*J Forensic Sci,* January 2006, Vol. 51, No. 1
doi:10.1111/j.1556-4029.2005.00008.x
Available online at: www.blackwell-synergy.com

*Rod G. Gullberg,*[1] *M.P.A. and Barry K. Logan,*[2] *Ph.D.*

# Results of a Proposed Breath Alcohol Proficiency Test Program

**ABSTRACT:** Although proficiency test programs have long been used in both clinical and forensic laboratories, they have not found uniform application in forensic breath alcohol programs. An initial effort to develop a proficiency test program appropriate to forensic breath alcohol analysis is described herein. A total of 11 jurisdictions participated in which 27 modern instruments were evaluated. Five wet bath simulator solutions with ethanol vapor concentrations ranging from 0.0254 to 0.2659 g/210 L were sent to participating programs, instructing them to perform $n = 10$ measurements on each solution using the same instrument. Four of the solutions contained ethanol only and one contained ethanol mixed with acetone. The systematic errors for all instruments ranged from $-11.3\%$ to $+11.4\%$ while the coefficient of variations ranged from zero to 6.1%. A components-of-variance analysis revealed at least 79% of the total variance as being due to the between-instrument component for all concentrations. Improving proficiency test program development should consider: (1) clear protocol instructions, (2) frequency of proficiency testing, (3) use lower concentrations for determining limits-of-detection and -quantitation, etc. Despite the lack of a biological component, proficiency test participation should enhance the credibility of forensic breath test programs.

**KEYWORDS:** forensic science, breath alcohol, proficiency tests

Proficiency testing has long been recognized among analytical chemists as useful for evaluating instrumental, method, laboratory and program performance. Most forensic toxicology laboratories performing alcohol and drug analysis participate regularly in proficiency test programs designed to evaluate compliance with minimum analytical standards. Laboratories shown to comply with minimum performance standards generally have enhanced confidence in their analytical work, which can then be shared with their customers.

Although common in forensic toxicology laboratories, proficiency testing has not been routinely used among forensic breath alcohol test programs. There are several reasons for this, including: (1) the inability to provide uniform samples (e.g., homogeneous human breath) to several different programs in the same manner in which intoxicated subjects provide breath samples, (2) the lack of well-defined sites since breath test instruments are used in a variety of locations and environments, (3) the difficulty in having those actually performing forensic breath alcohol measurement (e.g., police officers in most cases) perform the proficiency tests, and (4) the fact that most programs use up to several hundred instruments dispersed throughout the jurisdiction. Even if breath samples from intoxicated subjects were to be preserved in some appropriate manner, they would not be introduced into the instruments in the same manner that a subject provides the sample, combining all of its sources of variation. Indeed, the breath-sampling component contributes the largest proportion to the total method variability (1). Despite these limitations, some elements of proficiency testing can be used in forensic breath test programs. Control standard solutions containing ethanol and water can be used to evaluate instrument performance, at least to a limited extent. Interfering compounds (e.g., acetone) can also be introduced to determine instrument response. Several statistical methods can be applied to evaluate the accuracy, precision, levels-of-detection (LOD), levels-of-quantitation (LOQ), linearity, etc. To our knowledge, no one has developed an interjurisdictional proficiency test program to evaluate some of these breath test analytical characteristics. Our purpose here is to describe the design, results, and limitations of a recently performed proficiency test program for breath alcohol instruments and suggest some improvements for future development.

[1] Washington State Patrol, Seattle, WA 98102.
[2] Washington State Toxicology Laboratory, Seattle, WA 98134-2027.

Received 6 Feb. 2005; and in revised form 4 Aug. 2005; accepted 20 Aug. 2005; published 26 Dec. 2005.

