Some Notes About Empty QC

Hassan bayat; CLS; Sina LAB; November 25, 2012

Recently I read the article “Empty QC” by Roy Midyett on the MLO website. I want to give some points about that; but let me give a short explanation about QC before the points.

An Example
Suppose two performances: making a road by a company, and a student studying to attend in an inter-schools physics race. And suppose we are responsible to assure that the results of these activities will be “well”. To do this, we have to go through some steps.

First, we have to ask our self: Are the performances being done well (i.e. has quality)?

To answer this question, the key is to know what the meaning of the “well” is. I.e. we have to know quality requirements.

Let’s suppose that in the case of making the road, “well” means that the performer finishes the work by 6 months; and “well” for the student is getting “A” mark in the race.

In the next step, we have to “evaluate” the performances for a while to see how the works are being run, and get knowledge about the characteristics of the performances. I.e. we have to do method validation. After this initial evaluation, we might get one of these conclusions:

1- No. This way that the company works, it can’t end the road on time. And in the case of the student: This way that this student studies, he can’t get “A” mark.
In this situation, the performers are failed. They should improve their performance and then be re-evaluated again, or be substituted with other performers for satisfying the intended quality.
2- Yes. This way that this company works, it can. And in the case of the student: This way that this student studies, s/he can.

Note when we conclude that the performances will be good in the future, we emphasize that if the performance is continued “the way is being done now”. Meaning that provided that the observed performance at present remains stable in the future, we can assure the quality of performance in the future.

What happens then? Is this the end?

No; it’s just the beginning! Now we have to sit down and think of different factors that could hurt “this way” of working; i.e. we have to find “risks” that threaten our stable performance. There are different risk factors that a company, and also a student, might face.

After thinking, and thinking, and thinking, and determining probable risk factors, we have to address ways to preventing them from happening. If, and it is a so big if, we can be completely sure that 1) we have known all the risks, and 2) we have provided impermeable obstacles so that the probability
of happening of those risks has reduced to zero, then: Congratulations! We have killed all of the spooks, and there is no other thing to do for assuring of quality. The observed quality during validation period will be presented in the future by complete sure.

But, can anybody in any situation be so sure?

Regardless of how completely we have performed risk assessment, and regardless of how robust obstacles we have provided, always some risk residues because we can’t fully know or provision all the risks and/or we can’t provide ideal guards. This residual risk demands that we to provide some kind of alarms so that when defects are close to happen or in a reasonable short time just after happening, the alarms ring and alert us.

The last step: We have to in a regular and systematic approach evaluate whole the activities, including the performance itself and the alarms, to find answers for questions like these:

- Is there any opportunity to improve the performance; and therefore lower the sensitivity of the alarms?
- Is there any opportunity to remove or reduce some of the residual risks; and again lowering the alarms’ sensitivity?
- Did the alarms ring when there were defects?
- How many times did the alarms ring while there were no defects?

Answers that we would get during these “Quality Assessments”, makes it possible to improve both the performance and the alarming system.

The process explained above can be categorized as:

- Determining the meaning of “well”
- Initial evaluation to find performance characteristics
- Judging whether the performance during “evaluation period” could satisfy the well criteria
- (In the case that performance passes the initial evaluation) Doing some works to be assure that “this kind of acceptable performance” will be stable and continued in the future, including:
  - Evaluating the risks that might deteriorate the performance
  - Find ways to obstacle them completely
  - And if any risk residues, assigning alarms that when defects are going to happen or just after happening could ring
- Periodic evaluation of the performance to find opportunities for improvement and reducing from residual risk; and also evaluating the capabilities of the alarming system for ringing when there were defects, and not ringing falsely when there were no defects.

**An Example From Laboratory**

The steps named above must be gone through for assuring quality of any performance including testing in the labs. Suppose an example: We want to assure quality of a new glucose kit (Note that we are only talking of quality related to the reliability issue). Let’s start:
Determining the meaning of “well” or “Quality Requirements”:

Reliability related quality has two parts:

1) How far are the patient’s results allowed to be away from the true values?
   Let’s use CLIA’s acceptability limits for assaying glucose; i.e. ±10% or TE_a = 10%.
2) How many of the results should be in the acceptability limits? Traditionally (and insufficiently!) it’s been accepted that at least 95% of the results should be in the acceptability limits.

In simple words: We would be testing glucose well, if at least 95% of the patient’s results are at most 10% away from their true values.

Initial evaluation to find performance characteristics:

Now we have to perform method validation experiences to find performance’s different characteristics, including bias (B) and CV. Suppose we get these: B = 1%, CV = 2%.

Judging whether the performance during “evaluation period” satisfies the well criteria:

To do this, we have to calculate that in what region is 95% of glucose results. It can be calculated simply by total error formula:

\[
TE = B + 2CV
\]

\[
TE = 1 + 2 \times 2 = 5
\]

This means that 95% of the glucose results produced by this kit are at most 5% away from true values. It seems acceptable.

