Journal of Analytical Toxicology 2012;36:153–161
doi:10.1093/jat/bks012

Article

Estimating the Measurement Uncertainty in Forensic Blood Alcohol Analysis

Rod G. Gullberg*

Clearview Statistical Consulting, 20119 61st Avaenue SE, Snohomish, WA 98296

*Author to whom correspondence should be addressed. Email: rod.gullberg@gmail.com

For many reasons, forensic toxicologists are being asked to determine and report their measurement uncertainty in blood alcohol analysis. While understood conceptually, the elements and computations involved in determining measurement uncertainty are generally foreign to most forensic toxicologists. Several established and well-documented methods are available to determine and report the uncertainty in blood alcohol measurement. A straightforward bottom-up approach is presented that includes: (1) specifying the measurand, (2) identifying the major components of uncertainty, (3) quantifying the components, (4) statistically combining the components and (5) reporting the results. A hypothetical example is presented that employs reasonable estimates for forensic blood alcohol analysis assuming headspace gas chromatography. These computations are easily employed in spreadsheet programs as well. Determining and reporting measurement uncertainty is an important element in establishing fitness-for-purpose. Indeed, the demand for such computations and information from the forensic toxicologist will continue to increase.

Introduction

Most toxicologists have a conceptual understanding of measurement uncertainty. However, there may be some confusion regarding exactly how to perform the computational estimates. The present paper is tutorial in nature and seeks to present examples employing hypothetical data for illustration purposes only. The purpose will be to illustrate a straightforward and well-established technique for computing, interpreting and communicating the measurement uncertainty associated with forensic blood alcohol analysis.

Many recent issues have prompted interest in measurement uncertainty. The pursuit of accreditation under what is probably the most prominent accrediting agency, the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB-International) for calibration and testing laboratories has been one significant force. The ASCLD/LAB-International accreditation program, as well as other accrediting agencies, is developed from the ISO/IEC 17025 program and requires, in part, that, "... all applicant and accredited laboratories in the ASCLD/LAB-International program to have completed estimating uncertainty of measurement for all reported 'measurements that matter'" (1). The ASCLD/LAB-International program does not designate how the uncertainty calculations are to be performed, only that a reasonable attempt be made. Some jurisdictions, moreover, have legislation requiring accreditation. Another factor bringing attention to the issue is the recent National Academy of Science (NAS) Report, *Strengthening Forensic Science in the United States: A Path Forward* (2). The NAS report states, "All results for every forensic science method should indicate the uncertainty in the measurements that are made..." The Daubert decision of the U.S. Supreme Court is another factor requiring one of four criteria for admissibility to be "... the technique's known or potential rate of error..." (3). Finally, the advent of *per se* statutory language in many jurisdictions, defining the driving offense as having a specified blood alcohol concentration (i.e., 0.08 g/100 mL or more), has resulted in focused challenge on the analytical results. Accordingly, these political, legal, financial and quality control considerations are likely to compel forensic toxicology programs to compute and report measurement uncertainty.

All measurements possess uncertainty. There are limitations in understanding, technology and procedures that all propagate through to the reported result. Measurement uncertainty should not be a cause for concern, as long as it is understood, estimated appropriately and demonstrated to be fit-for-purpose. Although a reported result has analytical meaning even without an estimate of uncertainty, its informative value is significantly enhanced when including such an estimate. Driving while intoxicated cases, particularly in a *per se* context, rely nearly exclusively on blood or breath alcohol results to prove the offense. Consider the illustration in Figure 1, in which the measure of blood alcohol concentration (BAC) yields a result near the *per se* limit of 0.080 g/100 mL. An appropriate uncertainty computation can result in a significant probability that the true BAC is less than the legal limit. Such information, it seems, would be relevant for the trier of fact to make an informed decision.

Measurement uncertainty is defined as a "... parameter characterizing the dispersion of the quantity values being attributed to a measurand..." (4). The key concept here is dispersion, or variation, of replicate measurement results all estimating the true property (concentration) of interest (the measurand). Basically, the uncertainty is a symmetrical interval around the measurement result (or the mean of replicate results) within which the true value is expected to lie with some level of probability. The uncertainty interval is most commonly quantified by the standard deviation (SD), also known as the standard uncertainty, or some multiple of it. The interval illustrated in Figure 1, for example, may be a 99% confidence interval, meaning that 99% of such sampling and computed intervals will bracket the true BAC. The concept of an interval, quantified by the standard deviation, is important to keep in mind when considering measurement uncertainty. Measurement uncertainty is commonly referred to as measurement error in the analytical and metrological literature. This should not cause alarm amongst forensic toxicologists. Indeed, this does not mean blunder or mistake, but rather, analytical and procedural limitations that still produce fit-for-purpose results. Finally, computing measurement uncertainty will rely heavily on some very basic statistics, simply a tool for studying variation.

A careful reading of the ASCLD/LAB-International policy on uncertainty reveals that no specific approach to performing

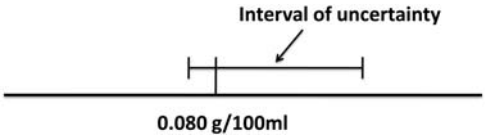

Figure 1. Illustrating an interval of measurement uncertainty for a hypothetical mean blood alcohol concentration near the critical *per se* limit.

