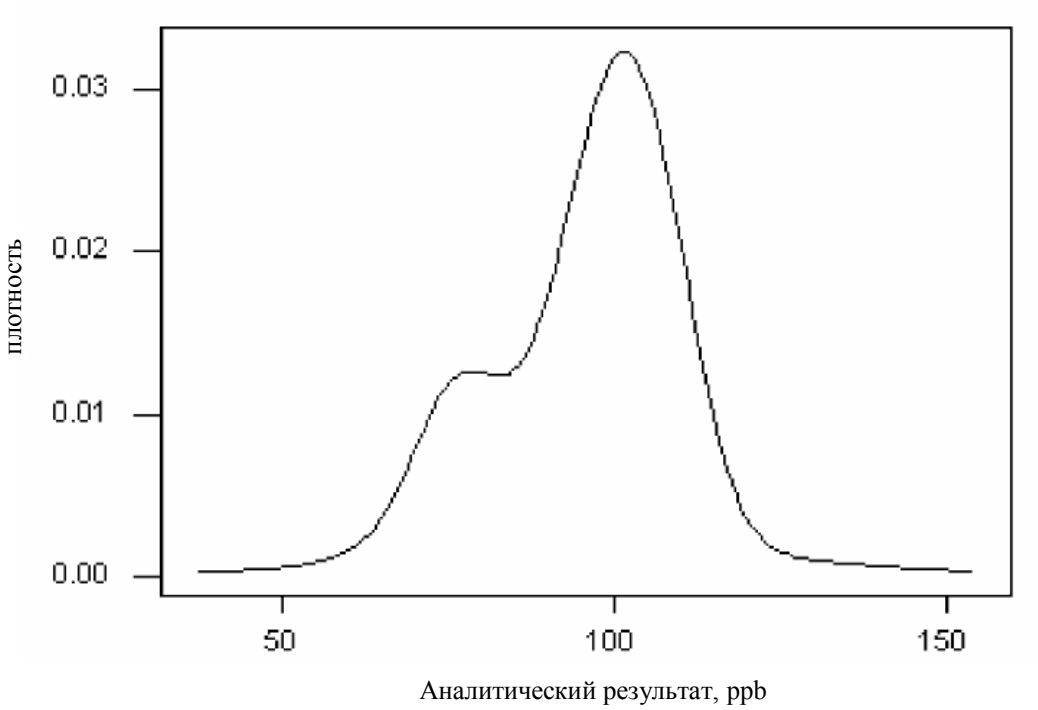
Предварительное значение σ_p, полученное на основании робастного среднего, равно 7,71 ppm, но робастное стандартное отклонение значительно больше, чем это, так что существует веское основание для подозрения смешанного распределения. Ядерная плотность была представлена с использованием ширины интервала *h*, равной 0,75×7,71 = 5,78.

Эта функция плотности имеет моды 78,6 и 101,5 ppm, с стандартными ошибками, оцененными с помощью процедуры бутстрап (bootstrap) как 13,6 и 1,6 соответственно. Отношение мод составляет 101,5/78,6 = 1,29, что близко к отношению 1,35 относительных молекулярных масс, и подтверждает предположение о том, что главная мода является верной, а второстепенные моды представляют участников, которые неверно представили ранее использовавшуюся форму взвешивания.

Поэтому согласованное значение определяется как 101,5 с неопределенностью 1,6 ppm. Пересчитанное целевое значение, основанное на этом согласованном значении, составляет 8,1. Так как неопределенность 1,6 меньше, чем 0,3×8,1 = 2,43, то неопределенность признается приемлемой и согласованное значение может быть использовано в качестве приписанного значения.

Библиография

1. Analytical Methods Committee. “Robust statistics—how not to reject outliers: Part 1 Basic concepts”, *Analyst* **114**, 1693 (1989).
2. International Organization for Standardization. *ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons*, Geneva, Switzerland (2005).



Приложение 4: Оценивание z-показателей в долгосрочной перспективе: суммарные показатели и графические методы

В то время как единичный *z*-показатель включает ценную информацию о функционировании лаборатории, рассмотрение набора или последовательности z-показателей обеспечивает более глубокое понимание. Кроме того z-показатели, собранные в течение некоторого времени (для единичной комбинации испытываемых/анализируемых материалов) влияют на неопределенности участника. Как статистические, так и графические методы могут быть подходящими для анализа совокупности z-показателей. Тем не менее, при интерпретации комбинированных статистик, требуется должная осторожность во избежание некорректных выводов. Следует особенно подчеркнуть, что использование комбинированного показателя, полученного на основании z-показателей, относящихся к ряду различных аналитов не рекомендуется; это имеет очень ограниченный диапазон допустимого применения и приводит к утаиванию случайных или постоянных проблем с отдельными аналитами. Кроме того, имеется склонность злоупотреблять этими показателями некомпетентными специалистами.

А4.1 Комбинированные показатели

Следующие два типа комбинированных показателей статистически обоснованны и могут быть использованы для отдельных участников для оценки последовательности z-показателей [z_1, z_2, …z_i, …z_n], полученных для единичной комбинации аналита, испытываемого материала и метода.

Измененная с помощью масштаба сумма z-показателей

$$S_{Z,rs} = \sum_i z_i \Big/ \sqrt{n}$$

может быть интерпретирована таким же способом, как и единичный z-показатель, т. е. ожидается, что его значения будут сцентрированы вокруг нуля с единичной дисперсией. Эта статистика имеет полезное свойство демонстрировать постоянное смещение или тренд, так что последовательность результатов [1,5; 1,5; 1,5; 1,5] приведет к статистически значимому значению $S_{Z,rs}$, равному 3,0 не смотря на то, что ни один единичный показатель

не является значимым. Однако, он может маскировать два больших z-показателя с противоположным знаком, таких как имеющихся в последовательности [1,5; 4,5; -3,6; 0,6].

