Calculation and Verification of Blood Ethanol Measurement Uncertainty for Headspace Gas Chromatography*

Jason H. Sklerov† and **Fiona J. Couper**

Washington State Patrol, Forensic Laboratory Services Bureau, Toxicology Laboratory Division, 2203 Airport Way South, Seattle, Washington 98134

Abstract

An estimate was made of the measurement uncertainty for blood ethanol testing by headspace gas chromatography. While uncertainty often focuses on compliance to a single threshold level (0.08 g/100 mL), the existence of multiple thresholds, related to enhanced sentencing, subject age, or commercial vehicle licensure, necessitate the use of an estimate with validity across multiple specification levels. The uncertainty sources, in order of decreasing magnitude, were method reproducibility, linear calibration, recovery, calibrator preparation, reference material, and sample preparation. A large set of reproducibility data was evaluated (n = 15,433) in order to encompass measurement variability across multiple conditions, operators, instruments, concentrations and timeframes. The relative, combined standard uncertainty was calculated as ±2.7%, with an expanded uncertainty of ±8.2% (99.7% level of confidence, k = 3). Bias was separately evaluated through a recovery study using standard reference material from a national metrology institute. The uncertainty estimate was verified through the use of proficiency test (PT) results. Assigned values for PT results and their associated uncertainties were calculated as robust means (x*) and standard deviations (s*) of participant values. Performance scores demonstrated that the uncertainty estimate was appropriate across the full range of PT concentrations (0.010–0.370 g/100 mL). The use of PT data as an empirical estimate of uncertainty was not examined. Until providers of blood ethanol PT samples include details on how an assigned value is obtained along with its uncertainty and traceability, the use of PT data should be restricted to the role of verification of uncertainty estimates.

Introduction

Measurement uncertainty is defined as the "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand" (1). International standards on this topic are extensive (2–6) and have become enmeshed in the standard operating procedures of many disciplines engaged in calibration and testing activities. To date, there have been relatively few measurement uncertainty publications specifically targeting the forensic toxicologist's needs and most of those have only recently been published (7–14). Measurement uncertainty is one component of ISO 17025 (15) a document which describes the general requirements for demonstration of laboratory competence. Increasingly, these requirements are being utilized by organizations that accredit forensic laboratories.

The results of a forensic examination are often used to show compliance to a legal specification limit, but more frequently to demonstrate when a limit has been exceeded. For forensic toxicologists measuring blood alcohol concentration (BAC), it may be necessary to consider multiple specification limits depending upon on the circumstances of a particular case. For motor vehicle operation, compliance to a BAC specification of less than 0.080 g/100 mL may apply, whereas compliance to a lower BAC is required for commercial vehicle operators (0.040 g/100 mL) or underage drivers (0.020 g/100 mL). In jurisdictions with stricter sentencing for elevated BAC, a specification limit of 0.15 or 0.20 g/100 mL may apply. It is at these threshold levels where estimating measurement uncertainty becomes most relevant.

The output of such estimates is the production of a range of values which encompass the true BAC with an associated confidence interval. Take, for example, a BAC reported as "0.089 ± 0.0073 g/100 mL (k = 3, 99.7% level of confidence)". The range of values encompassing the true BAC spans from 0.0817 g/100 mL to 0.0963 g/100 mL. When reported in this way a Forensic Toxicologist may render the opinion that the BAC exceeded 0.08 g/100 mL with a statistical confidence of at least 99.7%. Multiple BAC thresholds may require individual uncertainty estimates at each level as these tend to vary by concentration. Alternatively, a relative uncertainty estimate may be obtained for use across the entire reporting range. In this work a relative estimate of BAC measurement uncertainty was produced using

* The opinions expressed in this publication are those of the authors and may not reflect those of the Washington State Patrol.

† Author to whom correspondence should be addressed. Email: Jason.Sklerov@wsp.wa.gov.

Type A and Type B evaluations and a bottom-up approach.

Type A evaluations of uncertainty are those done by statistical analysis of a series of measurements, and Type B evaluations are those by non-statistical means. An example of a Type A evaluation would be the variance obtained from repeated measurements of a quality control sample while a Type B example would be the uncertainty appearing on the calibration report from equipment used in sample preparation. The process of developing an uncertainty estimate is referred to colloquially as either bottom-up or top-down. In the top-down, or empirical approach, a source of combined variances is identified and taken as representative of the uncertainty of the entire measurement process. Interlaboratory comparison studies are prime candidates for use in top-down estimates as they include variances from different operators, instruments, calibrations, environmental conditions, or materials. The bottom-up approach takes a detailed look at each potential source of variance in a measurement, regardless of its significance, and quantifies its contribution to the overall uncertainty. All uncertainty sources are then added together to form a combined standard uncertainty. This was the approach in the present work. As previously noted (12), by investigating uncertainty components individually a laboratory is able to improve the measurement process by reducing variability where it is most pronounced. Regardless of the approach taken, any estimate of BAC uncertainty should undergo some type of verification to ensure it has not been underestimated.

Proficiency test data was used to verify the correctness of the uncertainty estimated in this work. Performance scores were calculated for whole blood proficiency test results using the ratio of the laboratory BAC measurement error to the BAC uncertainty estimate. Several shortcomings were identified in how current proficiency testing schemes report results and associated traceability to their customers.

