

TECHNICAL NOTE**CRIMINALISTICS**

Joli Bregu,¹ B.S.; Danielle Conklin,¹ M.S.; Elisse Coronado,¹ M.S.; Margaret Terrill,¹ M.S.F.S.; Robin W. Cotton,¹ Ph.D.; and Catherine M. Grgicak,¹ Ph.D.

Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis^{*,†}

ABSTRACT: Determining appropriate analytical thresholds (ATs) for forensic DNA analysis is critical to maximize allele detection. In this study, six methods to determine ATs for forensic DNA purposes were examined and compared. Four of the methods rely on analysis of the baseline noise of a number of negatives, while two utilize the relationship between relative fluorescence unit signal and DNA input in the polymerase chain reaction (PCR) derived from a dilution series ranging from 1 to 0.06 ng. Results showed that when a substantial mass of DNA (i.e., >1 ng) was amplified, the baseline noise increased, suggesting the application of an AT derived from negatives should only be applied to samples with low levels of DNA. Further, the number and intensity of these noise peaks increased with increasing injection times, indicating that to maximize the ability to detect alleles, ATs should be validated for each post-PCR procedure employed.

KEYWORDS: forensic science, minimum distinguishable signal, minimum discernible signal, forensic DNA analysis, analytical threshold, signal to noise

In dealing with fluorescent emission from DNA fragments amplified via polymerase chain reaction (PCR) methods, it is of interest to determine limits of detection in order to select among methods and optimize procedures. To determine the limits of detection for a given analytical procedure, it is necessary to determine a minimum distinguishable signal (MDS)—the signal at which a peak can reliably be distinguished from noise. The MDS may be considered a relative fluorescence unit (RFU) or analytical threshold (AT) for forensic purposes. Additionally, SWGDAM Interpretation Guideline for Autosomal STR Typing 3.1.1.2 states the AT should be established on signal-to-noise considerations and should not be used for purposes of avoiding artifact labeling, which may result in loss of data (1).

A review of the forensic and analytical literature revealed a number of expressions and terminology, where a number of definitions were inconsistent or ambiguous. For example, one finds terms such as lower limit of detection, limit of detection, MDS, minimum discernible signal, detection sensitivity, sensitivity, limit of determination, limit of purity, and limit of quantitation (2–9). To compound the problem, statements akin to “detection limit at 99.7% confidence” are used without consideration of the distribution. Typically, the normal or Gaussian distribution is

assumed namely because it is well defined, easily applied to analytical applications and has a plethora of formulae and tables ready for use. However, the normal distribution may not be the norm in chemical analysis. For example, technical limitations to the equipment and/or software may preselect data such that asymmetric distributions are observed. If this is the case, the correct 99.87% confidence (as opposed to the erroneously stated 99.7%) no longer applies and approaches 89% as per Tschebyscheff’s inequality or the “Law of Large Numbers” (10).

Because traditional analytical measurement is interested in the ability to quantitate a trace element or molecule, complications from other components need also be considered. As a result, the figures of merit calculated for a given method should include all steps of the analytical procedure tested. These figures include the limit of detection, sensitivity, precision, bias, dynamic range, and selectivity (11). The fact that forensic DNA analysis utilizes an intermediary amplification step may complicate matters. For example, if nonspecific amplified product is considered noise, the reliability of determining an accurate MDS from analysis of the instrument’s baseline may be decreased.

This work represents a detailed examination into the meaning and application of analytical “figures of merit” to forensic DNA analysis. It is not intended as a comprehensive statistical review of the various methods of calculating these figures. Comprehensive reviews on the subject are available, and the readers are referred to articles by Kaiser (10), Currie (4–6,12), Mocak et al. (8), and Ingle (13). Implications to the determination and use of MDSs as the analytical (RFU) threshold are discussed. Multiple methods of determining thresholds were examined, where each was analyzed with and without common assumptions. Implications to the nonlabeling of “true” alleles and the false

¹Biomedical Forensic Sciences, Boston University School of Medicine, 72 E. Concord Street, Boston, MA 02118.

*Presented in part at the 21st International Symposium for Human Identification, October 11–14, 2010, in San Antonio, TX.

†Supported in part by Boston University School of Medicine Biomedical Forensic Sciences Program and NIH 2008-DN-BX-K158.

Received 15 Dec. 2010; and in revised form 17 Nov. 2011; accepted 4 Dec. 2011.

labeling of noise were assessed. Sensitivities and signal-to-noise comparisons were examined in the context of forensic short tandem repeat (STR) typing. Last, evaluation of the signal-to-noise determined whether ATs derived from negatives are representative of samples amplified with common DNA input amounts, or whether thresholds need be derived from methods that utilize signal from a set of samples containing DNA.

Materials and Methods

The procedures followed are in accordance with the ethical standards of the Institutional Review Board. All reagents were purchased from Sigma Aldrich (St. Louis, MO) unless stated otherwise.

Organic Extraction

High molecular weight DNA was extracted from 750 μL of whole blood from four individuals using phenol/chloroform extraction and alcohol precipitation. The DNA was dissolved in 50 μL of TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) at 56°C.

Quantitation—Ultraviolet-Visible Spectrophotometry

Visible spectroscopy (Genesys 10S; Thermo Scientific, Rockford, IL) was used to determine the concentration of each DNA sample where it was assumed that one absorbance unit corresponded to 50 ng/ μL of double-stranded DNA. A_{260}/A_{280} was assessed to ensure there was minimal protein contamination. As the Genesys 10S is a single-beam instrument, a blank sample containing the dilution buffer was run prior to the sample and the background was subtracted. The scan ranged from 400 to 200 nm. Each sample was run in triplicate, and the mean concentrations were used to prepare the samples for amplification. The average and two standard deviation values of A_{260}/A_{280} over all four samples were 1.8 ± 0.2 .