Сумма квадратов z-показателей

$$S_{ZZ} = \sum_i z_i^2$$

может быть интерпретирована, как распределение χ_n^2 для сцентрированных вокруг нуля z-показателей с единичной дисперсией. Эта статистика имеет преимущество в том, что она позволяет избежать сокращения больших z-показателей противоположного знака, а также менее чувствительна к малым смещениям.

Обе эти комбинированные статистики нуждаются (например, с помощью робастных методов или фильтров) в защите от полученных ранее выбросовых показателей, которые имеют длительное постоянство. Обе статистики (если применены робастные методы) могут иметь отношение к неопределенности измерения следующим способом. Если *z*-показатели основаны на соответствии назначению и, следовательно, считаются случайными *N*(0,1), то значительно высокие уровни комбинированных статистик показывают, что неопределенность измерения участника больше, чем критерий соответствия назначению, указанный в программе проверки квалификации.

А4.2 Графические методы

Графические методы обобщения набора z-показателей могут быть столь же информативны, как и комбинированные показатели, и их труднее неверно истолковать. Могут использоваться контрольные карты Шухарта (с предупреждающими и регулирующими границами $z = \pm 2$ и $z = \pm 3$ соответственно). Составные одномерные символьные графики [1], такие как показано ниже, дают четкое представление и особенно полезны, когда рассматриваются показатели из группы аналитов, определенных общим методом. Вручную составленные графики можно быстро обновлять и управлять ими, так же как и теми, которые составлены с помощью компьютера.

Контрольная карта (рис. А4.1) показывает направленными вверх значками (символами) *z*-показатели, которые больше ноля и направленными вниз символами те показатели, которые меньше нуля. Малые символы представляют случаи, когда $2 \le |z| < 3$ и большими – случаи, когда $|z| \ge 3$. Проиллюстрированные данные сразу же показывают несколько заслуживающих внимание особенностей. Результаты из тура 11 являются слишком низкими, демонстрируя наличие определенных ошибок в процедуре, в то время как аналит 7 дает слишком часто высокие результаты, демонстрируя постоянную проблему с этим конкретным аналитом. Оставшиеся результаты являются достаточно согласованными с соответствием цели, которая в среднем приводит к 5 % z-показателей, представленных малыми символами.

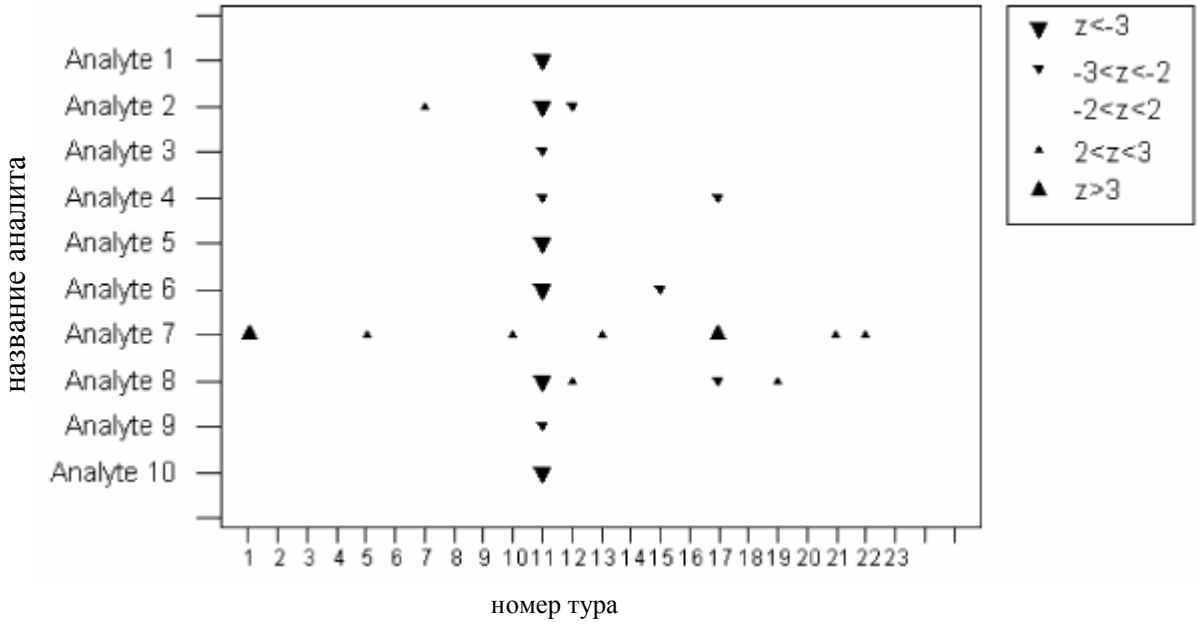


Рисунок А4.1 – Контрольная карта для z-показателей

J-диаграмма (также известная как «диаграмма зон») является даже более информативной, так как комбинирует возможности карты Шухарта и карты кумулятивных сумм. Она это делает с помощью накопления значений специальных *J*-показателей, приписываемых последовательным результатам по обе стороны от нулевой линии. Это позволяет обнаружить постоянные смещения, а также резкие большие изменения в аналитической системе. Обычные правила для преобразования *z*-показателей в *J* являются следующими.

Если $z \geq 3$, тогда $J = 8$.

Если $2 \leq z \leq 3$, тогда $J = 4$.

Если $1 \leq z < 2$, тогда $J = 2$.

Если $-1 < z < 1$, тогда $J = 0$.