Experimental

Materials

Two-hundred-proof ethanol was purchased from either Spectrum Chemicals and Laboratory Products (Gardena, CA) or AAPER Alcohol and Chemical (Shelbyville, KY). Aqueous ethanol reference materials were purchased from Restek Corporation (Bellefonte, PA) and SRM2893 ethanol standard reference material was purchased from the National Institute of Standards and Technology (NIST, Gaithersburg, MD). The internal standard, *n*-propanol, was obtained from Mallinckrodt Baker (Phillipsburg, NJ) and sodium chloride was obtained from Fisher Scientific (Fair Lawn, NJ). Whole blood proficiency samples were obtained from the College of American Pathologists (CAP, Northfield, IL) and the Wisconsin State Laboratory of Hygiene (WSLH; Madison, WI).

Evidence sampling and dilution with internal standard was performed using a Hamilton MicroLab 500 series diluter (Reno, NV). Analyses were performed using one of four separate systems consisting of an Agilent/Hewlett Packard headspace autosampler (7694 or G1888, Palo Alto, CA) coupled to an Agilent/Hewlett Packard 6890/6890N gas chromatograph (GC). Samples were analyzed in duplicate on chromatography systems containing either a J&W DB-ALC1 (30 m × 0.53-mm i.d. × 3 µm) or DB-ALC2 capillary column (30 m × 0.53 mm × 2 µm; Palo Alto, CA). All glassware used for the volumetric production of ethanol calibrators was ASTM Class A. Precision measurements of the Hamilton diluter and gravimetric measurements of calibrator repeatability were conducted using Mettler Toledo top-loader and analytical balances traceable to NIST and SI units (PL602-S, BB2400, AT261, Columbus, OH).

Methods

Testing was performed according to a laboratory-developed and validated method. The validated linear range of the method was from 0.010 to 0.50 g/100 mL. Values greater than 0.40 g/100 mL were diluted and re-analyzed. As this constituted less than 0.8% of the lab's ethanol-positive caseload, the uncertainty of dilution was not examined. Briefly, 0.2 mL of sample was diluted with 2 mL of internal standard (*n*-propanol in 1% NaCl solution) using a Hamilton diluter. Diluted samples were dispensed into 10-mL headspace vials and sealed with butyl rubber septa and aluminum crimp seals. After sample heating, a measured volume of headspace was transferred to the GC and the volatiles separated on a capillary GC with proprietary stationary phase before passing into a flame-ionization detector.

Quantitation was by internal standard, multi-point calibration with aqueous calibrator concentrations of 0.079, 0.158, and 0.316 g/100 mL. The instrument was re-calibrated with each testing batch. Aqueous negative and positive controls were routinely assayed throughout each batch and the positive controls contained ethanol at concentrations of 0.04, 0.10, and 0.20 g/100 mL. Proficiency test material was tested and stored according to the instructions of the proficiency test provider.

Ethanol-positive samples are tested on two headspace GC instruments with differing column selectivity for volatile compounds. Evidence is re-sampled for the second test. Precision tolerance for reporting is agreement to within 0.01 g/100 mL of the mean. The mean value is reported to two significant figures in units of g/100 mL.

Results and Discussion

The measurand is defined as the concentration of ethanol in whole blood (BAC) and is derived from Equation 1:

$$BAC = \frac{BAC_1 + BAC_2}{2} \cdot f(r) \cdot f(cur) \cdot f(\bar{R}_m) \cdot f(Cal) \cdot f(CRM) \cdot f(SP) \qquad \text{Eq. 1}$$

where $BAC_{1,2}$ are duplicate measurement results, *f(r)* is the correction factor related to method reproducibility, *f(cur)* is the correction factor for the target quantity obtained from a linear calibration, $f\bar{R}_m$ is the correction factor for method recovery, *f(Cal)* relates to calibrator preparation, *f(CRM)* relates to the assigned value for certified reference materials and *f(SP)* is the

correction factor for the precision in sample preparation. The uncertainties $u(x_i)$ from each of these inputs were estimated and combined into a standard uncertainty for the BAC, $u(BAC)$, as described in Equation 2:

$$u(BAC) = \sqrt{u(r)^2 + u(cur)^2 + u(\bar{R}_m)^2 + u(\text{Cal})^2 + u(\text{CRM})^2 + u(\text{SP})^2} \quad \text{Eq. 2}$$

Long-term precision uncertainty was evaluated through the use of quality control data for the aqueous ethanol controls. Although some authors have used duplicate analyses of case samples (7), the variance between duplicate results doesn't encompass a long enough timeframe to include all environmental and instrumental effects that would be represented by the quality control data pool. Validation of the resultant combined and expanded uncertainty against whole blood proficiency test materials would be used to justify this approach.

Reproducibility uncertainty: *u(r)*

Results of quality control measurements constituted the largest data pool for this work. Measurements of three aqueous ethanol controls with assigned values of 0.04, 0.10, and 0.20 g/100 mL were made under reproducibility conditions. The number of measurements recorded for each control was 5329, 6293, and 3811, respectively. These measurements were obtained over a three-year time period and included measurement variability due to environmental conditions, operator technique, individual instrument variation and variation between four instruments. Results due to obvious, documented errors such as the omission of internal standard or misidentification of the material were omitted. All other results were included to obtain the maximum variability.

The standard deviations (SD) and relative standard deviations (RSD) were calculated at each concentration. RSDs at each level (0.0238, 0.04 g/100 mL; 0.0160, 0.10 g/100 mL; 0.0145, 0.20 g/100 mL) were pooled using equation 3.

$$u(r) = RSD_{pooled} = \sqrt{\frac{(n_1-1)\cdot RSD_1^2 + (n_2-1)\cdot RSD_2^2 + (n_3-1)\cdot RSD_3^2}{(n_1-1)+(n_2-1)+(n_3-1)}} \quad \text{Eq. 3}$$

The resulting pooled RSD of 0.0187 was applied as the reproducibility uncertainty $u(r)$.

