

Accuracy-Based Programs for ‘bigger dividends,’ better care

Valerie Neff Newitt

July 2021—Proficiency testing is the bedrock of good laboratory performance, but Accuracy-Based Programs are equally important, which is why members of a CAP committee are hoping for 10 labs from each peer group to participate.

“If we can get enough people to step up to the plate, we can make sure that the methods that are FDA cleared in the United States are accurate. And then traditional proficiency testing will pay even bigger dividends,” says Gary L. Horowitz, MD, chair of the CAP Accuracy-Based Programs Committee and professor of pathology, Tufts University School of Medicine, and chief of clinical pathology, Tufts Medical Center.

The CAP’s Accuracy-Based Programs do what proficiency tests can’t: verify the accuracy of test results against a gold standard. Proficiency testing provides a check on procedural methodology and results reliability as compared with peer laboratories. “While it is reassuring to know you match your peers, sometimes an entire peer group could be getting inaccurate results,” Dr. Horowitz says.

“That’s the whole point,” says committee vice chair Andrew N. Hoofnagle, MD, PhD, professor of laboratory medicine and head of the Division of Clinical Chemistry, University of Washington. “The whole peer group can be skewed, different than either the rest of the peer groups or, more importantly, the reference measurement procedure, the gold standard, if there is one.”

It is difficult to find human specimens with the range of concentrations needed to assess all relevant concentrations of all tests. Proficiency testing providers therefore use materials that simulate human specimens, adding materials to obtain the range of concentrations needed, Dr. Horowitz notes.

“When PT material is made, it begins with basic plasma or serum, but that starting material is dialyzed, preservatives are added, and then it’s spiked with calcium or cystatin C or transaminases and/or other materials to get the concentrations desired. At each of those steps the material becomes less like genuine human specimens,” he explains. “There are occasions when it is okay, but most of the time it doesn’t work. These specimens do not react exactly like genuine human specimens, due to a phenomenon referred to as matrix effect.”



Dr. Harry

Matrix effects are caused by something nonspecific in the sample that changes the value of whatever the laboratory is trying to measure, says Brian Harry, MD, PhD, a member of the Accuracy-Based Programs Committee and assistant professor of pathology and medical director of special chemistry, University of Colorado School of Medicine. “We don’t know specifically what they are, but we do know they exist and can disrupt measurements.”

Two peer groups using a given test might get exactly the same results on human specimens but different results on the proficiency material. “Vitamin D testing provided a good example of this,” Dr. Horowitz says. “Using regular PT Survey materials, there were twofold differences in responses. We’d get a value of 50 by one method in one peer group, and 120 on the same specimen using another method in another peer group. But when we sent real human serum, the numbers agreed between the two groups. So even with tests that have FDA clearance and

existing documentation indicating they are good tests, labs were getting values that appeared to be wildly different on traditional proficiency testing material when in reality the tests provided comparable results on real human specimens.”

“When you agree with your peer group and the peer group is inaccurate compared to the reference method, it’s not a mark against you as a laboratory. It’s a mark against the manufacturer,” he says.

Accuracy-Based Programs run only on genuine human samples that exhibit essentially no matrix effects. Commutability and reference measurement procedure are the two terms to keep in mind, Dr. Hoofnagle says. “Commutability means samples behave like actual patient samples in each assay. Reference measurement procedure means we have some gauge on the truth—the actual value of the concentration of the analyte in the sample. That’s what sets Accuracy-Based Programs apart from traditional PT, which uses samples that I call concocted, pretend. They do not represent actual human biology.”

The CAP obtains specimens for the Accuracy-Based Programs in several ways. For some tests, such as testosterone, cortisol, and A1C, Dr. Horowitz says, “it’s relatively easy to get specimens from individuals with a range of concentrations—men versus women, morning versus evening, individuals whose diabetes is controlled to varying degrees. In other cases, like creatinine, we have proved we can add creatinine to normal serum to get high concentrations without introducing matrix effects.” These matrix-effect-free specimens are commutable and can be used to compare different methods to one another and to reference methods, he says. “Then we can assess whether the method used in a lab is generating truly accurate results.”

These assessments drive industrywide improvement. In general, he says, “things work pretty well.” But when a problem is detected, the CAP shares the data with manufacturers, and they can then improve the affected assays accordingly.

