

## Программа «ALTEY» для контроля качества клинических лабораторных анализов

Балаховский И.С., д.м.н.,  
действительный член Академии космонавтики им. К.Э. Циолковского.

Чтобы выводы доказательной медицины были действительно доказательными исходные данные должны быть достоверны. Такие, как возраст, заболеваемость или смертность непосредственно наблюдаются, поэтому достоверны a priori, другие – рост, масса тела достоверны потому, что точность обычных методов измерения во много раз выше возможного биологического разброса. Однако бывает и так, что доказательность выводов зависит от того, насколько точно проведены измерения. К этой категории относятся данные лабораторной медицины, или, как у нас принято говорить, клинических лабораторных исследований. Они почти всегда исходно оцифрованы, поэтому могут быть непосредственно использованы для вычислений. Но это ещё не означает, что они достоверны. Надо знать, насколько результаты одной лаборатории сопоставимы с данными другой, а также отражает ли обнаруженное изменение реальные биологические процессы или же возникли вследствие методической погрешности. Ошибки тут могут быть как в ту, так и в другую сторону – может быть сделан вывод о различии, которого на самом деле нет или пропущен реальный эффект. Строгий, научно обоснованный контроль качества лабораторных исследований необходим для обеспечения преемственности оказания медицинской помощи и осуществления «голубой мечты народного здравоохранения» - единой электронной истории болезни, доступной всем медицинским учреждениям.

Перед врачами всегда стоит по существу философский вопрос, – а что, собственно говоря, измеряется в лаборатории? Реально ли в живом организме существуют индивидуальные молекулы тех веществ, которые фигурируют в заглавии метода, или же это всего лишь абстракция, условные обозначения каких-то комплексов или отношений? [1]

Сейчас не место для глубокого обсуждения, и мы затронем этот вопрос только в том размере, который необходим для понимания методов статистической оценки качества лабораторной работы – т.е. того, как отличить реальные сдвиги состояния организма от ошибки измерения. Для этого существуют национальные и интернациональные институты, как бюджетные, так и коммерческие. Используемые ими методы близки к методам биометрии и метрологии, поскольку опираются на те же положения теории вероятностей, но все же имеются и особенности. Некоторое время назад московская компания ООО «Группа Алтэй» выпустила компьютерную программу для контроля качества клинических лабораторных исследований, пояснения к которой **Методическое пособие Altey Laboratory Quality Control** опубликовано на сайте [http://www.lab-universum.com.ua/files/Quality\\_Control\\_Ref\\_Book.pdf](http://www.lab-universum.com.ua/files/Quality_Control_Ref_Book.pdf)

Этот документ, во многом повторяющий некоторые официальные и полуофициальные документы [2-6], производит странное впечатление, как по используемой терминологии, так и по существу. Ниже мы постараемся проанализировать это. Поскольку код программы неизвестен, трудно сказать, действительно ли она реализует все нелепое, что написано в методическом пособии.

### **Общие принципы статистического контроля и терминология.**

Прежде чем перейти к существу вопроса, несколько слов об общих принципах статистического контроля качества лабораторных анализов и используемой терминологии. Это должно помочь понять суть проблемы тем, кто не работает

ежедневно в лаборатории. Существуют две группы способов оценки точности – одни основаны на обработке больших массивов результатов анализов, выполненных пациентам, другие на анализе специально приготовленных образцов, так называемых «контролей» (их иногда называют также «стандарты»), состав которых известен «аттестован». Сравнивая полученные результаты с аттестованными значениями, оценивают качество работы аналитической системы. Выводы автоматически распространяют на все анализы пациентам, которые были выполнены в тот же день или в той же аналитической серии. Поскольку на полное совпадение рассчитывать не приходится, нужны критерии того, какое различие допустимо. Это и есть научная проблема.

Давно прошло время, когда каждая лаборатория самостоятельно готовила реактивы, калибровочные и контрольные материалы. В наши дни их производят промышленные предприятия, а лаборатории покупают готовые наборы. Это, безусловно, упрощает работу и улучшает её качество, но одновременно создаёт новые проблемы. Биологический материал это сложная система, в которой всегда есть так называемые интерферирующие вещества, либо препятствующие развитию аналитической реакции, либо непомерно усиливающие её. Исследуемый параметр – аналит – чаще всего вещь в себе, невозможно непосредственно измерить его абсолютную концентрацию. Можно только сравнить разные образцы и, взяв один из них за эталон, построить функцию, связывающую величину сигнала прибора с концентрацией. Эта функция называется калибровочным графиком, использованные образцы калибраторами, а весь процесс калибровкой. Естественно, что результаты анализов в разные дни и в разных лабораториях будут сопоставимы, только если калибраторы одинаковы. Это большая проблема, но решают её не лаборатории, а изготовители реактивов путём использования калибраторов разного уровня. Самый высокий это международный эталон, по нему калибруются эталоны, используемые промышленностью, внизу иерархии те, которые используются в повседневной работе лабораторий. Система в целом называется «прослеживаемостью эталонов»[6]. Таким образом, в лаборатории есть, или, во всяком случае, должны быть, два типа специальных препаратов заводского изготовления, которые принято называть «материалами». Это калибровочные материалы для калибровки («калибраторы») для выполнения калибровки и контрольные материалы («контроли») для повседневного контроля правильности результатов. В принципе это препараты одного плана и взаимозаменяемы, но настоятельно рекомендуется, чтобы они были разных изготовителей, что обеспечивает объективность проверки.