Если $-2 < z \leq -1$, тогда $J = -2$.

Если $-3 < z \leq -2$, тогда $J = -4$.

Если $z \leq -3$, тогда $J = -8$.

J – показатели из последовательных туров накапливаются (суммируются) до тех пор пока $|J| \geq 8$, который является слишком удаленным от пределов регулирования, и тогда начинаются процедуры исследования. После любого такого отклонения и когда последовательные значения *J* имеют противоположные знаки, «накопитель» обнуляется сразу (то есть, перед любым дальнейшим суммированием).

Несколько примеров кумулятивного эффекта смещения видны на рисунке А4.2 (который иллюстрирует для сравнения те же результаты, что и на рисунке А4.1). Например, аналит 3 в турах 1 – 4 получает z-показатели, равные 1,5; 1,2; 1,5 и 1,1 соответственно, преобразуя их *J*-значения, равные 2, 2 ,2 и 2, которые суммируются до 8 в 4-ом туре и запускают процедуру исследования. Подобные примеры следует рассмотреть для аналита 7.

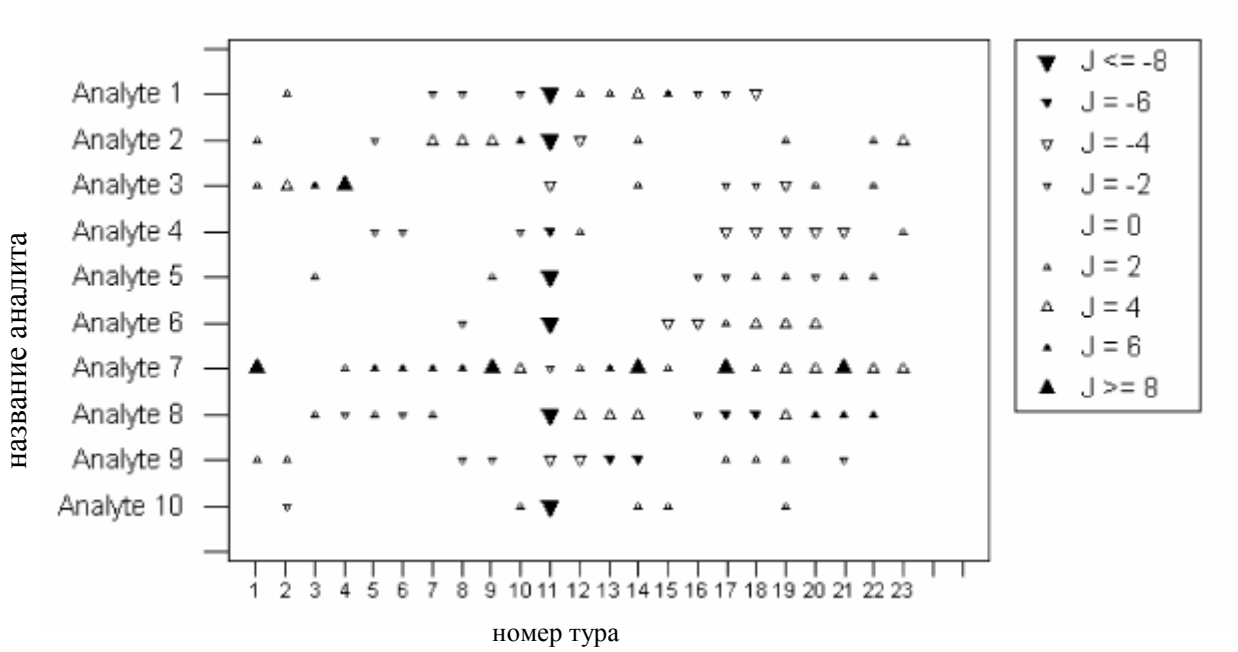


Рисунок А4.2 – *J*-диаграмма для *z*-показателей (те же данные, что и на предыдущем графике)

Библиография

1. M. Thompson, K. M. Malik, R. J. Howarth. «Multiple univariate symbolic control chart for internal quality control of analytical data» Anal. Comm. 35, 205-208 (1998).
2. Analytical Methods Committee. «The J-chart: A simple plot that combines the capabilities of Shewhart and cusum chart, for use in analytical quality control», AMC Technical Briefs: № 12. <www.rsc.org/amc/>.

Приложение 5: Метод валидации на основании результатов программ проверок квалификации

Целью программы проверки квалификации является проверка точности участвующих лабораторий. Участники могут свободно выбрать метод анализа и, в общем, использовать множество методов (или разновидность одного «метода»). В результате обычно нет возможности валидировать метод с помощью проверки квалификации. Однако, валидация метода становится возможной, если имеется достаточное количество участников программы проверки квалификации, которые используют один и тот же определенный метод анализа. Такая возможность, если используется должным образом, может рассматриваться в качестве недорогой альтернативы или признания в качестве совместного исследования, которое является известным [1] (но дорогим) экспериментом для валидации межлабораторным методом. (Совместные исследования обычно стоят 30 000 € для исследования одного метода.)

Программы проверки квалификации, однако, отличаются от совместных исследований проектом, окончательными результатами и выводами, а именно:

- Часто, только один материал для испытаний (или малое число) посылается в любом одном туре проверки квалификации в отличие от совместных исследований, где используются как минимум пять. Поэтому необходимо собрать данные из многих туров в течение периода времени, охватывающего возможно несколько лет, чтобы обеспечить достаточной информацией процесс валидации. (В этом контексте важно помнить, что, строго говоря, мы не валидируем «метод» отдельно (изолировано). То что мы валидируем так это метод применительно к определенным аналитам и определенным диапазонам

испытываемых материалов и концентраций аналита. Поэтому, не все туры в сериях могут быть использованы для задачи валидации.)

