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

Ron Chapman, MD, MPH
Director & State Health Officer

EDMUND G. BROWN JR.
Governor

September 14, 2012

TO: Forensic Alcohol Analysis Laboratories

SUBJECT: Assigned Values and Acceptable Ranges for July 2012 Proficiency Test in Forensic Alcohol Analysis

The Department has recently received several requests for more detailed descriptions of the statistical analysis of its proficiency test data. In response, we have provided the attached summary of the descriptive statistics for the July 2012 proficiency test. 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 a graphical summary 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. 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.

During the last 10 years, there have been significant developments in the statistical analysis of proficiency test data. There are now several International proficiency testing standards[1]. A number of new statistical techniques have been proposed for the evaluation of participant data. These techniques are intended to reduce the impact of outlier results, skewness, heavy tailing, and multi-modality on the descriptive statistics with the ultimate aim of providing the

[1] The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories (*IUPAC Technical Report*). The report is based on *International Organization of Standardizations (ISO)* guides, ISO/IEC 17043:2010 Conformity assessment -- General requirements for proficiency testing and ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons

best estimates of the central tendency of the reported results and the dispersion of these data. The Department has evaluated the use of some of these newer statistical procedures on the data obtained from California laboratories. Out of these evaluations, one clear conclusion is that the truncation of reported results to two decimal places and the use of truncated results to evaluate the proficiency test data is not appropriate. The use of the truncated data was originally proposed by the laboratories based on the reporting requirements of the regulations [cf. Title 17 §1220.4 (b)]. The reporting requirements are in fact consistent with the Vehicle Code per se and presumptive blood alcohol limits, which are expressed to the second decimal place. However, the truncation step is not appropriate for the evaluation of proficiency test data. In metrological terms, truncation can be viewed as introducing a significant amount of readability uncertainty. The current methods used to determine alcohol concentrations are certainly precise enough to make the third decimal place meaningful. For the present, the Department will continue to use its current evaluation procedures, with an eye toward modifying these procedures in the future. In the meantime, participants are advised to use the three decimal place results when evaluating their methods for bias or unduly large random variations.

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. With this technique, the proficiency test data are converted to a standard normal form. This is accomplished by dividing the error or difference in a reported result from the consensus value[2] by a standard deviation of the data. The primary advantage of *z*-scores is that they make all proficiency test results directly comparable regardless of concentration. A laboratory's performance can be easily interpreted from a *z*-score. Generally a score between -2 and +2 ($|z| \leq 2$) is considered satisfactory or acceptable. A *z*-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.

Various techniques have been proposed for determining the normalizing standard deviation. The Harmonized Protocol recommends the use of a "fitness-for purpose" based standard deviation for proficiency assessment (σ_P). Here, fitness-for purpose is defined as the standard uncertainty that is most appropriate for the application of the results of the analysis. The Department has determined a value for σ_P 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 [cf. Title 17 §1220.1.(a)(1)]. The Department found that a reasonable value for σ_P expressed as a relative standard deviation is 2.5%. The standard deviation can also be determined based on the data obtained from a

[2] Since the consensus value is a center of the statistical distribution, several measures of central tendency can be used (mean, median, robust mean, mode). If the distribution is roughly symmetrical and unimodal then all the central tendency characteristics are coincidental. The Harmonized Protocol recommends the use of the robust mean to determine the consensus value.

given proficiency test round. The Harmonized Protocol does not recommend this procedure noting that this can result in variations in the determinations of acceptable limits from test to test. In fact, the standard deviations for the data obtained in the July 2012 proficiency test were very close to the fitness-for purpose value determined by the Department.

The proficiency test results for the laboratory participants in the July 2012 test expressed as *z*-scores are summarized in Figure 3. Individual laboratories are identified by codes. An attachment to this letter provides your laboratory's code. The figure is provided for educational purposes only and was not used to formally evaluate a laboratory's performance.

For future laboratory-wide proficiency tests, the Department will provide summary statistics describing the assigned values and limits of acceptable performance.