В принципе к лабораторным анализам применимы те же правила метрологии, что и к любым другим измерениям, но есть и особенности. Это связано с тем, что в процессе исследования проба разрушается, потому измерение не может быть повторено, исследование другого образца того же материала дорого и обычно не практикуется. Лаборатория вынуждена оценивать достоверность одного единственного измерения и доверительный интервал в обычном понимании – как интервал с доверительной вероятностью, накрывающий истинное значение [7, стр.318], не может быть построен. Кроме того, метрологи чаще всего измеряют неслучайные величины, а в нашем случае измеряемый параметр – концентрация аналита у данного пациента, это случайная величина – у одного одна, у другого другая. Характер ее распределения обычно постоянен, поэтому можно оценить вероятность каждого возможного ответа. Это позволяет взглянуть на доверительный интервал не как на интервал с известной вероятностью накрывающий истинное значение, а как интервал вокруг найденного значения, в котором с заданной вероятностью находится истинное. Численно оба интервала могут быть равны. Чтобы избежать путаницы, мы предлагаем называть его медицинским доверительным интервалом.

Рецензируемый документ состоит из двух глав: «Статистические основы внутрилабораторного контроля качества количественных методов исследования» и «Методика проведения внутрилабораторного контроля качества на основе системы ALTYAY Laboratory QC». Первая содержит общие положения, вторая конкретные рекомендации по работе с программой. Естественно, что интерес представляет анализ главы 1. Чтобы сделать его более последовательным, рассмотрим сначала, как оценивается точность путём обработки массивов данных пациентов, а затем по контролям.

### **Контроль правильности по ежедневным средним.**

Оценка по данным пациентов надёжна и малозатратна, поскольку не нужна никакая дополнительная аналитическая работа, надо просто правильно обработать достаточно большой массив уже имеющихся данных.

К этой категории относится метод «Средних нормальных величин» (“Average of Normals” AON) предложенный Hoffman and Waid в 1965 г. [8]. Вариант которого предлагают использовать авторы программы.

Текст	Комментарий
<p><b>1.8. Метод контроля правильности по ежедневным средним</b></p> <p>Принцип метода состоит в ежедневном расчете средней арифметической всех результатов определения данного показателя, полученных в лаборатории за день.</p> <p>Последовательность действий:</p> <ol style="list-style-type: none"> <li>1. Ежедневно из полученных в течение дня результатов проводится расчёт ежедневной средней арифметической величины <math>\bar{X}</math>. Эта процедура повторяется в течение 20 дней.</li> <li>2. Далее из 20 ежедневных средних проводится расчёт общего среднего <math>\bar{X}_{общ}</math> и среднего квадратичного отклонения <math>S</math>.</li> <li>3. Рассчитываются контрольные пределы <math>\bar{X}_{общ} \pm 1S</math>, <math>\bar{X}_{общ} \pm 2S</math>, <math>\bar{X}_{общ} \pm 3S</math> и строится контрольная карта.</li> <li>4. После построения контрольной карты в лаборатории ежедневно рассчитывается <math>\bar{X}</math> из всех результатов каждого анализируемого показателя, и полученное значение наносится на контрольную карту в виде точки.</li> <li>5. Анализ контрольной карты проводится по выше описанным правилам Westgard.</li> </ol>	<p>Метод основан на очевидном факте – результат анализа случайная величина, кривая плотности распределения которой одинакова во все дни (если, конечно, контингент пациентов не меняется). Поэтому если средняя величина неожиданно возрастает или уменьшается, возникает сомнение в правильности результатов. Трудность практического применения этого метода в том, что, тяжёлые больные с очень высокими показателями (во много раз выше нормы) бывают не каждый день, средняя величина в эти дни значительно возрастает, что может быть ошибочно трактовано, как признак плохой работы. В данном случае средняя величина используется как критерий (далеко не лучший) характера распределения результатов.</p> <p>Раздел написан очень бестолково - п.5 рекомендует проводить анализ по правилам Westgard, что фактически невыполнимо, поскольку это правила для работы с контрольными анализами, а не со средними. Поэтому остаётся неясным как все же рекомендуется оценивать результаты.</p>

Оценку правильности анализов по характеру распределения результатов, следует решать, опираясь не на среднюю величину, а на критерий Колмогорова-Смирнова. Этот критерий позволяет оценить вероятность того, что две выборки сделаны из одной и той же генеральной совокупности. Если эта вероятность велика (больше чем 0,05) можно спать спокойно – аналитическая система сегодня работает также, как она работала раньше. Если же вероятность мала, надо искать причину. Если это произошло потому, что изменился контингент обследуемых, должны измениться многие параметры: соотношение мужчин и женщин, возрастной состав и т.д. В

противном случае, следует полагать, что методика хромает. Удивительно, что в 1965 г., когда критерий Колмогорова уже был во всех справочниках, нашлись люди, которые предложили оценивать изменение характера распределения по средней величине - косвенному, малоинформативному показателю. Но ещё удивительней, что через 50 лет эта же нелепость повторяется Российскими программистами ООО «Группа Алтэй», которые реанимируют метод АОН к тому же в безграмотном исполнении! Казалось бы, программисты должны быть знакомы с азами теории вероятности и понимать, какие критерии пригодны для оценки изменения характера распределения, какие нет.

