Panel discussion: how to monitor and minimize variation and mistakes

DAVID L. WITTE and MICHAEL L. ASTION

Panel II of the 1996 Clinical Chemistry Forum was a discussion initiated by brief formal presentations from Rosemary Bakes-Martin of the US Centers for Disease Control and Prevention (CDC), Cornelia Rooks of the US Food and Drug Administration (FDA), William Hamlin of the College of American Pathologists (CAP), Nathan Gochman of Beckman Instruments, and David Witte of this Forum Committee.

Initial Presentations

Rosemary Bakes-Martin, CDC. The CDC concept of quality control (QC) in 1996 includes the problematic phrase, “Run two levels of control per day.” The process for analytical QC should both assess the potential for error and monitor the process over time. In assessing the potential for error, laboratorians need to look at the test system, the environment, and the operator.

Alternatives need to be considered. Nontraditional ways to assure that the test system is under control, or is producing the correct result, must be considered. The manufacturer builds systems to monitor equipment and environment. A good training program assures that the operator knows what to do with these systems. Competency testing and quality assurance (QA) help ensure that the operator does the correct thing. No matter how training, competency, and QC are accomplished, it is important to monitor all three over time. The laboratory director probably is the best person to determine how often to monitor the test system, environment, and operator.

Because we at CDC are in a regulatory environment and there must be minimum standards, we believe there must be something done to monitor an analytical process at least weekly.

Cornelia Rooks, FDA. The FDA has been involved in the premarket evaluation of in vitro diagnostic devices for >20 years. Review is directed at establishing the safety and effectiveness of products before marketing through independent review of performance data and labeling. FDA’s review of QC is grounded in its unique labeling regulation and is directed at providing information on the proper understanding and use of both external QC samples and device components, so-called procedural or process controls. The FDA “Points to Consider” document for review of calibration and QC labeling for in vitro diagnostic devices was published in February 1996.

The FDA currently requires that package inserts of all devices contain information on the QC materials appropriate for a test system. In addition, any internal, electronic, reagent, or process control that is an integral part of the device must be clearly described and the nature of its use clearly explained. Recommended QC-specific rules, including the run frequency to use for assessing quality, are left to the discretion of the individual laboratory.

The FDA considers that QC procedures encompass the set of laboratory materials and processes to (a) monitor the performance of laboratory analytical systems with regard to reagents, instruments, and operators; (b) monitor the precision and accuracy of the test procedure; and (c) assure that proper testing conditions and instructions have been met. To comply with these requirements, package inserts must include information on QC.

QC of in vitro devices can be characterized in two very different but complementary components: external QC samples and device QC components (i.e., process or procedural controls), which can complement or augment external QC samples. The FDA requires that all device QC components be clearly identified and that the actions and limitations of each component be clearly outlined. For example, a color control line as part of a visual readout device can, depending on the design of the product, be used to assess nothing more than the proper fluid uptake, or it could assess the fluid uptake plus reactivity of key reagents or assess uptake plus reactivity plus the order in which reagents were used. The FDA requires a clear
statement of what each built-in QC feature does and how it contributes to the monitoring of device quality.

The FDA does not consider its review mission as establishing guidelines or standards for laboratory practice. Rather, we believe that the frequency of QC testing and the specific rules to assess the results of that testing will depend on various laboratory-specific factors. The FDA works to ensure adequate labeling to facilitate decisions about appropriate QC for a particular laboratory setting.

William Hamlin, discussing CAP statement to CLIA Congress in September 1996. The CAP Accreditation Program has always stated that the laboratory is responsible for designing, implementing, and monitoring the system(s) for quality (process) control for all the testing performed by that laboratory. Thus, a laboratory must define the quality (process) control methodology for each instrument/method being used. The Accreditation Program’s interest is that: (a) such definition be written, (b) the laboratory is carrying out the defined system, and (c) documentation demonstrates both that the process is being monitored and that appropriate corrective action promptly ensues when the monitoring system indicates the process is not within defined control standards.

The basic reason for this broad general approach is the difficulty, at best, of defining detailed and specific quality (process) control requirements to be applicable to the diverse array of instrument/method systems in use in all of the specialty/subspecialty areas in all laboratories. Any precise and specific definition of a quality (process) control system may be superseded by technological advances or innovative new control mechanisms that are superior to the predefined requirements.

As the Accreditation Program has evolved over 25 years, more specific quality (process) control limits have been added in response to federal and/or state regulatory requirements. Often these additions have not been consistent with the above described philosophy, but necessary for deemed status.

