State of California—Health and Human Services Agency
California Department of Public Health

KAREN L. SMITH, MD, MPH
Director & State Health Officer

EDMUND G. BROWN JR.
Governor

April 29, 2016

TO: Participants in the March 2016 Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Acceptable Ranges of Results for the March 2016 Proficiency Test in Forensic Alcohol Analysis

Attached is a summary of the descriptive statistics for the March 2016 proficiency test in forensic alcohol analysis. Included here are the target formulation values, the true values as determined by the Department's analyses, the peer-group or consensus values and the standard deviations, and graphical summaries of the distribution of participant results.

Historically, the Department has determined the acceptable limits of performance based on reported results that are within the range representing ±5% of the 99% confidence interval of the peer group mean, where the range has been truncated to two significant figures (Table 1). This range is described as the "Tier #2 interval." The Department also calculates a "Tier #1 interval," which represents the range of reported results that are within ±5% of the 95% confidence interval of the peer group mean where the range is based on the results reported to three significant figures. Tier #1 is expected to include those laboratories demonstrating a high degree of accuracy. The second, wider tier would include those laboratories not as close to the central tendency as the first tier, but still accurate and therefore adequately competent. Again, historically, the Department has used the wider second tier to evaluate the laboratories' results.

The IUPAC International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories (Harmonized Protocol) recommends the use of z-scores for evaluating proficiency test data. However, the Harmonized Protocol notes that that the interpretation of the z-scores is based on the normal distribution of reported results, in which case the z-scores can be expected to follow the standard normal distribution. As indicated in Table 2, the results for Pools 02256 and 02296 in this proficiency test were not found to be normally distributed. Accordingly, the use of z-scores may not be completely appropriate, but they still may be useful to identify outlier and/or warning level results. The expression for calculating a z-score is included in Table 2. Generally a score between -2 and +2 ($|z| \leq 2$) is considered satisfactory or acceptable. A score outside the range -3 to +3, inclusive ($|z| \geq 3$) is considered unsatisfactory or unacceptable and the laboratory must take corrective actions. Z-scores between -3 and -2 or +2 and +3 ($2 < |z| < 3$) are considered questionable and these two ranges should be used as warning limits. Scores within the warning limit ranges in two or more consecutive test events could be considered unacceptable.

The proficiency test results expressed as *z*-scores for the participants whose results were used to determine the peer group mean and statistics in the March 2016 test are summarized in Figure 4. Participants are identified by codes. An attachment to this letter provides codes for participants from your laboratory. The figure is provided for educational purposes only and was not used to formally evaluate a laboratory's performance.

Another approach for evaluating proficiency test data, which is non-parametric and does not require the data to be converted to a standard normal form, divides the test data at regular intervals or quantiles[1]. The quartile is a type of quantile: the first quartile (Q_1) is defined as the middle value between the lowest value and the median of the data set. The second quartile (Q_2) is the median of the data set. The third quartile (Q_3) is the middle value between the median and the highest value of the data set. The interquartile range (IQR), a measure of the dispersion of the data, is the difference between the upper and lower quartiles ($IQR = Q_3 - Q_1$). Boundaries (called fences) are set at $Q_1 - 1.5$ IQR (lower fence) and $Q_3 + 1.5$ IQR (upper fence) to identify potential outliers in the tails of the distribution. In Figure 3, the data from the two pools are presented as box and whisker or Tukey plots with the quartiles and fences shown. The median of the data is shown by a black line and the mean of the data is shown by a red line inside the box. These figures can be used by the participants to evaluate their data.

Sincerely,

Clay Larson, Chief
Abused Substances Analysis Section
Food and Drug Laboratory Branch

[1] Statistics and Chemometrics for Analytical Chemistry Sixth Edition, Miller and Miller (p. 158)

Statistical Data for March 2016 Proficiency Test in Forensic Alcohol Analysis

Table 1 CDPH Tier #1 and Tier #2 Acceptable Ranges

Pool	Peer Group Mean	Tier #1	Tier #2
#1	0.157	0.144 – 0.170	0.14 – 0.17
#2	0.276	0.259 – 0.293	0.25 – 0.29