## Methods

Thirteen jurisdictions using 27 different instruments participated in the proficiency-testing program. Five solutions containing ethanol in water or ethanol+acetone in water were prepared by the organizing laboratory (Washington State Toxicology Laboratory, Seattle, WA) and were placed in 500 mL plastic bottles and labeled with an identification number. The solutions, designed for use in wet bath simulator devices, were analyzed by gas chromatography to determine their ethanol and/or acetone concentrations and provide the reference value. Aliquots (200 μL) of the simulator solution were pipetted with an automated pipettor/diluter (Hamilton, Co., Reno, NV) mixed with 2 mL of internal standard solution (20 mg sodium chloride, 0.3 mL *n*-propanol, q.s. to 2 L), and dispensed into 10 mL headspace vials and sealed. Analysis was performed on an automated gas chromatograph with a headspace autosampler (Agilent, Palo, Alto, CA). The gas chromatographic conditions were as follows: 30 m DBALC1 megabore (0.53 mm i.d.), 30 m DB ALC2 megabore (0.53 mm i.d.), run isothermally at 40 and 37°C, respectively. Instruments were calibrated daily, and controls were run on every tenth injection. Control results must lie within 0.01 g/100 mL of the target value. Using a water/air partition coefficient for ethanol of 1.23 (2), the simulator solutions were determined to generate vapor reference values of: 0.0254 g/210 L ethanol, 0.0807 g/210 L ethanol, 0.1420 g/210 L ethanol, 0.2659 g/210 L ethanol and 0.0810 g/210 L ethanol with 0.0960 g/210 L acetone. The solutions were then distributed to the participating laboratories during October

2001. The participating laboratories were provided with written instructions for performing the testing protocol. They were to select one typical field evidential instrument on which to perform all analyses. They were to set their instruments to perform $n = 10$ measurements on each of the five simulator solutions. No information was obtained on how the jurisdictions may have calibrated their instruments—specifically regarding the water/air partition coefficient they may have used. They were not given instructions regarding the type or model of simulator device to use nor were they told to use the same simulator device for all analyses. A small fraction of the measurement variability can arise from the simulator and its configuration. For consistency, the participants were instructed to record their measurement results to three decimal places.

The participating programs were from the following jurisdictions: Sweden, Kentucky, Iowa, South Carolina, Nevada, Wyoming, Arizona, New York, Missouri, Oregon, Indiana, Wisconsin, and Washington. The four types of instruments used were: Intoxilyzer 5000 (CMI Inc., Owensboro, KY), BAC Datamaster (National Patent Analytical Systems Inc., Mansfield, OH), Intoximeter ECIR (Intoximeters, Inc., St. Louis, MO) and Alcotest 7110 (Drager Safety, Inc., Durango, CO).

Several statistical analyses were performed to evaluate the proficiency test results. These were designed to evaluate the systematic error, precision, and components of variance.

The objective of components-of-variance analysis is to determine the percentage of total measurement variance ($\sigma_T^2$) contributed by the between-instrument ($\sigma_{Between}^2$) and within-instrument ($\sigma_{Within}^2$) components in the equation: $\sigma_{Total}^2 = \sigma_{Between}^2 + \sigma_{Within}^2$. A random effects model is assumed for this analysis (3). Statistical analyses were performed using Microsoft Excel 2000 (Microsoft, Inc., Redmond, WA) and SPSS V 11.0 (SPSS, Inc., Chicago, IL). Graphic plots were generated using SigmaPlot V 2.01 (Systat Software, Inc., Point Richmond, CA).

## Results

Table 1 summarizes the results by pooling the data for all instruments. Although testing began with 27 instruments, only 25 instruments completed all levels of the test protocol. The results show that 15 of the 27 instruments did not detect the presence of an interfering substance on any of the ten tests performed on the solution containing the acetone.

Figure 1 plots the percent systematic error for the mean of $n = 10$ measurements performed by each instrument against the reference concentration. The solid circles represent those solutions containing ethanol only, while the open circles represent the one solution containing ethanol and acetone. For display purposes the reference value was set slightly higher for the solution with ethanol and acetone (open circles) only to separate them from the solid circle symbols and enhance interpretation. The systematic errors for the ethanol only solutions range from $-11.3\%$ to $+11.4\%$ while they ranged from $-11.1\%$ to 12.2% for the ethanol plus acetone solution. Although most of the larger systematic errors are seen to be at the lowest concentration, the largest negative error ($-11.3\%$) occurred at the highest concentration.