Amplification—AmpF/STR® Identifier®

Each of the four specimens was run in quadruplicate at each of the seven target quantities (i.e., mass of DNA input into the amplification in ng—4, 2, 1, 0.5, 0.25, 0.125, and 0.0625 ng) using Applied Biosystems (Foster City, CA) AmpF/STR® Identifier® PCR Amplification Kit (14), resulting in a total of 96 samples. Each of the samples was diluted in TE Buffer (Tris-EDTA, 10^{-4} M). To amplify each sample in quadruplicate, the appropriate amount of master mix and DNA was combined in a volume large enough for quadruplicate amplifications. A volume of 25 μL of the master mix/DNA mixture was aliquoted into four wells on the same plate. PCR amplification was carried out according to the manufacturer's protocol, with an additional 60-min extension during thermal cycling that was performed on Applied Biosystems GeneAmp® PCR System 9700 using 9600 emulation mode. Positive and negative amplification controls were also run and showed expected results (data not shown). Additionally, baseline data on 31 negative samples, consisting of both extraction and amplification controls, run over a period of 14 months using multiple kits were collected for further analysis.

Capillary Electrophoresis—3130 Genetic Analyzer

Fragment separation was accomplished using Applied Biosystems 3130 Genetic Analyzer and a mixture containing appropriate amounts of HiDi (highly-deionized) formamide

(8.3 $\mu\text{L}/\text{sample}$) and GeneScan™-600 LIZ® Size Standard (0.7 $\mu\text{L}/\text{sample}$) (Applied Biosystems). A volume of 9 μL of that mixture and 1 μL of sample, negative or ladder, was added to the appropriate wells. Samples hereafter referred to as “run blanks,” which consisted of 0.7 μL of GeneScan™-600 LIZ® Size Standard and 8.3 μL of formamide were also run. Additionally, the DNA samples (containing 2–0.0625 ng from each contributor amplified in quadruplicate) were run and analyzed. The samples were put on a heating block at 95°C for 3 min and snap-cooled at –20°C for 3 min. Two-, 5-, and 10-sec injections at 3 kV were performed on each of the samples and run according to the manufacturers recommended protocol (14).

Comparison of Six Methods to Determine Analytical Thresholds

The sample name, allele, peak size, and peak height for each sample were exported from GeneMapper® ID v 3.1 (Applied Biosystems) into Microsoft® Excel 2007 (Redmond, WA) or Igor v. 6.1.2.1 (WaveMetric Inc, Lake Oswego, OR) for analysis. The following are descriptions of each method used to calculate ATs:

- Method 1 utilized a total of 31 run blanks (i.e., formamide and LIZ®) or 31 negatives. An RFU threshold of 1 was used, and all peaks within ± 2 bases from the internal lane standard were removed. The average and standard deviation for all run blanks and negative signals were calculated and substituted into the following equation.

$$AT_{M1} = \bar{Y}_{bl} + k s_{Y,bl} \quad (1)$$

- AT_{M1} is the AT using Method 1, \bar{Y}_{bl} is the average blank signal, $s_{Y,bl}$ is the standard deviation of the blank signals, and k is a numerical factor chosen in accordance with the confidence level desired. The Gaussian or log-normal least squares fit was performed on the blanks and negatives with Igor v. 6.1.2.1.
- Method 2 is the percentile rank method. Here, the percentage of noise peaks below a given value was calculated to determine the proportion of baseline, which is expected to be below said value. For example, for the 2-sec green channel, there were a total of 5317 baseline points > 0 RFU labeled for the 31 blank samples, where 3.87% of them were ≤ 1 RFU (i.e., 206/5317) and 31.72% were ≤ 2 RFU (i.e., [206 + 1481]/5317), etc. An example of how such a method may be employed is shown in Table 1.
- Method 3 (as in Method 1) relies on the average and standard deviations of the noise peaks calculated from the 31 blank or negative samples and is expressed as the following equation,

$$AT_{M3} = \bar{Y}_{bl} + t_{\alpha,v} \frac{s_{Y,bl}}{\sqrt{n_{bl}}} \quad (2)$$

where AT_{M3} is the AT derived using Method 3, \bar{Y}_{bl} and $s_{Y,bl}$ have the same definitions as in Method 1, $t_{\alpha,v}$ is the critical value obtained from the t -distribution for a given confidence interval (one-sided), and n_{bl} are the number of negative samples.

- Method 4 is similar to Methods 1 and 3 in that it also relies on utilizing the mean and standard deviation of blank signal and is expressed as follows,

$$AT_{M4} = \bar{Y}_{bl} + t_{\alpha,v} (1 + 1/n_{bl})^{1/2} s_{Y,bl} \quad (3)$$

where AT_{M4} is the AT using Method 4, \bar{Y}_{bl} and $s_{Y,bl}$ have the same definitions as in Methods 1 and 3, $t_{\alpha,v}$ is the critical

TABLE 1—An example of the percentile ranks of noise of 31 run blanks that may be used to calculate the analytical threshold (AT) for the green channel, 2-sec injection. The AT calculated using this method is chosen at a percentile rank of >99%.