Table 2 Summary of Test Pool Data

Parameter	POOL 1 (02256)	POOL 2 (02296)
Pre-distribution Data	Target Value 0.16% True Value[2] 0.157 Standard Deviation[2] 0.0013	Target Value 0.28% True Value[2] 0.276 Standard Deviation[2] 0.0012
Descriptive statistics	Mean 0.157 Adjusted Mean[3] 0.157 Standard Error[4] 0.0004 Median 0.158 Standard Deviation 0.0026 Minimum 0.152 Maximum 0.163 Count 38	Mean 0.277 Adjusted Mean[3] 0.276 Standard Error[4] 0.0009 Median 0.277 Standard Deviation 0.0054 Minimum 0.263 Maximum 0.283 Count 37
Descriptive statistics (box plot)	Q1 (25%) 0.155 Q3 (75%) 0.159 IQR 0.004 Lower Fence 0.149 Upper Fence 0.165	Q1 (25%) 0.271 Q3 (75%) 0.280 IQR 0.009 Lower Fence 0.258 Upper Fence 0.294
Histogram	Figure 1	Figure 2
Normal distribution?[5]	No (p=0.036)	No (p=0.022)
Box Plot (SigmaPlot)	Figure 3	Figure 3
Robust mean, X*[6]	0.158	0.276
Robust standard deviation, σ_{rob}	0.0018	0.0056
Fitness-for-purpose standard deviation, σ_p[7]	0.0042	0.0067
Consensus value (X_a) determined as Mode ($\mu_{1/2}$) of Gaussian Kernel distribution	0.1578	0.2772
Uncertainty of the consensus value, X_a , S.E.[8]	0.0005	0.0012
$X_a \pm$ S.E.	0.1578 ± 0.0005	0.2259 ± 0.0012
z-score	$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$

[2] Based on CDPH's Headspace Gas Chromatographic Method
[3] Mean determined from participant data after the removal of outlier(s)
[4] Standard Error of the Mean
[5] Shapiro-Wilk test used at 0.05 significance level.
[6] Robust average of the results reported by the participants was calculated using Algorithm A in Annex C of ISO 13528:2005.
[7] The Department has determined a value for σ_p of 2.5% based on the uncertainties associated with the reported results on recent tests together with the 5% accuracy and precision standard of performance requirements set forth in the regulations.
[8] Determined as Standard Error of Mode using bootstrap simulation technique with bandwidth of 0.75*σ_p