Figure 2 plots the standard deviation estimates against the means for each instrument performing $n = 10$ measurements at each of the four concentrations containing ethanol-only solutions. These estimates corresponded to the systematic error estimates shown in Fig. 1 (solid circles). As expected, the larger standard deviation estimates occurred at the highest concentration. The

TABLE 1—*Summary of all simulator test results with data pooled according to instrument type.*

| Type and Number of Instruments* | Reference† (g/210 L) | Mean (g/210 L) | Systematic Error Range (%) | Standard Deviation Range (g/210 L) | Coefficient of Variation (%) |
|---|---|---|---|---|---|
| BAC Datamaster | | | | | |
| 17 | 0.0254 | 0.0253 | −9.4 to 11.4 | 0–0.0016 | 0–6.1 |
| 17 | 0.0807 | 0.0806 | −9.3 to 5.2 | 0.0004–0.0023 | 0.5–2.8 |
| 16 | 0.1420 | 0.1392 | −10.2 to 3.8 | 0.0003–0.0026 | 0.2–1.9 |
| 15 | 0.2659 | 0.2612 | −11.3 to 2.9 | 0–0.0025 | 0 to 1.0 |
| 7‡ | 0.0810§ | 0.0857 | −9.7 to 12.2 | 0.0003–0.0021 | 0.3–2.8 |
| Intoxilyzer 5000 | | | | | |
| 6 | 0.0254 | 0.0244 | −9.8 to −1.6 | 0–0.0009 | 0–3.6 |
| 6 | 0.0807 | 0.0809 | −1.4 to 2.4 | 0.0005–0.0013 | 0.6–1.6 |
| 6 | 0.1420 | 0.1430 | −2.1 to 2.5 | 0–0.0013 | 0–0.9 |
| 6 | 0.2659 | 0.2680 | −3.4 to 2.1 | 0.0007–0.0024 | 0.3–0.9 |
| 4‡ | 0.0810§ | 0.0796 | −11.1 to 4.4 | 0.0005–0.0011 | 0.6–1.3 |
| Intoximeter ECIR | | | | | |
| 3 | 0.0254 | 0.0267 | −2.0 to 2.9 | 0.0003–0.0008 | 1.2–2.9 |
| 3 | 0.0807 | 0.0829 | −3.5 to 7.1 | 0.0005–0.0010 | 0.6–1.3 |
| 3 | 0.1420 | 0.1418 | −3.2 to 4.6 | 0.0009–0.0014 | 0.6–1.0 |
| 3 | 0.2659 | 0.2707 | −3.5 to 5.0 | 0.0010–0.0034 | 0.4–1.3 |
| 3‡ | 0.0810§ | 0.0814 | −1.0 to 2.7 | 0.0010–0.0015 | 1.2–1.8 |
| Alcotest 7110¶ | | | | | |
| 1 | 0.0254 | 0.0232 | −8.7 | 0.0004 | 1.8 |
| 1 | 0.0807 | 0.0778 | −3.6 | 0.0008 | 1.0 |
| 1 | 0.1420 | 0.1375 | −3.2 | 0.0010 | 0.7 |
| 1 | 0.2659 | 0.2585 | −2.8 | 0.0015 | 0.6 |
| 1‡ | 0.0810§ | 0.0792 | −2.2 | 0.0010 | 1.3 |

*Ten measurements performed on each instrument.

†Vapor ethanol estimates determined from solution concentrations measured by headspace gas chromatography and assuming a water/air partition coefficient at 34°C of 1.23.

‡Includes only those instruments not detecting the presence of the acetone.

§Solution also containing 0.096 g/210 L vapor acetone.

¶Only the infrared results considered.

