Implementation of a Blind Quality Control Program in Blood Alcohol Analysis

Jackeline Moral[1], Callan Hundl[1], Dayong Lee[1], Maddisen Neuman[1,2], Aimee Grimaldi[1], Maria Cuellar[3] and Peter Stout[1,*]

[1]Houston Forensic Science Center, 500 Jefferson Street, 13th Floor, Houston, TX 77002, USA, [2]Center for Statistics and Applications in Forensic Evidence, 613 Morrill Road, Ames, IA 50011, USA, and [3]University of Pennsylvania, Philadelphia, PA 19104, USA

*Author to whom correspondence should be addressed. Email: pstout@houstonforensicscience.org

Abstract

Declared proficiency tests are limited in their use for testing the performance of the entire system, because analysts are aware that they are being tested. A blind quality control (BQC) is intended to appear as a real case to the analyst to remove any intentional or subconscious bias. A BQC program allows a real-time assessment of the laboratory's policies and procedures and monitors reliability of casework. In September 2015, the Houston Forensic Science Center (HFSC) began a BQC program in blood alcohol analysis. Between September 2015 and July 2018, HFSC submitted 317 blind cases: 89 negative samples and 228 positive samples at five target concentrations (0.08, 0.15, 0.16, 0.20 and 0.25 g/100 mL; theoretical targets). These blood samples were analyzed by a headspace gas chromatograph interfaced with dual-flame ionization detectors (HS-GC-FID). All negative samples produced 'no ethanol detected' results. The mean (range) of reported blood alcohol concentrations (BACs) for the aforementioned target concentrations was 0.075 (0.073–0.078), 0.144 (0.140–0.148), 0.157 (0.155–0.160), 0.195 (0.192–0.200) and 0.249 (0.242–0.258) g/100 mL, respectively. The average BAC percent differences from the target for the positive blind cases ranged from −0.4 to −6.3%, within our uncertainty of measurement (8.95–9.18%). The rate of alcohol evaporation/degradation was determined negligible. A multiple linear regression analysis was performed to compare the % difference in BAC among five target concentrations, eight analysts, three HS-GC-FID instruments and two pipettes. The variables other than target concentrations showed no significant difference ($P > 0.2$). While the 0.08 g/100 mL target showed a significantly larger % difference than higher target concentrations (0.15–0.25 g/100 mL), the % differences among the higher targets were not concentration-dependent. Despite difficulties like gaining buy-in from stakeholders and mimicking evidence samples, the implementation of a BQC program has improved processes, shown methods are reliable and added confidence to staff's testimony in court.

Introduction

Accredited forensic laboratories are required to complete proficiency tests (PT) on an annual basis. PTs, or interlaboratory testing, are generally purchased from an accredited or approved external vendor. Multiple laboratories across the country participate in a PT program and analyze the same samples. The consensus results from all participants are disseminated to the participating laboratories and applicable accrediting bodies. These results are then evaluated by the laboratory to determine if the PT was successfully completed. At the Houston Forensic Science Center (HFSC), analysts are aware that they are being assigned a PT and must analyze them as real casework to the greatest extent possible. However, external proficiencies are different than typical case samples in packaging, paperwork, analytes and other subtleties. While an important part of quality assurance programs, 'open' or declared PTs are limited in their use for testing

the performance of the entire system. Declared PTs are beneficial for the laboratory to measure results and processes against other laboratories in the country but are typically completed once per year per analyst for the Toxicology section as part of the accreditation requirement. In this environment, PTs become regulatory tools rather than educational tools designed to improve quality (1).

In their 1973 and 1975 studies, LaMotte et al. (2) found significant differences in urine drug screening test results between declared PTs and blind PTs. The difference in the attention given to the declared PTs versus the blind PTs seemed to account for the differences in the results. Because declared PTs have been shown to receive biased treatment (2), laboratories are recommended to implement an internal (intralaboratory) blind quality control (BQC) program (3). While the value of declared PTs are not negated, laboratories wanting a more constant and unbiased way to continually improve and monitor the performance of their processes should use an internal BQC program. Only in this way will the analyses receive no special attention and thus truly reflect the treatment received by actual samples (3).