Judging about the future of the assay:

Though the assay passed method validation, but this “passed mark” is just for the “past” time of validation; what can we judge about using this kit in the future? This bigger judgment demands having knowledge of risks, ways to remove or mitigate them, alarms that can ring when there are defects, and also practicability and expenses of risk mitigation and alarm assigning. Then we can judge whether the performance would be stable in the future, i.e. continued “this way”, or not.

Evaluating the risks that might deteriorate the performance:
As we know there are a lot of risks for our assays’ qualities: Electrical power instability, different abilities of operators, minor silent fail of devices, reagent deterioration, and so on.

Find ways to obstacle them:
Fortunately, we are equipped with some powerful barriers in front of defects: Electricity stabilizers, SOPs, device function checks, refrigerated analyzers, and so on.
And if any risk residues: Assigning alarms that when defects are going to happen or just after happening would ring:

By sure, unfortunately, some risk residues. Here we need alarms.

What kind of alarms can help us?

Statistical Quality Control (SQC) is the alarming system that can help us in many times. So important to note is that SQC is not just a “2SD ring”. It has different rings suitable for different level of performances; the better the performance, the easier and cheaper the alarms.

Where to buy the rings? Believe or not, a lot of them are free! We can get a lot SQC rings with detailed instructions and accessories by no charge in the Westgard website.

Can we find suitable SQC for every acceptable performance?

For a lot of the performances: Yes; but not for any acceptable assay! If we refer to Westgard and present our performance, we might be offered one of these example options:

- Your assay’s performance is so excellent that you can control it with just the least trouble and expanse; 1:4SD ring with 2 controls per run is enough for that.

- Though your assay’s performance is not excellent, but good enough. To control it you need more demanding SQC and also more “non-statistical” controls; take 1:3SD/2:2sd with 2 control per run and program more preventive maintenance, function check and so on.

- Your assay’s performance is marginal. Unfortunately controlling it would be so hard. You need multirole SQC using 6 controls per run. Also assign a very qualified operator, increase non-statistical QC activities, and, and, and.

And sometimes we may have such a conversation:

- Why! My assay’s TE is 9, smaller than TEa; why can’t I use SQC?

- Though your assay is acceptable, but it has so poor performance and is so close to the acceptability borders that remains no space for an alarming system.

- So, what can I do to control it?

- You have to use as hard as possible SQC just as a minor help; instead you have to mainly establish QC of this assay on the non-statistical activities. You’ll have a so hard time with this assay!

- Is there any easier option?

- Yes. Not to say change, but THREW AWAY this fragile assay and use a better assay!
After completing this step, we can judge about to run this assay in the lab or not. If judged to run, then we have to write QC program compromising of statistical and non-statistical actions.

- **Periodic evaluation:**
  In a regular program, e.g. every 3 months, we have to do Quality Assessment to see:
  - Is it possible to reduce bias and/or CV?
  - Is it possible to remove or reduce some residual risks?
  - Was our SQC alarm able to aware us from whole, or at least most, of the problems? Is it necessary and/or possible to change the current SQC with a more demanding one that is more sensitive?
  - How many times did SQC ring falsely? How much time, effort, and money were wasted because of these false alarms? Is it necessary and/or possible to change the current SQC with a more relaxed while less sensitive one?

Through these quality assessments we can have a comprehensive picture of our activity, and when needed and/or possible, improve the performance of the assay or the “quality of the quality control”. After any improvement, we have to go to method validation and run the cycle from the first.

Some Points about the “Empty QC” article

Midyett complains “statistical tools used are entirely arbitrary. We have decided that 2SD is a good tool to use to monitor QC, but there is nothing inherent in nature that says this is so”. Of course, using 2SD rule as unique alarming device for all assays, irrespective of how good or bad is the performance, is absolutely incorrect. As Professor Westgard has evangelized since long time before, any specific performance needs its specific QC that is based on the quality requirements and performance characteristics; what he calls “Doing right QC right”. Depended to the situation: 2SD might be suitable for one performance; might be so excessive QC giving many false rejections with performances that are good and better than good; and even might be insufficient and ineffective in finding defects with performances that are poor; situation where SQC is not helpful at all.

2SD rule is one of the stringent SQC rules with a high false rejection probability. If the QC program is planed using right QC planning, 2SD rule will be devoted just to week performances, and not to today’s 4th and 5th generation of analyzers.

**We Use 2SD Rule Without Any False Rejections, And Also Without Any True Alarms. Why Is It Such An “Empty QC”?**

May be we are doing “wrong QC wrong”. If we are using 2SD limits without knowing what the quality requirements for that specific test is, and without method validation, we have accepted that any level of performance is well. It’s a critical mistake. For example, we’ve put in run a glucose assay with a big CV and bias so that a lot of the patient’s results are more than 20% away from their true values. Then, irrespective of the bias, we have used that bad SD of the assay to draw 2SD limits. After that, we are comparing our bad performance at present with our bad performance in the past! What should we expect from such a QC? It’s not surprising that when physician judgment and patient’s clinic show that
the glucose result is incorrect, our control results are in the limits; the patient’s clinic is out of control while our QC is in control!