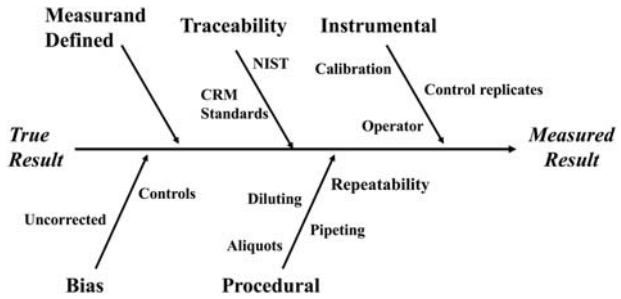

Figure 2. An illustration of a cause-and-effect diagram for blood alcohol measurement showing the key components that contribute to measurement uncertainty.

these calculations is required or suggested. Indeed, many valid computational approaches exist. Slight changes in protocol or components of traceability will also change the computational details. Toxicologists must consider the details of their own programs to incorporate the appropriate elements and statistical procedures. Many very useful resources exist that further explain and illustrate these concepts (5–7).

Determining the Measurement Uncertainty

Determining measurement uncertainty generally consists of the following steps:

(1) Define the measurand. While this may seem obvious, it is a very important step (8). The measurand in blood alcohol analysis is generally the concentration (mass/volume) of ethanol in whole venous blood collected from a living person at a specific point in time. This precludes postmortem, arterial, capillary and serum, which would be defined separately.
(2) Identify the major components contributing to blood alcohol measurement uncertainty.
(3) Quantify the contribution of each major component as SDs (i.e., standard uncertainties).
(4) Statistically combine the contribution for each major component and compute the combined uncertainty.
(5) Compute the expanded uncertainty using an appropriate coverage factor.
(6) Report the results as the best estimate, along with the expanded uncertainty or a confidence interval.

A BAC measurement is the product of several influential components, all contributing different levels of importance. Figure 2 illustrates a cause-and-effect diagram providing a useful tool for identifying the influential components. The traceability component should include a certificate of analysis establishing traceability to the National Institute of Standards and Technology (NIST) for the aqueous ethanol controls purchased from a vendor. The instrumental component will include the variation found in the replicate measurement of controls on the gas chromatograph (GC). The bias existing in the GC will also be determined from the replicate measurement of controls. This illustrates how bias estimates resulting from replicate measurements of traceable controls can take on a random characteristic. Finally, the procedural component will include all pre-analytical elements following receiving the blood sample within the laboratory and prior to actual analysis. This includes pipetting, diluting, dispensing, sealing and identifying. The actual sampling of blood from the subject is another important pre-analytical component that should ideally be included as well. However, this is not routinely Included in forensic toxicology, and will be commented on later. Figure 2 is only an illustration. Other elements or arrangements may influence the quantitative measurement of BAC for a particular laboratory. The purpose of Figure 2 is simply to provide a tool for analysts to carefully consider and document all elements in their program that might influence their measurement process. Moreover, it may be determined that some of the elements considered are not significantly influential.

Traceability is a very important component that will exist in every quantitative measurement program. Traceability is defined as the "... property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty" (4). Certificates of analysis are important documents for establishing traceability and documenting uncertainty. However, their informative value can vary greatly between manufacturers. The certificates should be reviewed carefully and the manufacturer consulted to determine exactly the interpretation of reported values. Traceability needs to be established to the level of the controls used during the actual blood analytical runs. These controls may be purchased materials or in-house developed materials; either way, their traceability to NIST certified materials must be established.

The metrology literature will describe the components of uncertainty as either Type A or Type B (9). Type A components include those for which there exist actual experimental data from which the SD (standard uncertainty) can be computed. Performing replicate controls on the GC from which the standard deviation can be computed is an example of a Type A component. Type B components include those for which there is no experimental data, but rather literature or other authoritative sources. The certificate of analysis would be an example of a Type B component. Statistically, we will treat both the same in computing the combined uncertainty.

Mathematical Foundation

The BAC measurement is considered a function of several influencing components:

$$Y = f(X_1, X_2, \cdots, X_k) \quad (1)$$

where X_1, $X_2, \ldots,$ X_k represent various influence or predictor components. Moreover, some of the X_i components may themselves be functions of other influencing terms. Our objective is to estimate the SD of Y, given the SD associated with each of the input components. These components may or may not

correspond exactly to the elements illustrated in Figure 2. If the exact form of the measurement function f in Equation (1) is known then the general equation for the propagation of error derived from the Taylor series can be employed: (10)

$$S_Y = \sqrt{\left[\frac{\partial Y}{\partial X_1}\right] S_{X_1}^2 + \left[\frac{\partial Y}{\partial X_2}\right] S_{X_2}^2 + \cdots + \left[\frac{\partial Y}{\partial X_k}\right] S_{X_k}^2} \qquad (2)$$

Equation (2) simply computes the SD of Y (S_Y) by finding the partial derivatives of Y with respect to each term, multiplying by the variance of each term and summing these overall terms. Equation (2) also assumes independence amongst all input variables. If some of the terms are not independent, but rather correlated, then a covariance term would need to be included:

$$S_Y = \sqrt{\begin{array}{l}\left[\frac{\partial Y}{\partial X_1}\right] S_{X_1}^2 + \left[\frac{\partial Y}{\partial X_2}\right] S_{X_2}^2 + \cdots + \left[\frac{\partial Y}{\partial X_k}\right] S_{X_k}^2 \\ \quad + 2\left[\frac{\partial Y}{\partial X_1}\right]\left[\frac{\partial Y}{\partial X_2}\right] \rho S_{X_1} S_{X_2}\end{array}} \qquad (3)$$

where ρ = the correlation coefficient between X_1 and X_2 and S_{X1} and S_{X2} are the respective SDs. In many cases, the exact form of the measurement function f is not known. Despite this, many assume the measurement function to have the following multiplicative form:

$$Y = f_{X_1} \cdot f_{X_2} \cdot f_{X_3} \cdots f_{X_k} \qquad (4)$$

where each factor f is assumed to be equal to one. With this and the assumption of independence, Equation (2) reduces to the following relative uncertainty form:

$$\frac{S_Y}{Y} = \sqrt{CV_{X_2}^2 + CV_{X_2}^2 + \cdots + CV_{X_k}^2} \qquad (5)$$

The multiplicative model, along with its assumption of independence, seems reasonable in forensic blood alcohol analysis and will be employed in the following example.