To provide continuous quality improvement and real-time assessment of the laboratory processes, HFSC incorporated a BQC intralaboratory program as a supplement to declared PT tests. In the 2009 National Academy of Sciences report, blind proficiency testing was recommended, but not required, by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board as a more precise test of a worker's accuracy (4). The 2017 revision of the *General Requirements for the Competence of Testing and Calibration Laboratories* standard published by the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC 17025) requires laboratories to monitor the validity of their results (5). To maintain compliance with the ISO/IEC 17025 standard, HFSC must meet all requirements listed in the document. The ISO/IEC 17025 clause 7.7.1 lists out several quality controls that a laboratory may implement to conform to this requirement. The BQC programs fulfill both 7.7.1 j) intralaboratory comparisons and 7.7.1 k) the testing of blind samples.

The intent of the present BQC program is to continually test the performance and adequacy of HFSC's processes from evidence intake to reporting. The results of this continual testing are routinely provided to our stakeholders to help demonstrate results are accurate and reliable. The BQC program is facilitated by HFSC's Quality Division, which is organizationally separate from laboratory operations reporting directly to executive management. These individuals are intentionally detached from casework operations to maintain objectivity for the evaluation of BQC results. Unknown quality controls are prepared and introduced into the system by personnel not connected with the actual testing. The identity of the samples must remain unknown to the analysts throughout the entire process, and therefore the results should be checked by a person not performing actual testing (3).

This article presents performance of BQCs submitted to the HFSC toxicology section for blood alcohol analysis from September 2015 to July 2018. The results of these submissions were reported from October 2015 through August 2018.

Materials and Methods

Study design

To implement this BQC program, the quality division first reviewed the workflow of the toxicology section. Most request types received by the section are for driving while intoxicated (DWI) cases. The quality division studied the analytical process for blood alcohol analysis, from intake of evidence to reporting of the results, to ensure BQC samples moved smoothly through the process. Next, common blood alcohol concentrations (BACs) received were determined and used to create samples that mimicked normal casework.

BQCs were created using two separate blood sample sets. The first sample set, containing four different concentrations of alcohol (ethanol), was manufactured on June 25, 2015, and submitted for analysis as BQCs from September 3, 2015, through June 27, 2017. The second blood sample set contained three different alcohol concentrations and was manufactured on April 20, 2017. This sample set was submitted as BQCs from July 7, 2017, through July 20, 2018. Supplemental Table I lists the sample sizes and concentrations.

HFSC established a goal of introducing 14 blind cases per month. This represented roughly 5% of the monthly completed work. The 5% goal sought to balance introducing enough blind cases that analysts have a real likelihood of routinely handling blind cases and a manageable cost for the organization.

Sample preparation

HFSC purchased blood samples from Research Triangle Institute (RTI) International. These samples were prepared in bulk by volumetric addition of ethanol to whole blood fortified with 2% potassium oxalate. RTI aliquoted ~7 mL of the fortified blood into 10 mL grey-top vacutainer tubes, which contained 100 mg of sodium fluoride and 20 mg of potassium oxalate. The samples were analyzed in triplicate by a reference laboratory by headspace gas chromatograph interfaced with dual-flame ionization detector (HS-GC-FID). The first sample set was analyzed by Clinical Reference Laboratory and the second set was analyzed by MedTox Laboratories. The samples were shipped on ice and maintained refrigerated at ~4°C after receipt by HFSC. The quality division placed two blood tubes with fabricated case information and paperwork into typical toxicology collection kits used by submitting agencies.

To mimic a real toxicology case, information must include the agency's case number, subject name, date of birth and driver's license number, as well as submitting officer, collector name, incident date and time, offense type and location. Collaboration with the Houston Police Department (HPD) allowed the quality division to gain access to the police department's records and evidence management systems, which produces the agency's incident numbers and maintains other case-related information. The subject's name and date of birth were created using Fake Name Generator (https://www.fakenamegenerator.com/). The driver's license consisted of a string of eight digits resembling common Texas driver's license numbers. The incident date and time were dependent on the date that the samples were submitted to the laboratory. An offense location was chosen from within the Houston city limits using an online map to find an actual street address. In a genuine case, the submission form is typically filled out by both the officer and the collector. HFSC Quality Division personnel created a handwriting distinction between these two individuals and attempted, to the best of their ability, to disguise their own handwriting on this form. The tubes were placed within the toxicology collection kits, sealed with security tape, and submitted to the HPD Property Room where submission and storage of real evidence takes place.