RFU Signal of Blanks	No. of Observations	Percentile Rank	AT _{M2} at Rank >99%
1	206	3.87	
2	1481	31.73	
3	1884	67.16	
4	1161	89.00	
5	453	97.51	6
6	110	99.59	
7	18	99.92	
8	3	99.98	
9	1	100	

value obtained from the *t*-distribution for a given confidence interval (one-sided), and n_{bl} are the number of negative samples. In both Methods 3 and 4, a confidence interval is calculated, where Method 4 takes into account the variance of the difference between average calculated mean and individual blank signals (see Mocak et al. [8] for a complete discussion), that is, the term $(1 + 1/n_{bl})$ expresses the correction for the uncertainty of the true and calculated mean blank signal. Therefore, as n_{bl} approaches infinity, the term approaches 1.

- Method 5 is known as the upper limit approach (8) and utilizes the upper boundary of the signal versus concentration confidence band, thereby taking into account the uncertainty of the regression line (i.e., error in the signal with respect to concentration) and therefore the uncertainty of the *y*-intercept. In this case, it is the upper limit of the *y*-intercept, which is utilized as the AT.
- Method 6 is a simplified version of Method 5 and assumes the *y*-intercept is not different from the mean blank signal. A way to determine the validity of this approach is based on whether the *y*-intercept $\pm (1 - \alpha)100\%$ contains the mean blank signal (i.e., \bar{Y}_{bl}). If the mean blank signal is included in the *y*-intercept band, the following relationship can be used to determine the AT

$$AT_{M6} = b + t_{(n-1,\alpha)}S_Y \quad (4)$$

where AT_{M6} is the AT calculated using Method 6, b is the *y*-intercept, $t_{(n-1,\alpha)}$ is the *t*-statistic at a chosen significance, n is the number of points in the plot, and S_Y is the standard error of regression.

A linear least squares weighted regression between RFU signal and input mass of the positive samples was performed. Igor v. 6.1.2.1 was used for Method 5, and a Microsoft[®] Excel 2007 template made by this laboratory (15) was used to calculate the “simplified” least squares weighted regression parameters for Method 6. In both cases, the mass range was limited to samples amplified from 0.0625 to 1 ng.

Results

ATs vary from laboratory to laboratory and can range anywhere from the very low to the extremely high (i.e., 25–200 RFU). These values have a large impact on the ability of the analyst to interpret low-level samples, particularly if they are present within a mixture. For example, Fig. 1 shows a 1:19 male/male mixture amplified at 2 ng using the AmpF/STR[®]

Identifiler[®] Kit (5-sec injection). At a threshold of 30 RFU (indicated by the gray line), the entire minor component (indicated by solid peaks) is labeled. At a threshold of 55 RFU, “drop-out” begins to occur. At high RFU thresholds (i.e., 150 RFU), only three of the seven peaks corresponding to the minor component are labeled. At a 30 RFU threshold, the entire minor component is labeled, but so are nonallelic peaks. Although some of these may be characterized as known artifact, such as bleed-through, -A, or spikes, the majority cannot. Therefore, Fig. 1 illustrates the need for forensic DNA laboratories to ensure all discernable profile peaks are labeled and interpreted appropriately. This in part can be accomplished by determining the MDS.

Analytical Thresholds Calculated Using Negative Samples

One of the goals of any analyst is to report the detection of very small quantities of analyte given a specific method or procedure. There are a number of ways to accomplish this where the classical approach is based on the International Union of Pure and Applied Chemists (IUPAC) and American Chemical Society (ACS) definitions, which in large part were in response to Kaiser’s discussion on *Quantitation in Elemental Analysis* (10). In this approach, a number of blanks (i.e., 20–30) are run and the mean and standard deviation of the blank signals are used to determine the fluctuations of noise for each analytical procedure. For the signal to be accepted as true—and not just an accidentally high level of baseline noise—a multiple k is used in Eq. (1) such that only signal above this level is considered reliable. The factor k is chosen based on the desired level of confidence, and as such, a discussion regarding its value is warranted. Unless there is a normal distribution of baseline signal, it would be incorrect to state that when establishing an AT, a value of $k = 3$ gives 99.87% confidence (one-sided) and therefore represents “genuine” signal and not randomly high levels of baseline noise. Figure 2 represents typical electropherograms from a run blank injected using 2-, 5-, and 10-sec injection times (green fluorophore shown). Qualitatively, it was observed that the baseline between injections seemed to remain constant—albeit bleed-through from the LIZ[®] did increase with injection time. Figure 3 shows a histogram of the noise signal (green channel) obtained when 31 negatives (amplified samples containing no DNA) were run using a 10-sec injection time. Both normal and log-normal fits via the least squares method, where the bin size and number of bins were chosen via the method detailed by Sturges (16,17), were performed. Not only were the correlation coefficients closer to one for the log-normal fit, but residual plots showed the residuals were randomly distributed throughout. This was not the case for the normal fit which showed large positive residuals near the left side of the fit, indicating the distribution is better described as log-normal. Further, the cumulative frequency versus each labeled noise peak was plotted for each color channel and injection time and did not exhibit as S-shaped curve as is expected with normally distributed data (not shown). Therefore, using the log-normal results, the maxima for the 2-, 5-, and 10-sec injections were determined to be 2.56 ± 0.03 , 2.60 ± 0.03 , and 2.64 ± 0.03 RFU (the \pm error is two times the standard deviation originating from the fit and not the standard deviation of the blank measurements) and the width of the curves were 0.51 ± 0.01 , 0.50 ± 0.01 , and 0.50 ± 0.01 , respectively. Each injection time resulted in noise distributions, which were not significantly different from each other, indicating injection time had an insignificant impact on noise derived from running negative samples.