As Midyett says using 2SD “arbitrarily” is “empty QC”. But, if we use QC wrong that’s our fault not QC. Some people, at the top Professor Westgard, from long years ago have worked in the QC field and empowered the field a lot. There are different QC rules suitable for different levels of performance. “Arbitrary” use of QC, especially 2SD rule’ belongs to far past.

**What’s Right Interval of Assaying Control Materials?**

“We have decided that running QC every eight hours is reasonable in most cases. ... [Even running QC material before and after every sample] wouldn’t suffice”.

It’s the method’s stability issue. What interval do we offer to an apparently healthy young person for repeating his or her glucose test? What about a diabetic patient that has been accepted to hospital with hyperglycemia?

Doing right QC needs considering the stability of the assays. When planning QC program for a specific assay, we have to refer to the history of the assay, consider whether the device is a new fresh analyzer or a sick one, is the device in its ideal situation or it’s being used after a repair, and so on; and based on these factors decide suitable interval for testing controls.

As Midyett says “running QC every eight hours” is not suitable for all assays. The same way that testing glucose “every eight hours” is not suitable for all patients; suitable for some patients, excessive for some, and insufficient some others.

**Primary Methods**

“Primary Methods are methods against which other secondary methods are measured and are tests which are inherently simple and/or rely solely on the expertise of the trained tech performing them, and as such cannot be quality controlled. They are the source of QC itself”.

It seems that two definitions are combined: Primary reference methods and simple methods.

Not needed to explain that all primary reference “methods against which other secondary methods are measured” aren’t “inherently simple and/or rely solely on the expertise of the trained tech performing them”; and also not all “cannot be quality controlled”.

Even simple methods that “rely solely on the expertise of the trained tech performing them” aren’t always done by well trained people. Here is where we need some form of QC to assure that these primary activities are done correctly in the hands of operators other than experts.

Also simple methods such as manual cell counting, manual ESR, and manual hematocrit need QC. There are different risks for these assays; it’s possible that the operator doesn’t fix cover glass firmly on the hemocytometer slide, multiplies the number by an incorrect factor, ESR stand is somewhat tilted, the temperature is incorrect, or some part of the hematocrit device has been deteriorated. If, and just if, after risk assessment we assign procedures that assure us all the risks are prevented completely, then
we don’t need any QC. But if any risk residues, though simple risks such as mentioned ones, we have to hire some QCs.

“There is nothing more fundamental than the manual hematocrit itself that could be used to QC those manual readings”. It is a calibration issue not QC: Basic fundamental assays are the reference for calibrating other assays, but even these assays could and should be quality controlled to assure that their performance remains stable during the time.

**Quality Assessment**

“It may, however, fit one definition of insanity—doing the same thing over and over again and expecting different results—we run a QC procedure month after month, year after year, and expect it to do something different than it did in the preceding months and years.”

That’s right. We never have to continue without assessing the past to improve future. If we do quality assessment, when needed, we will substitute our QC program with a better one, and then we can “expect it to do something different than it did in the preceding months and years”.

**No Doubt; When QC is Out, The Results Shouldn’t Be Released!**

If “The ER doctor needs that result badly”, it doesn’t mean the doctor need bad results. We have to be sure of the quality of the results; especially when the results are needed badly, because they are going to be used for making more critical decisions. If “the QC will not make any difference in whether or how we report the result”, it’s the reason that our QC program is wrong and must be changed with a correct one. It’s the time to have an urgent Quality Assessment to repair QC, not reason to deny the usefulness of QC at all. If our fire alarm doesn’t work correctly, we have to repair it get or change it with a suitable one not ignore fire alarms.

**When Can’t Jump High: Lower the Bar**

“We know some change is possible. CAP, for instance, combined BANDS with SEGS when it couldn’t get a consensus from customers.”

Unfortunately, this is kind of lowering the bar; accepting the customers’ abilities as well performance and not trying to improve the quality via improving customers’ abilities. No doubt, differentiating BANDS from SEGMS is critical for patents’ care. If different customers, meaning labs, give highly different results for this test, it’s a bad news for clinicians and patients; It’s needed something be done for those real customers, not the acceptability criteria be relaxed to make service providers, meaning us, happy.

On the other hand, this is an example of the tests that “rely solely on the expertise of the trained tech performing them”. As this experience shows, such simple tests are performed by operators with a quality that is poorer than performance of experts. We need to use QC programs that could assure all operators perform differential counts as well as experts do.

**The End Words**
May be sometimes we do wrong and arbitrary QC, what Midyett exactly calls “empty QC”, but it’s our fault, not the QC. QC’s toolbox is not so empty; there are lots of correct and scientific tools. And, If there is any incompleteness, it’s up to “the industry, meaning us,” to try to fill those holes.

I agree that sometimes it’s so hard to do right QC, nevertheless, the same way that Midyett says, “to echo a theme from the classic movie Jurassic Park, just because we can do QC doesn’t mean we should”, I believe that: Just because we can’t do right QC doesn’t mean we shouldn’t.