Example

Assume an individual's BAC is measured in duplicate with results of 0.082 and 0.083 g/100 mL, yielding a mean result of 0.0825 g/100 mL. The mean is employed because it is the best unbiased estimate of the true population mean (μ) with the minimum variance. The mean is also rounded to one extra digit because it is more reliable than the individual measurements (11). The following measurement model will be assumed:

$$C_{corr} = \frac{C_0 R}{\overline{X}} \cdot f_{dilutor} \cdot f_{Calib} \qquad (6)$$

where C_{corr} = the corrected BAC results; C_0 = the mean of the original measurement results; R = the traceable reference control value; $\overline{X}$ =the mean results from measuring the controls; $f_{dilutor}$ = the correction factor for the dilutor; f_{Calib} = the correction factor for the linear calibration.

It is further assumed that the duplicate BAC results were obtained between runs on the same GC instrument with aliquots prepared by the same analyst. While this example assumes duplicate analyses, the same approach [Equation (6)] can be used for a single determination where C_0 is simply the single value. In addition, eight measurements were assumed performed on this GC throughout the day with a commercially purchased control having a reference value of 0.100 g/100 mL and a combined uncertainty of 0.0002 g/100 mL, as noted on the certificate of analysis (12, 13). The ratio of R to X-bar in Equation (6) will quantify the proportion of bias and correct the raw result (C_0). Ideally, these controls should be near in concentration to the subject results. The mean and SD for these controls in our example were 0.1020 and 0.0010 g/100 mL. The bias in our example has been corrected for through the use of Equation (6). Methods accounting for bias have been widely discussed in the analytical literature. If the bias is small relative to the combined uncertainty and not statistically significant, some would advocate ignoring it (14). Others advocate reporting it separately and not including its uncertainty as part of the overall combined uncertainty (15). Another approach is simply to use the mean bias observed from the controls and compute the SD employing the uniform distribution, resulting in a less conservative estimate than using the maximum observed bias. Because the same dilutor is assumed to have been used to prepare the controls as well as the blood samples, there is no correction for any bias in the dilutor. It is simply assumed that $f_{dilutor} = 1$. The uncertainty in the dilutor from its certificate of calibration, however, is accounted for, which indicates $n = 10$ gravimetric measurements combined with the density to yield a mean volume of 10.106 μL with $u = 0.049$ μL. Finally, a term accounting for the linear calibration of the GC instrument ($f_{Calib} = 1$) is added. The instrument is assumed to have been calibrated with a linear five point calibration curve generated by the use of traceable control standards and the following calibration function:

$$Y = a + bX \qquad (7)$$

where Y = instrument response; X = known control concentration values; a and b are model parameters.

The initial objective is to estimate the parameters a and b by least squares regression. Some goodness-of-fit criteria, such as R^2, can be employed before proceeding. The function should not be forced through zero (making $a = 0$), because this can bias the estimate of the slope b. The final objective in developing a calibration curve is to estimate the true value of a future unknown concentration (X) given some instrument response (Y). Therefore, the inverse of Equation (7), the analytical function, is found according to:

$$X = \frac{Y - a}{b} \qquad (8)$$

Our objective is to determine the uncertainty in X. The parameters a and b, however, are correlated. The parameter a can be eliminated by solving for it according to $a = \overline{Y} - b\overline{X}$ and then substituting into Equation (8):

$$X_0 = \frac{Y_0 - (\overline{Y} - b\overline{X})}{b} \Rightarrow \quad X_0 = \frac{Y_0 - \overline{Y}}{b} + \overline{X} \qquad (9)$$

where $X0$ = a future single estimate of concentration; $Y0$ = a

future single instrument response; $\overline{Y}$ = the mean of all instrument responses during calibration; $\overline{X}$ = the mean of all control samples used during calibration.

Equation (9) reveals that X_O is a function of three random variables: Y_O, $\overline{Y}$, and b; $\overline{X}$ = is assumed without error. Solving for the uncertainty in X_O by the method of error propagation (linear terms in the Taylor series) yields:

$$u_{X_0} = \frac{S_{Y|X}}{b}\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(Y_0-\overline{Y})^2}{b^2\sum_{i=1}^{n}(X_i-\overline{X})^2}} \qquad (10)$$

where $S_{Y|X}$ = the standard error of the estimate from the regression of Y on X in developing the calibration curve; b = the slope of the calibration curve; m = the number of measurements used to estimate X_O; n = the number of measurements used to generate the calibration curve.