Для оценки точности по колебаниям среднего данных пациентов, существует и другие методы, но они в рецензированном материале не фигурируют, поэтому и мы их не рассматриваем.

### **Внутрисерийные и межсерийные погрешности, воспроизводимость и правильность**

Недостаток метода оценки качества работы по пробам пациентов очевиден – он даёт лишь относительные данные - имеется ли дефект аналитической системы, абсолютная величина погрешности остаётся неизвестной. Чтобы её узнать, надо сделать две вещи 1) проанализировать специальные контрольные образцы, состав которых известен, и 2) обработать данные, опираясь на представление о природе возможных погрешностей.

Всякий аналитик, приступая к работе, имеет гипотезу о том, какие ошибки возможны, даже если он и не думает об этом. Кто бы стал делать измерение, если бы считал, что никакой закономерности нет, любая ошибка, и, стало быть, любой результат возможны? Представления авторов программы по этому вопросу изложены в пунктах 1.1.1 и 1.1.2.

Текст	Комментарий
<p><b>1.1.1. Воспроизводимость</b></p> <p>Воспроизводимость – качество измерений, отражающее близость друг к другу результатов измерений, выполняемых в различных условиях (т. е. степень совпадения результатов повторных измерений одной и той же пробы).</p> <p><b>1.1.2. Правильность</b></p> <p>Правильность – качество измерений, отражающее близость к нулю систематических погрешностей в результатах.</p> <p><i>Правильность измерений характеризует наличие систематических погрешностей.</i></p> <p>Систематическая погрешность – это составляющая погрешности измерения, остающаяся постоянной или закономерно изменяющаяся при повторных измерениях одной и той же величины.</p>	<p>Авторы различают воспроизводимость и правильность, ошибочно полагая, что первая зависит от случайных ошибок, а вторая от систематических.</p> <p>Если лаборатория соблюдает правила выполнения анализов, регулярно и тщательно выполняет калибровки, постоянная систематическая ошибка должна быть исключена на этапе калибровки. Если, тем не менее, она присутствует, значит либо калибровка плохо сделана, либо калибровочный материал не соответствует контрольному. Вторая причина более вероятна, тогда надо решать, кому верить, какой материал правильный. Но, мир не совершенен, нельзя требовать, чтобы концентрации калибровочного и контрольного материалов всегда абсолютно соответствовали друг другу, возможен и такой вариант, когда правила работы допускают определённую погрешность на этом этапе. В этом случае надо оговорить её максимальный предел. Тем самым систематическая ошибка (смещение или параметр В по терминологии авторов ) отражает в первую очередь качество калибровочных и контрольных материалов, хотя, конечно, может быть и следствием небрежной работы.</p>

На самом деле правильность – т.е. то, насколько результат анализа отражает реальную концентрацию аналита, зависит от существа метода, а не от того, как он выполнен. Лаборант, выполняющий анализ по определённым правилам, строит калибровку и отвечает за неё, а не за правильность калибратора. Это не означает, что можно использовать негодные калибровочные материалы, но надо понимать, что практическая лаборатория может улучшить правильность, только внедрив, более совершенный метод анализа. Случайные ошибки бывают межсерийные и внутрисерийные. Межсерийная ошибка или погрешность, постоянна для всех анализов данной серии, но меняется от серии к серии, внутрисерийная каждый раз своя. Различие между результатами повторных анализов, выполненных одновременно, всегда меньше, чем, если их делают в разные дни (в разных аналитических сериях). Поскольку систематические ошибки (смещение) должны быть устранены на стадии калибровки, главный источник неточности это случайные межсерийные погрешности. Надо заметить, что судя по всему, авторы понимают это различие, поскольку в словаре приведены термины внутри- и межсерийной воспроизводимости, а также внутри- и межсерийной вариации, Видимо, воспроизводимость и вариация это одно и то же. С нашей точки зрения обилие схожих по смыслу терминов неоправданно, поэтому мы будем всегда говорить только об ошибке или погрешности. Обычно межсерийная погрешность больше внутрисерийной, она вызвана в первую очередь изменением свойств реактивов. Но это лишь грубая схема, поскольку нет гарантии, что межсерийная погрешность постоянна на протяжении нескольких часов, пока выполняется данная серия анализов.

Фактически воспроизводимость, как она определена в п. 1.3, это и есть точность – тот параметр, который в конечном итоге определяет качество работы. Тут два вопроса - какая нужна точность, и как её оценивать. Очевидно, что чем выше точность, тем труднее её достигнуть, в тоже время она должна быть достаточной, чтобы можно было обнаружить спонтанные колебания исследуемого параметра. В своё время Тонгс [3] разделил все возможные биологические колебания (как пишут авторы документа «вариабельность») на две группы – межиндивидуальные и внутрииндивидуальные, понимая под первыми различия между разными здоровыми людьми, а под вторыми колебания во времени у одного и того же индивидуума. По умолчанию предполагается, что материал для исследования всегда берётся в стандартных условиях - утром, натошак и т.д. Тонгс высказал пожелание, чтобы точность анализа была не хуже, чем 1/8 внутрииндивидуальных колебаний. Этот же принцип перекочевал и в рецензируемый документ.