Linear calibration uncertainty: *u*(*cur*)

The uncertainty of the linear calibration was estimated according to the indirect calibration model (16) in which the uncertainty of the blood ethanol concentration x is determined from the ratio of ethanol peak area to internal standard peak area y according to:

$$x = \frac{y - \alpha}{\beta} \quad \text{Eq. 4}$$

where α is the intercept and β is the slope. The residual standard deviation (S_R) of the calibration points was calculated according to Equation 5.

$$S_R = \sqrt{\frac{\sum_{i=1}^{n}[A_i - (\alpha + \beta \cdot c_i)]^2}{n-2}} \quad \text{Eq. 5}$$

In this equation A_i is the measured response of the ith calibration point, c_i is the concentration of the ith calibrator, and n is the number of measurements for the calibration. The standard uncertainty for the BAC calculated from the linear calibration $u(cur)$ is then obtained using:

$$u(cur) = \frac{s_R}{\beta}\sqrt{\frac{1}{p} + \frac{1}{n} + \frac{(c_0 - \bar{c})^2}{s_{xx}}} \quad \text{Eq. 6}$$

where p is the number of measurements to determine c_0, and c_0 is the concentration of ethanol in the reference material, c is the mean value of the calibrator concentrations and S_{xx} (the sum of squared deviations in concentration) is given by Equation 7.

$$Sxx = \sum_{i=1}^{n}(c_i - \bar{c})^2 \quad \text{Eq. 7}$$

Repeatability data (n = 40) from method validation at the 0.080 g/100 mL level was used to estimate the $u(cur)$ for each headspace GC instrument. The results were expressed as relative standard uncertainties (RSU) relative to the 0.080 g/100 mL reference material. The RSUs ranged from 0.0076 to 0.0187. The maximum value was applied for all further estimates.

Recovery uncertainty: $u(\bar{R}_m)$

An evaluation of measurement bias was conducted through recovery studies. Bias is an estimate of the systematic error of a method, and a method must be corrected for any recognized, significant bias before uncertainty can be estimated (1,2). For the treatment of measurement bias a three-step process is followed (17). First a reference material or spiked sample undergoes replicate testing to establish its mean observed concentration $\bar{c}_{obs}$. The method recovery $\bar{R}_m$ is calculated as the ratio of the observed concentration to the reference concentration c_{ref} (Equation 8).

$$\bar{R}_m = \frac{\bar{c}_{obs}}{c_{ref}} \quad \text{Eq. 8}$$

Additional recovery correction factors may apply depending upon 1. the availability of suitable reference material; 2. variability due to multiple sample matrices; or 3. the influence of measurand concentration on recovery from spiked materials. A full description of these recovery factors and their use is described elsewhere (17,18).

The second step requires testing $\bar{R}_m$ for significance. If the method recovery is found to differ significantly from 1 within the limit of its uncertainty then a correction factor may be required to compensate for systematic effects. The uncertainty of the method recovery $u(\bar{R}_m)$ is calculated

using Equation 9.

$$u(\bar{R}_m) = \bar{R}_m \cdot \sqrt{\left(\frac{1}{p}\right) + \left(\frac{s_{obs}}{\bar{c}_{obs}}\right)^2 + \left(\frac{u_{ref}}{c_{ref}}\right)^2} \quad \text{Eq. 9}$$

where p is the number of measurements, is the standard deviation of the mean, and u_{ref} is the uncertainty of the reference material. The significance is tested using Equation 10.

$$t = \frac{|1 - \bar{R}_m|}{u(\bar{R}_m)} \quad \text{Eq. 10}$$

If the value of t is greater than the coverage factor (k) used to calculate expanded uncertainty, then significant bias is assumed to exist. The third step is deciding the course of action based upon the calculation of t (19).

Under repeatability conditions, 10 measurements were made of an aqueous ethanol, standard reference material (SRM2893; NIST). Matrix differences due to aqueous versus whole blood samples are reduced, if not eliminated, with the use of an internal standard and a high sample dilution factor (20). The study was conducted in triplicate. The mean, standard deviation, method recovery, and its uncertainty were calculated along with the t value for each study. Results appearing in Table I demonstrate that method recoveries did not differ significantly from 1 when using a coverage factor of 3. This two-sided coverage factor, at an approximate level of confidence of 99.7%, was based upon the calculation of the effective degrees of freedom as described.

Reconciliation of bias may be done in one of three ways. When bias is insignificant, no correction of the measurement is required but $u(\bar{R}_m)$ is added into the combined standard uncertainty estimate. When bias is significant either a correction factor is applied to the measurement result, or no correction is applied and the standard uncertainty is enlarged to ensure the range encompasses the true value. Because no significance was found in this estimate, the maximum recovery uncertainty for the three studies (0.0046) was incorporated into the overall uncertainty budget.

Table I. Recovery Study Evaluation of Bias

Replicate #	Study 1 (g/100 mL)	Study 2 (g/100 mL)	Study 3 (g/100 mL)
1	0.08020	0.07911	0.07870
2	0.08005	0.07952	0.07903
3	0.08026	0.07967	0.07972
4	0.08004	0.07981	0.07937
5	0.08014	0.08050	0.07956
6	0.07992	0.08067	0.07952
7	0.08042	0.08039	0.07991
8	0.08045	0.08024	0.07946
9	0.08029	0.08031	0.07963
10	0.08049	0.08006	0.07925
c_{obs} (g/100mL)	0.08023	0.08003	0.07942
s_{obs} (g/100mL)	0.00006	0.00016	0.00011
R_m*	1.0018	0.9993	0.9916
$u(R_m)$	0.0046	0.0046	0.0046
t†	0.38	0.15	1.8

* c_{ref} (u_{ref})= 0.08009 (±0.00037) g/100 mL.
† coverage factor k = 3 applied for significance test.