- Программы проверки квалификации редко требуют представления от участников в отчетах повторных результатов, поэтому нет возможности получить оценку стандартного отклонения повторяемости из результатов проверки квалификации. (Это не большая потеря – это просто интересная информация для лабораторий для оценки их собственных стандартных отклонений повторяемости.)
- В программах проверки квалификации нет гарантии, что одни и те же лаборатории будут участвовать в программе в различных турах.
- В совместном исследовании участники выбираются на основании возможной компетенции. В проверках квалификации не следует предполагать всеобщую компетенцию.

С учетом этих отличий результаты проверки квалификации, ограниченные участниками, использующими четко определенный протокол метода, можно разумно использовать для оценки стандартного отклонения метода [2]. Для получения желаемого результата следует обязательно использовать методы робастной оценки в комбинации с экспертной оценкой. Если два или более четко определенных метода используются достаточным количеством участников, то возможно оценить любое отклонение между методами по всему диапазону концентраций [3, 4] с помощью оценки функциональной зависимости [5, 6].

Библиография

1.W. Horwitz (Ed.). “Protocol for the design, conduct and interpretation of method performance studies”, *Pure Appl. Chem*. **67**, 331–343 (1995).

2.Paper CX/MAS 02/12 of the Codex Committee on Methods of Analysis and Sampling. Validation of Methods Through the Use of Results from Proficiency Testing Schemes, Twenty-fourth Session, Budapest, Hungary, 18–22 November 2002, FAO, Rome.

3.P. J. Lowthian, M. Thompson, R. Wood. “The use of proficiency tests to assess the comparative performance of analytical methods: The determination of fat in foodstuffs”, *Analyst* **121**, 977–982 (1996).

4.M. Thompson, L. Owen, K. Wilkinson, R. Wood, A. Damant. “A comparison of the Kjeldahl and Dumas methods for the determination of protein in foods, using data from a proficiency test”, *Analyst* **127**, 1666–1668 (2002).

5.B. D. Ripley and M. Thompson. “Regression techniques for the detection of analytical bias”, *Analyst* **112**, 377–383 (1987).

6.Analytical Methods Committee. “Fitting a linear functional relationship to data with error on both variables”, AMC Technical Brief No 10. <www.rsc.org/amc/>.

Приложение 6: Как участники должны реагировать на результаты проверки квалификации

A6.1 Общее введение

Участие в программах проверки квалификации будет напрасным до тех пор, пока участник полностью не воспользуется результатами каждого тура, избегая любой неверной интерпретации. Проверка квалификации является в первую очередь инструментом самопомощи, который позволяет участникам определять непредвиденные источники погрешностей в своих результатах. Однако проверка квалификации не предназначена для установления их причин. Следовательно, она только помогает участнику, который уже использует валидированные методы и имеет систему внутреннего контроль качества (IQC) для постоянных (рутинных) операций. В таких условиях неожиданный плохой результат в проверке квалификации указывает на несоответствие либо метода валидации или IQC или, что более вероятно, и того и другого. (Соответствующая система IQC будет обычно указывать на проблему с анализом за долго до того как была выполнена проверка

квалификации. Существует очевидная связь между функционированием участника в проверке квалификации и эффективностью внедрения системы IQC [1].)

Очень важно избегать недоразумений, когда использование показателей проверки квалификации выходит за рамки чисто научной области и используется, например, для аккредитации или в рекламных материалах лаборатории. Участник должен принимать во внимание статистическую природу z-показателя при его интерпретации.

Следующие рекомендации могут помочь участникам при выполнении правильной интерпретации и использовать z-показатели. Они приводятся практически неизмененными по отношению к Техническому отчету АМС с согласия Королевского общества химиков [2].

А6.2 Проверка квалификации и аккредитация

Проверка квалификации является настолько эффективной при определении неожиданных проблем при выполнении аналитических работ, что участие в программе (если она существует), как правило, требуется в качестве предварительного условия при аккредитации аналитической работы. Эксперты по аккредитации ожидают увидеть документально оформленную процедуру, относящуюся к соответствующему реагированию лаборатории на любые результаты, которые показали недостаточную точность.

Такая процедура должна включать следующие возможности:

- определение соответствующего критерия для определения момента начала действий по исследованию и/или устранению недостатков;
- определение исследовательских и процедур по устранению недостатков, которые будут использоваться и план их применения;
- запись результатов испытаний и заключений, полученных в таких исследованиях; и
- запись последующих результатов, показывающих эффективность любых процедур по исправлению.

Этот раздел содержит рекомендации, которые позволят удовлетворить эти требования и продемонстрировать, что эти требования были удовлетворены.

А6.3 Процедуры и документация

Участники должны ввести в действие документированную процедуру для исследования и обращения с плохими z-показателями. Лучшая практика для участника будет зависеть от того, как именно организована программа проверки квалификации. Система может принимать форму блок-схемы или дерева принятия решений, основанных на рассуждениях, обсуждаемых ниже и конкретных потребностей участника. Однако, в процедуру должны быть однозначно включены возможности использования профессионального суждения.

Программы проверки квалификации в области химии обычно включают набор критериев для оценки соответствия, который применим ко всем областям применения. Даже если задан критерий «соответствия назначению», он может подходить или не подходить для индивидуальной работы участника для отдельного потребителя. Этот фактор необходимо учитывать, если участник устанавливает систему реагирования на показатели, полученные в каждом туре программы. Основные возможности рассматриваются ниже.