Instrumentation has evolved from multiple-channel continuous-flow analyzers to discrete analyzers. Continuous-flow was typically controlled with two samples per run. Discrete analyzers, however, no longer have a defined run; when, then, are the controls to be tested?

Technology has moved to unit instrument testing that is transported to the patient and contains reagent/control packages (unit) used singly for a single analytical procedure. Many of these systems include unique self-contained control mechanisms. In such situations, there are no runs, and assay of control samples with each process not only is very expensive but also provides no information about the next reagent unit. Perhaps the periodic (weekly, monthly) use of control samples to validate the acceptability of a specific lot of units is more appropriate.

Regulatory agencies and other accrediting organizations should return to less-specific requirements that encompass flexible general principles of quality (process) control, focusing on leaving the laboratory responsible for the definition and implementation of a quality (process) control plan that incorporates the following features:

1) Definition, by the laboratory, of a quality (process) control system for each analyte and (or) analyte system, based on the method used and the level of precision and accuracy required.
2) The QC system should be based on clinical laboratory scientific literature and should be able to detect analytical problems throughout the intended range of measurement.
3) There must be evidence (documentation) that the QC system is routinely utilized and monitored.
4) The laboratory must document that corrective action is taken when results deviate beyond the predefined acceptable limits.

Nathan Gochman, Beckman Instruments. Manufacturers and laboratories are operating on two parallel pathways. The manufacturers are trying to improve their instruments by following all the good manufacturing principles [1]. Laboratories are pressured by cost constraints that result in reduced manpower, substitution of lesser-trained personnel, and less resources available for training. There is insufficient confluence between how instruments are evolving and how laboratories are changing their skill mix and adjusting to the reduction in resources. As a result, the necessity for guidelines for QC are more important than ever.

David Witte, Laboratory Control, Ltd. The complexity of the testing process contributes to common cause variation (our CVs) and special cause variation (mistakes or blunders). Both sources of variation can lead to laboratory results that are unacceptable or misleading in the care process (the nonconforming results). One potential definition of analytical quality could be: “Fewer than 10 misleading inaccurate results per million [106].”

Our quality assessment, control, assurance, and improvement systems must be consistent with the complexity of our processes and reduce nonconforming results arising from both variation and mistakes [2]. Simultaneously, we must understand the impact of laboratory results on the outcomes of care processes if we are to define the size of deviation necessary before a laboratory result is judged as nonconforming. Quality tolerance limits can then be specified to design management systems to prevent an excessive frequency of nonconforming results. These global issues are unlikely to be resolved by critical and statistical thinking alone, but will require creative thinking by all stakeholders.

Current nonconformity rates are reportedly as high as 500–2000 per million for routine cuvette chemistry analyses and HIV testing [3]. This compares with 0.4 plane crashes per million flights [4]. Electronics manufacturers have stated nonconformity rate goals of 3 per million.
Laboratory technology continues to improve. Can we expect measurement systems with SD values one-twelfth of the quality tolerance limit? If so, what is the role of the traditional statistical process control style of QC? Will our QC rules become 1–6s, more complex rules, or will we go to zero inspection?

Will we be able to design rigorous processes with redundant feedback so that mistakes will be detected before they hamper the care process? Will medicine be able to adopt some of the error-proofing systems of the aviation industry [4]?

The future may include many innovations in quality management. Statistical process control will evolve to match new technology that has lesser variation. Information systems will provide better feedback through known processes such as limit check (against preset values), delta check (against previous values), and intercheck (against other clinical and laboratory values). Errors will be “absorbed” by the new redundancies and feedbacks in the care process and corrected before any adverse outcomes can occur.

The quality discussion should be framed around how to reduce nonconforming, nonusable, or misleading laboratory results by using new quality schemes. The impact of quality schemes on both the number of nonconforming results and the care process outcomes need to be measured. Laboratorians are challenged to engage in the real-time science of quality improvement [5].

Panel Discussion

Ron Feld, University of Iowa Hospitals and Clinics: Dr. Gochman, would you address industry standards for reagent lot-to-lot variation? It is rumored the industry standard is 5% lot-to-lot.

Dr. Gochman: Lot-to-lot variation in reagents varies according to the reagents. For enzyme reagents, ±10% might be acceptable. Cholesterol or glucose, however, require tighter limits.

Manufacturers must build in a storage factor, given that reagents are not delivered at once. A factor of a few percent is typical, allowing for deterioration during the shelf life of the reagent under specified conditions. The manufacturer may say (e.g.) that a reagent needs to be transported and stored under refrigeration. If reagents are stored in an environment without adequate temperature control, the supplier’s specifications may cover a part of that deterioration.