Calibrator uncertainty: *u(Cal)*

The uncertainty in the approximation of calibrator concentration *u(Cal)* was estimated. An example for the volumetric preparation of the 0.079 g/100 mL calibrator is shown in Table II. Uncertainties due to glassware tolerance, environmental conditions, ethanol purity, the inexact value for ethanol density, and variations in flask filling and observer discrimination were combined. The relative standard uncertainty for calibrator preparation was 0.0033.

CRM uncertainty: *u(CRM)*

Aqueous certified reference materials in use for quality assurance purposes were supplied with the manufacturer's certificate of analysis. The certificate contained traceability information, the assigned value of the material and the uncertainty of the assigned value. For purposes of combining the CRM uncertainty, it must first be reduced from expanded to standard uncertainty. This was done by dividing the relative expanded uncertainty (0.0058) by the manufacturer-provided coverage factor, k = 2, to produce a CRM relative uncertainty *u(CRM)* of 0.0029.

Sample preparation uncertainty: *u(SP)*

A semi-automatic, dual-syringe liquid processor is used for the sampling of calibrators, controls, and specimens and their simultaneous dilution with internal standard solution. The manufacturer's specification for precision varies by the stroke length of the syringe with maximum precision equal to 0.2% at a syringe stroke length of 30%. For this work, a stroke length at 80% of nominal is used. Annual calibration by the laboratory's gravimetric procedure is used to verify this precision is met or exceeded for all liquid processors. A rectangular distribution is assumed for sample preparation and the relative uncertainty was calculated as $0.002/\sqrt{3} = 0.0012$.

Combined standard uncertainty: *u(BAC)*

The relative standard uncertainties for the six sources were combined through the root sum of squares method of Equation 2:

$$u(BAC) = \sqrt{u(r)^2 + u(cur)^2 + u(\bar{R}_m)^2 + u(\text{Cal})^2 + u(\text{CRM})^2 + u(\text{SP})^2} =$$

$$\sqrt{(0.0187)^2 + (0.0187)^2 + (0.0046)^2 + (0.0033)^2 + (0.0029)^2 + (0.0012)^2}$$

$$u(BAC) = 0.0272$$

Effective degrees of freedom v_{eff} and coverage factor k

The effective degrees of freedom v_{eff} used to select a coverage factor were calculated using the Welch-Satterthwaite formula. This formula is used to calculate the appropriate coverage factor when a normal distribution cannot be confirmed or when Type A components of uncertainty are based on fewer than 10 repeated observations (6). Although it can be assumed that a normal distribution applies here based on the number of observations for each data set, nevertheless, the effective de-

Table II. Uncertainty in the Volumetric Preparation of the 0.079 g/100 mL Ethanol Calibrator

$$[Cal] = \frac{V_{EtOH} \times P_{EtOH} \times d_{EtOH}}{V_{final}}$$

$u(V_{EtOH})$ Uncertainty in volume of ethanol dispensed

$$u(Pipette) = \frac{0.006\ \text{mL}}{\sqrt{6}} = 0.002\ \text{mL}$$

Tolerance of ASTM Class A, 1-mL volumetric pipette (20°C) = ±0.006 mL; assuming a triangular distribution

$$u(Temp) = \frac{0.003\ \text{mL}}{\sqrt{6}} = 0.002\ \text{mL}$$

Thermal expansion coefficient of ethanol (20°C) = $7.5 \times 10^{-4}\ C^{-1}$ × 1 mL sampling of ethanol × 4°C temperature variation = ±0.003 mL; assuming a rectangular distribution

$$u(V_{EtOH}) = \sqrt{u(Pipette)^2 + u(Temp)^2} = \sqrt{0.002\ \text{mL}^2 + 0.002\ \text{mL}^2} = 0.003\ \text{mL}$$

Repeatability of pipette filling not factored as flask filling repeatability is dominant

$u(P_{EtOH})$ Uncertainty in the purity of ethanol

$$u(P_{EtOH}) = \frac{0.001}{\sqrt{3}} = 0.00058$$

Purity on supplier's Certificate of Analysis = 99.9%; 0.999 ± 0.001; assuming a rectangular distribution

$u(d_{EtOH})$ Uncertainty in the density of ethanol

$$u(d_{EtOH}) = \frac{0.001\ \text{g/mL}}{\sqrt{3}} = 0.00058\ \text{g/mL}$$

Density of ethanol (20°C) = 0.789 g/mL; tolerance of a standard density meter = ±0.001 g/mL; assuming rectangular distribution

$u(V_{final})$ Uncertainty in the final volume of the calibrator

$$u(Flask) = \frac{0.3\ \text{mL}}{\sqrt{6}} = 0.12\ \text{mL}$$

Tolerance of ASTM Class A, 1000mL volumetric flask (20°C) = ±0.3 mL; assuming a triangular distribution

$$u(Temp) = \frac{0.84\ \text{mL}}{\sqrt{3}} = 0.48\ \text{mL}$$

Thermal expansion coefficient of water (20°C) = $2.1 \times 10^{-4}\ C^{-1}$ × 1000 mL final volume × 4°C temperature variation = 0.84 mL variance; assuming a rectangular distribution

$$u(Rep) = 0.75\ \text{mL}$$

Repeatability of filling and observing volume level; obtained from weighing 1000 mL flask 10 times and calculating standard deviation

$$u(V_{final}) = \sqrt{u(Flask)^2 + u(Temp)^2 + u(Rep)^2} = \sqrt{0.12\ \text{mL}^2 + 0.48\ \text{mL}^2 + 0.75\ \text{mL}^2} = 0.90\ \text{mL}$$