С этого места документа начинается совершенно недопустимое жонглирование терминами и понятиями. Свойства  $\sigma$  - корня квадратного из дисперсии, без всякого основания и объяснения переносятся на CV - коэффициент вариации – т.е. среднюю квадратичную, делённую на среднюю концентрацию. Уже это одно делает программу ООО «Группа Алтэй» непригодной для работы.

Текст	Комментарий
<p><b>1.3.1. Требование к воспроизводимости лабораторных результатов</b></p> <p>Содержание или активность ряда компонентов в организме, может значительно меняться у здоровых людей в течение суток или в течение месяца. Для большинства компонентов не существует точно определённых ритмов, и имеют место случайные вариации вокруг определённых значений, что и является индивидуальной характеристикой организма. Такая флуктуация</p>	<p>Чтобы, руководствуясь этими советами , разработать количественные критерии точности, надо, как минимум, знать величину внутрииндивидуальных колебаний. Неплохо было бы также выяснить, зависят ли они от пола и возраста? Это серьёзная научная проблема, которая в данном документе не получает решения. В статистической науке нет такого понятия «среднее значение коэффициента вариации», усреднение ведь можно проводить по разному. Следовало бы привести алгоритм вычисления и</p>

<p>вокруг определённого значения компонента называется <b>внутрииндивидуальной биологической вариацией</b>.</p> <p>Внутрииндивидуальная биологическая вариация устанавливается следующим образом. Выбирается определённое количество здоровых людей. Отбираются у них пробы крови в течение примерно 10 недель ежедневно. Затем одновременно все пробы анализируются, и на основе полученных результатов вычисляется среднее значение коэффициента вариации <math>CV_{\text{внутриинд}}</math> и диапазон его изменения.</p>	<p>показать сходимость результата. А что такое «диапазон изменения CV»? Как он вычисляется? И, вообще, при чем тут коэффициент вариации?</p>
--	--

Текст	Комментарий
<p><b>1.3.2. Выражение для дисперсии при расчёте общей (групповой) биологической вариации</b></p> <p>Биологическая вариация имеет два компонента: внутрииндивидуальную и межиндивидуальную, соответственно, можно записать выражение для дисперсии при расчёте общей (групповой) биологической вариации:</p> $S_b^2 = S_{\text{внутриинд}}^2 + S_{\text{межинд}}^2 \quad (1.7)$ <p>где <math>S_{\text{внутриинд}}^2</math> – дисперсия при расчёте внутрииндивидуальной биологической вариации, <math>S_{\text{межинд}}^2</math> – дисперсия при расчёте межиндивидуальной биологической вариации.</p>	<p>Тут появилось и ещё несколько раз используется термин межиндивидуальная биологическая вариация, но нигде не сказано, что это такое и как её вычислять.</p> <p>Формула (1.7) верна только, если внутри - индивидуальные колебания у всех людей одинаковы, что весьма сомнительно.</p> <p>Сопоставив тексты параграфов 1.3.1 с 1.3.2 и 1.3.3 можно понять, что авторы хотят получить две величины</p> <ol style="list-style-type: none"> <li>1) дисперсию результатов многократных повторных анализов, выполненных одному и тому же лицу.</li> <li>2) дисперсию результатов большого числа анализов выполненных разным лицам.</li> </ol> <p>Первая величина используется для оценки разброса, вторая смещения.</p>
<p><b>1.3.3. Требование к правильности лабораторных результатов</b></p> <p>Рабочая Группа Европейского Комитета рекомендует в качестве требования для правильности использовать следующее соотношение:</p> $B \leq 0,25 * \sqrt{CV_{\text{внутриинд}}^2 + CV_{\text{межинд}}^2} \quad (1.8)$ <p>т. е. максимально допустимое аналитическое смещение в % при незначительной невоспроизводимости должно определяться соотношением (1.8).</p> <p><b><i>B</i></b> – это целевое (т. е. математически ожидаемое при бесконечном числе исследований) значение аналитического смещения</p>	<p>Формула (.1.8) получена совершенно недопустимым методом – путём замены оценок дисперсий на «оценки» коэффициентов вариации. Просто средние квадратичные разделили на средние. Конечно, бумага все терпит, но полученный показатель ни о чем не говорит – в разных сериях опытов он будет разным, соответственно и допустимая величина смещения (т.е. ошибки) будет меняться.</p> <p>В этом разделе вводится нормируемый параметр <i>B</i> – смещение, т.е. систематическая ошибка. Кроме того нормируется и другой параметр – случайная ошибка (воспроизводимость).</p>

Нет никакой необходимости в двух нормируемых параметрах точности - разбросе и смещении. Вполне достаточно одного - погрешности измерения, которая характеризуется расстоянием между истинным значением и результатом измерения. Пациенту ведь важен не механизм возникновения неточности, а её размер. Не надо

забывать, что разнонаправленные ошибки могут компенсировать друг друга, поэтому оба параметра могут быть вне пределов, но их сумма допустима.

