Using historical PT data to establish interlaboratory comparison limits for clinical test parameters in a new international PT scheme

Sue Styles, MS[1] and Anne Gore, MT(ASCP), MBA[2]

[1]American Proficiency Institute and [2]LGC Clinical Diagnostics

American Proficiency Institute

LGC CLINICAL DIAGNOSTICS

INTRODUCTION

Comparison limits for clinical (medical) proficiency testing (PT) have been a subject of research for years. Robust standard deviations (SDs) derived from participant data are used as PT limits in many fields. However, robust SDs do not always align with the clinical utility of reported results. Very precise medical tests can produce small standard deviations, which generate unsatisfactory z-scores for some suitable results. These "unsatisfactory" results would lead to the same patient care decisions as results with "satisfactory" z-scores. Conversely, tests with low precision can have such high standard deviations that almost any result receives a satisfactory z-score, even if some of the results would cause a different decision to be made regarding patient care.

OBJECTIVE

Many clinical proficiency testing organizations have recognized the shortcomings of standard deviations and are using other types of comparison limits. In the United States, many comparison limits are set by regulation. The Clinical Laboratory Improvement Amendments of 1988 (CLIA) required the use of participant standard deviation for many analytes, and percentages or fixed amounts for some analytes. A 2022 revision of CLIA shifted to percentage limits for most analytes. Before the revision was published, our group worked to determine relevant comparison limits for a new international PT scheme where the CLIA limits would not be mandated. We reviewed limits being used around the world and found more differences than expected. With no clear international consensus and without the constraints of regulation, we sought to identify the best type of limit for each analyte.

METHODS

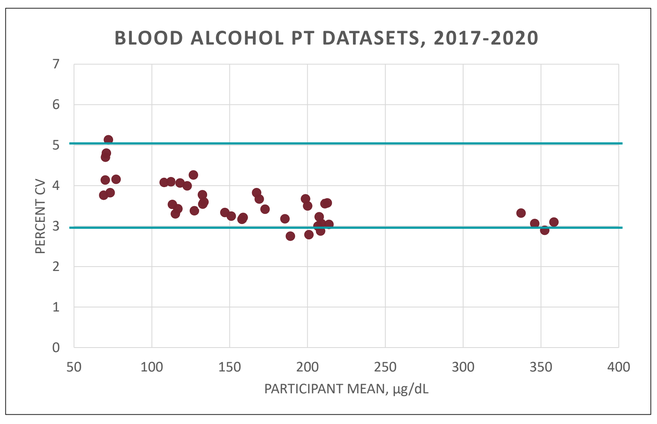
Our background research included historical and proposed U.S. PT limits, limits used by international providers, studies on biologic variation and method performance, and historical data from the American Proficiency Institute (API) PT scheme with thousands of participants testing the most common analytes. We found that percentages are the most common type of clinical PT comparison limits used internationally, but they are not suited for all analytes. We then used the historical API data to model relationships between concentration and the variability of results for each analyte to determine the most consistent and clinically relevant comparison limit for each analyte.

RESULTS

Four distinct dataset types were identified. Where the spread of results (e.g., coefficient of variation or CV%) is consistent across concentrations, percentage limits can provide consistent feedback to PT participants. Where variation increases or decreases with concentration, standard deviation is more suitable and simpler to implement. Where variability is low or clinical utility requires very precise results, a fixed amount such as 4 mmol/L is best for identifying results that may need corrective action. Finally, some analytes are most precise at a middle range and become less precise at low or high concentrations. For these analytes, regression formulas based on historical datasets may provide a comparison limit that best fits the concentration of a particular PT sample.

PERCENTAGE LIMIT – BLOOD ALCOHOL

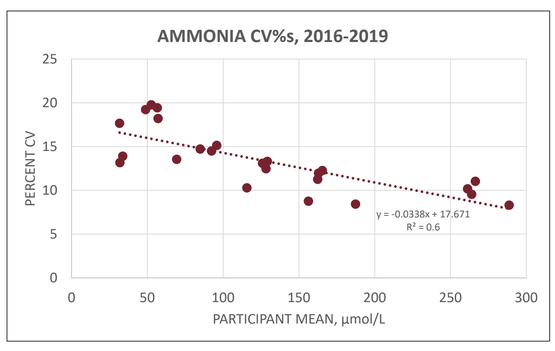
Blood alcohol is an example of an analyte with a fairly consistent CV of 3-5% across a range of concentrations. A consistent CV can be multiplied by 3 to create a comparison limit of approximately 3 standard deviations. Using a consistent limit instead of participant SDs (which vary between rounds) helps participants better compare z-scores across multiple rounds.

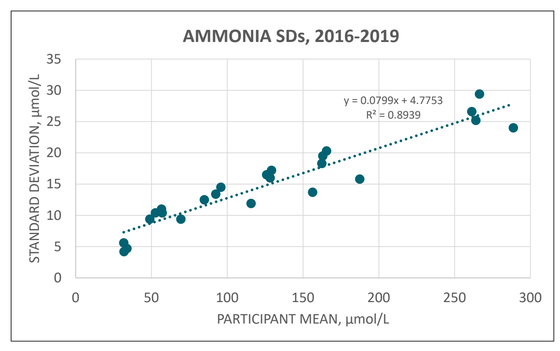


This data is from a common method with 100 participants per round. However, since other methods have higher CVs and other providers use values above 15%, 20% seemed more suitable than a value like 15%. In the new scheme, 20% has assessed results appropriately.

SD LIMIT – AMMONIA IN BLOOD

Ammonia CVs for participants testing our samples decline with concentration. According to our data, applying some PT providers' comparison limits of +/- 20% would only accept results within 1 standard deviation at low concentrations, and just under 3 SDs at the highest concentration recently offered. That would not be a useful limit for our laboratories. In contrast, a regression equation shows the relationship between concentration and SD is very strong, with an R^2 value of 0.89 (max. 1.0).

