STATEMENT TO THE CLINICAL LABORATORY IMPROVEMENT ADVISORY COMMITTEE
November 7, 2024

Range of Values for Proficiency Testing Samples

Mr. Chairman, members of the Clinical Laboratory Improvement Advisory Committee, my name is Sue Harmer and I am President of the American Proficiency Institute (API). Established in 1991, API is one of the nation's largest, accredited proficiency testing providers, serving over 20,000 hospital and physician office laboratories. Thank you for this opportunity to speak with you today as you consider issues related to the determination of clinically relevant ranges of values for proficiency testing samples.

My comments focus on how to ensure proficiency testing providers are covering the appropriate range of targets. Simply put, there are two ways to calculate the target value for each analyte tested – "All-Participants" means and "Peer Group" method-specific review.

With the All-Participants measure, the target value is an average of all the various methods used by all proficiency test participants for a particular analyte. The average is then used to determine the range of acceptable results for each analyte in a sample. However, since one or more large peer groups may recover significantly lower or higher than other methods, the All-Participants mean may be biased.

Alternatively, for a Peer Group measure, data is grouped and examined by reagents or instruments for a particular analyte tested. This measure is more feasible for certain analytes, especially the newly regulated cardiac markers. Peer Group measures are often used instead of All-Participants because proficiency testing results may be subject to a matrix effect. An example of a matrix effect is when the presence of substances in a sample, other than the intended analyte, may affect the quantification of the intended analyte. A matrix effect can impact each methodology differently but those using the same methodology are affected similarly, which adds to the importance of grading by peer group.

Thirty-plus years ago, All-Participants means was written as the default for CLIA proficiency testing measurement, but Peer Group measures were permitted when necessary. Today, Peer Groups measurement is the default for proficiency testing providers as the importance of matrix effects has been acknowledged and the sheer number of methods makes it impractical to review and document bias or matrix effects for every method.

The July 11, 2022, final rule on proficiency testing regulations recognizes this by stating, "peer grouping is generally the way that target values are set for most analytes." The resulting §493.2, now includes a definition of peer group and a broader use of peer groups in a revised definition of target value. The target value definition now states, "If the peer group consists of 10 participants or greater," the target value is the "mean of all participant responses after removal of outliers;" "the mean established by a definitive method or reference methods;" or "the mean of a peer group…." All-Participants means would be required when a peer group consists of fewer than 10 participants.

To highlight how the measurement method impacts the averages and range of values obtained, take a look at prothrombin time. Results from different prothrombin time methods are not comparable, and this difference is magnified when testing proficiency testing samples. The first chart below shows the All-Participants averages for all 2023 proficiency testing samples from API.

Test Event	All-Participant Means				
2023 1st Test Event	11.4	11.8	19.5	32.0	39.7
2023 2nd Test Event	11.0	11.1	11.9	28.8	45.7
2023 3rd Test Event	11.0	11.1	11.7	11.9	45.7

This All-Participants data was reviewed to determine whether API had offered challenges covering a reference range of 9.4-12.5 seconds. Using the data above, it appears we did not challenge the low end of the reference range. If the data is grouped by method (as in the chart below), you can see that method bias in the All-Participants data obscures how low the results were for many participants.

The chart below shows the spread of values reported by different methods on one sample with an All-Participants mean of 11.0.

Peer Group (# participants)	Lowest Result	Mean	Highest Result
Reagent A (1250)	8.7	10.3	11.5
Reagent B (25)	10.1	10.8	11.3
Reagent C (600)	9.8	10.9	12.0
Reagent D (30)	10.0	11.2	12.1
Reagent E (500)	12.0	12.9	14.7

This data shows three peer group means were lower than 11, and two were higher – one significantly higher. The significantly higher group represented 24% of all participants, and raised the overall mean so much that it appeared a low target had not been provided. In fact, over half the participants, those using Reagent A, had a target value of 10.3, and approximately half of those participants reported values lower than 10.3. The lowest acceptable results reported in that Reagent A peer group were lower than the requested All-Participants reference range.

The All-Participants data at low or high levels can be skewed, not meeting the 80% consensus requirement for evaluating participant results. In the instance above, Peer Group data should be used to determine whether the reference range for prothrombin time is covered by a proficiency testing program.

Another technical matter impacting proficiency samples addresses how some targets may not be achievable in manufactured samples. A leading provider of hematology samples, known for its expertise in cell stabilization, explained the technical limitations of proficiency samples for white blood cell (WBC) differentials in a 2022 open letter: "Proficiency products are manufactured from multiple donors. When the donor cells are stabilized and pooled, they become very uniform in ratios for the differential. Due to the overall similarity of donors that make up the stabilized cellular material, the differential can only be altered slightly (15-40% lymphocytes, 50-75% neutrophils) to maintain manufacturable product. Stabilized proficiency materials can produce anomalous results when differential populations are varied at extreme (abnormally high or low) levels… it is not achievable for WBC differentials in stabilized proficiency products to represent the full range of values expected in patient results."

To address these, and other, issues related to the range of values for proficiency testing samples, we offer the following suggestions for your consideration:

1. For a more impactful data review, it may be useful to know which analytes are best summarized through Peer Group data. It is reasonable to recommend that proficiency testing providers list up front each analyte for which Peer Group data is preferred.

2. There should be recognition that proficiency testing results vary from results on patient samples, due to matrix effects. This is common when using a manufactured sample as pooled plasma is frequently used, there may be stabilizers or preservatives present, or multiple analytes may be targeted in the sample.

3. To better ensure a broad range of targets, it would helpful if proficiency testing providers were provided with the desired ranges for each analyte further in advance. The current annual approval process allows limited time for research and development activities that might allow the desired reference ranges to be met. Providing the desired ranges for each analyte in advance would also allow proficiency testing providers an opportunity to provide feedback on challenges that may arise from manufactured samples.

On behalf of API, thank you for your consideration of these comments. I would be pleased to answer questions or provide additional details.

American
Proficiency
Institute