### Допустимые отклонения результатов контрольных анализов.

Допустим, что авторы программы правы - внутрииндивидуальная вариация всегда одинакова и её оценкой служит дисперсия, величина которой  $\sigma_{\text{вн}}^2$  известна. В этом случае величину допустимой ошибки анализа также можно выразить через критическую величину дисперсии ( $\sigma_{\text{кр}}$ ), приняв, например, что сумма дисперсий внутри- и межсерийной погрешностей должна быть по крайней мере в 4 раза меньше дисперсии внутрииндивидуального разброса

$$\sigma_a^2 < \sigma_{\text{кр}}^2 = \sigma_{\text{вн}}^2/4 \quad \text{или} \quad \sigma_a < \sigma_{\text{кр}} = \sigma_{\text{вн}}/2$$

Вопрос состоит в том, как проверить выполняется ли это условие, в конкретной лаборатории. Для этого авторы программы рекомендуют один и тот же контрольный материал анализировать 10 или 20 раз в разные дни, по этим данным вычислить среднее и среднее квадратичные  $\bar{X}_{10}$ ,  $\bar{X}_{20}$ ,  $S_{10}$  и  $S_{20}$  соответственно. На первый взгляд кажется, что если  $S_{10}$  или  $S_{20}$  меньше  $\sigma_{\text{кр}}$  то требование выполнено. На самом деле ответ сложнее, потому что  $S_{10}$  и  $S_{20}$  всего лишь оценки, сама дисперсия  $\sigma$  может быть меньше или больше. Как же, зная  $\sigma_{\text{кр}}$ , вычислить  $S_{\text{кр}}$  - максимально допустимое среднее квадратичное выборки, которое в документе называется предельно допустимым значением (ППЗ)? Правильный ответ основан на распределении хи-квадрат ( $\chi^2$ ) и предполагает, что случайные погрешности измерения, т.е. сумма внутрисерийных и межсерийных погрешностей, имеет нормальное распределение. Хи-квадратом называется сумма квадратов  $n$  чисел, случайно выбранных из бесконечного, распределённого по нормальному закону, множества:

$$\chi_n^2 = \frac{\sum (x_n - \bar{X})^2}{\sigma^2} = (n-1) \frac{S^2}{\sigma^2}$$

В нашем случае выборка делается из генерального множества всех возможных результатов повторных анализов данного контрольного материала (авторы рецензируемой программы называют это «целевым значением»). Из приведённой формулы следует что  $S_{\text{кр}}$  - максимально допустимое среднее квадратичное выборки

$$S_{\text{кр}} = \sigma_{\text{кр}} \sqrt{\frac{\chi_n^2}{n-1}}.$$

Подобно нормальному распределению, хи-квадрат не может быть выражен через элементарные функции, но его функция распределения известна и представлена в виде таблицы, в ней указаны вероятности, с которыми сумма квадратов выборки из  $n$  чисел меньше заданного значения или равна ему. Так, если выборка состоит из 10 анализов ( $n=10$ ) то хи-квадрат с вероятностью 0,05 равен 3,33 или меньше, если  $n=20$  то с такой же вероятностью хи-квадрат равен 10,12 или меньше. Из определения хи-квадрата следует

$$\sigma^2 = \frac{\sum (x_m - \bar{X})^2}{\chi_n^2} = \frac{S_n^2}{\chi_n^2} (n-1)$$

При прочих равных – одинаковом числе точек и вероятности,  $\chi^2$  и  $\sigma^2$  связаны обратной зависимостью -  $\sigma^2$  уменьшается при увеличении  $\chi^2$  и наоборот. Хи-квадрат случайная величина, деление на неё в общем случае сложно. Чтобы облегчить задачу, фиксируем вероятность, например, принимаем, что она 0,05 или меньше. По таблице находим, что в этом случае  $\chi_n^2$  - сумма квадратов 10 чисел, случайно выбранных из стандартной формы нормального распределения, с вероятностью 0,05 равна 3,33 или меньше, это значит что с вероятностью 0,95 эта сумма больше 3,33:

если  $P \chi_n^2 > 3,33 \approx 0,95$  то с той же вероятностью

$$\sigma = S_{10} \sqrt{\frac{n-1}{\chi_n^2}} \leq S_{10} \sqrt{\frac{9}{3,33}} = S_{10} * 1,644$$

Требования к точности выполнены когда  $\sigma \leq \sigma_{кр}$ , поэтому должно выполняться

неравенство  $S_{10} < \frac{\sigma_{кр}}{1,644} = 0.6083 \sigma_{кр}$ . В этом случае с вероятностью 0.95 дисперсия

меньше заданной величины. Если же в контрольной серии было 20 анализов, то в

случае успеха  $S_{20} < \frac{\sigma_{кр}}{1,3702} = 0.7298 * \sigma_{кр}$ .

Посмотрим теперь, как авторы программы решают эту задачу.