Specific values for the terms in Equation (10) are assumed while the uncertainty is solved for according to:

$$u_{X_0} = \frac{(0.0025)}{(1.01)}\sqrt{\frac{1}{2}+\frac{1}{5}+\frac{(0.0825-0.1548)^2}{(1.01)^2(0.042)}} = 0.0023g/100mL$$

This result expresses the uncertainty resulting from the statistical least squares estimation only, and not in the calibrants or procedures used. Incorporating our assumed values into Equation (6) yields a corrected BAC of 0.0809 g/100 mL as follows:

$$C_{corr} = \frac{(0.0825)(0.100)}{(0.1020)}\cdot 1\cdot 1 = 0.0809g/100mL$$

Finally, an uncertainty function generated from a large number of duplicate results is used to estimate the uncertainty for the method (16, 17). Ideally, this would be generated from data produced in the jurisdiction computing the uncertainty. Different jurisdictions can have largely different method uncertainty estimates due to differences in instrumentation, protocols and decision rules. Some jurisdictions, for example, perform their duplicates using two different GC instruments with different columns. Some jurisdictions might perform duplicates on only a fraction of their total cases. Others might allow for a longer time between duplicate analyses. Each of these procedural and analytical differences can yield differences in the total method uncertainty and needs to be developed within each laboratory reflecting their own protocols. Figure 3 illustrates an uncertainty function generated from a jurisdiction performing duplicate BAC determinations and is used here for illustration only. Each point on the plot represents the pooled SD determined from duplicate BAC results having means within a specified concentration bin. The SD is determined from the duplicates within the bin according to:

$$SD = \sqrt{\frac{\sum_{i=1}^{k} d_i^2}{2k}} \qquad (11)$$

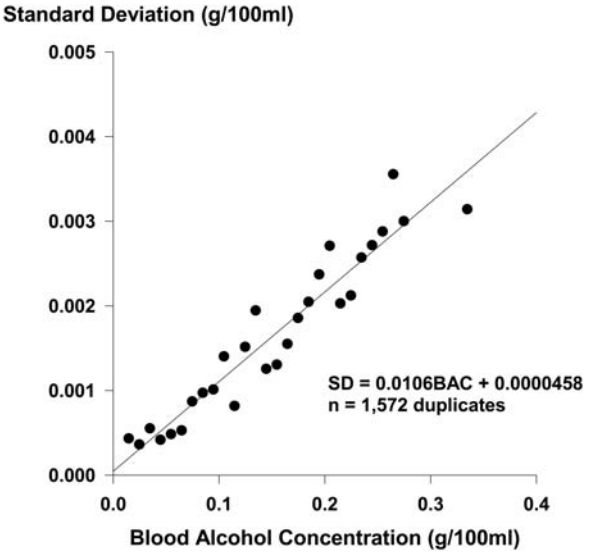

Figure 3. Plot of pooled standard deviation estimates for a single observation against concentration. The linear regression equation from the uncertainty function represents the total method variability as a function of concentration.

where d_i = the difference between duplicate results for the ith sample; k = the total number of duplicate samples within the bin.

From the data in Figure 3, a linear regression model is generated from which we estimate the SD at a specific future concentration. The uncertainty function in Figure 3 was developed with data obtained from living subjects. Uncertainty functions for postmortem cases should be developed separately from postmortem data, which will necessarily have much larger variability. These uncertainty functions will include the variability contributed by sample aliquoting, pipetting, crimping, calibrating, diluting, analytical, reporting and any other procedural features of the analysis. Separately including the uncertainty estimates from these sources, along with the uncertainty function estimate, would amount to double counting and thus overestimate the combined uncertainty. Therefore, including only the uncertainty function estimate will generally sufficiently represent the total method uncertainty. Moreover, these uncertainty functions would include the variability associated with actual blood samples collected from the many individuals by many phlebotomists. The concentration-dependent uncertainty revealed in Figure 3 should also be kept in mind when establishing duplicate test agreement decision rules. Where duplicate blood alcohol results are not available, the use of proficiency test data or historical quality control data could also be used for generating the total method uncertainty component. For our example, the model generated in Figure 3 is used from which we obtain a total method uncertainty of 0.0008 g/100 mL at a mean concentration of 0.0809 g/100 mL. Assuming independence, Equation (4) is employed to estimate our combined uncertainty as follows:

$$\frac{S_Y}{\overline{Y}} = \sqrt{CV_T^2 + CV_A^2 + CV_D^2 + CV_M^2 + CV_C^2} \qquad (12)$$

where CV_T^2 = uncertainty due to traceability; CV_A^2 = analytical uncertainty due to the GC replicates; CV_D^2 = uncertainty due to

the dilutor; CV_M^2 = uncertainty due to the total method; CV_C^2 = uncertainty due to the linear calibration.

Introducing our previously assumed values, the combined standard uncertainty is determined:

$$\frac{S_{\bar{Y}}}{0.0809} = \sqrt{\left[\frac{0.0002}{0.100}\right]^2 + \left[\frac{\frac{0.0010}{\sqrt{8}}}{0.1005}\right]^2 + \left[\frac{\frac{0.049}{\sqrt{10}}}{10.106}\right]^2 + \left[\frac{\frac{0.0009}{\sqrt{2}}}{0.0809}\right]^2 + \left[\frac{\frac{0.0023}{\sqrt{5}}}{0.0824}\right]^2} \qquad (13)$$

$$S_{\bar{Y}} = 0.0012 g/100 mL$$

If single determinations were being used, the value under the radical sign in the fourth term would be 1 rather than 2. The estimate from Equation (13) is actually the standard error of the mean, because the number of measurements has been included where appropriate. The traceability component from the certificate of analysis does not include n because it is a Type B uncertainty, commonly assumed to have an infinite number of degrees of freedom. (8) Jurisdictions performing duplicates on different instruments with different columns will need to have two terms for CV_A^2 in Equation (12), one for each separate instrument, because their analytical properties may differ. A multiplicative measurement model with independence has been assumed in this example. This is reasonable because the largest component of uncertainty (the total method) shows uncertainty to be proportional to concentration, a property of a multiplicative model (18). In addition, one can perform any number of measurements on the subject sample and simply enter that value (rather than two) into the radical of Equation (13). Increasing n will slightly reduce the combined uncertainty.