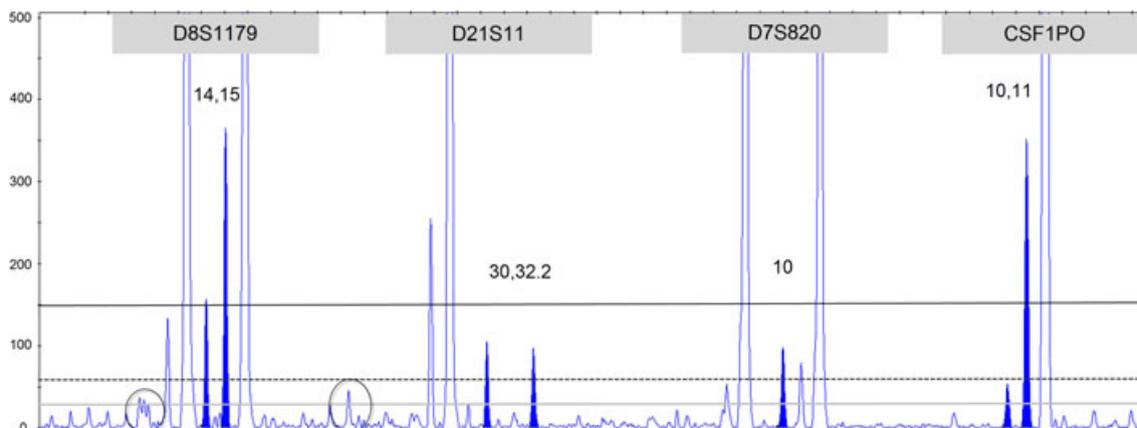


FIG. 1—Illustration of RFU threshold on the ability to interpret DNA data. Electropherogram of the blue channel of a 1:19 male:male mixture amplified using a 2-ng target. Black line (—) represents an analytical threshold (AT) of 150 RFU, dotted line (- -) represents an AT of 55, and the solid gray line is representative of a 30 RFU threshold. Noise peaks above 30 RFU are circled and the nonoverlapping alleles of the minor component are labeled.

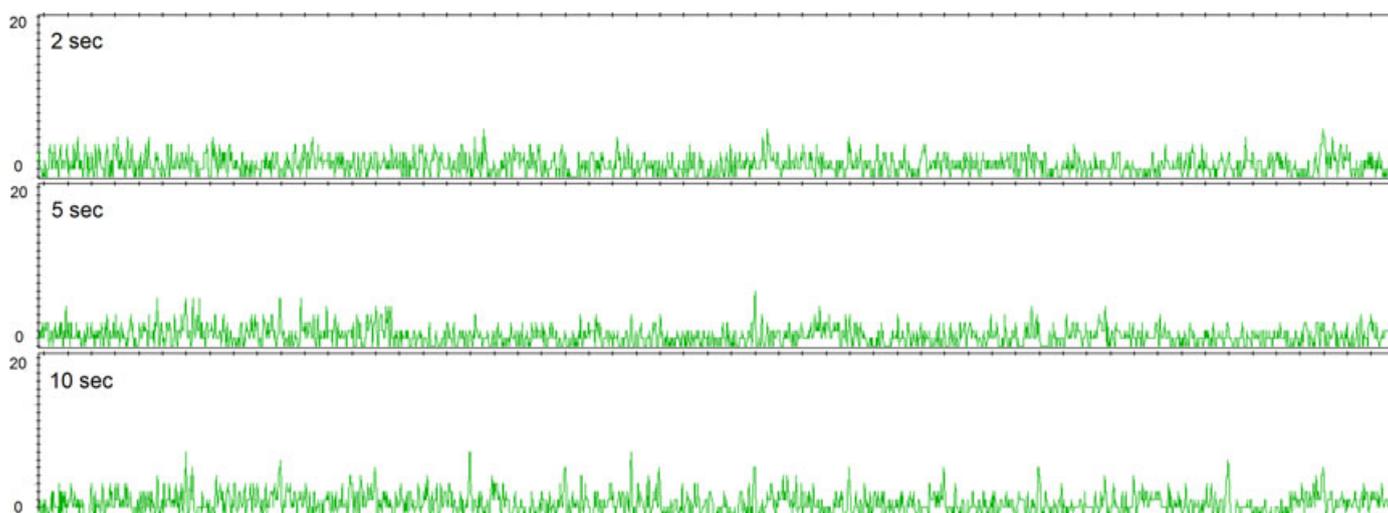


FIG. 2—Representative green channel electropherograms from a run blank consisting of 9.3 μ L formamide and 0.7 μ L LIZ[®] 600 size standard. Each blank was injected using 2-, 5-, and 10-sec injection times.