Текст	Комментарий
<p><b>1.4. Методика расчета биологически обоснованных предельно допустимых значений (ПДЗ) аналитических погрешностей</b></p> <p>В реальных условиях проведения внутрилабораторного контроля качества осуществляется исследование не бесконечно большого числа проб контрольного материала, а лишь по одной – две пробы в каждой серии. При построении контрольной карты используются результаты исследований контрольных проб в первых 10 и 20 аналитических сериях.</p> <p>Для расчёта <i>предельно допустимого значения коэффициента вариации результатов, полученных в n аналитических сериях</i> (показателя воспроизводимости исследований) используется формула</p> $CV_n = CV \sqrt{\frac{n-1}{\chi^2}}, \quad (1.9)$ <p>где n – число аналитических серий, CV – целевое (т. е. математически ожидаемое при бесконечном числе исследований) значение коэффициента общей аналитической вариации в %, <math>\chi^2</math> – критерий Пирсона с уровнем значимости 0,05 и числом степеней свободы (n-1).</p>	<p>В разделе 1.4 использована нестандартная терминология, но, видимо, CV это квадратный корень из дисперсии, делённый на среднее генеральной совокупности т.е. <math>\sigma/M</math>, а <math>CV_n</math> это средняя квадратичная 10 или 20 контрольных анализов, делённое на их среднюю т.е. <math>S/X_{cp}</math>. В обычном написании</p> $\frac{S}{X_{cp}} = \frac{\sigma}{M} \sqrt{\frac{n-1}{\chi^2}}$ <p>Если, по счастливой случайности, среднее контрольных анализов равно генеральному среднему (<math>X_{cp} = M</math>), то формула приобретёт вид, похожий на (1.9)</p> $S_{кр} = \sigma_{кр} \sqrt{\frac{n-1}{\chi_n^2}}.$ <p>Эта формула содержит ошибку (или опisku?) Как было показано выше, умножать <math>\sigma_{кр}</math> надо на <math>\sqrt{\frac{\chi_n^2}{n-1}}</math>. Среднее квадратичное выборки обязательно должно быть меньше критической (предельно допустимой) дисперсии. Иначе будет пропущено много заведомо высоких результатов.</p>

Очевидно, что предельная величина  $S_{кр}$  должна быть меньше  $\sigma_{кр}$ . Ведь S это оценка, которая может быть и меньше и больше оцениваемой величины. Если оцениваемая величина меньше оценки это нас в любом случае устраивает, если она больше, то должен быть какой-то резерв.

Как это часто бывает при использовании статистических методов, граница между «хорошим» и «плохим» размыта. Всегда можно ужесточить требования, тогда будет мало хороших и много плохих результатов, или, наоборот, снизить их, и будет мало плохих и много хороших. Но всегда есть пограничная или серая зона, в которой примерно одинаково и тех и других. Оптимально граница должна быть проведена так, чтобы «серая» по формальным признакам зона, включала бы в функциональном отношении полноценные результаты.



Разделы 1.3 и 1.4, как и 1.5 содержат ещё ряд спорных положений, но рассмотренного достаточно, чтобы, не останавливаясь на них, исключить возможность пользования программой.

### **Контрольные правила Ветгарда.**

Раздел 1.6 «Оценка результатов исследования контрольных материалов с использованием контрольных правил» концептуально неверен, к тому же даёт размытые, практически невыполнимые формулировки.

Чтобы оценить результат контрольного исследования или измерения – вычисляют вероятность того, что оно случайно выбрано из генеральной совокупности всех правильно сделанных измерений. Если она мала – обычно меньше 0,05 считается, что такой результат не мог получиться случайно, стало быть, методика хромает. Понятно, что достоверность заключения увеличивается, если одновременно выполнено несколько контрольных анализов, но тут возникают два вопроса:

- где взять генеральную совокупность всех правильно выполненных анализов и
- как вычислять вероятность, когда несколько контролей.

Авторы рецензируемой программы обе проблемы решают неудачно, что делает ее практически неработоспособной.

В качестве эталона – генеральной совокупности правильно выполненных анализов - рекомендуется брать результаты контролей первых 10 или 20 аналитических серий (иногда это называется «установочная серия»). По умолчанию принимается, что распределение нормальное, что, вообще говоря, было бы неплохо проверять. Это типичный пример планирования от достигнутого. Недостаток в том, что если в установочной серии получились данные, более точные, чем это необходимо, лаборатория вынуждена все время «гнаться за собственной тенью» расходуя силы и ресурсы на достижение ненужного результата. При серьезной работе надо исходить не из достигнутого, а из объективно необходимого.

Когда выполняется только один контрольный анализ, вероятность того, что он относится к данному эталонному пулу, оценивается по площади под кривой Гаусса. Часто принимают, что при качественной работе результат контрольного анализа должен укладываться в диапазон  $M \pm 2S$  аттестованного значения, это означает, что при самой тщательной работе 4,54% будут необоснованно признаваться негодными. Число ложных отклонений можно значительно уменьшить, если выполнять два контрольных анализа, в этом случае вероятность что оба будут за пределами  $M \pm 2S$  всего 0,2% . Сложность использования двух и большего контролей связана с тем, что совместная вероятность не может быть оценена по кривой нормального распределения. Тут нужны более сложные вычисления, основанные на распределении хи-квадрат или, иногда, распределения Релея (7, стр. 200). Ведь надо оценивать не каждый контроль отдельно, а все результаты как единое целое. В ручном режиме это делать очень хлопотно, но вполне по силам ЭВМ.

Когда выполняется несколько контрольных анализов, очень важно правильно решить какие контроли к какой серии относятся. Так, например, два контрольных анализа были сделаны утром, в начале работы, а два другие в конце. Если существует уверенность, что за все это время условия анализа не изменялись, можно считать, что была одна аналитическая серия, качество которой оценивается в зависимости от вероятности, с которой могли одновременно получиться все 4 результата контролей. Если же такой уверенности нет, надо считать, что было две аналитических серии по два анализа в каждой. Разумеется, результаты оценок будут другими.

Располагая компьютером, не сложно точно вычислить вероятность любой комбинации результатов контрольных анализов и тем самым оценить качество работы. Без компьютера это практически невозможно сделать быстро, поэтому были разработаны различные «правила», чтобы хотя бы приблизительно оценить эту вероятность. Они сводятся к тому, что количественные данные заменяются качественными – отклонение на  $1,59S$  заменяют на «больше  $1S$ » или вместо  $0,75S$  берут «меньше  $1S$ » и т.д. в результате много полезной информации теряется, получается что-то вроде балльной оценки.