Table I is the uncertainty budget showing the percentage that each component contributes to the total combined uncertainty in this example. The table also reveals the assumed form and distribution for each component. Also revealed is the estimated proportion that each component has contributed to the total uncertainty. This is useful to identify where effort should be made to reduce total uncertainty. It also can reveal those components that contribute so little that they can be ignored. The Guide to the Expression of Uncertainty in Measurement (GUM) recommends that those terms contributing less than one-third of the maximum contributing component can be reasonably ignored. (5) Others recommend that components contributing less than one-fifth of the total can be safely ignored. (8) Table I reveals that the calibration component actually contributed the largest proportion of combined uncertainty in this example. This may not reflect the results from most laboratories. Moreover, the small contribution of the traceability, instrumental and dilutor (each $\leq$5%) suggests that these elements could be ignored. While perhaps justified in considering these as negligible, it may be forensically prudent to retain the terms, acknowledge them and monitor their continuing contribution over time.

Table I
Uncertainty Budget for the Illustrated Example

Source	Type	Distribution	Standard uncertainty	Percent*
Traceability	B	Normal	0.0002 g/100 mL	2%
Instrument	A	Normal	0.0010 g/100 mL	5%
Dilutor	B	Normal	0.049 mL	1%
Total method	A	Normal	0.0014 g/100 mL	26%
Calibration	A	Normal	0.0044 g/100 mL	66%
Combined uncertainty			0.0012 g/100 mL	
Expanded uncertainty (k = 2.576)			0.0031 g/100 mL	
99% confidence interval			0.0778 to 0. 0840 g/100 mL	

*Percent of contribution to total combined uncertainty.

Our uncertainty is now expressed as an expanded uncertainty (U), simply a multiple of the standard uncertainty (SD). This is commonly done with a confidence interval such as $\bar{Y} \pm kS_{\bar{Y}}$, where k is a coverage factor for a specified level of probability and $S_{\bar{Y}}$ is the SD of the mean. A common value for the coverage factor is $k = 2$, which generates an approximate 95% confidence interval. ASCLD/LAB's most recent policy on measurement uncertainty requires that the coverage probability be at least 95% (19). A 99% confidence interval may be more appropriate forensically, in which case, $k = 2.576$. A 99% confidence interval for our preceding example would be: $0.0809 \pm 2.576\ (0.0012) \rightarrow 0.0778$ to $0.0840 g/100 mL$. Because the mean and the standard uncertainty are random variables and vary from sample to sample, so are the interval end points. Therefore, the correct interpretation of a confidence interval is that for every 100 similarly constructed intervals around the mean of duplicate results, approximately 99 of them will include the true population mean μ. This results from a classical statistical view that claims that the true mean μ is a fixed but unknown quantity. The confidence interval is also based on several assumptions, including:

(1) The blood measurement results are normally distributed;
(2) All computed standard uncertainties are valid estimates;
(3) The estimate of the method uncertainty component is probably larger than necessary because it involves a large number of subjects, analysts, calibrations and time;
(4) The method of confidence intervals will be robust even for non-normal distributions (e.g., will also include the population mean approximately 99% of the time); and
(5) Because the population mean (μ) is a fixed but unknown quantity, 99% of the confidence intervals computed from duplicate samples obtained from the subject will include μ the confidence interval expresses the uncertainty due to sampling variability only, not from any bias in the experimental design or performance.

These assumptions should be understood when interpreting measurement and uncertainty results.

Considerable thought should be given to how the results along with their uncertainty are reported. People unfamiliar with computational details (juries) will be making critical decisions based on this information. One example is:

> The duplicate whole blood alcohol results were 0.082 and 0.083 g/100ml with a corrected mean measurement result of 0.0809 g/100ml. An expanded combined uncertainty of 0.0031 g/100ml assuming k = 2.576 and a normal distribution was generated from all principle components

contributing to the uncertainty. An approximate 99% confidence interval for the true mean blood alcohol concentration is 0.0778 to 0.0840 g/100ml.

This can certainly be refined within each jurisdiction considering their unique procedural, analytical and legal contexts. A figure similar to Figure 1 could also be provided to give a visual assessment of the uncertainty relative to critical concentrations. The key is to include all important elements and assumptions so they are clearly and fully interpretable, particularly by non-scientists. Although a truncated lower result can be presented as the reported result or prosecution result, all uncertainty computations should be performed using the mean of replicates. Forensic toxicologists should be prepared to explain these principles of measurement uncertainty in clear and unequivocal terms for the jury.

The preceding example shows that the lower 99% confidence limit falls below the critical level of 0.080 g/100 mL. This was intentional to illustrate the next computation where there may be the interest in estimating the probability that the true mean μ exceeds the 0.080 g/100 mL limit. First, consider the following form for expressing the confidence interval showing the probability that μ is bracketed by the upper and lower limits:

$$P\left[\overline{Y} - Z_{(1-\alpha/2)}\, S_{\overline{Y}} \le \mu \le \overline{Y} + Z_{(1-\alpha/2)}\, S_{\overline{Y}}\right] = \pi \qquad (14)$$

Because the objective is to determine the probability that μ exceeds the lower limit, Equation (14) is rewritten as follows:

$$P\left[\overline{Y} - Z_{(1-\alpha/2)}\, S_{\overline{Y}} \le \mu \le \infty\right] = \pi \qquad (15)$$

The lower limit expressed in Equation (15) is now set equal to 0.080 g/100 mL and $Z_{(1-\alpha/2)}$ is solved for:

$$\overline{Y} - Z_{(1-\alpha/2)}\, S_{\overline{Y}} = 0.080 \Rightarrow 0.0809 - Z_{(1-\alpha/2)}(0.0012) = 0.080$$
$$\Rightarrow Z_{(1-\alpha/2)} = 0.75$$

Next, our probability statement is rearranged and the value for $Z_{(1-\alpha/2)}$ is introduced:

$$P\left[\overline{Y} - Z_{1-\alpha/2}\, S_{\overline{Y}} \le \mu\right] = P\left[\frac{\overline{Y} - \mu}{S_{\overline{Y}}} \le Z_{1-\alpha/2}\right] = P\left[Z \le Z_{1-\alpha/2}\right]$$
$$= P[Z \le 0.75] = 0.7734$$

From the standard normal tables, $P(Z < 0.75) = 0.7734$, corresponding to the probability that the individual's true mean BAC exceeds 0.080 g/100 mL. This information may be of relevance in either civil or criminal proceedings.