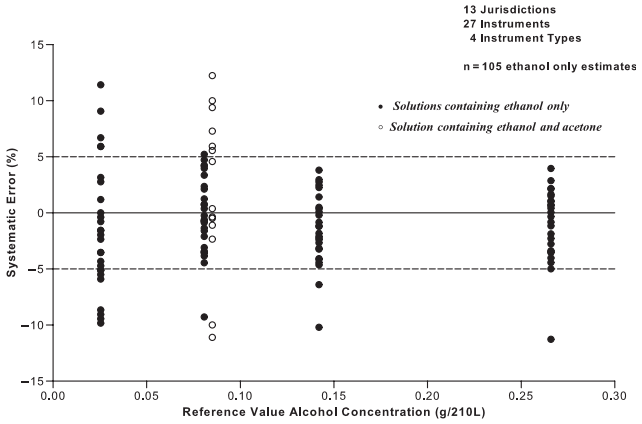


FIG. 1—*Percent systematic error for each instrument being tested at each of the four ethanol only concentrations along with that for the solution containing ethanol and acetone.*

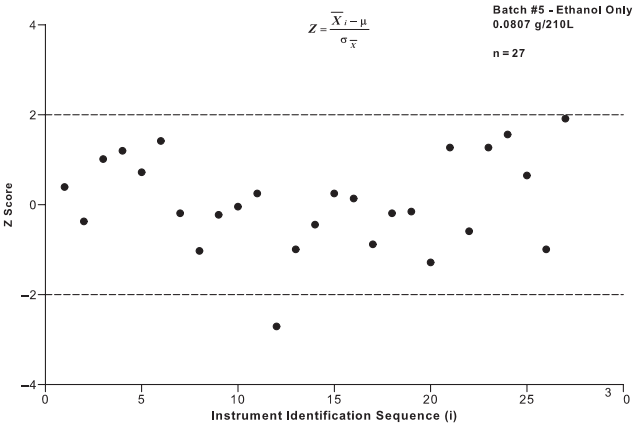


FIG. 3—*Plot of Z-scores for each instrument along with ± 2σ limits resulting from the replicate measurement of an ethanol only solution having a reference value of 0.0807 g/210 L.*

standard deviation estimates ranged from 0.0 to 0.0034 g/210 L—all below the National Highway Traffic Safety Administration (NHTSA) standard of 0.0042 g/210 L (4).

Figure 3 shows the Z-scores plotted against an instrument identification number for the ethanol only solution with a concentration of 0.0807 g/210 L. Z-score plots generated for the other concentrations appeared essentially the same. Z-scores are useful for identifying outliers. The one instrument falling outside the Z-score limits of ± 2 was an instrument having a systematic error of − 9.3%. This instrument would not have complied with the NHTSA standard for accuracy.

Table 2 summarizes the results of the components-of-variance analysis. The concentration reference values are shown along with the percent of the total measurement variation contributed by the between-instrument ($\sigma^2_{\text{Between}}$) and within-instrument ($\sigma^2_{\text{Within}}$) components. The between-instrument component clearly dominates when measuring simulator standards by contributing at least 79% to the total variance. This reflects the high within-instrument precision characterizing each of these current generation instrument types. Figure 4 plots the mean ± 2 SD for each instrument measuring the solution with a reference concentration of 0.1420 g/210 L. Figure 4 illustrates graphically the results found in Table 2, revealing the greater variation between instruments relative to that within instruments. The mean results among the different instruments are observed to vary more than the data within each instrument as indicated by the error bars. Table 2 also reveals that the percentage of total measurement variance contributed by the between-instrument component increases with concentration. This may reflect instrument calibration differences being amplified at higher concentrations. A plot such as the one on Fig. 4 is also useful for identifying general method bias (5). Among all four ethanol concentrations, the maximum general method bias was − 1.5%, occurring at the concentration shown in Fig. 4.

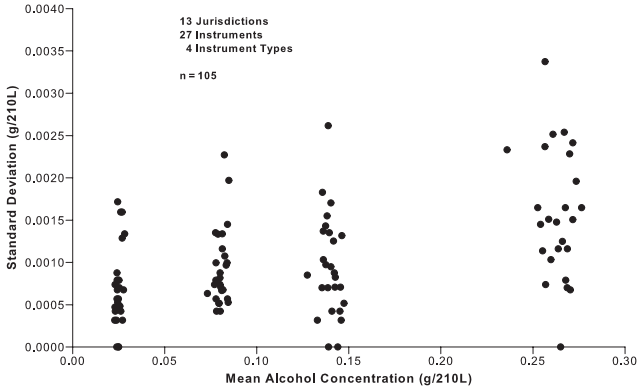


FIG. 2—*Standard deviations resulting from each instrument's replicate measurements at each of the four ethanol-only concentrations.*