Sample submission and evaluation

The submission of the kits to the property room was an imperative step because it allowed the kit to follow the same chain of custody as

normal cases as well as to receive the same item barcodes from the submitting agency. These kits were brought back to HFSC by HFSC evidence technicians and prepared for analysis in the same manner as real cases. During analysis, the analyst was not aware if they were working a BQC or a real case. If at any point during analysis the analyst detected a BQC, they reported it to the CEO and the quality division. The indicators were discussed between the analyst and the quality division and used to improve the blind nature of the next case. The CEO incentivized and rewarded any analyst who discovered a BQC during analysis. This helped maintain the efficacy of the BQC program and ensured that the cases remained blind. Once the analysis was completed, the quality division reviewed the reported results to determine if they were satisfactory. The BQC was deemed satisfactory if the reported value ± the uncertainty of measurement encompassed the theoretical target.

Uncertainty of measurement (UM) components for HFSC alcohol analysis included method reproducibility in % relative standard deviation (%RSD), uncertainty of certified reference materials used as calibrators, pipette variability in %RSD from external calibration and duplicate analysis variability. Uncertainty was reported at a 99.73% confidence interval (k = 3). The alcohol UM value was recalculated yearly as more quality control sample and periodic pipette calibration data accumulated. The UM was also recalculated after method changes (between September 2015 and May 2016, hydrogen was used as the carrier gas instead of helium, and the 50 µL of sample with 500 µL of internal standard was used rather than 100 and 1000 µL, respectively), and the introduction of a new headspace system (Headspace Instrument 1 was removed and Headspace Instrument 3 was added to service in May 2016). The UM values reported during the time period for this study were 9.091% (September 2015–May 2016), 9.182% (May 2016–August 2016), 9.628% (August 2016–February 2017), 8.987% (February 2017–February 2018) and 8.954% (February 2018-present).

Standards and reagents

Aqueous-mixed volatile standards containing ethanol, methanol, isopropanol and acetone at concentrations of 0.010, 0.025, 0.050, 0.100, 0.200 and 0.400 g/100 mL along with aqueous ethanol standard at a concentration of 0.500 g/100 mL were purchased from Cerilliant Corporation (Round Rock, TX). Aqueous low and high ethanol standards at 0.080 and 0.400 g/100 mL, respectively, were also purchased from Cerilliant and Lipomed (Cambridge, MA). An aqueous low-mixed volatile standard containing ethanol, methanol, isopropanol and acetone was prepared at 0.0192 g/100 mL by dilution of Cerilliant mixed volatile standard at 0.400 g/100 mL. Whole blood low and high ethanol standards (~0.08 and 0.2 g/100 mL, respectively) and whole blood low and high mixed volatile standards containing ethanol, methanol, isopropanol and acetone at ~0.03–0.08 g/100 mL and 0.08–0.15 g/100 mL, respectively, were obtained from Cliniqa Corporation (San Marcos, CA). n-Propanol internal standard was prepared at 0.01% v/v using n-propanol from EMD Chemicals Inc. (Gibbstown, NJ) in deionized water obtained from Millipore Direct-Q 3 UV water purification system (Burlington, MA).