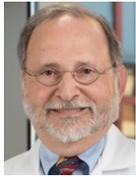
“For the CAP, the great-grandparent of Accuracy-Based Programs is hemoglobin A1C,” Dr. Horowitz says. “For many years it has used commutable material and results have been compared to a reference-based measurement procedure. A1C is a huge success story because when data became available from hundreds, if not thousands, of labs, manufacturers with problems fixed their methods.” The methods used today are far more accurate and precise, he says. “We wouldn’t be where we are today without that Accuracy-Based Survey. Now everyone can reliably use the same cut-points for making a diagnosis of diabetes and for determining what is considered good or bad control. It made such a dent in the field when people actually saw performance using commutable material and the reference values for it.”

A1C was unique, he adds, in that it was easy to find patients with normal values and various degrees of elevation because diabetes is prevalent. “It wasn’t hard to find materials to cover the range of values we needed. So the CAP was able to send out real commutable materials for that Survey. Newer Surveys are the grandchildren of A1C, but clearly it has been a lot more difficult to get commutable materials and the ranges we want.”

The Accuracy-Based Programs available now are as follows: Accuracy-Based Glucose, Insulin, and C-Peptide; Accuracy-Based Testosterone, Estradiol; Accuracy-Based Lipids; Accuracy-Based Vitamin D; Accuracy-Based Urine; Harmonized Thyroid; Hemoglobin A1c GH5 (five challenge), GH2 (three challenge); Hemoglobin A1c Accuracy Calibration Verification/Linearity; and Creatinine Accuracy Calibration Verification/Linearity.

At a meeting in March, members of the Accuracy-Based Programs Committee discussed what other analytes should be measured in the Accuracy-Based Programs. “ABP doesn’t develop material for everything that’s measured in the lab,” Dr. Harry says, “but it does focus on things that are critical, whether it be because technologies produce different results, because an analyte is particularly hard to measure because of its chemistry, or because it is something that affects many patients.”

Dr. Horowitz makes a strong case for the use of Accuracy-Based Programs for labs using laboratory-developed tests rather than FDA-cleared assays.



Dr. Horowitz

“If you’re using a manufacturer’s FDA-cleared assay and run it the way the manufacturer has directed, the burden of making sure that assay is good is on the manufacturer,” he notes. “You have fewer validation studies to do than when you develop your own test. In the case of LDTs—and almost all the LC-MS assays are LDTs—each one is a little different. They use different columns, different transitions, different reagents. You don’t have a peer group. So the only way you can know your test is accurate is to use commutable samples and compare results to a reference method. Those using LDTs, in particular, should be using Accuracy-Based Surveys.”

Harmonization is a related issue of concern to the committee. In the absence of a reference method, “it’s impossible to know what the true value is,” which is needed to establish accuracy, Dr. Horowitz said in a CAP podcast on the Accuracy-Based Programs. But with commutable matrix-matched specimens, whether different methods get the same results can be determined, “and if all the results are the same, or harmonized, that’s a good thing. And once a reference method and reference materials are developed, we can then assess accuracy. If different methods are not harmonized using commutable specimens, we would hope that the reference intervals are accordingly different,” he told listeners of the podcast.

“Accuracy-Based Programs can, at a minimum, provide insight into whether various test methods are harmonized and get comparable results on genuine human specimens,” he said in an interview.

TSH is one of many analytes for which no reference method is available. “We know there are differences in TSH measurements between methods,” Dr. Horowitz says. “They’re not harmonized; we do not know what the true value is. We’re hoping the reference intervals are different to reflect those differences.” Cystatin C, too, was a problem. “But based on a commutable specimen, one of the peer groups that was different has recalibrated, and now cystatin C is harmonized. We don’t know what the true value is, but at least all the major methods are getting the same value.”

D-dimer is under discussion between the Accuracy-Based Programs Committee and the CAP’s Hemostasis and Thrombosis Committee. “We don’t know if everybody’s getting the same results,” Dr. Horowitz says. “What we do know is that for D-dimer in particular, two different sets of units are being used across the country,” with some using a cutoff of 500 and others using 250 because of the differing units. “At the very least we want to get everybody on the same units and to see whether these assays are harmonized by using commutable materials.” The goal is to test the tests to make sure they’re accurate. “Let’s make sure doctors can use them interchangeably,” he says, “because people are much more mobile than they used to be.”