Sincerely,

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

Enclosures:

Statistical Data for July 2012 Proficiency Test in Forensic Alcohol Analysis
CDPH July 2012 Proficiency Tests Lab Code(s)

Statistical Data for July 2012 Proficiency Test in Forensic Alcohol Analysis

CDPH Tier#1 and Tier #2 Acceptable Ranges

Pool	Peer Group Mean	Tier #1	Tier #2
#1	0.139	0.130 – 0.148	0.12 – 0.14
#2	0.210	0.197 – 0.223	0.19 – 0.22

Suimmary of Test Pool Data

Parameter	POOL 1 (06112)	POOL 2 (06182)
Pre-distribution Data	Target Value 0.14% True Value[1] 0.139 Standard Deviation[1] 0.0009	Target Value 0.21% True Value[1] 0.207 Standard Deviation[1] 0.0013
Descriptive statistics	Mean 0.139 Adjusted Mean[2] 0.139 Standard Error 0.0003 Median 0.139 Standard Deviation 0.0034 Minimum 0.129 Maximum 0.146 Count 128	Mean 0.209 Adjusted Mean[2] 0.210 Standard Error 0.0005 Median 0.210 Standard Deviation 0.0054 Minimum 0.192 Maximum 0.221 Count 132
Histogram	Figure 1	Figure 2
Normal distribution?	YES	NO
Robust mean, X*	0.139	0.210
Robust standard deviation, σ_{rob}	0.0032	0.0036
Fitness-for-purpose standard deviation, σ_p	0.0035	0.0052
Consensus value (X_a) determined as Robust mean (X*)	0.139	0.210
Standard error of consensus value, X_a , S.E.	0.0003	0.0005
$X_a \pm$ S.E.	0.139 ± 0.0003	0.210 ± 0.0005
z-score	$z = \frac{X - X_a}{\sigma_p}$	$Z = \frac{X - X_a}{\sigma_p}$

[1] Based on CDPH's Headspace Gas Chromatographic Method
[2] Mean determined from participant data after the removal of outlier(s)