Blood alcohol analysis

Ethanol and other volatiles in blood were quantified using Agilent 7697A headspace sampler equipped with 7890B gas chromatograph interfaced with dual-flame ionization detectors (HS-GC-FID). All calibrators, controls and blood case samples were allowed to come to room temperature and mixed by rocking prior to sampling. Using a Hamilton Microlab 600 dual pipettor-dilutor, 100 µL of the calibrators, controls and case samples along with 1000 µL of internal standard were aliquoted into 20 mL HS vials. The vials were capped, crimped and placed onto the autosampler for analysis. All case samples were analyzed in duplicate. Samples were aliquoted in the original order of the batch and then in reverse order of the batch. The 7-level aqueous Cerilliant standards ranged 0.010–0.500 g/100 mL for ethanol (the six-level aqueous Cerilliant standards ranged 0.010–0.400 g/100 mL for other volatiles) were used to generate linear calibration curves. Two air controls (empty HS vials) were inserted at the end of each batch sequence. After the highest calibrator, a negative control (deionized water mixed with internal standard) was injected, followed by a set of aqueous and whole blood controls prior to case samples. After every 10 case samples, a whole blood ethanol control was injected. Subsequently, another set of aqueous and whole blood controls was injected, followed by the water control (deionized water left open for the duration of the aliquoting process to monitor potential environmental contamination, with internal standard added at the end of sampling) and air controls. Four results were obtained for each case sample: two results of the first aliquot from FID1 and FID2 and another two results from the second aliquot from FID1 and FID2. The average result of the first and second aliquots from FID1 was reported.

Restek RTX-BAC Plus 1 (0.32 mm × 30 m × 1.80 µm) and RTX-BAC Plus 2 (0.32 mm × 30 × 0.60 µm) fused-silica capillary columns were connected to FID1 and FID2, respectively, via the splitters and restrictors. Helium was used as the carrier gas at 7 mL/min. The GC run time was 4 min with the vial equilibration time of 7 min. The HS loop temperature was set at 70°C and the transfer line temperature at 90°C. The oven temperature at 40°C was held for 4 min and increased to 50°C during the post-run. The inlet temperature was set at 110°C using a 10:1 split ratio injection. FID maintained the temperature at 250°C and used hydrogen at 30 mL/min, air at 400 mL/min and nitrogen as the makeup gas at 25 mL/min. The method was validated for linearity, sensitivity, accuracy, precision, carryover, endogenous and exogenous interferences, dilution integrity and autosampler stability. The limits of quantification for all analytes (ethanol, methanol, isopropanol and acetone) were 0.010 g/100 mL. Bias, within-run imprecision and between-run imprecision for ethanol analysis were ≤ 5.59% difference from the target, ≤ 1.54 %CV and ≤ 1.19 %CV, respectively. Further details on instrument settings and validation study outcomes can be found in the alcohol batch files and method validation packages posted on the HFSC eDiscovery site (https://records.hfscdiscovery.org/).

Statistical analysis

Statistical analysis was performed using the results from the positive samples with R version 3.5.2 (2018-12-20). Diagnostic tests performed to test the fit of the linear model included: a plot of residuals against fitted values, a Scale-Location plot of sqrt (|residuals|) against fitted values, a Normal Q-Q plot, a plot of Cook's distances versus row labels, a plot of residuals against leverages and a plot of Cook's distances against leverage/(1-leverage). The estimation of alcohol evaporation/degradation over time was performed using a linear model fit of percent differences versus age of sample in days, as well as a nonlinear model fit using local polynomial regression fitting, using the R function loess, both of which yielded the same null result. The linear regression was performed using the function lm in R, and the multi-factor analysis of variance (ANOVA) was performed

Table I. Linear Regression Model Results for the % Difference Between Reported and Target BACs

	Difference in % difference between target and reported concentrations
Target: 0.15 g/100 mL	−3.912**
Target: 0.16 g/100 mL	−13.017***
Target: 0.20 g/100 mL	−9.753***
Target: 0.25 g/100 mL	−8.999***
Analyst: 2	−0.295
Analyst: 3	1.192
Analyst: 4	2.272
Analyst: 5	2.347
Analyst: 6	−2.076
Analyst: 7	−0.909
Analyst: 8	1.142
Pipette: B	0.024
Instrument 2	−1.397
Instrument 3	−2.079
Constant	24.547***
N	228
R^2	0.354
Adjusted R^2	0.311
Residual Std. Error	6.960 (df = 213)
F Statistic	8.324*** (df = 14; 213)

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Reference categories are as follows: target, 0.08 g/100 mL; analyst, 1; pipette, A; and instrument, 1.

by applying the ANOVA function to the linear model object. The regression was considered significant for $P < 0.05$, and the *P*-value ranges were expressed using asterisks (∗) described in the caption for Table I. The percent difference was calculated by subtracting the expected theoretical value from the reported value, dividing by the theoretical value and then multiplying by 100 for each point.

