COLLEGE of AMERICAN PATHOLOGISTS

Educational Discussion: Whole Blood Glucose Meters Performance

The College of American Pathologists believes that proficiency testing is not simply a regulatory requirement but also a means to encourage, and document, improvement in the field. At the beginning of 2015, the Chemistry Resource Committee implemented changes to the grading criteria for glucose meter performance, from within 20% (or 12 mg/dL, whichever was greater) of the peer group mean to within 12.5% (or 12 mg/dL, whichever was greater) of the peer group mean.

This decision was prompted not only by the current CLSI POCT12-A3 guideline "Point-of-Care-Blood Glucose Testing in Acute and Chronic Care Facilities" for professional use but also by empirical data that, over the past ten years, the coefficient of variation on our Surveys has decreased significantly. For example, on a sample with a glucose concentration of ~400 mg/dL, the three peer groups with the largest number of participants in 2004 (representing 87% of all participants) had CVs of **10.5%, 4.5%, and 3.8%**. In 2015, the three largest peer groups (representing 90% of all participants) had CVs on a comparable specimen of **5.0%, 4.4%, and 2.6**%. Perhaps equally notable, **none of the most commonly used meters in 2004 were among the most commonly used meters in 2015**.

At first blush, it would appear that such a change in grading criteria would result in widespread failures in the field. Some would point out that a change from 20% to 12.5% represents a decrease of (20.0-12.5)/20.0 = 32.5%. However, we do not make such changes without careful consideration. We reviewed our proficiency testing data for several prior years, performing trial grading with the proposed new criteria, and we found that the overall pass rate remained well over 98% and that the pass rates for virtually all of the peer groups also remained well over 98%. As an example, the table below shows the pass rates from the 2014 WBG survey (old criteria) and from the 2015 WBG survey (new criteria) for the major peer groups, which represent ~90% of all participants:

2014 Survey Results, with "old" grading:

	Specimen 1		Specimen 2		Specimen 3	
	mean value (mg/dL)	pass rate	mean value (mg/dL)	pass rate	mean value (mg/dL)	pass rate
Manufacturer A	402.73	99.8%	156.31	99.7%	68.82	99.4%
Manufacturer B	369.35	99.6%	207.11	95.8%	99.61	95.7%
Manufacturer C	309.42	99.7%	167.33	99.7%	78.06	99.6%
Overall		99.6%		98.9%		98.7%

2015 Survey Results, with "new" grading:

	Specimen 1		Specimen 2		Specimen 3	
	mean value (mg/dL)	pass rate	mean value (mg/dL)	pass rate	mean value (mg/dL)	pass rate
Manufacturer A	411.02	99.0%	155.56	98.5%	64.56	99.2%
Manufacturer B	379.52	99.7%	205.11	99.6%	94.70	99.5%
Manufacturer C	304.86	97.9%	166.08	98.5%	73.55	99.2%
Overall		98.9%		98.9%		99.2%

In other words, thanks to major advances introduced by manufacturers over the past several years, current glucose meters perform substantially better than their predecessors, whose performance was used to establish the prior grading criteria of 20% or 12 mg/dL. If newer meters can, and do, perform much better than their predecessors, shouldn't our grading criteria reflect that reality?

You will notice that the peer group mean values are substantially different from each other. This is a reflection of "matrix effects" – the reality that proficiency testing specimens are not the same as genuine specimens from patients. (It is worth noting that the CAP Accuracy-Based Proficiency Testing Surveys are notable exceptions. For those Surveys, the materials are indeed genuine human specimens.) The Chemistry Resource Committee has spent considerable time and devoted considerable effort to finding commutable materials for whole blood glucose, but we have not been successful. Glucose in whole blood is a particularly challenging sample, in that genuine red blood cells metabolize glucose, decreasing its concentration over time. To prevent this metabolism, one needs to remove the red blood cells (in which case the sample is no longer whole blood) or inactivate the enzymes responsible (in which case the sample no longer acts exactly like a real human whole blood).

Thus, we recognize that peer groups may get different mean values for the proficiency testing material, which is why we grade each peer group against its own mean value. We have seen no evidence, though, that differences in imprecision among peer groups can be ascribed to matrix effects. Perhaps even more important, the empirical data (>98% pass rate) indicate that the new criteria are not so stringent that they pose a significant burden. Put differently, if the mean peer group value for a proficiency specimen whole blood glucose is 300 mg/dL, should we grade as acceptable values as low as 225 and as high as

375 (a range of 150 mg/dL!), when the reality is that current meters achieve values between 262 and 338 more than 98% of the time?

We think the new criteria reflect the reality that glucose meter performance has improved significantly. And we applaud the achievements of the manufacturers who have made those improvements as well as the participants in our surveys who have documented the reality of those improvements in the field. Kudos to all!

Gary L. Horowitz MD, FCAP
Chemistry Resource Committee