Текст	Комментарий
<p><b>Правило <math>1_{2S}</math></b>: одно из контрольных измерений выходит за пределы <math>\left[ \bar{x} \pm 2S \right]</math>.</p> <p><b>Правило <math>1_{3S}</math></b>: одно из контрольных измерений выходит за пределы <math>\left[ \bar{x} \pm 3S \right]</math>.</p> <p><b>Правило <math>2_{2S}</math></b>: два последних контрольных измерения выходят за пределы <math>\left[ \bar{x} \pm 2S \right]</math>.</p> <p><b>Правило <math>R_{4R}</math></b>: два измерения одной серии по разным сторонам коридора <math>\left[ \bar{x} \pm 2S \right]</math>.</p> <p><b>Правило <math>4_{2S}</math></b>: 4 последних измерения выходят за пределы <math>\left[ \bar{x} \pm S \right]</math>.</p> <p><b>Правило <math>10_X</math></b>: все 10 последних измерений больше или меньше среднего.</p>	<p>Чтобы судить о вероятности события надо знать общее число испытаний.</p> <p>Из текста нельзя понять, сколько контрольных измерений должно быть в серии. А это очень важно потому, что вероятность «срабатывания» правила зависит не только от результатов, которые попали «под него», но и всех других данной серии. Так, если один из контролей вышел за пределы <math>2S</math>, а второй <math>0,5S</math> то совместная вероятность <math>0,02814</math>, а если второй <math>1S</math> то совместная вероятность в два раза меньше <math>0,01447</math>.</p> <p>Тут целый спектр формулировок:</p> <ul style="list-style-type: none"> <li>- одно из контрольных измерений,</li> <li>- два измерения одной серии,</li> <li>- четыре последние измерения,</li> <li>- десять последних измерений.</li> </ul> <p>Как это понимать? Сколько же анализов в серии? Может быть какие-то серии объединяются?</p>

Программа рекомендует начинать с проверки укладывается ли контрольный результат в коридор  $M \pm 2S$ , если укладывается, значит, его вероятность больше  $0,0456$ , работа выполнена хорошо, данные могут быть приняты. Если не укладываются, то проводятся новые тесты. Если один из них не прошёл, значит, вероятность, мала и все результаты анализов бракуются. В таблице приводятся вероятности, которые служат основанием для отклонения

Вероятности результатов контрольных анализов, которые по правилам Вестгарда служат основанием для отклонения всей серии.

Название	Число контролей	Правило (отклонение от среднего)	Отклоняются результаты, вероятность которых
$1_{3S}$	1	Одно контрольное измерение $>3S$	0.00270
$2_{2S}$	2	Два контрольных измерения $>2S$	0.00207
$R_{4S}$	2	Одно контрольное измерение $<2S$ другое $>2S$	0.00104
$4_{1S}$	4	Четыре контрольных измерения $>1S$	0.01014
$10_X$	10	Десять измерений по одну сторону от среднего	0.00098

Как видно из таблицы, одни критерии более жёсткие, другие мягче. Так согласно правилу  $10_X$  отклоняются результаты, вероятность которых меньше  $0,1\%$ , а правило  $1_{3S}$  более жёсткое, оно отклоняет, если вероятность  $0,27\%$ . Данные этой таблицы несколько приукрашивают реальность, поскольку предполагает, что было

выполнено ровно столько контролей, сколько необходимо для проверки контрольного правила, как оно описано. Это нереально, так как заранее неизвестно потребуется ли дополнительная проверка, для которой нужен дополнительный материал, а это может сказаться на всей статистике.

Только из далёкого докомпьютерного времени могли переключаться «Контрольные правила Вестгарда», для работы с которыми нужен ручной труд. Намного разумнее было бы фиксировать число контролей и написать несложную программу, которая сразу выдавала бы вероятность данной комбинации результатов контролей.

### **Контроль по дубликатам.**

Раздел 1.8. «Метод контроля правильности по ежедневным средним» и раздел 1.9. «Метод контроля воспроизводимости по дубликатам» такие же реликты каменного века обработки данных, когда единственным «средством автоматизации» были деревянные конторские счёты. О методе ежедневных средних AON- «Average of Normals» уже было сказано.

В разделе 1.9 вместо средней квадратичной используется абсолютная величина разности, видимо раздел переписан из старого руководства, того времени когда не было компьютеров и вычисление средней квадратичной было проблемой. Правильное решение следующее.

Мы принимаем, что погрешности распределены по нормальному закону с математическим ожиданием 0 и неизвестной нам дисперсией  $\sigma^2$ . Повторные анализы можно представить в виде  $X=M+a\sigma$  и  $Y=M+b\sigma$ , где  $a$  и  $b$  случайные Гауссовы числа. Разности  $X-Y=(a-b)\sigma$  также распределены нормально с дисперсией  $2\sigma^2$  и средним 0.

Величина  $S^2 = \frac{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 \dots}{n-1}$  есть наилучшая оценка  $2\sigma^2$ . Отсюда средняя

квадратичная дубликатов  $\sigma = \sqrt{\frac{S^2}{2}} = \sqrt{\frac{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 \dots}{n-1}}$ . Её и следует

использовать для построения положительной и отрицательной границ 95% и 99% квантилей.