In general, manufacturers’ specifications are quite tight for lot-to-lot variation and are specified for each analyte. Other compelling factors, however, can introduce problems outside of the manufacturers’ complete control (e.g., shipping, storage).

Carl Garber, Corning Clinical Laboratories. Some systems have excellent stability, reliability, and consistency within lot: How does the Panel feel about defining for QC a run as the life of a lot of reagents? Intense QC measurements would be required to qualify the next lot. Then, once that lot has been established and qualified as being acceptable, its use could be allowed without frequent retesting.

Dr. Bakes-Martin: The definition of a run is an evolving concept. I propose that our concept of having to define a run is flawed. We need to decide at what sources of variability do laboratorians see possible introduction of error? I think most laboratories would see a change in lot number as a source of variability.

Dr. Witte: Dr. Bakes-Martin, you indicated that perhaps a minimum standard from the CDC’s or CLIA’s point of view might be at least weekly QC. In light of Dr. Garber’s suggestion of a stable material (e.g., an immunoassay-based system having calibration for 3 months), would you still suggest weekly liquid QC?

Dr. Bakes-Martin: I presented a two-pronged approach to quality control. First, look for the potential for error; second, monitor over time. Both steps need to occur. Monitoring over time has validity even if you feel there is no error in the system. We are proposing that some system be in place to look at data weekly.

Dr. Witte: If the manufacturers have the most information on stability and also on the frequency and potential for mistakes, how would we devise a system for a laboratory to acquire information regarding past and current performance of products in the field? Could Company X accumulate data about an assay lot from the field and then inform the customer of the error frequency for that lot?

Dr. Gochman: We have estimates of problem rates. Products are used by literally thousands of customers. When we hear a lot with a projected life of 12 months is failing at 10 months, that report becomes part of our complaint-reporting process. If we confirm the finding, we notify all customers not to use that lot beyond 10 months or some other safe date. However, that does not imply that all other lots or all future lots would have that decreased stability. Thus, manufacturers do update information about stability and the useful life of a product in the field.

Dr. Witte: Dr. Gochman, how could we extend this thought to the unit-testing systems, where it is not a reagent component that somehow fails its stability claims, but maybe some digital device that causes a higher “mistake” frequency. How would a laboratory director learn that company X was selling unit-dose devices that had a error frequency of 350 per million instead of the 10 per million that were expected?

Dr. Gochman: My old analogy for this problem is, How do you know the flashbulb is any good?—you flash it and
see. But that doesn’t work, of course. Statistical QC is necessary for determining these things.

**Dr. Garber:** Perhaps we give too much attention to things that we can detect because we know how to determine a mean, an SD, and some QC limits. We may be deluding ourselves into a sense of false security because we are so focused on detecting changes in the process variation that we are not paying attention to all the mistakes that can happen.

Can we strike a new balance? Can we de-emphasize running a number of controls a day and refocus some of our energy on mistake-proofing the other factors that are not detectable with QC samples?

**Dr. Gochman:** Laboratorians worry about what they can control. Perhaps we should spend more time outside of the laboratory worrying about having the right sample from the right patient, and whether that sample has been properly handled.

**Deanna Bowlds, Methodist Hospital:** I would like further discussion from the regulators on the preanalytical phase of analysis, which was estimated as the source for 45% of reported errors (I think 80% is more likely). The preanalytical phase is becoming more and more critical and our control is lessening because reengineering is impacting specimen collection and delivery. No regulations say you can’t have a broom in one hand and a syringe in another, and, believe me, this is happening.

Also, what about continuous monitoring devices, which are coming on the scene more quickly than regulations can keep pace with?

Lastly, general hospital laboratories do a very good job. But tests of moderate complexity are being done in physicians’ office laboratories (POLs), where the medical director’s only qualification is that before CLIA ’88, he or she had been doing it for a long time. There’s nothing that makes these “analysts” competent at it. What do you do about QC for POLs, in which voluntary QC compliance may be low?

**Dr. Bakes-Martin:** Preanalytical/postanalytical concepts are present in almost any QA/QC program in the laboratory, and in any published guidelines. I agree that some laboratories have a lower level of expertise and will only do the minimum. With any regulation process, one is setting the minimum standards and risking that some will go towards the minimum.

Regarding continuous-monitoring devices, I agree (not necessarily speaking as a government representative now) that we need to look at a different focus.

**Ms. Rooks:** The FDA reaches the consumer by requiring truth in labeling. We are pushing manufacturers to put any information in their package insert that addresses preanalytical error so that the end user is aware of whatever problems exist.