The observation that noise is not normally distributed raised concerns regarding the establishment of the AT via Method 1. Therefore, the noise was analyzed using a percentile rank method (Method 2). Table 2 shows that ATs calculated via Method 1 result in values with high percentile ranks (all above 99.28%), suggesting there is still a high confidence that noise will not be labeled if the AT is calculated via Method 1. However, given the number of baseline peaks per sample and their distribution, laboratories may choose to utilize a k of 4 in lieu of 3. For example, as shown in Table 3, the AT_{M1s} eliminated at least 99.28% of the background noise from analyses as shown by the percentile rank for the red channel of the 10-sec injection for the negatives. This suggests that if an AT_{M1} with a k of 3 was utilized, the confidence level that noise will not be labeled is closer to 99% than to 99.9%. Because an electropherogram contains *c.* 150 baseline points >1 RFU, a confidence level of *c.* 99% would suggest that two of every three electropherograms are expected to result in one labeled baseline peak. This may be considered too high for some forensic applications. In contrast, a confi-

dence level of at least 99.9% suggests that approximately one of every seven electropherograms will have a randomly high noise peak above AT. Other nonallelic peaks that need to be considered are artifacts, which result due to the chemistry of the PCR and/or the equipment. Typically, nonallelic signal originating from stutter, -A, disassociated dyes, and incomplete spectral separation is reproducible and can be characterized as artifact. Table 2 also shows that although the ATs did not differ between injection times within a color, they did differ between colors and is the result of the differences in sensitivities between the dyes.

It is also noted that Method 2 may also be utilized to determine the AT. A laboratory may choose to determine the percentile ranks as described earlier, and based on the desired level of confidence, choose the RFU at which a predetermined percentage of baseline signals fall below (i.e., 99% or 100%). Although this method has not been accepted by IUPAC or the ACS, it is presented here because it affords the laboratory the ability to determine an AT for negative samples without the need to consider the type of distribution. Additionally, it does not require

additional software or significant statistical or programming expertise to implement (Table 3).

In the laboratory, several steps are performed prior to DNA detection via fluorescence, and therefore, it was of interest to examine whether a set of negatives analyzed utilizing the two

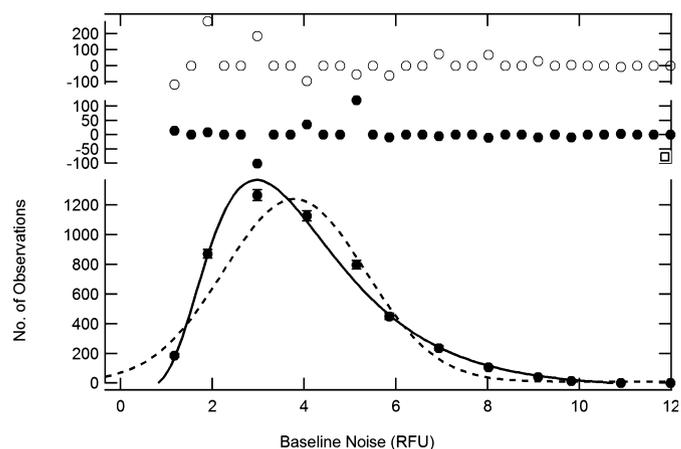


FIG. 3—Histogram of the baseline signal (green fluorophore) obtained when 31 negatives were run on a 3130 Genetic Analyzer using a 10-sec injection and the corresponding normal (- -) and log-normal (-) “best-fit” curves. The residuals resulting from the normal (o) and log-normal (•) fits. The error bars are the square root of the number of observations and were used for weighting during the fit.

methods described earlier resulted in ATs similar to the run blanks. The resulting ATs calculated via Method 1 for a set of 31 negatives are summarized in Table 2. As with the run blanks, injection times did not have a significant impact on ATs between injection times, but ATs were different between color channels. Although the ATs calculated between run blanks and negatives are not substantially different, utilization of the negatives rather than run blanks is more representative of forensic samples and is recommended.

In addition to the method cited by Kaiser (10), other methods and formulae that attempt to describe the point at which signal can reliably be distinguished from baseline have been proposed. One such method was adopted by IUPAC in 1995 (18) and is largely based on Currie’s work (4–6) (Method 3). The other method is described in detail by Mocak et al. (8) (Method 4). Table 3 includes the ATs calculated using Methods 3 and 4 using 31 negatives. Because it was previously shown that injection times did not significantly impact the \bar{Y}_{bl} and $s_{Y,bl}$, nor did these values need to be obtained by a more complicated method, the average mean and standard deviation of 31 negatives injected for 5 sec were used to calculate these ATs. It should also be noted that by utilizing the *t*-distribution and Eqs (2) and (3), a normal distribution is once again assumed. Because the data are better described as following a log-normal distribution, the confidence intervals chosen to determine $t_{\alpha,v}$ will be lower than expected. The ATs derived from all methods are summarized in Table 3 for comparison purposes.

TABLE 2—Analytical thresholds (ATs) calculated using $AT_{M1} = \bar{Y}_{bl} + ks_{Y,bl}$ (Eq. [1], $k = 3$) and the corresponding percentage of noise peaks that fall below AT_{M1} (i.e., percentile rank).

Color Channel	Injection Time (s)	AT_{M1} Calculated Using Eq. (1) and Run Blanks (RFU)	Percentile Rank of the AT_{M1} Value for Run Blanks	AT_{M1} Calculated Using Eq. (1) Amplification Negatives (RFU)	Percentile Rank of the AT_{M1} Value for Amplification Negatives
Blue	2	5	99.91	7	99.90
	5	6	99.93	7	99.72
	10	6	99.32	8	99.92
Green	2	7	99.92	9	99.82
	5	7	99.94	9	99.57
	10	7	99.90	10	99.23
Yellow	2	12	99.96	12	99.56
	5	12	99.93	14	99.31
	10	12	99.72	16	99.34
Red	2	12	99.91	12	99.64
	5	13	99.51	13	99.74
	10	13	99.91	15	99.28

TABLE 3—Analytical thresholds (ATs) calculated using Methods 1 ($AT_{M1} = \bar{Y}_{bl} + ks_{Y,bl}$), 2 (percentile rank), 3 ($AT_{M3} = \bar{Y}_{bl} + t_{\alpha,v} \frac{s_{Y,bl}}{\sqrt{n_{bl}}}$), and 4 ($AT_{M4} = \bar{Y}_{bl} + t_{\alpha,v} (1 + 1/n_{bl})^{1/2} s_{Y,bl}$). Thresholds were derived by utilizing the mean and standard deviation of 10-sec injection data of 31 blanks and 31 amplification negatives with a $t_{\alpha,v}$ of 2.46 (99% confidence, one-sided).