$u(Cal)$ Uncertainty in the volumetric preparation of the calibrator

$$u(Cal) = [Cal] \times \sqrt{\left(\frac{u(V_{EtOH})}{V_{EtOH}}\right)^2 + \left(\frac{u(P_{EtOH})}{P_{EtOH}}\right)^2 + \left(\frac{u(d_{EtOH})}{d_{EtOH}}\right)^2 + \left(\frac{u(V_{final})}{V_{final}}\right)^2}$$

$$u(Cal) = \sqrt{\left(\frac{0.003\ \text{mL}}{1\ \text{mL}}\right)^2 + \left(\frac{0.00058}{0.999}\right)^2 + \left(\frac{0.00058\ \text{g/mL}}{0.789\ \text{g/mL}}\right)^2 + \left(\frac{0.90\ \text{mL}}{1000\ \text{mL}}\right)^2} = 0.0033 = \textit{relative standard uncertainty}$$

grees of freedom were calculated using the following formula:

$$v_{eff} = u(BAC)^4 / \Sigma_{i=1}^{N} \frac{u_i^4(y)}{v_i} \quad \text{Eq. 11}$$

where $u_i(y)$ are individual uncertainty contributions and v_i are the effective degrees of freedom for each uncertainty source. All Type B sources had well defined limits of tolerance so v_i = infinity, reducing their effective contributions to zero.

By this approach ν_{eff} was calculated as 173, corresponding to a coverage factor k = 3 with an approximate confidence interval of 99.7%. The relative expanded uncertainty with a coverage factor of 3 is equivalent to $0.0272 \times 3 = 0.0816$, or an expanded uncertainty *U(BAC)* of ±8.2% of the BAC measurement.

Verification of the estimate: charting PT bias

The expanded uncertainty estimate for BAC testing was evaluated against the laboratory's performance on proficiency testing (PT). Over a three-year period 126 ethanol-positive, whole blood PT samples were received from two PT providers. This total includes replicate submissions of PT cycles for evaluation of multiple analysts.

Percent biases between the laboratory's individual PT results and their respective consensus values were graphed as absolute bias versus concentration in the manner of Van Eenoo et al. (21) (Figure 1). The expanded uncertainty estimate (99.7%) was added to this graph as a threshold level to evaluate compliance between the laboratory's PT reporting variability and the maximum variability predicted by the calculated uncertainty. Comparative threshold levels corresponding to expanded uncertainties of 99% (k = 2.576) and 95% (k = 2) were also plotted.

Three distinct regions are present in Figure 1. The area bounded by 0.040 and 0.190 g/100 mL display absolute biases below 4% with the large majority below 3%. This suggests that the expanded uncertainty is substantially overestimated within this concentration range. The region below 0.040 g/100 mL shows a tendency toward higher variability with several PT samples in excess of the 95% confidence level and one (0.0267 g/100 mL) in excess of the 99.7% level. For PT concentrations above 0.190 g/100 mL, the variance is more evenly distributed with the majority below 4%. For the nine PT results in Figure 1 where bias exceeded 4%; six had negative signs and three positive; five were 0.036 g/100 mL or less; and four were 0.204 g/100 mL or greater.

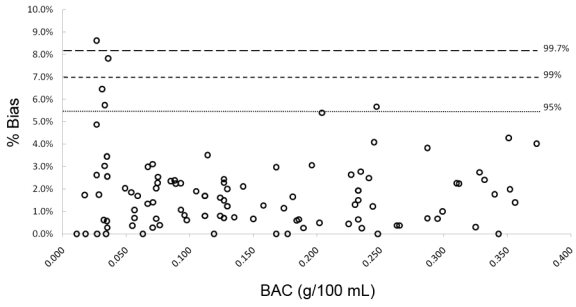

Figure 1. Expanded uncertainty evaluation using whole blood ethanol proficiency test data. (Absolute bias of PT results plotted against PT concentration.)

The utility in overestimating uncertainty lies in its application to the lower and upper limits of a method's dynamic range. If the method were modified to include routine sample dilution, then a re-examination of the expanded uncertainty would likely support a much lower, relative estimate for concentrations between 0.040 and 0.380 g/100 mL. Such a modification would require increased testing to identify those samples that qualify for pre-treatment. This would be required in approximately 34% of the ethanol-positive cases for this laboratory in a given year. The uncertainty associated with dilution would originate from the observed precision of the diluter, pipette or other equipment used in the dilution. Alternatively, an uncertainty estimate could be made specific to the lower end of the dynamic range, but this would apply to a small minority of ethanol-positive cases. Neither approach lends itself to use in a high-volume testing lab.

It is important to note that the consensus values used in Figure 1 are simply group averages of participant results rather than assigned values with associated uncertainties. ISO 13528 describes the five ways a PT coordinator may assign a value to a PT test (22): 1. through calculation of the formula used to prepare the PT sample; 2. through use of a CRM; 3. by within-laboratory measurements of the PT sample calibrated against a CRM; 4. calculated as a robust mean x^* of PT results from expert laboratories; or 5. calculated as the robust mean from participant laboratories. The source of assigned values for blood alcohol PT samples is usually not identified although at least one program does provide this information (23). A laboratory makes the assumption that the consensus value is the simple mathematical average of all participant results with some treatment applied to identify and eliminate outlier data. As such, the graphical uncertainty treatment in the present case adopts the same assumption.