**Dr. Hamlin:** I also think >40% of the “mistakes” made in the clinical laboratory are a direct consequence of preanalytical error. That situation is going to get worse before it gets better, given the current move to make everything cheaper. For example, using laboratory personnel to collect specimens is being dispensed with in favor of using personnel who are already on location but who may not have appropriate training.

You can have every analytical or process control mechanism in place, but if you get bad samples or if the people you send the results to don’t use them, misuse them, or send them on to the wrong place, then the impact on the patient is no better than if you had made an analytical mistake.

Regarding POLs or small laboratories operating with relatively unsophisticated people, I am also the CAP representative on the Board of Governors of COLA, the accrediting program started jointly in 1986 or 1987 by the American Medical Association, internists, family practitioners, and pathologists. Because we realized the clinical laboratory world would never affect laboratory improvement in those small facilities by attacking them, we tried a joint program to initiate a voluntary laboratory improvement program—which is having an impact. Over the last 6 years, proficiency testing performance data for accredited vs nonaccredited POLs are dramatically improved. I can say that because back in the 1960s, when CAP’s laboratory accreditation program was implemented (prior to Medicare, CLIA 67, and CLIA 88), proficiency testing performance by all of your laboratories or their predecessors was not exactly overwhelming. But there was evolving improvement as the laboratories became involved in that program—and it is continuing. I think the experience with COLA will be the same. One of COLA’s biggest problems now is that their subscription rate is not increasing, primarily because many physicians’ offices no longer do laboratory work but send it out; only those truly interested in doing it right are surviving. Unlike 30 years ago, I have a little more confidence that we will move towards a fairly equal playing field, where the same set of standards applies regardless of the site of testing—a very fundamental and important principle.

**Robyn Hawkins, Bio-Rad Laboratories:** Does everybody in this room really say that control of analytical error is so good that we’re just going to leave QC recommendations to the manufacturer? Do clinical laboratories really believe that they can reliably turn out results by running controls once every 3 or 6 months? This has not been our daily experience. To me, the CAP proficiency data just don’t show that everybody is this good analytically. I understand that some of the new instrumentation has great analytical promise, but we are not there now.
Dr. Hamlin: I think that’s inappropriate to leave QC systems solely to the manufacturers. What we’re saying is that the QC system has got to be defined by the laboratory. It can’t be defined by the manufacturer, it can’t be defined by CAP, and it can’t be defined by CLIA ’88 or any other regulations because no matter what we define that might satisfy your needs, it is not going to satisfy somebody else’s needs.

Running an accreditation program, we’ve had these checklist issues come up for many, many years. Often, we may have tended to use terms like “periodic,” “appropriate,” and so on. Having been involved in this program, I’ve been virtually assaulted by people who say, “What do you mean by “periodic” and “appropriate”? You’ve got to give us the specifics.” But there’s no one right answer: It has got to be done by the laboratory for its systems and its people, the level of education they have, and so on.

Mr. Garber: I appreciate Dr. Hawkins’ comments made about our relative lack of concern about analytical error. I think the QC policies and programs really should be based on the technological capabilities; one QC program requirement cannot apply across all technologies.

Laboratories rarely make an analytical error. As Dr. Cembrowski commented, of all the major errors maybe only 7% are analytical. Is our focus on common causes of variability wrong as we talk about quality for tomorrow? When we investigate proficiency test errors or patients’ reports errors, 95% of them are mistakes, not analytical errors.

Dr. Hamlin: The bulk of the errors we find out about and investigate are mistakes. The fact that these are occurring must be incorporated into the QC plan or system.

Ann Belanger, Joint Commission on Accredited Healthcare Organizations: I was really interested in Dr. Witte’s data on unacceptable results. Like Dr. Hamlin, I was struck by the apples and oranges quality to it. Plane crashes, I would suggest, might be an outcome; bad cuvettes are part of a process. I would be very interested to see the data for the plane crashes. It can’t be defined by the manufacturer, it can’t be defined by CLIA ’88 or any other regulations because no matter what we define that might satisfy your needs, it is not going to satisfy somebody else’s needs.

What about the need for outcome measurement of a laboratory’s data, and should clinical usefulness be the issue at this point?

Dr. Witte: I agree about the apples and oranges comparison. The “error” rate of 300 per million for cuvette chemistries means that 300 times out of a million results, we would have observed a difference as great as 10 SD. There was no outcome attributable to that because all these results were from method evaluation studies and none was reported. Even so, I suspect that a 10 SD error would at least generate some confusion if not a frank change in outcome.