Type of Sample	Color Channel	\bar{Y}_{bl} Mean Negative Signals	$s_{Y,bl}$ Standard Deviation of the Negative Signals	AT_{M1} Calculated Using Method 1 (Percentile Rank)	AT_{M2} Calculated Using Method 2 at 99th Percentile Rank	AT_{M3} Calculated Using Method 3 (Percentile Rank)	AT_{M4} Calculated Using Method 4 (Percentile Rank)
Blanks	Blue	2.40	1.11	6 (99.32)	6	3 (86.96)	6 (99.32)
	Green	3.14	1.14	7 (99.90)	6	4 (87.93)	6 (99.32)
	Yellow	6.17	1.82	12 (99.72)	11	7 (78.05)	11 (99.41)
	Red	6.37	1.89	13 (99.91)	11	8 (86.62)	12 (99.84)
Amplification negatives	Blue	3.08	1.55	8 (99.92)	8	4 (85.52)	7 (98.99)
	Green	3.96	1.81	10 (99.57)	9	5 (82.79)	9 (99.24)
	Yellow	6.23	3.17	16 (99.34)	13	8 (87.37)	15 (99.34)
	Red	6.43	2.70	15 (99.28)	15	8 (83.02)	14 (98.98)

Analytical Thresholds Calculated Using Samples Containing DNA

Figure 4 shows the baselines of three types of single-source samples, where (panel *a*) represents a sample with high input (i.e., 2 ng), (panel *b*) represents baseline of a negative, and (panel *c*) represents a low-input (i.e., 0.06 ng) sample, prepared using 1 μ L of product and injected for 5 sec. Qualitatively, it is observed that the baselines are not similar between samples, and the baseline in Fig. 4*a* is substantially “noisier” than those of the other two samples. To assess the probability of erroneously labeling noise, analysis of average allelic signal to “noise” peak—which could not be characterized as known artifact (i.e., bleed-through, stutter, -A, spike, etc.)—was performed and is depicted in Fig. 5. Loci exhibiting off-scale data were not included. Figure 5 represents the average RFU signal (normalized for diploidy) at a given locus plotted against all “noise” peaks $>AT_{MI}$. It is observed that as the injection, and therefore peak height, of the largest allele increases so does the probability of labeling amplification noise. Additionally, it is observed that the majority of noise peaks $>AT_{MI}$ originate from samples that were amplified using >1 ng of DNA and gets worse with increasing injection time, suggesting the input of substantial masses of DNA into the PCR is the reason for this increased noise. As a result, if a laboratory were to utilize an AT derived from methods that use negative baseline signals—where the ATs are *c.* 10–15 RFU—a significant amount of “noise” would be labeled. As a result, mixture interpretation of low-level minor components would become increasingly more difficult as the signal from the major contributor becomes more prominent.

Therefore, another approach commonly found within the analytical literature was assessed for forensic purposes. In this method, the *y*-intercept and the error derived from the relationship between mass and signal are employed (19). In the case of forensic DNA analysis, if the amplification efficiency of the target amplicon is 100% at the endpoint cycle number of 28 and it is assumed the portion of the product and the RFU is directly proportional to the concentration of amplified product at 28 cycles, the following linear relationship is obtained,

$$RFU = \phi C_{j,0} 2^{28} + A \quad (5)$$

where the *A* is the *y*-intercept (and may be expected to be 0 if the proportionality is unbiased) and the slope is 2^{28} multiplied by a proportionality constant (ϕ). Therefore, if optimal conditions are met, the samples contain accurate DNA concentrations, and the approximation that PCR efficiency does not change between concentrations is valid, a plot of $C_{j,0}$ versus RFU should result in a straight line. The resultant linear parameters may then be employed to determine the methods’ figures of merit. In this approach, the calibration plot demonstrates the relationship between the amount of target DNA and the expected signal where the *y*-intercept represents the “expected” signal when no DNA is present. Ideally, the \bar{Y}_{bl} of Methods 1, 3, and 4 would be indistinguishable from the *y*-intercept and recommendations to measure both the blank and its variance using both approaches and to stipulate consistency have been proposed (12). Additionally, the 1997 IUPAC Commission on Electroanalytical Chemistry described the use of calibration curves as an appropriate method for determining ATs (8). Figure 6 shows the average RFU values obtained for a set of 16 samples amplified using an input of 2–0.0625 ng for the green color channel. The error bars represent the two standard deviation spread from the mean signal for the 16 samples amplified at that target. Based on Fig. 6, it is observed that as DNA input increased, so did the variation, suggesting a weighted regression analysis would provide better estimates of the linear parameters. Additionally, the 2 and 4 ng (data not shown for 4 ng sample) points begin to curve down and do not maintain linearity, suggesting these two points are outside the linear range. Therefore, the application of the upper limit approach to the determination of ATs in forensic DNA analysis should take these two factors into consideration. As such, to obtain the upper limit of the *y*-intercept, a weighted linear regression between 0.0625 and 1 ng was performed with Igor Pro v. 6.1.2.1 using a 99% and 95% (one-sided) confidence and the results are summarized in Table 4. As Eq. (5) assumes the PCR efficiency remains *c.* 1 for low-level samples, an examination into the validity of that assumption is warranted. The error bars in Fig. 6 are two standard deviations from the mean of signal for all alleles at a given target within a color. As a result, the standard deviations are expected to increase with increasing signal as