Verification of the estimate: performance statistics

For one of the PT programs (WSLH), all participant results were reported anonymously at the conclusion of each testing cycle. With the full complement of data available, the robust mean x^* and standard deviation s^* for each PT sample were calculated according to Algorithm A of ISO 13528. The robust mean was used as the reference value for the PT sample and the reference value uncertainty was calculated from the robust standard deviation in the following way.

$$u(x^*) = 1.25 \times s^* / \sqrt{p} \quad \text{Eq. 12}$$

The validity of the laboratory's expanded uncertainty was evaluated through the use of the E_n score,

$$E_n = \frac{x - x^*}{\sqrt{U(BAC)^2 + U(x^*)^2}} \quad \text{Eq. 13}$$

where x is the laboratory's PT result and $U(x^*)$ is the expanded uncertainty of the PT reference value (k = 3). The E_n score is similar to a z-score in that the measurement accuracy is expressed in relation to variance. Unlike a z-score, which uses the

standard deviation of the PT results as the variance, E_n relates accuracy on a PT result to the expanded uncertainty contributions from the laboratory's measurement and assigned value. As such it is less representative of performance within the scheme of the PT program as representative of the individual laboratory's performance against their own uncertainty estimate (24).

The robust statistics were calculated for 56 ethanol-positive, whole blood proficiency samples. The BAC expanded uncertainty estimate was evaluated through the production of E_n performance scores. The critical boundary of E_n is ±1. An underestimated uncertainty will frequently produce E_n values greater than 1 and an overestimated uncertainty will repeatedly produce E_n values significantly less than 1.

Figure 2 shows E_n scores plotted as a function of robust mean. The results are evenly distributed around zero with maxima of 0.282 (0.0849 g/100 mL) and –0.725 (0.2479 g/100 mL). There are limitations to this approach in that the number of participants for these PT's do not routinely exceed $N = 12$. Without some uniformity between laboratory test methods, small-scale PT programs may produce excessively large standard uncertainties (standard deviations) for the assigned values and generate E_n scores indicating a poor uncertainty estimate by a laboratory.

For large-scale PT programs ($N > 200$), such as the College of American Pathologists' (CAP) AL1, the full cadre of participant results are not published in the summary report. The assigned value and its standard uncertainty are taken as the consensus mean and standard deviations, respectively without the application of robust statistics to downweight the effect of data furthest from the mean. E_n scores calculated for these PT results showed a larger range of values with a potential concentration dependency (Figure 3). The maxima for these PT scores were 0.930 (0.0267 g/100 mL) and –0.969 (0.0358 g/100 mL).

Conclusions

The measurement uncertainty for BAC testing was estimated. The largest uncertainty sources were from method reproducibility and the calculation of BAC from a multi-point, internal standard calibration curve. Reproducibility data was drawn from a large pool of quality control measurements over a three-year period. This was done to include as many sources of variability (operators, instruments, environmental conditions, etc.) as possible and may be considered an empirical (top-down) estimate of the uncertainty. Similarity in the values obtained for pooled reproducibility $u(r)$ and calibration $u(Cal)$ uncertainty and the apparent overestimation of the expanded uncertainty across the majority of the reporting range suggest that the uncertainty attributed to the calibration may already be represented in the reproducibility estimate. Taking the pooled RSD from reproducibility data as the sole source, an expanded uncertainty of ±5.6% (k = 3) would set an upper bound on bias in Figure I that encompassed all but one PT result from the concentration range 0.040–0.380 g/100 mL. Re-evaluation of any such estimate is appropriate when changes arise in the analytical method, for example when a CRM is replaced, or when the estimate's fitness for purpose changes.

It is interesting to note that the ±8.2% uncertainty estimate is within the typical tolerance for ethanol quality control accuracy. Variability for ethanol control and calibrator results is commonly held to ±10% in a quality assurance (QA) program (25,26). An aberrant control result is a clear sign that the method is not in statistical control, and the re-testing of all or a portion of case samples is necessary. For laboratories not currently calculating blood alcohol uncertainty but operating with similar QA, it is likely that the estimate produced in this work is reflective of what would be expected for their uncertainty estimate.

A method recovery study was included to determine the presence of, and make corrections for, any significant bias. In the case of ethanol testing, sufficiently characterized CRM was available with appropriate traceability and uncertainty so additional recovery factors (matrix, spiking) were not included. This was predicated on the understanding that the 10-fold dilution of sample with a NaCl solution produces a sample similar enough in composition to an aqueous CRM that headspace sampling recoveries are identical (27). When matrix effects on recovery have been noted, even in the presence of internal standard, they have been measured at sample compositions of 20% (w/w) protein content, a level that would not be encountered with the dilution utilized here (20).

No significant measurement bias was identified for this method. This is not surprising given the fact that a properly selected internal standard is used to correct for any systematic effects should they be present. The recovery uncertainty was

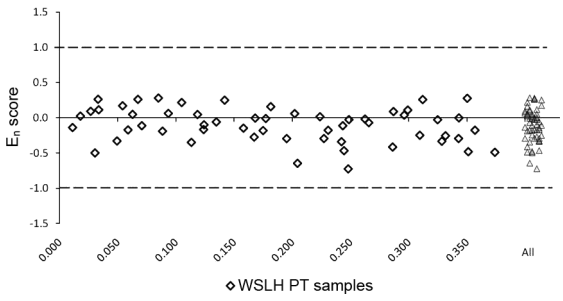

Figure 2. Expanded uncertainty evaluation using performance scores on WSLH proficiency testing.

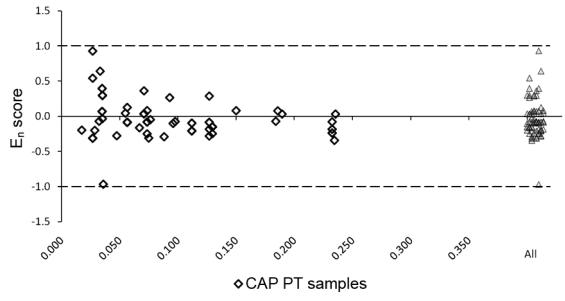

Figure 3. Expanded uncertainty evaluation using performance scores on CAP proficiency testing.

instead added to the final combined estimate to account for any unrecognized matrix effects.