Figure 1

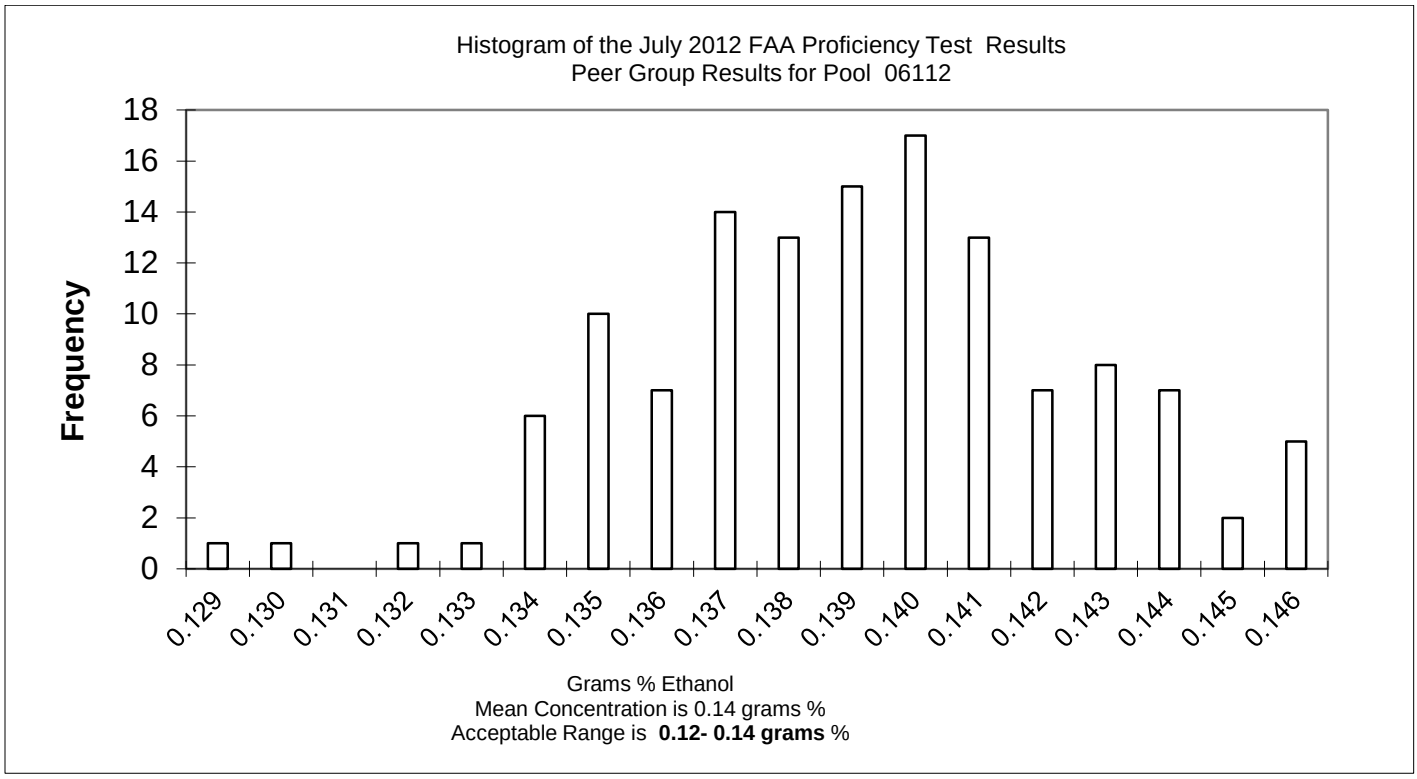


Figure 2

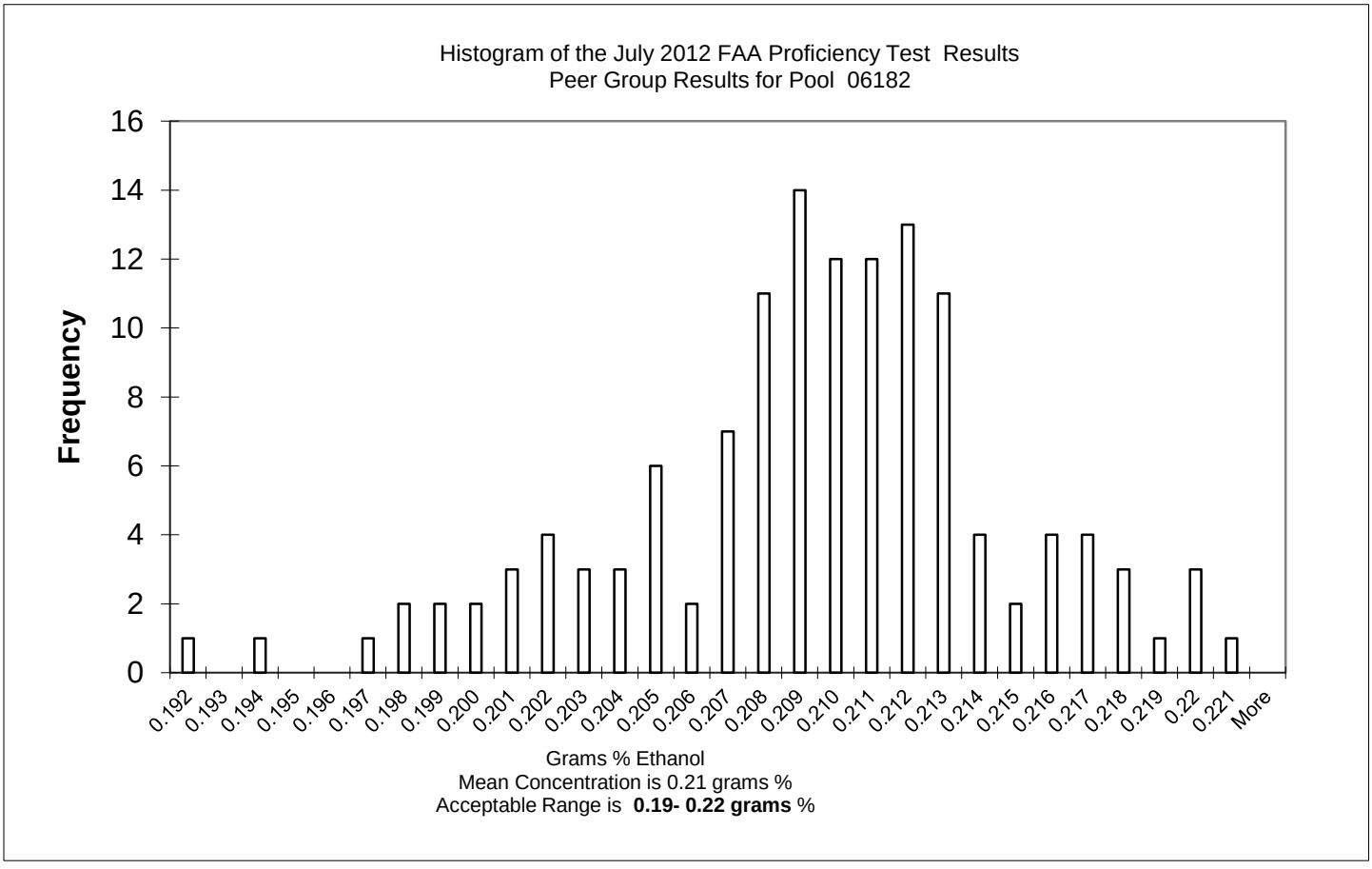
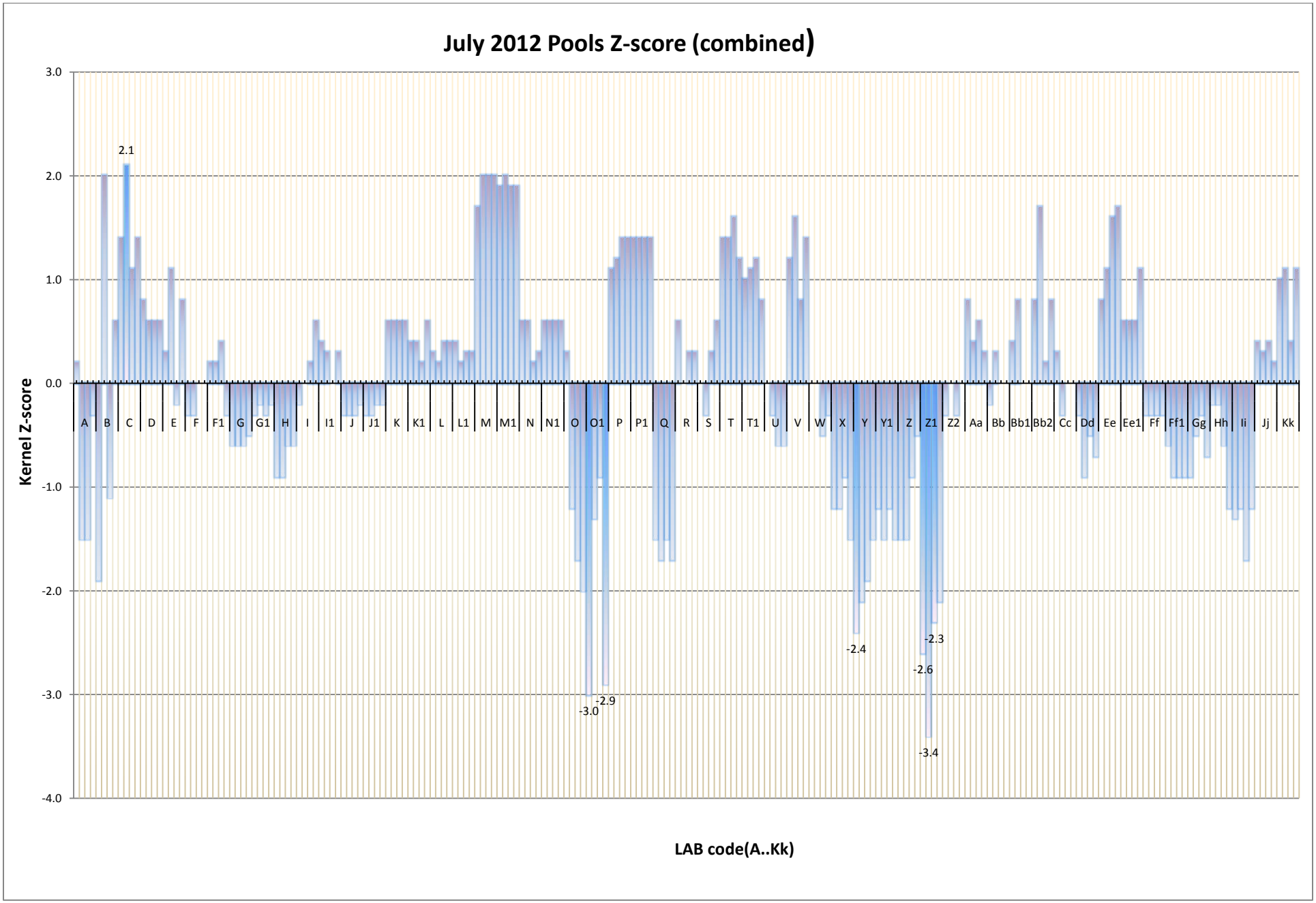


Figure 3


July 2012 Pools Z-score (combined)
Kernel Z-score
LAB code(A..Kk)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
2.1
-3.0
-2.9
-2.4
-2.3
-2.6
-3.4
A B C D E F F1 G G1 H I I1 J J1 K K1 L L1 M M1 N N1 O O1 P P1 Q R S T T1 U V W X Y Y1 Z Z1 Z2 Aa Bb Bb1 Bb2 Cc Dd Ee Ee1 Ff Ff1 Gg Hh Ii Jj Kk