## Discussion

The data presented here are actually the result of the third attempt to perform an interjurisdictional proficiency test program. Several things were learned from the first two attempts that were adjusted for in the present study. During the first study, 18 different simulator solutions were prepared with four being delivered for testing on 25 different instruments. Therefore, not every solution was tested on each instrument. In addition, most instruments measured two solutions containing acetone. Another problem resulted when some of the solutions froze during shipment—yielding low results due to the loss of ethanol because of different freezing temperatures for ethanol and water. These experiences resulted in the following adjustments for the present

TABLE 2—*Percentage of between- and within-instrument components of variance determined for each of the four solution concentrations.*

| Concentration* (g/210 L) | Between-Instrument† $\hat{\sigma}^2_{\text{between}}$ (%) | Within-Instrument† $\hat{\sigma}^2_{\text{between}}$ (%) |
|---|---|---|
| 0.0254 ($n = 27$) | 79.7 | 20.3 |
| 0.0807 ($n = 27$) | 89.3 | 10.7 |
| 0.1420 ($n = 26$) | 95.2 | 4.8 |
| 0.2659 ($n = 25$) | 96.8 | 3.2 |

*Reference concentration of the solution as determined by gas chromatography.

†The components of variance were determined according to:

$$\hat{\sigma}^2_{\text{within}} = \text{MS}_{\text{within}}, \quad \sigma^2_{\text{between}} = \frac{\text{MS}_{\text{between}} - \hat{\sigma}^2_{\text{within}}}{n}$$

where MS is the mean square and $n$ the number of measurements within each instrument.

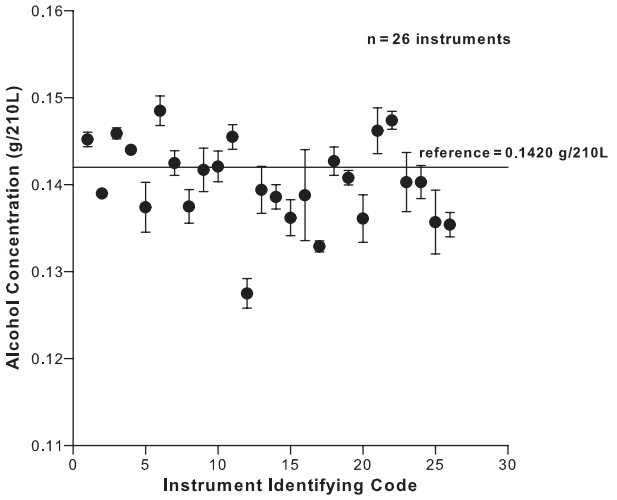


FIG. 4—*Plot of the mean ± 2 SD for each instrument measuring the ethanol-only solution having a reference value of 0.1420 g/210 L.*

study: (1) shipment of solutions during a time of the year when freezing would be unlikely, (2) having all instruments measure all solutions representing the relevant forensic concentration range, (3) use only one solution containing acetone that was close to the typical threshold of 0.01 g/210 L ethanol equivalent, and (4) further clarify the protocol for the participating laboratory.

Acetone has been generally considered the only endogenous volatile organic compound (VOC) that could even remotely be considered a potential interfering substance in forensic breath alcohol analysis. The acetone concentration used in the present study was 0.096 g/210 L vapor, corresponding to 457 µg/L. Because of the different absorbance of acetone relative to ethanol at the selected infrared frequencies, this concentration yields approximately 0.01 g/210 L ethanol equivalent, the typical threshold in most evidential infrared breath test instruments. Previous work has shown that approximately 600 µg/l of acetone is required to yield a 0.01 g/210 L ethanol equivalent (6). Moreover, a study analyzing the blood for acetone of 500 drunk drivers in Sweden revealed the maximum to be 61.9 mg/L (7). This would theoretically yield a breath acetone concentration of approximately 206.3 µg/l (assuming a blood/breath measurement ratio of 300:1), well below that necessary to yield a 0.01 g/210 L ethanol equivalent. Since the acetone concentration used in this study was close to the 0.01 g/210 L ethanol equivalent threshold used by most instruments, it was not surprising that 15 of the 27 instruments failed to detect its presence. Although some of the systematic errors observed in Fig. 1 exceed 10% in the presence of acetone, they are explainable for instruments that do not subtract the contribution from acetone below 0.01 g/210 L ethanol equivalent. Adding 0.009 g/210 L from undetected acetone along with 0.003 g/210 L from being calibrated 4% high at the 0.080 g/210 L level will result in a systematic error of +15%.