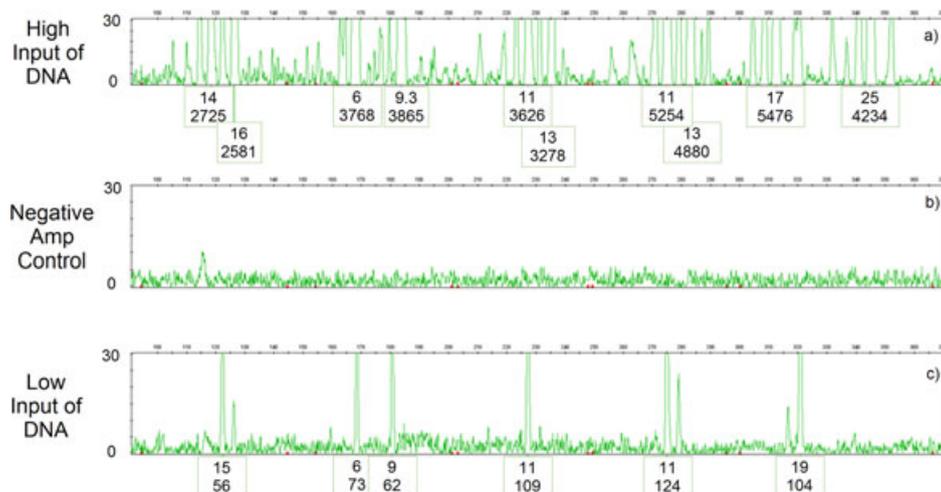


FIG. 4—Electropherograms (green channel) and baselines of DNA samples amplified at (a) 2 ng of DNA, (b) 0 ng of DNA (i.e., negative amplification control), and (c) 0.06 ng DNA. Each was amplified using the manufacturer’s recommended protocol and injected for 10 sec.

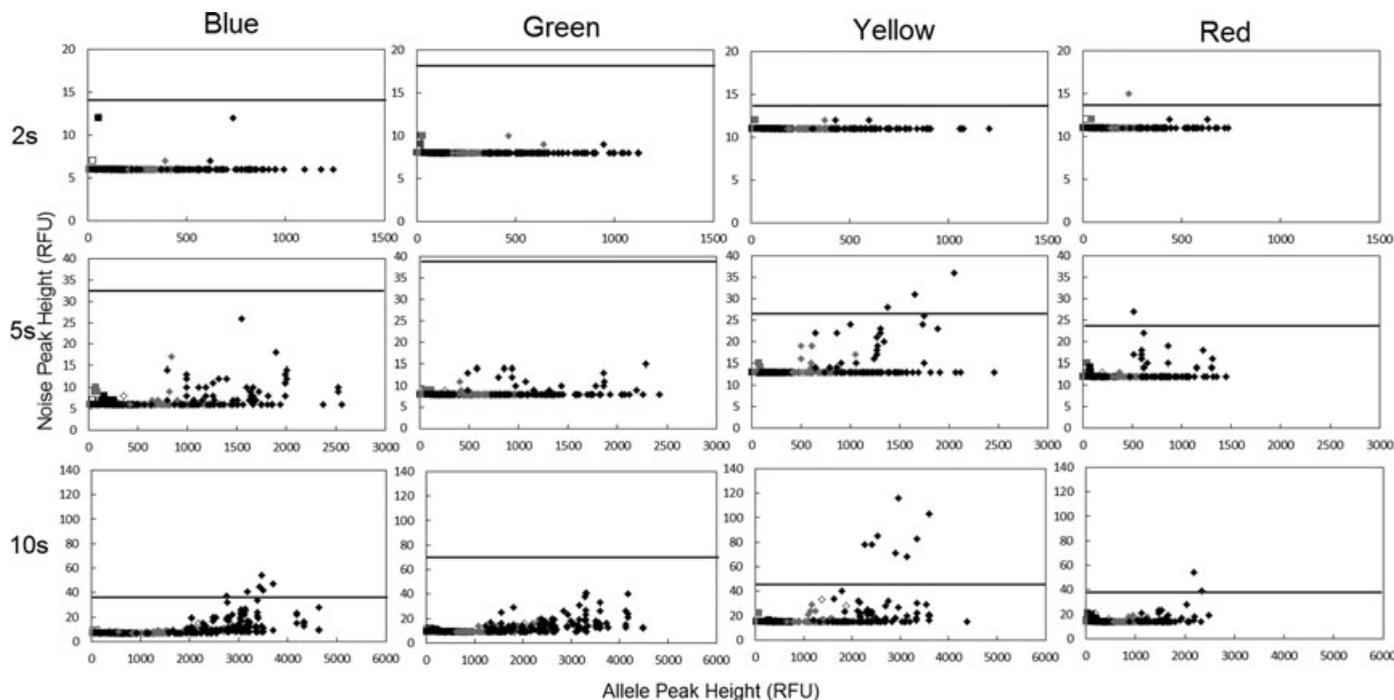


FIG. 5—Height of noise peaks greater than AT_{M1} plotted against the average peak height of the known alleles. A total of 96 single-source samples were amplified using 0.06 (\square), 0.13 (\blacksquare), 0.25 (\blacksquare), 0.5 (\diamond), 1 (\diamond), and 2 (\diamond) ng of DNA and run at 2-, 5-, and 10-sec injections. Loci with offscale data were not included. (—) indicates the AT_{M6} at 99% confidence (one-sided).