Dr. Hoofnagle agrees there is a need for greater industry standardization that might be achieved through efforts such as Accuracy-Based Programs and harmonization. He illustrates the point by offering examples of problems discovered in his own laboratory.

One assay he was running for LDL and HDL cholesterol on a specific platform was greatly skewed. “Even though there’s a standardization program at the CDC, LDL and HDL cholesterol numbers remain a disaster. It’s really bad,” Dr. Hoofnagle says. “One instrument that I have gives patients better numbers in terms of cardiac risk because the HDL results we get are higher than every other peer group. It’s possible that a patient would decide against taking a statin if their specimens were measured in our laboratory on that platform. If they drive half a mile away and go to a different laboratory, a different platform will give them a different number and their HDL will be lower and their LDL will be higher.”



Dr. Hoofnagle

Another disturbing situation he encountered involved testosterone testing. “I used an immunoassay by a manufacturer for testosterone for many years,” Dr. Hoofnagle says. “Our providers kept saying, ‘Andy, you are not getting the right result; your answers are wrong.’ I said, ‘That’s okay because I’ve changed my reference range. My reference range is shifted lower because the results are lower than what we see on average across the industry.’ And they said, ‘That doesn’t matter. There’s a cutoff now in the literature. All of the patients are looking at that cutoff, so when an older man walks into my office and they get a test on your platform, they may fall below that number.’ And I say to them, ‘It’s okay, you’re in the reference range,’ but they point to that number on the paper and say, ‘No, I’m below that number. You have to give me testosterone.’ It makes the patient-provider conversation more complicated than it needs to be.”

Dr. Hoofnagle says he heard this from his providers repeatedly over the years. Then the Accuracy-Based Program for testosterone became available. “Now, looking over seven years, we see we have had this bias of about 25 to 30 percent every year compared with the reference measurement procedure. That’s where this accuracy-based effort is so amazing. This wasn’t concocted material. These were actual human specimens and every single one of them was wrong. I brought this to the manufacturer’s attention on three different phone calls. And I said, ‘This is bad. You need to fix this.’”

If laboratories don’t use the Accuracy-Based Programs it could be because they know they meet the requirements of accreditation with traditional proficiency testing Surveys or because fewer analytes are in the accuracy-based specimens, Dr. Horowitz says. “For the regular proficiency test, you could have 40 or 50 different analytes in it, but it is not commutable. So it’s a very small number of tests that are in the Accuracy-Based Surveys. Ideally, we would do this for all proficiency testing, but it’s prohibitively expensive and very hard to get commutable samples that would achieve everything.”

Dr. Horowitz is hopeful that the CAP will get a representative sample of laboratories from each peer group to use Accuracy-Based Programs. “All we need is 10 labs from each peer group to participate. Then we will be able to say, ‘Peer group one uses XYZ method for testosterone, and when they participated in the testosterone Accuracy-Based Survey their values were right on target with the reference method.’ Then anyone else that uses that specific FDA-cleared method, running it the way the manufacturer directs, can say, ‘This is great. I’m on the mean for my peer group and I know my peer group is accurate based on the Accuracy-Based Survey.’ So if you’re close to the mean value for your peer group and your peer group shows good performance in the Accuracy-Based Programs, you know you’re good.”

Participation in Accuracy-Based Programs helps the entire field take care of patients, Dr. Hoofnagle says. “It’s saying, ‘I want the entire field to get the right answer; therefore I am going to be in ABP.’ It is our job as the CAP to help labs give patients the right answers every time. I would plead, if I could, with every lab director to consider ABP as a way to improve health care.”

Identifying manufacturer methods that behave differently from other methods in the field is itself a compelling reason to participate, Dr. Harry says. The most important reason, he says, is to drive better patient care.

“Most physicians look at a number on a paper or in a computer and never question that number or what it means or how it came about. They take it at face value,” Dr. Harry says. “So it’s the clinical pathologist’s job to make sure that number is the most accurate number we can provide.” Doing so is how physicians and others practicing laboratory medicine treat their patients. “We don’t see our patients in the clinic,” he says. “We have to make sure they get good care because we provide their doctors with good data.”□

Valerie Neff Newitt is a writer in Audubon, Pa.