Results

The results from 317 BQC samples were reported between October 10, 2015, and August 10, 2018. The age of the samples ranged from 84 to 746 in days from the date of manufacture to the date of report. The positive target concentrations included 0.08, 0.15, 0.16, 0.20 and 0.25 g/100 mL in addition to 89 negative samples. All 89 negative samples were reported as 'no ethanol detected' results, and all 228 positives were reported with satisfactory results. The ground truth for alcohol concentration for the BQC samples could not be definitively determined. However, two target values that approximate the ground truth were provided by the manufacturer. The first value was the theoretical target concentration at which the manufacturer prepared the BQC samples. The second value was the analytical result provided by the reference laboratory. The first value does not take into account the potential alcohol loss during the manufacturing process. The second value does not account for the potential differences in analytical methods and procedures between HFSC and the reference laboratory. Table II lists the means (ranges) of the reported BAC for the aforementioned target concentrations and compares % difference of the mean BQC concentrations obtained by HFSC from the theoretical target and the analytical result obtained by the reference laboratory at each concentration level. The % difference in mean for BAC results between HFSC and the reference laboratories were larger (−9.6 to 8.3%) than the % difference between mean BAC results of HFSC and the theoretical targets (−6.3 to −0.4%). This suggests the interlaboratory variability was more significant than the variability introduced from the manufacturing process. The theoretical target concentrations and reported concentrations are plotted in Figure 1. The deviation from the target is visually minimal as shown by the data points closely scattered around the slope of 1.

Effect of alcohol evaporation/degradation

Alcohol evaporation/degradation (oxidation) over time during storage was evaluated to determine its effect on the reported BQC results. The reported concentrations decreased over time, which suggested that alcohol evaporation/degradation caused the reported concentrations to be lower than the manufacturer's target concentration. The theoretical target concentrations were used to calculate % difference to evaluate the effect of alcohol evaporation/degradation and to determine the relationship via a linear regression model:

$$\text{Percent difference}_i = \beta_0 + \beta_1 \text{ Age of sample}_i. \quad (1)$$

The estimates from the model were $\hat{\beta_0} = 1.66$ and $\hat{\beta_1} = -8.52 \text{ x } 10^{-5}$, where $\hat{\beta_0}$ is statistically significant at the 0.05 level. However, $\hat{\beta_1}$, the alcohol evaporation/degradation value, was not statistically significant. This is depicted in Figure 2, where the thin solid line represents a simple linear regression line from Equation 1 with a virtually zero slope (−8.52 × 10–5), and the thick solid curve is a nonlinear loess fit. Since both lines were almost horizontal (i.e., $\hat{\beta_1}$ is minimal), they indicated a randomly distributed error and that the alcohol evaporation/degradation from the BQC samples had a negligible effect on the reported results.

Statistical model

The variables used for statistical analysis included the theoretical target concentrations of the BQC samples, the BACs reported by HFSC, the analyst's identifier (1–8), pipette (A or B) and instrument used (1, 2 or 3). A multiple linear regression analysis was performed to determine what variables affected the variability of the percent errors in measurement, i.e. the difference between the reported concentrations from the analyst and the theoretical target from the manufacturer, all divided by the theoretical target from the manufacturer. The linear regression model used was

$$\begin{aligned}\text{Percent difference}_i &= \beta_0 + \beta_1 \text{ Target level}_i + \beta_2 \text{ Analyst}_i \\ &\quad + \beta_3 \text{ Pipette}_i + \beta_4 \text{ Instrument}_i + e_i, \quad (2)\end{aligned}$$

where *e* is the error and the subscript *i* refers to each observation in the given dataset and ranges from 0 to $N = 228$.

After fitting the models, diagnostic plots were created to confirm that the main assumptions required to perform a linear regression were satisfied. Results indicated normal residuals, no extreme outliers and symmetrical error around zero (Table I and Supplemental Table II). Supplemental Table II presents the same model as an ANOVA.