А6.4 Влияние критерия, на котором основаны количественные показатели

А6.4.1 Программа проверки квалификации использует критерий соответствия назначению

Простейший случай возникает, если программа предлагает критерий соответствия назначению σ_p в качестве стандартной неопределенности и использует ее для расчета z-показателей. В этом случае важно реализовать, чтобы σ_p определялась заранее организаторами программы для описания их представления соответствия назначению: оно полностью не зависит от результатов, полученных участниками. Значение σ_p определяется так, что оно может рассматриваться, как стандартное отклонение. Поэтому если результаты не смещены и распределены нормально, и стандартное отклонение участника от серии к

серии σ равно σ_p, то z-показатели будут распределены как $z \sim N(0, 1)$, так что в среднем примерно 1 из 20 z-показателей будут попадать в диапазон ± 2 и только около 3 на 1000 случаев будут лежать вне диапазона ± 3.

Однако, немногие (если такие имеются) лаборатории точно выполняют эти требования. Для несмещенных результатов, если стандартное отклонение от серии к серии σ меньше σ_p, тогда меньше точек, чем определено выше, попадут вне соответствующих границ. Если $\sigma < \sigma_p$, то большая часть попадает вне границ. На практике большинство участников оперируют в условиях $\sigma > \sigma_p$, но производимые результаты также включают смещение большего или меньшего размера. Такие смещения часто включают основную часть общей погрешности результата, и они всегда приводят к увеличению части результатов, попадающих вне границ. Например, в лаборатории, где $\sigma = \sigma_p$, смещение величины равное σ_p будет приводить к увеличению части результатов лежащих вне границ $\pm 3\sigma_p$ в восемь раз.

С учетом этих выводов безусловно полезно записывать и интерпретировать z-показатели для конкретного типа анализа в форме контрольной карты Шухарта [3] или других контрольных карт (смотри также Приложение 4). Если функционирование участника согласуется с целью, z-показатель будет попадать за пределы диапазона ± 3 очень редко. Если такое случается, то было бы более разумно предположить, что аналитическая система производит серьезное смещение, чем произошла очень необычная случайная ошибка. Такой инцидент будет демонстрировать, что лаборатории необходимо предпринять меры для устранения проблемы. Два последовательных z-показателя, которые попали между 2 и 3 (или между -2 или -3) могут быть интерпретированы тем же самым способом. На самом деле, могут использоваться все обычные правила для интерпретации карт Шухарта (например, правила Вестгарда [3]).

В дополнение к использованию карт Шухарта может быть полезной проверка z-показателей на наличие долгосрочных смещений, а также с использованием карт кумулятивных сумм или *J*-карт (Приложение 4). Эти проверки смещений не являются строго необходимыми: Если z-показатели участников практически полностью отвечают требованиям критерия соответствия назначению, небольшое смещение может быть не важно. Однако, как мы сказали выше, наличие даже небольшого смещения будет приводить к увеличению части результатов, выходящих за пределы регулирования, и поэтому возможно следует его устранить. Участник, который решает проигнорировать смещение должен сказать об этом в отчете по исследованию. Другими словами, участник должен сделать понятным для оценщиков по аккредитации, что решение проигнорировать смещение является намеренным и обоснованным, а не случайным.

А6.4.2 Программа проверки квалификации не использует критерий, основанный на соответствии назначению

Некоторые программы проверки квалификации не оперируют понятием «соответствия назначению». Провайдер проверки квалификации сам рассчитывает показатель из результатов участников (т. е. без внешних ссылок на действующие требования). Чаще участник может обнаружить, что критерий соответствия назначению, используемый провайдером, является неприемлемым для основных видов работ, выполняемых в лаборатории. Участникам таких программ потребуется рассчитать свои личные показатели, основанные на соответствии назначению. Это может быть осуществлено способом представленным ниже.

Участник должен согласовать с потребителем определенный критерий соответствия назначению в форме стандартной неопределенности σ_{ffp}, и использовать для расчета модифицированного z-показателя, определяемого по формуле:

$$z_L = (x - x_a) / \sigma_{ffp}$$

для замены принятого z-показателя (смотри раздел 3.5.4). Приписанное значение x_a может быть получено из программы самостоятельно. Если имеется несколько потребителей с различными требованиями по точности, то возможно получение нескольких показателей,

полученных на основании одного результата. С этими показателями можно обращаться точно таким же способом, как рекомендовано выше для z-показателей, то есть обычными типами контрольных карт. Так как приписанное значение аналита неизвестно участнику во время анализа, критерий соответствия назначению должен определяться как функция от концентрации аналита *c*, как показано в разделе 3.5.

А6.5 Как исследовать плохой z-показатель

Исследование плохого z-показателя обычно неразрывно связано с IQC [3]. В обычных обстоятельствах участник проверки квалификации узнает о плохом z-показателе через несколько дней или недель после выполнения анализа. В рутинных анализах, однако, любые существующие проблемы, влияющие в целом на ход измерения, должны обнаруживаться немедленно с помощью процедур IQC. Причина проблемы должна исправляться немедленно. Выполнение анализа на материале для проверки квалификации должно быть проанализировано и предположительно более точный результат представляется для программы проверки квалификации. Так неожиданно плохой z-показатель показывает, что или (a) система IQC является неадекватной или (b) материал проверки квалификации, только один из испытываемых в серии материалов, являлся причиной проблемы. Участники должны рассматривать обе эти возможности.

А6.5.1 Недостатки в системе IQC

Общий недостаток IQC состоит в том, что материал IQC плохо соответствует типичному материалу для испытаний. Материал IQC должен быть насколько возможно ближе отображать типичный материал для испытаний, в отношении матрицы, компартментализации (пространственного разделения), видообразования, концентрации аналита. Только тогда поведение материала IQC может быть полезным руководством в целом для выполнения анализа. Если материалы для испытаний сильно отличаются по любому из этих аспектов, то желательно использовать более одного материала IQC. Например, если концентрация аналита значительно отличается в материалах для испытаний (скажем, на два порядка по величине), то следует рассмотреть два различных материала IQC с концентрацией, равной предельным значениям обычного диапазона. Особенно важно избегать использования простого стандартного раствора аналита, в качестве заменителя IQC для материалов со сложной матрицей.