Этот раздел основывается на тех же принципах, что и раздел 1.6 «Оценка результатов исследования контрольных материалов с использованием контрольных правил» и претензии к нему такие же. Надо оценивать достоверность не каждой отдельной пары дубликатов, а сразу всех пар данной серии, обязательно надо принимать во внимание и нормальные результаты, т.е. те, которые не вышли за пределы 95% или 99% квантилей. В противном случае оценки будут в лучшем случае весьма приблизительными.

### **Заключение.**

Создаётся впечатление, что комментарий «Статистические основы внутрилабораторного контроля качества количественных методов исследования» к одноименной программе ООО «Группа Алтэй», представляет собой компиляцию нескольких документов, главным образом Отраслевого стандарта 91500.13.0001.2002 «Правила проведения внутрилабораторного контроля качества количественных клинических лабораторных методов исследований с использованием контрольных материалов». Авторы, видимо, не очень хорошо понимают задачу и, как это ни удивительно, возможности вычислительной техники. Контроль качества в любой сфере деятельности основывается на оценке вероятности того или иного события. Однако в обсуждаемой программе для этого ни разу не используется процессор, а все оценки делаются «на глазок» по контрольным правилам. Компьютер используется

только для рисования графиков и лишь изредка для самых простых вычислений. Даже если отвлечься от многочисленных ошибок и неточностей, такой подход иначе, как тупиковым не назовешь.

Идея взять за единый эталон точности размер внутрииндивидуальных колебаний выглядит как вполне здравая, непонятно только осуществима ли она. Ведь размер этих колебаний неизвестен и вычислительный механизм для его оценки не разработан. В некоторых документах рекомендуется опираться на величины внутрииндивидуальных колебаний, которые были опубликованы 10 лет назад группой испанских исследователей [9], но в России никто их не проверял. Помимо того, что этот источник в России практически недоступен, возникает вопрос можно ли на него полагаться? Насколько данные обоснованы и пригодны ли для использования в совершенно других условиях?

С нашей точки зрения, главный концептуальный недостаток программы состоит в том, что делается попытка найти универсальный критерий, не исследовав природу возможных погрешностей и их размер. Разумней было бы сначала разработать надёжные способы измерения величины ошибок, а затем уже решать какие ошибки и в каком случае допустимы. Поскольку природа ошибок может быть разной, возможны разные способы оценки их величины, нельзя исключить и возможность разработки новых подходов.

В свое время мы предложили три разных алгоритма оценки вероятных погрешностей по данным контрольных анализов [10]. Выбор метода и достоверность оценок зависят от объёма имеющейся информации. Если много контрольных анализов, выполненных в разные дни, можно использовать метод совместного распределения контрольных и опытных проб. Он основан на решении следующей задачи – параметры нормального распределения неизвестны, но известны  $n$  экземпляров случайной величины, найти параметры распределения  $n+1$  экземпляра. Если контрольных анализов мало - всего 2-5, подходит модифицированный метод Стьюдента. Он позволяет найти доверительный интервал единичного измерения, в то время как оригинальный даёт доверительный интервал математического ожидания. Наконец, когда имеется только один результат контрольного анализа и известна дисперсия внутрисерийной погрешности, можно воспользоваться теоремой Байеса.

#### Литература

1. Балаховский И.С. Границы нормы и точность лабораторных анализов. Клиническая лабораторная диагностика 2007, №10, стр. 47-54.
2. Управление качеством клинических лабораторных исследований. Нормативные документы. Москва Лабпресс 2000 г.
3. Гаранина Е.Н. Качество лабораторного анализа. Москва. Лабинформ. 1997. стр. 41.
4. Меньшиков В.В., редактор. Качество клинических лабораторных исследований. Москва. Аналитика 2002.
5. Меньшиков В.В., редактор. Стандартизация в клинической лабораторной медицине. Москва. Международный общественный благотворительный фонд милосердия и здоровья. 2005.
6. Отраслевой стандарт 91500.13.0001.2002 . «Правила проведения внутрिलाбораторного контроля качества количественных клинических лабораторных методов исследований с использованием контрольных материалов».
7. Вентцель Е.С. Теория вероятностей. 7 издание. Москва. Высшая школа 2001 г. Стр.318.
8. Gembrowski G.S., Elliot P.C., Westgard J.O. Assessment of “Average of Normal” Quality Control Procedures and Guidelines for Implementation/ Am. J.Clin.Pathol. 1984, **81**: 492-400.

9. Sebastian –Gembaro M.A./Liron-Hemmandez F.J., Fuenntes-Arderin X. Intra and inter Individual Biological Variability Data. Eur.J. Clin. Chem. Clin. Biochem. 1997, 35, 11, 845-852

10. Балаховский И. С. Введение в статистический контроль качества клинических лабораторных анализов (в редакции февраля 2009 г.)

<http://www.clinlab.ru/win/Qualit/intr.htm>

<http://www.clinlab.ru/win/Qualit/Intrroduction-2009.pdf>

Несколько сокращённый вариант «Клиническая Лабораторная диагностика» 2009 г., № 10, стр. 30-37 и №11, стр. 33-42.

Балаховский И.С. [\*\*igbalakh@yandex.ru\*\*](mailto:igbalakh@yandex.ru)

Москва, тел. 495-432-33-20