None of the differences between analysts, pipettes or instruments were statistically significant ($P > 0.2$). Each of the target concentration levels was statistically significantly different from each other. Furthermore, the 0.08 g/100 mL target showed a significantly larger % difference than higher target concentrations (0.15–0.20 g/100 mL); however, the % differences among the higher targets were not concentration dependent. In fact, 0.16 g/100 mL had the lowest differences

Table II. Summary of BQC BACs Reported by HFSC Compared to the Theoretical Target and the Mean Values from the Reference Laboratory, which Analyzed BQC Samples in Triplicate Before Sending the Samples to HFSC

Theoretical target, g/100 mL	Total sample size (n)	Mean concentration from reference laboratory, g/100 mL	HFSC mean concentration (range), g/100 mL	% Difference HFSC mean from theoretical target	% Difference HFSC mean from reference laboratory mean
0	89	0	N/A*	N/A*	N/A*
0.08	50	0.083	0.075 (0.073–0.078)	−6.3	−9.6
0.15	45	0.133	0.144 (0.140–0.148)	−4.0	8.3
0.16	48	0.163	0.157 (0.155–0.160)	−1.9	−3.7
0.20	43	0.185	0.195 (0.192–0.200)	−2.5	5.4
0.25	42	0.241	0.249 (0.242–0.258)	−0.4	3.3

*The negative samples did not contain ethanol and therefore the analysts reported them as 'no ethanol detected'; these are shown as N/A.

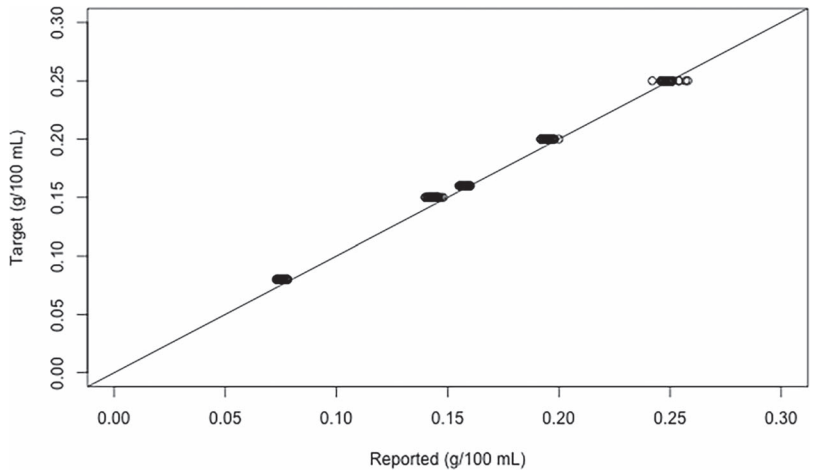

Figure 1. Reported BACs from 228 BQC samples were compared to the manufacturer's theoretical target concentrations. The solid line is the point at which the two concentrations are equal.

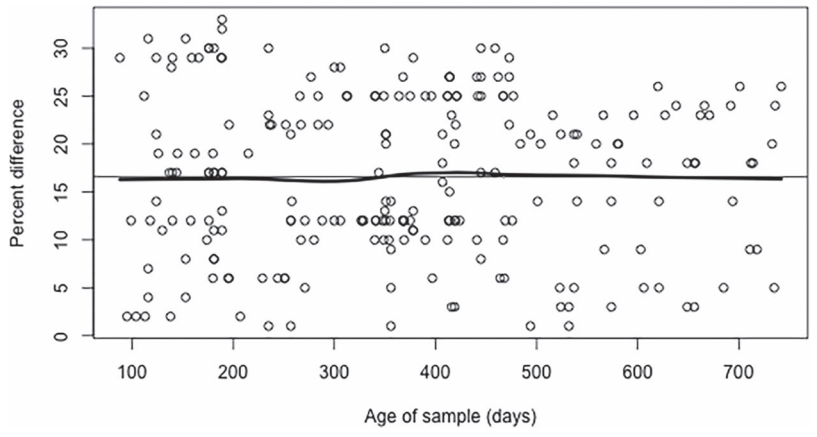

Figure 2. Percent difference (between the reported BACs and the theoretical targets of 0.08, 0.15, 0.06, 0.20 and 0.25 g/100 mL) versus the age of the tested samples. The thin solid line represents a simple linear regression line from Equation 1 with a virtually zero slope ($-8.52 \times 10{-}5$) indicating randomly distributed errors. The thick solid curve is the nonparametric loess fit, which also forms a nearly horizontal line, suggesting a minimal effect of alcohol loss over time.