One may also be interested in determining some decision limit above the critical level of 0.080 g/100 mL, above which there will be some high level of confidence that the true mean μ exceeds 0.080 g/100 mL. This is the calculation of what is called a "guard band," as illustrated in Figure 4. To be 99.5% confident that the true mean exceeds 0.080 g/100 mL, the measured mean would need to have a value greater than 0.080 + 2.576(0.0012) = 0.0831 g/100 mL. Some jurisdictions simply subtract a constant value or percentage from the mean results to determine a level for prosecution, equivalent to employing a guard band. However, it would be important to compute this correction value based on the combined uncertainty, as illustrated previously.

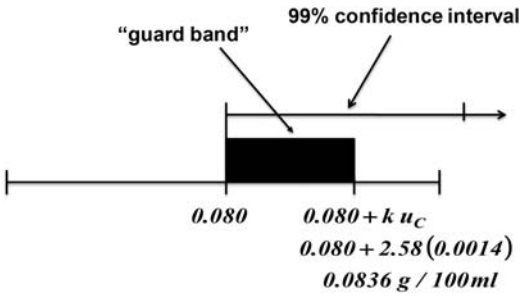

Figure 4. Illustration of a guard band that identifies a limit above which measurement results will yield a high probability that the true result exceeds the critical *per se* limit.

The example presented here made several assumptions that may not be relevant for some laboratories. The intent was simply to illustrate the computations and general approach. Moreover, the example did not include all possible sources of uncertainty. Other components might include, for example: (1) lack of linearity, (2) correlation between specific component uncertainties, (3) scale resolution and (4) uncertainty in the calibrants. Each laboratory must determine the uncertainty contributions from all relevant sources and decide whether they are significant or not. Following this, the general computational approach illustrated here could be employed.

Discussion

The uncertainty of measurement is becoming a more important concern in forensic toxicology. Its relevance will be in the quantitative demonstration of fitness-for-purpose—a necessary and important forensic consideration. Indeed, courts are becoming more familiar with the notion of uncertainty and analytical confidence that should accompany quantitative evidence (20). Forensic toxicologists are not alone regarding this issue. Clinical laboratories, driven by evidence-based medicine, are also facing increasing requests for uncertainty computations. The advantage of clinical laboratories, however, is that another sample can be obtained from the patient if necessary to confirm results. The forensic toxicologist lacks this luxury.

A straightforward and well-documented empirical approach to computing combined uncertainty in blood alcohol analysis has been presented. However, this has illustrated only one of many different approaches. Some other methods for estimating uncertainty include: (21).

(1) Other modeling approaches are available in which the measurement function is also known. Some have employed this in the determination of uncertainty with internally prepared alcohol control solutions (22, 23). With the gravimetric measurement of highly pure ethanol on balances calibrated with NIST traceable standards, traceability can be shown to the SI units of the kilogram and mole.
(2) Empirical approaches are also available in which the uncertainty components are combined statistically assuming

a multiplicative or additive model (21). Only the combined within-lab uncertainty was considered in our example. A between-lab component of uncertainty could also be determined, which would be larger, and may be of relevance for some programs.

(3) One could also use proficiency test results, as advocated by some (24), to employ between-lab reproducibility as an estimate of total method uncertainty. Proficiency test results reported by Wallace (24) suggest a combined uncertainty of approximately 0.004 g/100 mL for concentrations near 0.08 g/100 mL. A limitation of this approach, however, is that pooled results from all labs may not reflect the quality of a particular lab. In addition, labs are not generally selected at random and the results usually do not deal appropriately with bias (25).

(4) Similar to proficiency testing is collaborative test data. Generally, in collaborative testing, selected laboratories having a high level of quality and confidence are employed to validate a method. The reproducibility of this data can provide another source for determining uncertainty.

(5) Another guard band approach is that employed in the United Kingdom, in which 6 mg/dL [0.006 mg%; derived from three times an assumed SD of 2 mg/dL (0.002 mg%)] is subtracted from the mean of duplicate blood alcohol results below 100 mg/dL (0.1 mg%) and 6% is deducted from results over 100 mg/dL (0.1 mg%) (26). The UK also has case law recognizing such a practice (27). A similar practice occurs in Denmark, in which 0.1 g/Kg is subtracted from measured blood alcohol results to account for uncertainty and minimize the probability of the type I error (28).

(6) Sweden also employs a guard band approach in which the lower 99.9% confidence interval limit for the mean of the replicate results must exceed the legal per se limit (29).

(7) Another approach is simply reporting a pre-determined total allowable error: $TE_a = |bias| + ku_c$ (30). This would combine bias and the random error to yield an upper limit for combined measurement uncertainty. All measurements, therefore, can be provided with the assurance that they meet these criteria.

(8) There is also a Bayesian approach to estimating uncertainty (31). This involves combining prior information about the measurand (the prior distribution) with newly observed measurement results (the likelihood) and obtaining the posterior distribution for the measurand. A normal likelihood will yield a normal posterior distribution. A credible interval (the Bayesian analog to the classical confidence interval) is then computed within the posterior distribution which is interpreted as including the measurand with a specified level of probability.