Другая проблема может возникнуть, если система IQC направлена только на прецизионность между сериями и игнорирует смещение среднего результата. Такое смещение может появиться из-за проблемы несоответствия матрицы материала IQC обычному типу материала для испытаний (и, как следствие, материалу для проверки квалификации). Поэтому важно сравнивать среднее результатов с наилучшей возможной оценкой истинного значения для материала IQC. Получение такой оценки требует прослеживаемости во внешней вышестоящей (эталонной) лаборатории. Внешняя прослеживаемость может быть получена, например, сравнением с SRM, имеющем похожую матрицу, или подвергая материал, претендующий на материал IQC межлабораторному исследованию.

А6.5.2 Проблема с материалом для проверки квалификации

Если участник уверен в том, что система IQC очевидно не имеет смещение, то следует подозревать наличие проблемы, касающейся исключительно результата испытания материала проверки квалификации. Плохой результат может быть результатом ошибки, связанной с обращением с материалом для проверки квалификации (например, некорректное взвешивание или зарегистрированный объем). Это может быть легко проверено. И наоборот, неожиданная форма смещения (такая как ранее незамеченное влияние слишком малого извлечения) может быть обусловлено исключительно материалом проверки квалификации или измерительным процессом. Обоснованный вывод на данном этапе может заключаться в том, что материал проверки квалификации существенно

отличается от типичного материала для испытаний и приводит к тому, что полученная оценка z-показателя является неприменимой к аналитической задаче.

А6.5.3 Диагностические проверки

Плохой z-показатель только лишь указывает на наличие проблемы, но не диагностирует ее, так что участник обычно требует дальнейшей информации для определения причины плохого результата. В качестве первого шага участник должен перепроверить записи выполнения анализа, касающиеся материала проверки квалификации. Далее следует искать следующие особенности:

- систематические или случайные ошибки в расчетах;
- используются некорректное взвешивание или объемы;
- вышедшие из под контроля индикаторы из обычных карт IQC;
- необычно высокие результаты холостых проб; и
- плохая степень извлечения, и т. д.

Если эти действия ни к чему не приводят (не дают понимания), то требуются дальнейшие измерения.

Очевидным действием будет провести повторный анализ проблемного материала для проверки квалификации в следующей рутинной серии анализа. Если проблема исчерпана (т. е. новый результат приводит к приемлемому z-показателю), участник может приписать исходную проблему случайному событию неизвестной причины. Если плохой результат сохраняется, то требуется выполнить более расширенное исследование. Это может быть вызвано анализом серии результатов, касающейся материалов проверки квалификации из предыдущих туров программы и/или соответствующих CRM при их наличии.

Если плохой результат все еще получается для исследуемого материала проверки квалификации, но отсутствует для результата испытания другого материала для проверки квалификации и CRM, то плохой результат скорее всего является следствием наличия уникального свойства материала, возможной неожиданной влияющей величины или эффекта матрицы. Обнаруженный факт требует более расширенного исследования для определения причины влияния. В дополнение участнику может понадобиться модифицировать рутинную аналитическую процедуру для учета наличия влияния в будущих материалах для испытаний. (Однако, может быть известно, что обычные испытываемые материалы не будут никогда содержать влияющий фактор и тогда решить, что неблагоприятный z-показатель был неприменим к конкретной лаборатории.)

Если проблема является общей для результатов старых материалов для проверки квалификации и CRM, то возможно наличие недостатков в аналитической процедуре и соответствующих недостатков в системе IQC. Оба этих факта требуют внимания.

А6.5.4 Дополнительная информация на основании результатов измерения нескольких аналитов

Некоторые проверки квалификации включают методы, такие как атомно-эмиссионная спектрометрия, которые могут одновременно определять несколько аналитов из единичной порции для испытаний и единичной химической обработки. (Хромотографические методы, которые определяют несколько аналитов в быстрой последовательности могут также быть рассмотрены как «одновременные» в данном обсуждении.) Из результатов измерений нескольких аналитов материала для проверки квалификации иногда может быть получена дополнительная информация, которая является диагностической. Если все или многие аналиты имеют неудовлетворительные результаты и подвержены какому-то влиянию в одинаковой степени, то причина этого должна лежать в действии, которое влияет на всю процедуру, такое как ошибка взвешивания порции для испытаний или добавления внутреннего стандарта (образца). Если только один аналит подвержен неблагоприятному влиянию, проблема может лежать в калибровке для этого аналита или уникальном химическом аспекте этого аналита. Если значительное количество аналитов затронуто, то применяются те же факторы. Например, при элементарном анализе горных пород, если

группа аналитов приводит к низким результатам, то может быть полезным рассмотреть является ли это следствием неполного растворения одной из минеральных фаз, составляющих горную породу, в которой сконцентрированы эти элементы. Кроме того могут присутствовать спектрометрические изменения, вызванные изменениями в работе распылительной системы или самой плазмы, которые влияют на одни элементы больше, чем на другие.