All instruments evaluated during this proficiency test program appear on the NHTSA Conforming Products List (8). Since only one instrument of each model is generally evaluated by NHTSA, there is no assurance that each instrument of that same model would comply. Of the 25 instruments completing all aspects of the present study, only one would not have complied with the NHTSA standards for accuracy (4). All of the instruments would have complied with the NHTSA standards for precision.

Determining which component contributes most to total measurement variability—between-instrument or within-instrument—is a useful result of proficiency testing (9). The appropriate statistical method to use is the analysis of the components-of-variance in which we assume that both the instruments and the measurement results are random variables (the random effects model). Table 2 along with Fig. 4 summarizes these results. Replicate simulator tests result in an unusually high precision and, therefore, much of the total variance is determined by the between-instrument differences relative to the within-instrument variation. Table 2 also reveals that the proportion of the between-instrument variance component increased with the concentration. There was clearly a larger span for the mean results at the 0.2659 g/210 L concentration (span was 0.0433 g/210 L) compared with the 0.0254 g/210 L concentration (span was 0.0054 g/210 L). Differences between instrument calibration methods and assumed water/air partition factors can account for some of this (2). In addition, the instruments may have varied in their linearity characteristics, yielding a broader range of responses as the concentration increased while still maintaining a relatively high precision. Clearly, different mean results between instruments combined with a high precision (small CV estimates) assign a greater proportion of total variance to the between-instrument component. This same partitioning of variance, however, would not likely occur when measuring replicates of human breath where the within-instrument variation (due to sampling) would be significantly larger compared with simulator results. No inferences regarding the partitioning of variance in a biological context can be made from the present study.

Several things were learned from this effort that could be included in future breath alcohol proficiency test programs. First, it is very important that a clear set of instructions and test protocol be sent to all participants. The instructions might include items such as: (1) set the instrument interference threshold to 0.01 g/210 L, (2) prime the simulator tubing with the vapor effluent prior to performing the first measurement, (3) note the date and time when each set of ten measurements were performed, (4) if you have questions about the protocol please call the coordinating laboratory, (5) record the type of simulator device used, (6) record the water/air partition coefficient assumed at calibration and the initial results, and (7) please suggest future protocol improvements, etc. Indeed, clear instructions enhance the interpretation and informative value of results. Another improvement might be to have all instruments perform 10 measurements on a water-only solution, ensuring acceptable blank sample results. Another element to be decided is the appropriate frequency of proficiency test participation. Although many analytical laboratories might participate in proficiency testing several times per year (10), it might be appropriate to begin with annual participation in breath test programs.

As legal breath alcohol limits are reduced (e.g., 0.02 g/210 L for those under age 21), estimates of the LOD and LOQ become important considerations. The present study was not designed to determine LOD and LOQ estimates where $LOD = 3S_0$ and $LOQ = 10S_0$ ($S_0$ = standard deviation estimated at zero concentration) (11). However, if we simply consider the SD estimates for the measurement results at 0.0254 g/210 L (the lowest concentration evaluated), which ranged from 0 to 0.0016 g/210 L, we find that the maximum LOD and LOQ estimates would be 0.0048 and 0.016 g/210 L, respectively. Although these estimates are expected to increase on using actual breath test results, they nevertheless provide a rough estimate of the instrumental LOD and LOQ.