(9) Some have suggested using a more sophisticated technique known as kernel regression (32). This is another statistical approach, requiring specialized software, that models measurement variation as a function of concentration.

(10) Finally, there is the legal approach. Some jurisdictions have case law stating that measurement uncertainty is not to be considered (33).

Reasonable arguments can be made for all of these approaches because they all employ combined standard uncertainties in one form or another. Moreover, the references provided here will illustrate other equally valid methods for such computations. Each jurisdiction must necessarily and carefully consider their analytical and procedural details and develop a cause-and-effect diagram with an uncertainty budget from which their computations can be developed. Finally, most procedures can be easily incorporated into spreadsheet programs for ease and accuracy of computation.

Programs differ in many analytical and procedural details that determine the components contributing to combined uncertainty. For this reason, accrediting agencies will generally not require a specified algorithm for computing uncertainty. Kristiansen's informative paper (28), for example, illustrates a program that reports BAC in mass/mass units, requiring density and water fraction of blood to be accounted for in the combined uncertainty. Another consideration of some jurisdictions may be the reported scale resolution. This was not considered in our preceding example, because its contribution is negligible for three digit results, generally less than 1%. Some programs may use their internally prepared controls with which to monitor within-run precision and bias. These will have different traceability and uncertainty components. Although aqueous ethanol controls are most commonly employed, some may choose to use prepared or commercially available spiked blood controls. However, these matrix differences are also thought to contribute negligible effects (28). Others may wish to include the uncertainty in calibrants. These were not included here because bias was assumed to be best monitored by the within-day traceable controls following an assumed single calibration procedure. Neither did our example include any environmental or temperature considerations. This was based on assuming that the controls were subject to the same conditions within the same day. In addition, the method variability component of Figure 3 would involve replicate measurement conditions over nearly a year's time, contributing many subcomponent sources of variation. The purity and properties of the internal standard, assumed to be negligible because it is included with the controls and with the method repeatability, was also not included in our example, but may be available from certificates of analysis. Clearly, there are a number of unique program considerations with no single correct way to include all components, other than well-reasoned consideration. Moreover, one must be careful not to double-count sources of uncertainty by including them more than once in the uncertainty budget. Finally, the selected components and computational procedures should be documented in program policy so the determination of uncertainty will be unequivocal.

Equation (9) reveals the influence of the number of measurements on combined uncertainty. Because blood samples are often preserved, they can be reanalyzed in certain cases with a larger number of measurements. For the example presented here, performing three instead of two measurements will reduce the combined uncertainty by approximately 12%. Performing four instead of two will reduce the uncertainty by approximately 20%. If reanalysis of the sample is performed, it would be important to consider any reduction in concentration due to time and also necessary to re-compute the appropriate

uncertainty contributions from each component because they may have changed.

Several pre-analytical components potentially contributing to uncertainty were also not considered in our example (34). While thought to have minimal contribution to uncertainty, one pre-analytical component is that of sampling. Because a phlebotomist typically performs a single venipuncture, there is no estimate for the variability associated with repeated venipuncture sampling. A study by Jones (35) involved drawing samples from each arm in $n = 12$ subjects. The variability observed by Jones at 90 min after the start of drinking was SD = 0.0016 g/100 mL during the alcohol elimination phase. This seems rather too large to represent sampling variability alone and probably also includes the variation due to perfusion/distribution differences between the two arms as well as an analytical component. If one sample was drawn from the subject and the sampling uncertainty of 0.0016 g/100 mL was added to Equation (13), a combined uncertainty of 0.0020 g/100 mL would be obtained, with the sampling component contributing approximately 62% to the total uncertainty. Repeated sampling from the same arm in a short time interval would likely contribute much less variability. Indeed, an uncertainty model like that of Figure 3, based on a large number of duplicate collected blood samples, would also reveal its concentration dependence.

Conclusions

For many reasons, forensic toxicologists can anticipate an increasing demand to compute and report the measurement uncertainty associated with their blood alcohol analyses. This trend should not be regarded with apprehension, but rather appreciated because it will ensure fitness-for-purpose and enhance the confidence in the reported results. This effort will yield greater credibility with customers and ensure that they are better informed. Although there are several documented approaches to computing uncertainty, a simple bottom-up model has been presented. The coefficient of variation based on the combined uncertainty in forensic blood alcohol analysis is approximately 1–3%. While this seems very fit-for-purpose, no attempt has been made to define exactly how this should be determined. Defining the fitness-for-purpose is a related issue that is presumably determined by the collaborative efforts of toxicologists and their customers so their needs and expectations are met. Determining and properly reporting measurement uncertainty will be critical to ensuring fitness-for-purpose.

Disclaimer

The opinions contained herein are exclusively those of the author and not necessarily those of any agency/organization.