А6.5.4 Подозреваемое в смещении приписанное значение

Большинство программ проверки квалификации используют в качестве приписанного значения согласованное значение от участников. На практике редко существует альтернатива. Однако, использование согласованного значения повышает вероятность того, что среди группы лабораторий в основном использующих смещенный аналитический метод, небольшое количество участников используют метод измерения, не имеющий смещение. Это меньшинство может привести к результатам, которые отклоняются от согласованного значения и создают «неприемлемые» z-показатели. На практике, такое случается редко, но об этом известно, особенно когда в проверках квалификации используются новые аналиты или материалы для испытаний. Например, большинство участников могут использовать метод, который подвержен необнаруженному влиянию, в то время как меньшинство обнаружили влияние и разработали метод, который исключает его.

Часто проблема является очевидной для задетых ею участников, так как они использовали метод, основанный на более глубоком понимании химических процедур, чем тот, которые использовали большинство участников. Но проблема не видна другим участникам или провайдеру программы. Если участник подозревает, что он находится в таком положении, то ему следует, выполнив шаги описанные выше, послать провайдеру проверки квалификации детали накопленных доказательств о том, что приписанное значение является неверным. Провайдер обычно имеет доступ к записям информации о методах, используемых участниками, и может быть в состоянии отреагировать на жалобу немедленно. Кроме того, провайдер может начать долгосрочное исследование проблемы, которое должно разрешить несоответствие в установленном порядке.

Библиография

1. M. Thompson and P. J. Lowthian. “Effectiveness of analytical quality control is related to the subsequent performance of laboratories in proficiency tests”, *Analyst* **118**, 1495–1500 (1993).

2. Analytical Methods Committee. “Understanding and acting on scores obtained in proficiency testing schemes”, AMC Technical Brief No 11. <www.rsc.org/amc/>.

3. M. Thompson and R. Wood. “Harmonised guidelines for internal quality control in analytical chemistry laboratories”, *Pure Appl. Chem*. **67**, 649–666 (1995).

Приложение 7: Руководство по проверкам квалификации для конечных пользователей данных

Эти вопросы и ответы основаны на непонимании, имеющемся у конечных пользователей аналитических данных. Интерпретация результатов проверки квалификации в аналитической химии должна выполняться в согласии с химиком-аналитиком.

Что такое проверка квалификации?

Проверка квалификации включает межлабораторную систему для регулярной проверки точности, которую могут достичь участвующие лаборатории. В обычном случае организаторы программы распределяют порции гомогенизированного материала каждому из участников, которые анализируют материал в обычных условиях и сообщают результат организаторам. Организаторы обобщают результаты и информируют участников о результатах, обычно в форме числовых показателей, связанных с точностью результата.

Какова разница между проверкой квалификации и аккредитацией?

Органы по аккредитации требуют от аналитических лабораторий участия в соответствующих проверках квалификации, если они существуют и доступны для участия, и демонстрации системы для обработки результатов такой проверки. Это только одно из многих требований аккредитации.

Какие типы образцов распределяются?

Распределяемые материалы являются как можно более близкими к тем материалам, которые обычно анализируются, так что результаты программы представляют возможности лабораторий, работающих в обычных повседневных условиях.

Для чего нужны проверки квалификации?

Основная цель проверки квалификации – помочь лабораториям определить и исправить любую неприемлемо большую неточность в их результатах. Другими словами, она создана в качестве системы самопомощи, чтобы сообщить участникам, необходимо ли им модифицировать свои процедуры. Проверки квалификации не идеально подходят для любых других целей, хотя их результаты с учетом соответствующих ограничений могут использоваться и объединяться с другой информацией для некоторых других целей.

Почему имеются неточности в аналитических результатах?

Все измерения приводят к неточностям, технически известным в измерительном сообществе как «погрешности». (Слово «погрешность» здесь не означает, что была сделана ошибка, а означает только то, что результат измерительного процесса изменяется.) Погрешности возникают из-за неизбежного изменения в физической или химической процедуре, выполняемой для получения результата. Измерения химической концентрации требуют более сложных процедур, чем обычные физические измерения, такие как длина или время. Достаточно просто измерить длину с точностью до шестого знака после запятой, но химические измерения могут очень редко осуществляться с точностью выше второго знака после запятой. Часто точность еще хуже, особенно если концентрации очень низкие, например, когда определяются остатки пестицидов в продуктах питания.

Достаточна ли имеющаяся точность?

Это зависит от использования. Некоторые анализы имеют очень высокую точность. Например, при определении коммерческой ценности партии золотого лома содержание золота должно определяться с как можно меньшей возможной погрешностью (менее чем одна тысячная доля), так как даже малая погрешность может стоить нескольких тысяч евро. При других применениях, например, при определении концентрации меди в почве точность в одну десятую возможна будет достаточной, так как не играет большой роли, если действительное значение составляет 20 или 22 ppm, когда требуется принять решение является ли уровень выше или ниже 200 ppm. Стоимость также следует принимать во внимание. Эмпирическое правило состоит в том, что увеличение точности измерения в два раза уменьшает возможность неверного (т. е. дорого) решения, но увеличивает стоимость анализа в четыре раза. Эти соображения известны как «соответствие назначению».

Как оценивают программы проверки квалификации точность отдельных лабораторий?

Большинство программ преобразует результат участника в «z-показатель». Этот показатель отражает две характерных особенности, (a) действительную достижимую точность (т. е. разницу между результатом участника и принятым истинным значением), и (b) решение организатора программы о том, соответствует ли точность своему назначению.

Как следует интерпретировать z-показатели?

z-показатели должны интерпретироваться на статистической (вероятностной) основе, и это требует знаний эксперта. Однако, применяют следующие простые правила:

- Показатель, равный нулю, подразумевает отличный результат. Это случается достаточно редко даже в компетентных лабораториях.
- Лаборатории, выполняющие критерий соответствия назначению, установленный в программе проверке квалификации, будут иметь в основном показатели, попадающие

между -2 и 2. Можно ожидать, что их значения иногда будут попадать за пределы этого диапазона, примерно 1 раз из 20, однако такой отдельный случай не играет существенную роль. Знак (т. е. + или -) показателя показывает положительную или отрицательную погрешность соответственно.