We need to deal with the impact of our QC system on what happens to patients. The reason we have these QC systems is so that we have results that are good enough for their intended use.

Milton Kelly, private clinical lab consultant: In a “transitional” phase for CLIA and POLs, I think “periodic” needs to be defined, because testing personnel in POLs are unsure what the term means—one a day, once a week, once a month, once a year?

Dr. Hamlin: What you say is true, but the reality is that facilities will begin adequate QA/QC plans when they understand what they are. I don’t know that writing specific requirements is going to either enhance or decrease that understanding.

Greg Cooper, Bio-Rad Laboratories: Suppose data generated by manufacturers either within their own facilities or through field testing might be reliable data for laboratories to use in making certain policy decisions. When the manufacturer submits data to the FDA under the 510(k) approval process, does the FDA critically review the study design that generated the data? Are you looking at the study site and whether the conditions make these data commutable to the laboratory environment at large? Does the FDA look at the population of laboratories that are submitting the data?

Ms. Rooks: We do look at the data critically. However, it depends on the analyte and type of submission as to whether we will do site testing. For the most part, the data we receive are generated in the manufacturer’s laboratory. When we do request site data for certain types of products, the sites have to be identified and, preferably, the persons running the test. We cannot control what is sent to us. We can only review what we see and we’re often not really sure how many laboratory sites or places the manufacturer went to get the data. We do not know how many test sites are used unless we do a bioresearch monitoring inspection of that company. Circumstances dictate whether we will require site testing.

Mary Burritt, Mayo Clinic: Returning to the issue of medical relevance, small shifts in bias or lot-to-lot variation can have clinically significant outcomes. For example, a slight shift in laboratory results for prostate-specific antigen can double the number of men who are sent for biopsy. Most people would consider that clinically significant.

Dr. Hamlin: You are correct. The 1992 Clinical Chemistry Forum dealt with that subject without implementing any major change in defining the medical relevance of the results we report. It is a difficult problem, requiring collaboration with clinicians.
Herbert Malkus, Yale-New Haven Hospital: How do I sell quality? How do I show that the laboratory has a great effect on the financial performance of the hospital even though I represent only 4% of my hospital’s budget?

Dr. Witte: We are selling things that do not happen—a very difficult sale. We are selling people who don’t have to have their INR tested every 2 weeks to monitor coumadin because we devised a system stable enough to test every 4–6 weeks instead. We're selling people who don't go inappropriately for prostate biopsies. We are selling prevention.

Nina Peled, I-Stat Corporation: Regarding the CDC conference at the end of September: There is a sense that industry is feeding government information about point-of-care unit testing devices but nobody is looking at it.

Dr. Bakes-Martin: I can guarantee you we’re looking at it.

Dr. Peled: Good. Also, there is a sense of frustration because there are no change in prescribed QC regulations.

Dr. Bakes-Martin: I totally understand the frustration with the timeframes to change regulations.

Ms. Rooks: I’ve been with the government for some 20 years; nothing happens rapidly. It’s gotten very complicated: A lot of lawyers are involved and it’s just not going to move that rapidly unless some fundamental changes in the political structure of these organizations are made.

Dr. Hamlin: I don’t think one could anticipate any greater speed within the government context.

Few data regarding electronic controls are published in the peer-reviewed scientific literature. If the data are there, why aren’t they published? Most scientific organizations rely on published data in the peer-reviewed scientific literature.

David Phillips, Boehringer-Mannheim: We have had the data, but when submitted for publication it was not a hot topic and was rejected.

Frank LaDuca, International Tech United Corporation: There are data published in peer-reviewed journals. Also, many users have data to present to inspectors.

Dr. Hamlin: Now get it into a format that is useful to influence people with.

Mr. LaDuca: Numerous publications describe data for point-of-care unit test devices, including the history of thousands of patients with use of that device under the guidelines of QC that were in place by the manufacturer. The system was used in a defined manner and the outcome was observed.

Dr. Hamlin: The outcome as compared with a standard system?

Mr. LaDuca: Yes.

Dr. Hamlin: Of course, that is the kind of data needed.

Mr. LaDuca: The American Journal of Clinical Pathology has published at least five or six papers on point-of-care coagulation tests that use alternative QC instead of traditional QC.

In summary, an openness to flexible quality-control regulations has been expressed, and all participants supported minimum requirements that included documented monitoring over time of both the analytical system and the operator—monitoring tailored to minimize errors known to have the highest potential for occurring in the particular analytical system. Participants also discussed the interrelationship of regulations, procedures, and local responsibility.

References