Although all but one of the participating instruments demonstrated forensic acceptability regarding accuracy and precision, caution must be exercised when interpreting the present results. The analytical properties of one instrument cannot be generalized

to all instruments of the same type. Nor should one infer from these results that all breath alcohol results in the jurisdiction will be valid and fit for purpose simply because the one instrument performed well. Forensic breath test programs must implement carefully designed protocols ensuring the accuracy, precision, and legal acceptability of each individual subject tested. A sound breath test protocol including trained operators, minimum pre-exhalation observation time, duplicate analyses, control standard testing, and printout of results are highly recommended and infer fitness-for-purpose on an individual subject's test result (12). Compliance with a proficiency test program can provide, however, further but limited information regarding an individual's breath test result and associated fitness for purpose. Indeed, proficiency testing, should be considered supplementary to an already carefully designed program combining quality instrumentation with sound testing protocol. Participation in proficiency testing, however, can enhance the program's credibility overall by providing relevant information for the court to consider when determining the weight of the evidence. Proficiency test participation may also become a necessary component of future breath test accreditation programs. Moreover, with an increased emphasis on quality control and the use of Six Sigma metrics in the analytical sciences, proficiency test results are often the basis for estimating the total allowable error for a method (13).

Along with enhancing the quality assurance of a sound breath test program, administrators must expect that it will be necessary to provide proficiency test results to the defense counsel. While providing the defense with additional material from which to argue the weight of the evidence, proficiency test participation and results should generally bolster the program overall. Such materials can easily be provided through Internet sources.

## Conclusions

Participation in proficiency testing can usefully augment an already sound breath test program. Although the present study was not able to include the important component of breath sampling from an intoxicated subject, it provides an initial approach by considering other important elements such as instrument accuracy and precision, potential interfering substance bias, components of variance, data analysis and interpretation of results. Future efforts can build on this by incorporating additional instrumental and program features that might include: (1) error detection mechanisms, (2) instrument performance in their normal field environment, (3) instrument performance over a longer time interval, (4) instrument compliance with local quality control standards, and (5) determining LOD and LOQ, etc. Clinical and forensic laboratories have a long history of participation in proficiency test programs and much can be learned and applied from their experiences. In an effort to meet increasing defense challenges in forensic breath alcohol testing, the time may be appropriate to enhance an already sound program with proficiency test participation.

## References

1. Gullberg RG, Logan BK. Reproducibility of within-subject breath alcohol analysis. Med Sci Law 1999;38:157–62.
2. Jones AW. Determination of liquid/air partition coefficients for dilute solutions of ethanol in water, whole blood, and plasma. J Anal Toxicol 1983;7:193–7.
3. Snedecor GW, Cochran WG. Statistical methods. 6th ed. Ames, IA: The Iowa State University Press; 1967.
4. National Highway Traffic Safety Administration. Highway safety programs; model specifications for devices to measure breath alcohol. Fed Regist 1993;58:48705–10.
5. Youden WJ. How to evaluate accuracy. Mater Res Standards 1961; 1268–71.
6. Dubowski KM, Essary NA. Response of breath-alcohol analyzers to acetone: further studies. J Anal Toxicol 1984;8:205–8.
7. Jones AW, Sagarduy A, Ericsson E, Arnqvist HJ. Concentrations of acetone in venous blood samples from drunk drivers, type-1 diabetic outpatients, and healthy blood donors. J Anal Toxicol 1993;17:182–5.
8. National Highway Traffic Safety Administration. Highway safety programs; model specifications for devices to measure breath alcohol. Fed Regist 1999;64:30097–100.
9. Mandel J. Repeatability and reproducibility. J Qual Tech 1972;4:74–85.
10. Lawn RE, Thompson M, Walker RF. Proficiency testing in analytical chemistry. Cambridge: The Royal Society of Chemistry; 1997.
11. Taylor JK. Quality assurance of chemical measurements. Chelsea, MI: Lewis Publishers, Inc.; 1987.
12. Dubowski KM. Quality assurance in breath-alcohol analysis. J Anal Toxicol 1994;18:306–11.
13. Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. Arch Pathol Lab Med 2000;124:516–9.

Additional information and reprint requests:
Rod G. Gullberg, M.P.A.
Washington State Patrol
811 East Roanoke
Seattle, WA 98102
E-mail: Rod.Gullberg@wsp.wa.gov