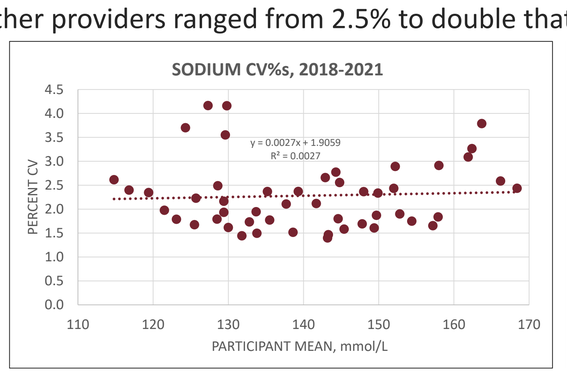


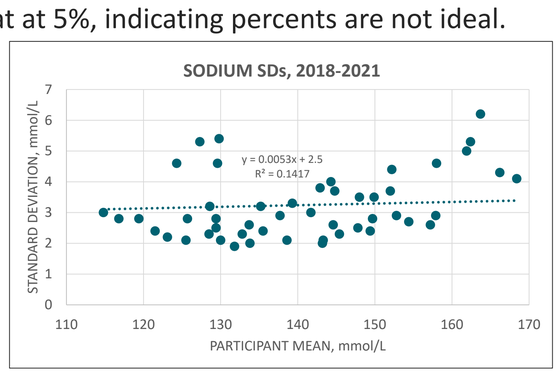


While we prefer to use a comparison limit that is consistent across rounds when it is possible and clinically relevant, standard deviations have been a reliable and useful limit for API for many years. The international scheme used robust SD as their comparison limit as well.

FIXED LIMIT – SODIUM IN BLOOD

Sodium has a very flat CV line, which initially appears suited to a percentage-based limit. However, sodium CV%s and SDs do not fit a line as well as those for blood alcohol or ammonia, with the R^2 (predictive) value for both trendlines near 0. In addition, the percents used by other providers ranged from 2.5% to double that at 5%, indicating percents are not ideal.

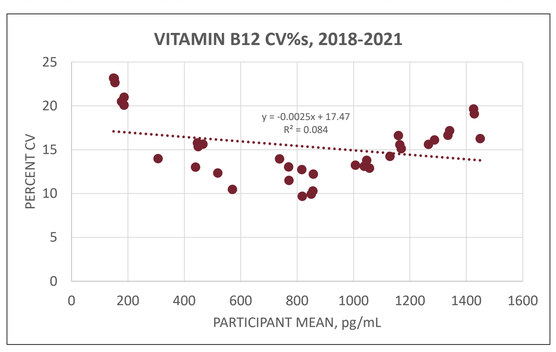


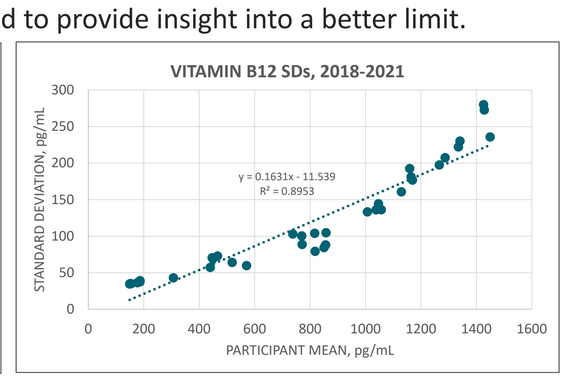


In patient care, tight control of sodium is critical. Percent limits often excluded good results at lower levels (around 110 mmol/L) and allowed dangerous results at higher levels (around 170). Possibly due to this limited range of concentration, the CLIA limit for sodium has been a fixed amount of +/- 4 mmol/L for many years. Our review of data for multiple methods confirmed this limit was the best one for sodium, and +/- 4mmol/L was used in the international scheme.

SLOPE FORMULA – VITAMIN B12

Vitamin B12 illustrates the importance of matching comparison limits to the data used in an actual PT scheme. The review of API data showed 25% was an appropriate limit for separating satisfactory PT results from questionable or unsatisfactory results. However, API has enough participants to separate data by method when determining assigned values and acceptance limits. In the smaller international scheme, 25% was not working as well, and robust participant SDs were not consistent enough to provide reliable performance assessment. API statistics for all combined methods were charted to provide insight into a better limit.





The "U" shape of the CV data shows why percent limits did not fit all-method data. Instead, the slope formula of the SD trendline calculates appropriate SD equivalents by concentration.

CONCLUSIONS

Where historical data on participant performance for an analyte is available, reviewing simple plots of CV or SD by concentration can help determine useful PT comparison limits. This is especially true when the historical data and the scheme for which the PT limits are being set have similar characteristics, such as the same sample material, similar participant test methods, and similar quality of laboratories. Caution should be exercised if the historical data is from a PT scheme using different sample material or if the laboratories are using different test methods. In these cases, the data may not reflect conditions experienced in the target PT scheme. In addition, consistent limits (rather than variable limits based on participant data) can help laboratories get more trending value from their z scores. We used historical information and a thorough review of limits in use around the world to determine the most consistent comparison limits possible for a new international clinical PT scheme launched in 2021.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the help of Shannon Mertz, MT(ASCP) and Leigh Akerley, MLS(ASCP)^{cm}, who helped collect and review data during this project.

CONTACT

Sue Styles, Quality & Compliance Manager
American Proficiency Institute
Traverse City, Michigan, USA
api-pt.com | sstyles@api-pt.com

American Proficiency Institute