Figure 1

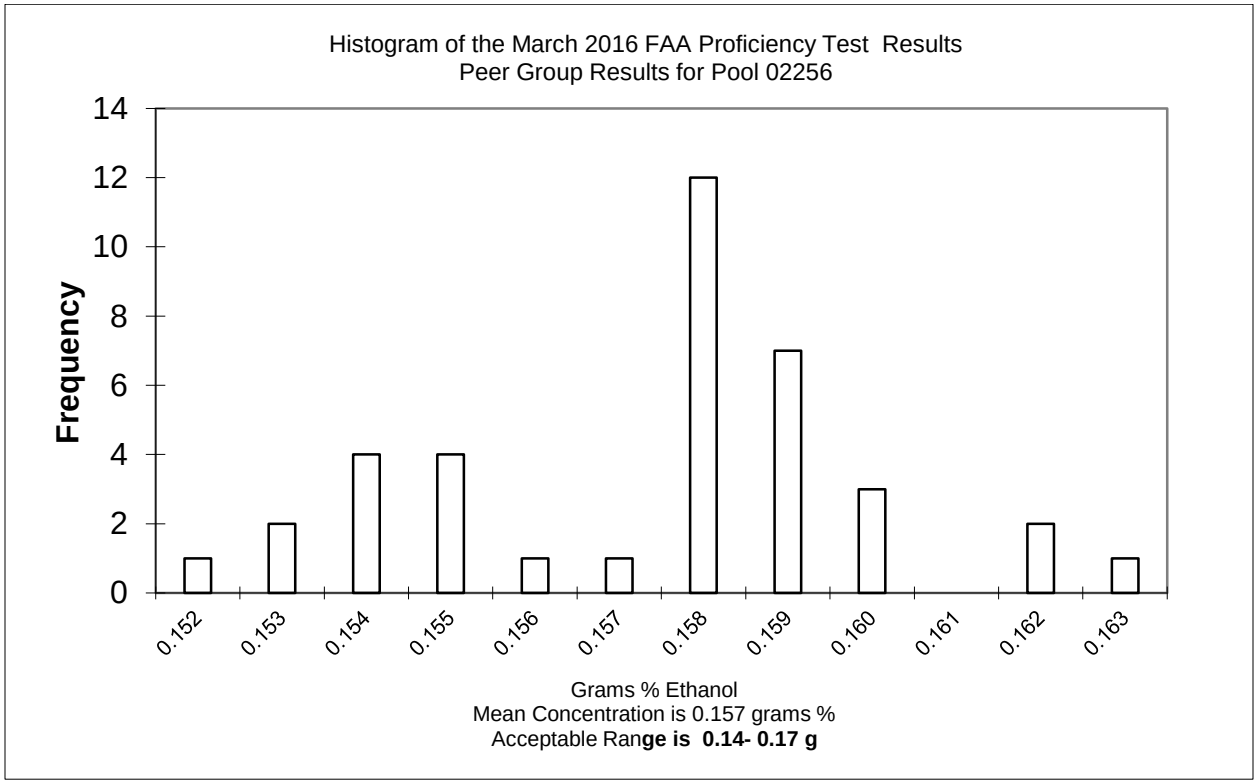


Figure 2

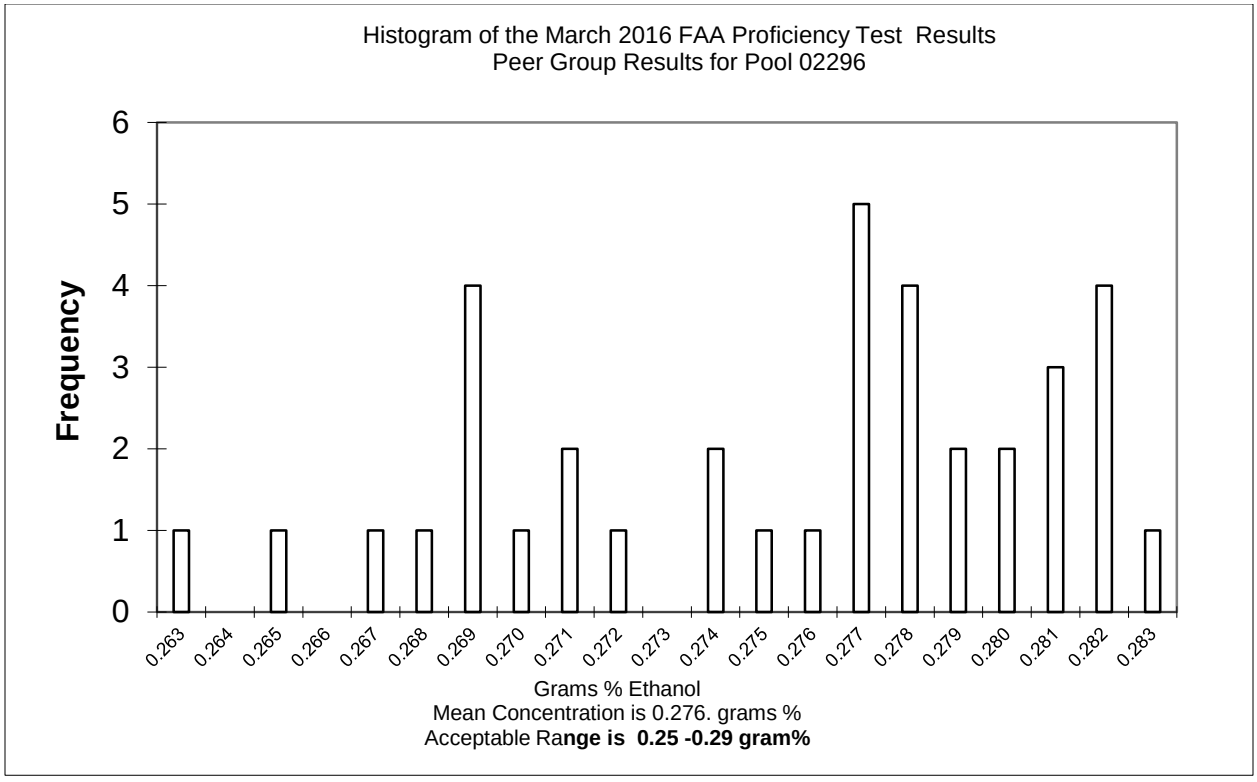


Figure 3 SigmaPlot analysis of pools 02256 & 02296

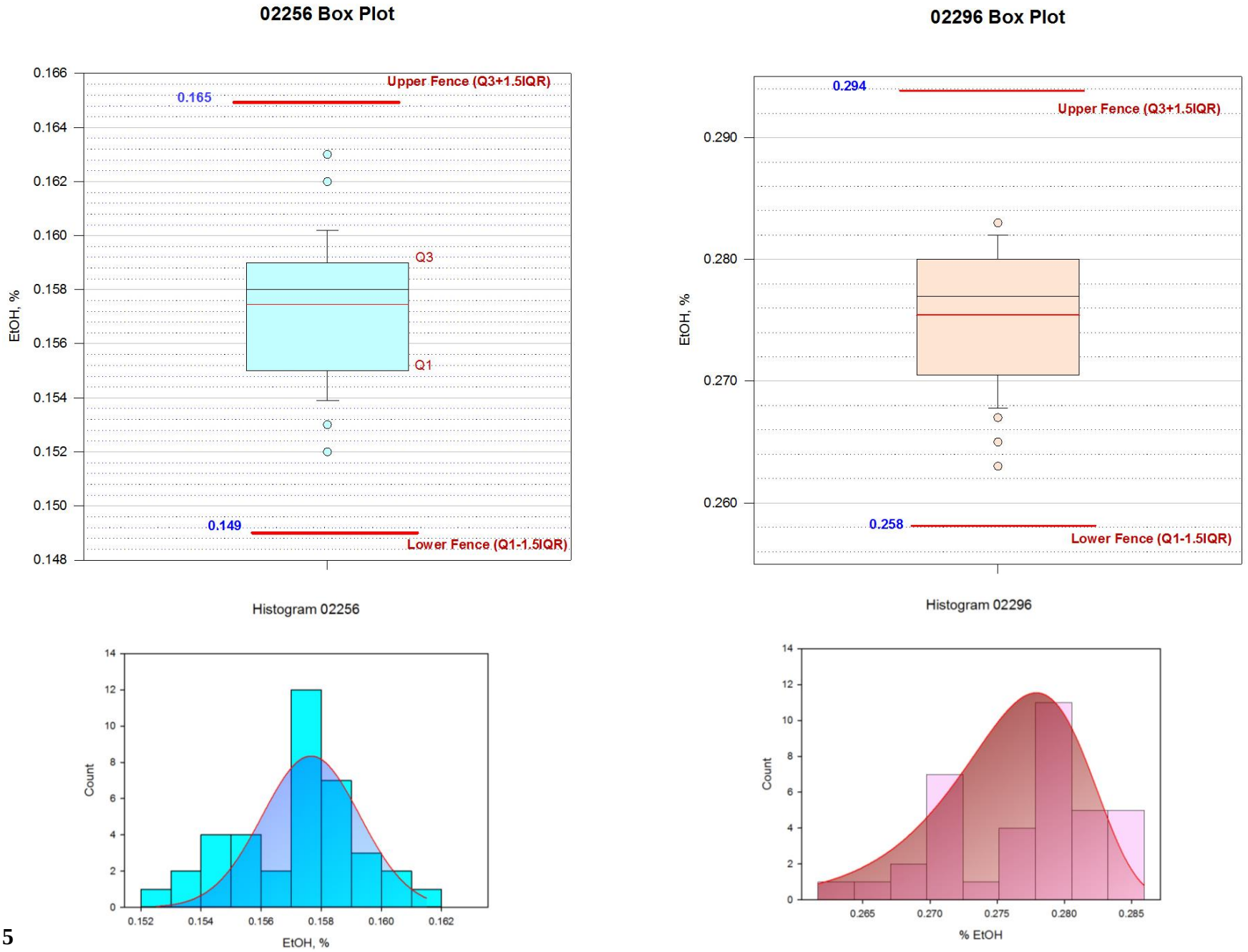
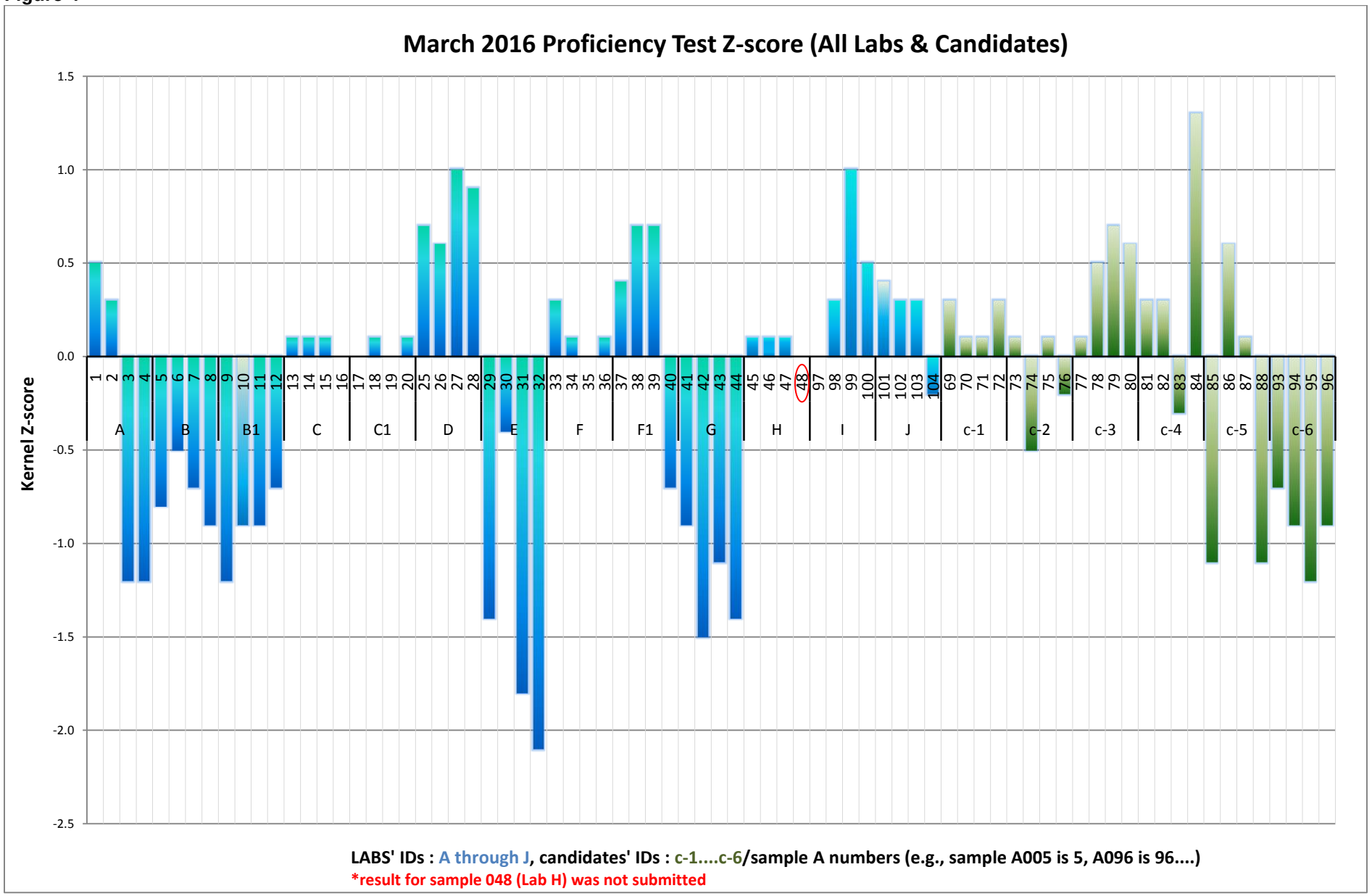


Figure 4


March 2016 Proficiency Test Z-score (All Labs & Candidates)
Kernel Z-score
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 97 98 99 100 101 102 103 104 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 93 94 95 96
A
B
B1
C
C1
D
E
F
F1
G
H
I
J
c-1
c-2
c-3
c-4
c-5
c-6
LABS' IDs : A through J, candidates' IDs : c-1....c-6/sample A numbers (e.g., sample A005 is 5, A096 is 96....)
*result for sample 048 (Lab H) was not submitted