The calibrator uncertainty, although a minor contributor to $u(BAC)$, required a fairly comprehensive evaluation of the steps involved in the preparation process. Its value is emblematic of what would be expected when high quality materials and well-categorized equipment are selected. The basic approach taken here could easily be adapted to gravimetric preparation. Alternatively, if a CRM was purchased for calibration with a reference value calculated farther up on the hierarchy of comparability, then the uncertainty of its assigned value would undoubtedly be lower.

The adoption of a consensus standard uncertainty (standard deviation) from a PT program is based on several assumptions. Ideally a laboratory would receive traceability information for the PT sample (CRM) in order to show comparability with its own results (24). This is usually not the case in forensic alcohol proficiency testing so traceability can only be assumed. The PT program coordinator should also provide information describing the assigned value and associated uncertainty for the PT material; however, this information is also absent in many forensic alcohol PT programs with the assigned value most often derived from the consensus mean of all participants with appropriate exclusion of outlier values (28). The assumption would also be made that tests for significance of bias have been conducted by participant laboratories and appropriate corrections have been applied.

Consensus values, from CAP, for example, are generated from multiple test methods. Even within the most widely used test method, GC, it is questionable whether there is enough interlaboratory consistency to call the method standardized. When a standard method or "well-recognized test method" is used by a laboratory, ISO 17025 states that measurement uncertainty is exchangeable between laboratories if defined limits on uncertainty sources have been specified. No such specifications are in evidence for GC methods of BAC testing and the inclusion of a sub-population of results from unrelated test methods, with their own ambiguity of uncertainty sources and limits, further confounds the issue. The adoption of a consensus standard uncertainty from a PT program is something that should be done with caution and, wherever possible, through adherence to the guidance in ISO 21748 (29).

Proficiency test programs for blood alcohol should publish traceability information, the assigned value of the PT material, the method for determining the assigned value, and its uncertainty. Accrediting bodies should insist upon this information when considering the endorsement of a PT provider. A harmonized protocol for proficiency testing of analytical chemistry laboratories has been in existence since 1993 and was recently updated to include measurement uncertainty (28). PT providers should encourage their customers to provide blood alcohol uncertainty estimates and methodology when submitting their results. As accrediting bodies are frequently receiving member laboratories' PT assessments directly from the provider, the collection and examination of this data will allow them to develop acceptability criteria for BAC measurement uncertainty just as they have developed standards for measurement accuracy and precision. A comparable treatment can be found in the anti-doping test community where interlaboratory comparisons are used to set maximum allowable uncertainty near a cutoff concentration (30).

Following the release of the U.S. National Academies' recommendations for forensic science laboratories, increased attention has been paid to forensic measurements (31). An evaluation of measurement uncertainty was recommended not only for the validation of a forensic method but also for inclusion in standardized reporting schemes. Blood alcohol determinations in particular were provided as one example of a measurement for which an uncertainty estimate is appropriate. For forensic testing laboratories accredited under ISO 17025, this topic is not unknown, but its adoption has been considerably slowed given its complexity and novel nature (32,33). At present the routine reporting of blood alcohol measurement uncertainty is not the norm in U.S. forensic toxicology laboratories. Compliance to a specification limit, such as BAC no greater than 0.080 g/100 mL for a vehicle operator or urine benzoylecgonine no greater than 150 ng/mL for Federal workplace drug testing, has historically been based upon the measurement itself rather than a range of measurements containing the true value within a stated confidence interval. The pace at which this practice changes will depend upon the institutional will of forensic professional organizations and accrediting bodies, the mandate of customers, and the availability of training and guidance documents on the topic.

Acknowledgments

We are grateful to Amanda Black, Dawn Cox, and Brianne O'Reilly of the Washington State Patrol's Toxicology Laboratory for the production of validation and recovery data used in this work.

References

1. International Vocabulary of Metrology-Basic and General Concepts and Associated Terms (VIM), Joint Committee for Guides in Metrology (JCGM: 200), BIPM, Sèvres, France, 2008.
2. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement, 1st ed. Joint Committee for Guides in Metrology (JCGM: 100), International Organization for Standardization, Geneva, Switzerland, 1995 (revised 1st ed. 2008).
3. Quantifying Uncertainty in Analytical Measurement—EURACHEM/CITAC Guide CG4, 2nd ed. S.L.R. Ellison, M. Rosslein, and A. Williams, Eds. EURACHEM/CITAC Working Group 2000.
4. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results—NIST Technical Note 1297, B.N Taylor and C.E. Kuyatt, National Institute of Standards and Technology, U.S. Government Printing Office, Washington, D.C., 1994.
5. The Expression of Uncertainty and Confidence in Measurement—UKAS M3003, 2nd ed. United Kingdom Accreditation Service, Middlesex, U.K., 2007.
6. Expression of the Uncertainty of Measurement in Calibration, EA-4/02, European Co-operation for Accreditation, EA, Paris, France, 1999.
7. J. Kristiansen and H.W. Petersen. An uncertainty budget for the measurement of ethanol in blood by headspace gas chromatog-