- Показатель вне диапазона от -3 до 3 будет необычен для лаборатории, работающей в соответствии с критерием соответствия назначению, и показывает, что требование точности не выполняется (по крайней мере в данном случае). Причина такого события должна быть исследована и устранена.

Какие ошибки обычно делаются при использовании z-показателей?

Важно не переусердствовать в интерпретации z-показателей. Это может произойти одним из следующих способов:

- Сравнение z-показателей между турами или между лабораториями следует осуществлять с большой осторожностью. Единичная лаборатория, работающая в соответствии с критерием соответствия назначению, обычно получает z-критерии в последовательных турах в диапазоне от -2 до 2.: Следующий набор [0,6; -0,8; 0,3; 1,7; 0,7; -0,1] был бы обычным. Небольшие спады и подъемы между показателями не показывают изменение в функционировании – они возникают случайно. Так 1,7 не «хуже», чем 0,3 и не показывает ухудшение функционирования.
- Из-за такой «естественной вариации» не следует составлять «турнирную таблицу» лабораторий, основанную на их z-показателях в туре. Не допустимо утверждать, что лаборатория имеющая значение 0,3 в единичном туре лучше, чем другая с показателем, равным 1,7.
- Суждения, основанные на средних z-показателях требуют осторожности. Средние значения z-показателей, полученные на нескольких различных аналитах не должны использоваться; они могут скрыть факт того, что один из аналитов приводит к плохому z-показателю. Средние значения показателей для одного и того же аналита по нескольким турам могут быть более полезными, но все-таки нуждаются в интерпретации эксперта.

Каковы ограничения проверки квалификации?

- Проверка квалификации должна выполняться в рамках комплексной системы качества в каждой лаборатории. Она не может быть использована для замены повседневной QC. Отдельно она не является ни достаточным способом валидации аналитических методов, ни обучения отдельных аналитиков.
- Проверка квалификации указывает участвующей лаборатории только на наличие проблем, если они существуют. Но не предоставляет никаких способов обнаружения этих (диагностирования) проблем, которые помогут их решить.
- Успех в проверке квалификации для одного аналита не показывает, что лаборатория также компетентна при определении других аналитов.

Условные обозначения

C	- статистика критерия Кохрена для повторных данных ($C = D_{\max}/S_{DD}$)
c	- критическое значение при проверке на достаточную однородность ($c = F_1\sigma_{all}^2 + F_2\, s_{an}^2$)
c_L	- произвольно определенный предел для представления в отчете
D_i	- разность i-ой пары повторных измерении при проверке на однородность
$D_{\max}$	- наибольшая разность из пар повторных измерений
f	- константа, используемая при определении стандартного отклонения для проверки квалификации
F	- отношение дисперсий выборок, используемых в F-проверке для равенства дисперсий
F_1, F_2	- критические значения для проверки однородности (Приложение 1)
H	- статистическая гипотеза
H_0	- статистическая нулевая гипотеза
h	- ширина интервала в ядерной плотности
J-показатель	- показатель, основанный на количестве удачных результатов по обе стороны от нулевой линии
l	- множитель для использования при определении верхних пределов приемлемой неопределенности для приписанного значения
m	- количество распределяемых единиц при проверке на однородность
$N(\mu, \sigma^2)$	- нормальное распределение со средним значением совокупности μ и дисперсией совокупности σ^2
n	- количество результатов
s_{an}	- экспериментальная оценка аналитического стандартного отклонения
s_{sam}	- экспериментальная оценка выборочного стандартного отклонения
S_{DD}	- сумма квадратов разностей при проверке однородности, основанной на повторных анализах
S_{ZZ}	- сумма квадратов z-показателей $S_{ZZ} = \sum_i z_i^2$
$S_{Z,rs}$	- промасштабированная сумма z-показателей $S_{Z,rs} = \sum_i z_i / \sqrt{n}$
t	- статистика Стьюдента
$u(x)$	- стандартная неопределенность x
V_x	- дисперсия сумм при проверке на однородность $V_s = \sum (S_i - \overline{S})^2 /(m-1)$
x	- результат участника
x_a	- приписанное значение
x_{max}	- предельно допустимая концентрация аналита
x_{true}	- истинное значение измеряемой величины
z	- z-показатель $z = (x - x_a)/\sigma_p$
z'	- модифицированный z-критерий, включающий неопределенность приписанного значения $z' = \dfrac{x - x_a}{\sqrt{u^2(x_a) + \sigma_p^2}}$
z_L	- модифицированный z-показатель, включающий критерий функционирования, специфический для лаборатории $z_L = (x - x_a)/\sigma_{ffp}$
μ_{cont}	- среднее значение совокупности результатов для материала, хранящегося в постоянно контролируемых условиях
μ_{expt}	- среднее значение совокупности результатов для материала при экспериментальных условиях при испытаниях на стабильность

μ_{rob} - робастное среднее

σ - стандартное отклонение совокупности, генеральной

σ_{all} - разрешенное стандартное отклонение

σ_{ffp} - стандартное отклонение соответствия назначению

σ_p - стандартное отклонение для проверки квалификации

σ_{rob} - робастное стандартное отклонение

σ_{sam} - (истинное) выборочное стандартное отклонение, т. е. вклад вариации между образцами в наблюдаемую дисперсию

χ_n^2 - распределение хи-квадрат с n степенными свободы

ζ - показатель «дзета», $\zeta = (x - x_a) / \sqrt{u^2(x) + u^2(x_a)}$