(i.e., best performance), followed by the 0.20 and 0.25, then 0.15 and then 0.08 g/100 mL target levels. This suggested that the source of the theoretical target level variability was likely from the manufacturing process rather than the systemic bias from the analytical method. The alcohol loss and potential error/bias during the manufacturing process appears to mask the possible loss during storage as the % difference was significantly affected by the target concentration levels but not by the age of samples. The data demonstrated that the contributions from different analysts, instruments and pipettes to the overall variability were not as significant as that from interlaboratory analytical differences and the effects of manufacturing and storage processes.

Discussion

Benefits

Submitting BQCs into regular toxicology casework has allowed for a continuous assessment of the case processing workflow. Fundamentally, BQCs provide continual monitoring of the entire laboratory system, not just analytical steps. Additionally, the presence of blind samples capitalizes on the Hawthorne Effect (6). Generically, the Hawthorne Effect is the tendency of people to alter their behavior when they know they are being monitored. While it is difficult to quantitate the impact on behavioral changes in analysts because of the presence of blind control samples, minimally the results of blind testing provide a set of data where potential analyst bias associated with open proficiency is controlled. When blind results routinely demonstrate low variability over time, between analysts, instruments and calibrations, greater confidence in the routine results of unknown samples can be inferred. Objective data can be provided to demonstrate that the expected answer is obtained whether proficiency is announced or is blinded. Furthermore, the BQC program provides analysts with objective data during court testimony that demonstrates that the procedures and methods used to analyze evidence samples are reliable and reproducible among all factors that could contribute to variability of results during analysis.

Prior to submission, the blind samples are stored in a refrigerator that is located in the BQC laboratory, only accessible by the quality division and HFSC evidence technicians, under the temperature conditions required by the manufacturer and HFSC policy. These blind cases are then submitted to the HPD Property Division to mimic a normal case submission. Since all DWI cases are automatically requested for alcohol analysis, the BQCs return to HFSC the same day they are submitted for analysis. Blind samples are stored at HFSC with real evidence samples after their submission. Therefore, the results between BQCs and real evidence samples can be compared to determine if storage conditions, such as drastic temperature changes related to refrigerator failures or evidence transportation from HPD to HFSC, influenced the BAC. This comparison is not possible with calibrators, standards and other quality control samples since they are required to be stored separately from evidence. Notably, for the BQCs in standard grey top tubes, refrigerated and handled as case samples, evaporation/degradation of ethanol was not a significant effect over 746 days from manufacture. After analysis, the BQCs are returned to the HPD property room, as typical with normal evidence. The BQCs are later retrieved from the property room by HFSC evidence technicians for long term storage in the BQC laboratory.

In 2016, a toxicology evidence refrigerator broke down causing the temperature to spike outside the storage temperature range (0–8°C). Evidence samples were left at room temperature for ~4 hours while being transferred to a functioning refrigerator. Of the 529 evidence samples in this refrigerator, 16 of these were BQC samples; 5 were negative, and 11 were positive. The reported results for the 11 positive samples were consistent with results obtained from blinds in the same sample set prior and after this incident. These blinds performed satisfactorily and therefore demonstrated that the stability of ethanol was not affected by the samples being at room temperature for 4 hours prior to their analysis.

Blind samples have provided multiple opportunities for process improvement. For instance, shortly after incorporating the BQC program in the toxicology section, a sample tube for outsourcing confirmatory testing broke during transit. Fortunately, this blood tube was a BQC. This provided an opportunity to evaluate and improve the shipment process. As a result, HFSC worked with a vendor to create a new version of the toxicology collection kit to prevent this from recurring.