raphy. *J. Anal. Toxicol.* **28:** 456–463 (2004).
8. R.G. Gullberg. Estimating the measurement uncertainty in forensic breath-alcohol analysis. *Accred. Qual. Assur.* **11:** 562–568 (2006).
9. G.N.W. Leung, E.N.M. Ho, W.H. Kwok, D.K.K. Leung, F.P.W. Tang, T.S.M. Wan, A.S.Y. Wong, C.H.F. Wong, J.K.Y. Wong, and N.H. Yu. A bottom-up approach in estimating the measurement uncertainty and other important considerations for quantitative analyses in drug testing for horses. *J. Chromatogr. A* **1163:** 237–246 (2007).
10. K. Gates, N. Chang, I. Dilek, H. Jian, S. Pogue, and U. Sreenivasan. The uncertainty of reference standards—a guide to understanding factors impacting uncertainty, uncertainty calculations, and vendor certifications. *J. Anal. Toxicol.* **33:** 532–539 (2009).
11. S. Lee, Y. Park, W. Yang, E. Han, S. Choe, M. Lim, and H. Chung. Estimation of the measurement uncertainty of methamphetamine and amphetamine in hair analysis. *Forensic Sci. Int.* **185:** 59–66 (2009).
12. S. Lee, H. Choi, E. Kim, H. Choi, H. Chung, and K.H. Chung. Estimation of the measurement uncertainty by the bottom-up approach for the determination of methamphetamine and amphetamine in urine. *J. Anal. Toxicol.* **34:** 222–228 (2010).
13. J. Wallace. Proficiency testing as a basis for estimating uncertainty of measurement: application to forensic alcohol and toxicology quantitations. *J. Forensic Sci.* **55:** 767–773 (2010).
14. R. Moroni, P. Blomstedt, L. Wilhelm, T. Reinikainen, E. Sippola, and J. Corander. Statistical modeling of measurement errors in gas chromatographic analyses of blood alcohol content. *Forensic Sci. Int.* **202:** 71–74 (2010).
15. General Requirements for the Competence of Testing and Calibration Laboratories ISO/IEC 17025, 2nd ed. International Organization for Standardization, Geneva, Switzerland, 2005.
16. Guide to the evaluation of measurement uncertainty for quantitative test results, Eurolab Technical Report No. 1/2006, European Federation of National Associations of Measurement, Testing and Analytical Laboratories, http://eurolab.online.be/pub/i_pub.html (accessed January 2011).
17. V.J. Barwick and S.L.R. Ellison. Measurement uncertainty: approaches to the evaluation of uncertainties associated with recovery. *Analyst* **124:** 981–990 (1999).
18. V.J. Barwick and S.L.R. Ellison. The evaluation of measurement uncertainty from method validation studies, Part 1: description of a laboratory protocol. *Accred. Qual. Assur.* **5:** 47–53 (2000).
19. G.E. O'Donnell and D.B. Hibbert. Treatment of bias in estimating measurement uncertainty. *Analyst* **130:** 721–729 (2005).
20. S. Strassnig and E.P. Lankmayr. Elimination of matrix effects for static headspace analysis of ethanol. *J. Chromatogr. A* **849:** 629–636 (1999).
21. P. Van Eenoo, P. Van Renterghem, C.H. Dimopoulou, F.T. Delbeke, and C.G. Georgakopoulos. Estimating measurement uncertainty in quantitative methods not based on chromatography for doping control purposes. *Drug Test. Anal.* **2:** 19–23 (2010).
22. Statistical Methods for use in Proficiency Testing by Interlaboratory Comparisons ISO 13528, 1st ed. International Organization for Standardization, Geneva, Switzerland, 2005.
23. General Protocol: Proficiency Testing Schemes, Laboratory of the Government Chemist (LGC) Standards Proficiency Testing, http://www.lgcpt.com/ (accessed February 2011).
24. I. Kuselman and A. Fajgelj. IUPAC/CITAC Guide: Selection and use of proficiency testing schemes for a limited number of participants-chemical analytical laboratories (IUPAC Technical Report). *Pure Appl. Chem.* **82:** 1099–1135 (2010).
25. SOFT/AAFS Forensic Toxicology Laboratory Guidelines, 2006 version, SOFT/AAFS Laboratory Guidelines Committee, Society of Forensic Toxicologists, Inc. and American Academy of Forensic Sciences, 1991 (revised 2006).
26. G.A.A. Cooper, S. Paterson, and M.D. Osselton. The United Kingdom and Ireland Association of Forensic Toxicologists forensic toxicology laboratory guidelines (2010). *Sci. Justice* **50:** 166–176 (2010).
27. B. Levine. *Principles of Forensic Toxicology*, 2nd ed. AACC Press, Washington, D.C. 2006, p 178.
28. M. Thompson, S.L.R. Ellison, and R. Wood. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. *Pure Appl. Chem.* **78:** 145–196 (2006).
29. Guidance for the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation ISO/TS 21748, 1st ed. International Organization for Standardization, Geneva, Switzerland, 2004.
30. WADA Technical Document TD2010DL Decision Limits for the Confirmatory Quantification of Threshold Substances May 8, 2010, Version 1.0, World Anti-Doping Agency, http://www.wada-ama.org/en/Science-Medicine/Anti-Doping-Laboratories/Technical-Documents/ (accessed February 2011).
31. Committee on Identifying the Needs of the Forensic Science Community, Strengthening Forensic Science in the United States: A Path Forward. The National Academies Press, Washington, D.C., 2009.
32. Estimating Uncertainty of Measurement Policy March 9, 2007, American Society of Crime Laboratory Directors / Laboratory Accreditation Board-International®, http://www.ascld-lab.org/documents/alpd3008.pdf (accessed February 2011).
33. Updated Approach to Uncertainty of Measurement Requirements in the ASCLD/LAB-International Accreditation Program for Testing Laboratories September 14, 2008, American Society of Crime Laboratory Directors / Laboratory Accreditation Board-International®, http://www.ascld-lab.org/documents/alpd3033.pdf (accessed February 2011).