References

1. ASCLD/LAB. Updated approach to uncertainty of measurement requirements in the ASCLD/LAB-International accreditation program for testing laboratories. http://www.ascld-lab.org/international/pdf/alpd3033.pdf Accessed May 2011
2. National Academy of Science. (2009) Strengthening forensic science in the United States: A Path Forward. National Research Council, The National Academies Press, Washington, DC.
3. Daubert, v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
4. International Organization for Standardization. (2008) International vocabulary of basic and general terms in metrology (VIM), 3rd edition. ISO, Geneva, Switzerland. http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008_pdf Accessed May 2011
5. International Organization for Standardization. (2008) Guide to the expression of uncertainty in measurement. ISO, Geneva, Switzerland. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf Accessed May 2011
6. EURACHEM/CITAC. (2000) Guide quantifying uncertainty in analytical measurement, 2nd edition. http://www.measurementuncertainty.org/mu/QUAM2000-1.pdf Accessed May 2011
7. Taylor, B.N., Kuyatt, C.E. (1994) NIST Technical Note 1297, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results,. National Institute of Standards and Technology, Physics Laboratory, Gaithersburg, MD. http://physics.nist.gov/Document/tn1297.pdf Accessed May 2011
8. Hibbert, D.B. (2007) Quality assurance for the analytical chemistry laboratory. Oxford University Press, New York, NY.
9. International Organization for Standardization. (2008) Guide to the expression of uncertainty in measurement. ISO, Geneva, Switzerland. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf Accessed May 2011
10. Ku, H.H. (1966) Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards*, **70C,** 916–921.
11. Devore, J.L. (1982) Probability and statistics for engineering and the sciences. Brooks/Cole Publishing, Monterey, CA.
12. Cerilliant Inc. Certificate of analysis for certified reference standards, Ethanol-100, http://www.cerilliant.com/products/searchCOA.aspx Accessed May 2011
13. Gates, K., Chang, N., Dilek, I., Jiam, H., Pogue, S., Sreenivasan, U. (2009) The uncertainty of reference standards—A guide to understanding factors impacting uncertainty, uncertainty calculations, and vendor certifications. *Journal of Analytical Toxicology*, **33**, 532–539.
14. Linsinger, P.J. (2008) Use of recovery and bias information in analytical chemistry and estimation of its uncertainty contribution. *Trends in Analytical Chemistry*, **27**, 916–923.
15. Thienpont, L.M. (2008) Calculation of measurement uncertainty—Why bias should be treated separately. *Clinical Chemistry*, **54**, 1587–1588.
16. Stowell, A.R., Gainsford, A.R., Gullberg, R.G. (2008) New Zealand's breath and blood alcohol testing programs: Further data analysis and forensic implications. *Forensic Science International*, **178**, 83–92.
17. Jones, A.W. (1991) Limits of detection and quantitation of ethanol in specimens of whole blood from drinking drivers analyzed by headspace gas chromatography. *Journal of Forensic Sciences*, **36**, 1277–1279.
18. Kristiansen, J. (2001) Description of a generally applicable model for the evaluation of uncertainty of measurement in clinical chemistry. *Clinical Chemistry and Laboratory Medicine*, **39**, 930–931.
19. ASCLD/LAB-International.ASCLD/LAB policy on measurement uncertainty. *ASCLD/LAB*, Garner, NC. http://www.ascld-lab.org/documents/AL-PD-3051.pdf (August 5, 2011).
20. Land, D.P., Imwinkelried, E.J. (2008) Confidence intervals: How much confidence should the courts have in testimony about a sample statistic? *Criminal Law Bulletin*, **44**, 257–274.
21. EUROLAB. (2007) *Technical Report 1/2007*, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation. http://www.eurolab.org/docs/technical%20report/Technical_Report_Measurement_Uncertainty_2007.pdf Accessed May 2011

22. Philipp, R., Hanebeck, O., Hein, S., Bremser, W., Win, T., Nehls, I. (2010) Ethanol/water solutions as certified reference materials for breath alcohol analyzer calibration. *Accreditation and Quality Assurance*, **15**, 141–146.
23. Sklerov, J.H., Couper, F.J. (2011) Calculation and verification of blood ethanol measurement uncertainty for headspace gas chromatography. *Journal of Analytical Toxicology*, **35**, 402–410.
24. Wallace, J. (2010) Proficiency testing as a basis for estimating uncertainty of measurement: Application to forensic alcohol and toxicology quantitations. *Journal of Forensic Science*, **55**, 767–773.
25. Analytical Methods Committee. (1995) Uncertainty of measurement: Implications of its use in analytical science. *Analyst*, **120**, 2303–2308.
26. Walls, H.J., Brownlie, A.R. (1985) Drink, drugs and driving, 2nd edition. Sweet and Maxwell, London, UK.
27. Walker v Hodgins (1984) RTR 34, (1983) Crim LR 555.
28. Kristiansen, J., Petersen, H.W. (2004) An uncertainty budget for the measurement of ethanol in blood by headspace gas chromatography. *Journal of Analytical Toxicology*, **28**, 456–463.
29. Jones, A.W., Schuberth, J. (1989) Computer-aided headspace gas chromatography applied to blood-alcohol analysis: Importance of online process control. *Journal of Forensic Sciences*, **34**, 1116–1127.
30. Westgard, J.O. (2010) Managing quality vs. measuring uncertainty in the medical laboratory. *Clinical Chemistry and Laboratory Medicine*, **48**, 31–40.
31. Kacker, R., Jones, A. (2003) On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent. *Metrologia*, **40**, 235–248.
32. Moroni, R., Blomstedt, P., Wilhelm, L., Reinkainen, T., Sippola, E., Corander, J. (2010) Statistical modeling of measurement errors in gas chromatographic analyses of blood alcohol content. *Forensic Science International*, **202**, 71–74.
33. State v. Downie, 569 Atlantic Reporter 2s 242 (N.J. Supreme Court 1990), State v. Rucker, 297 A.2d 400 (1972), State v. Lentini, 573 Atlantic Reporter 2d 464 (N.J. Appellate Division 1990).
34. Ashavaid, T.F., Dandekar, S.P., Keny, B., Bhambhwani, V.RF. (2008) Influence of blood specimen collection methods on various preanalytical sample quality indicators. *Indian Journal of Clinical Biochemistry*, **23**, 144–149.
35. Jones, A.W. (1989) Differences between capillary and venous blood-alcohol concentrations as a function of time after drinking, with emphasis on sampling variations in left vs right arm. *Clinical Chemistry*, **35**, 400–404.