Implementation challenges

The implementation of an intralaboratory comparison program that is blind to laboratory personnel presented many challenges. One of the biggest challenges was gaining buy-in from the stakeholders (e.g., submitting agency, district attorney's office, staff). It was essential for the stakeholders to understand the program and how it benefitted them. A successful BQC program provides confidence in the reliability of the test results and suitability of the laboratory processes. HFSC worked closely with the HPD to form important relationships to enable collaboration and gain buy-in for the program. Laboratories are dependent on submitting agencies' and requestors' participation for successful implementation of BQC samples.

Once buy-in from stakeholders was obtained, the next challenge was to create samples that mimicked evidence submitted to the laboratory. As previously mentioned, the BQCs were prepared using the same toxicology collection kits that are used in real cases. The blood tubes were prepared by an external vendor to look like tubes that would be created during a blood draw for a DWI case. In HFSC's experience, the single most challenging factor to mimic is handwriting on submission forms. Most cases discovered as BQCs by the analysts have been because the handwriting was 'too neat'. The quality division then started to disguise the handwriting as much as possible and even used their non-dominant hand to fill out the submission form.

At the start of the program several kits were submitted at one time and inevitably were assembled into the same batch for analysis with similar case numbers. These cases were quickly discovered by the analyst as blind cases because of the nearly consecutive case numbers. The BQC submission schedule was revised to include a smaller number of cases over a span of multiple days in order to spread out the cases amongst the batches. During the study period, less than 5% of the 317 BQCs submitted were discovered as blinds by the analysts.

The development of a BQC program has been presented from a practical point of view based on the needs of HFSC and the toxicology section. Despite early critiques of such a blind quality program being considered 'entrapment (7)', HFSC values the BQC program because it provides analysts with real-time feedback about their casework and allows HFSC to more quickly see potential problems in casework and take preventative measures rather than corrective actions. Toxicologists agree that seeing the results of their BQCs gives them confidence and encourages them to improve their procedures and internal quality control systems. HFSC encourages other forensic laboratories to consider implementing their own BQC program that is tailored to their specific needs.

Conclusion

Because toxicology analysis results have a direct impact on the lives of many people, it is vital that results are accurate and reliable on every sample tested; any false negative or positive could result in grave consequences, including, but not limited to, revocation of probation, imprisonment, and improper medical treatment (3). Thus, in addition to regular PTs as required to maintain accreditation, HFSC implemented a BQC program to assess laboratory procedures in real time.

Since implementing the blind program in the toxicology section in September 2015, HFSC has yet to receive an unsatisfactory result from a blood alcohol analysis BQC case. The results indicate no significant variation in blood alcohol results over time, between analyst, instrument or pipette. Therefore, we can conclude that the methods are reliable and produce accurate results. If a practical onsite program can be developed and put into use by other laboratories, PTs and a BQC program will become a reliable means of evaluating laboratories.

Acknowledgements

The authors thank the HPD and Martin Lopez of the HPD Property Division for their assistance in implementing the BQC program and the HFSC Toxicology section for willingly incorporating BQC samples into their operation.

References

1. Peterson, J.L., Lin, G., Ho, M. (2003) The feasibility of external blind DNA proficiency testing. I. Background and findings. *Journal of Forensic Sciences*, 48, 21–31.
2. LaMotte, L.C., Guerrant, G.O., Lewis, D.S., Hall, C.T. (1977) Comparison of laboratory performance with blind and mail-distributed proficiency testing samples. *Public Health Reports*, 92, 554–560.
3. Jain, N.C., Sneath, T.C., Budd, R.D. (1977) Blind proficiency testing in urine drug screening: The need for an effective quality control program. *Journal of Analytical Toxicology*, 1, 142–146.
4. Committee on Identifying the Needs of the Forensic Sciences Committee, National Research Council. (2009) *Strengthening Forensic Science in the United States: A Path Forward*. National Academy of Sciences: Washington, DC, https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf (accessed Feb 6, 2019).
5. ISO/IEC 17025 *General Requirements for the Competence of Testing and Calibration Laboratories*. International Organization for Standardization: Geneva, Switzerland, 2017.
6. McCambridge, J., Witton, J., Elbourne, D.R. (2014) Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, 87, 267–277.
7. Shoemaker, M.J., Sideman, L. (1977) Urine screens in monitoring methadone programs. *Journal of the American Medical Association*, 237, 869–887.