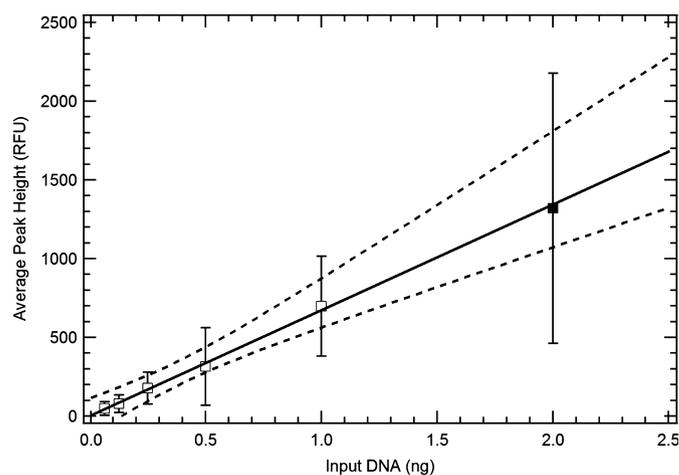


FIG. 6—Average peak height of known alleles (green channel) versus input DNA. Weighted linear regression was performed on points ranging from 0.06 to 1 ng (\square) of input. The 2-ng (\blacksquare) point was not included in the regression analysis to determine the analytical threshold. The 95% confidence bands are represented by the dotted lines. Error bars represent two standard deviations of the average signal within the color channel. Homozygote peaks were normalized for diploidy by dividing the RFU signal by two.

is depicted in Fig. 6. Although the standard deviation increases with signal, the relative standard deviation remains relatively constant between 0.0625 and 1 ng for all colors, at all injections, and range from *c.* 0.35 to 0.44—an indication that the error is growing with signal as opposed to some other phenomenon such as changes in PCR efficiency. Additionally, if the PCR efficiency was decreasing from 1 to a lower value, this would be indicated via the residuals. In this case, the residuals for 0.0625, 0.125, 0.25, 0.5, and 1 were 5.4, -18.5, 23.38, -37.5, and 24.87, respectively, showing the residuals do not correlate with target

($R^2 = 0.2$) again suggesting the assumption that efficiency at inputs <1 ng remains constant is applicable.

Recommendations for significance levels range from 0.05 to 0.001, where a one-sided confidence level of 99% ($\alpha = 0.01$) is recommended by Mocak et al., Glaser et al. and Clayton et al. (8,20,21). Confidence levels of 95% may also be reported. For laboratories that do not possess atypical data analysis tools, such as the one used in this work, a simplified approach to the one previously described may be implemented and calculated using readily available tools/freeware (15,22). This method can be used in cases when calibration models are represented by a line passing through \bar{Y}_{bl} (i.e., when net signal passes through the origin). These results are also summarized in Table 4. The ATs calculated from samples containing DNA are noticeably larger than those derived from negatives. For example, the ATs for the blue channel for a 5-sec injection using Method 1 and Method 5 are 6 and 60, respectively, and represent a 10-fold increase in ATs. Additionally, the ATs significantly differ between color, injection time, and desired confidence interval. The ATs derived from Methods 5 and 6 differ by *c.* 3-fold (at a given significance level), and therefore, a discussion of which Method should be used is necessary. Figure 5 shows the signal of noise peaks $>AT_{M1}$ for a given color and injection time. When ATs computed utilizing negatives are implemented, a significant number of noise peaks from 2 ng amplifications are labeled. The 1 ng samples also show an increase in the level and number of noise peaks labeled with injection time, where the number of noise peaks labeled for the 2-, 5-, and 10-sec injections was 4, 14, and 59, respectively. Additionally, a significant increase in the number of noise peaks above AT_{M1} starts when allele peak heights are >1000 RFU for the 2-sec injection. For the 5- and 10-sec injections, this increase is seen when allele peak heights are >1500 and >2000 RFU, respectively, indicating that amplification noise is observable above AT_{M1} after a certain input mass and injec-

TABLE 4—Analytical thresholds (ATs) calculated using Methods 5 and 6 and a calibration plot consisting of five points of 16 replicates. The confidence levels of 95% (i.e., $t_{n-1,\alpha} = 2.13$) and 99% (i.e., $t_{n-1,\alpha} = 3.75$) are shown. The y-intercepts and standard error of the weighted regressions are also included.

Color Channel	Injection Time (sec)	y-Intercept of Weighted Regression	S_y —Standard Error of Weighted Regression	AT _{M5} Using Igor Pro v. 6.1.2.1 (95% Confidence)	AT _{M5} Using Igor Pro v. 6.1.2.1 (99% Confidence)	AT _{M6} Using Excel Template (95% Confidence)	AT _{M6} Using Excel Template (99% Confidence)
Blue	2	0.94	3.38	30	56	9	14
	5	2.61	7.65	60	114	19	32
	10	5.49	8.43	101	190	32	38
Green	2	1.82	4.27	29	54	11	18
	5	2.55	9.66	57	107	24	39
	10	3.25	17.90	104	199	42	71
Yellow	2	1.03	3.04	26	49	8	13
	5	1.82	6.39	52	99	16	26
	10	1.81	11.43	89	170	27	45
Red	2	1.26	3.80	19	34	9	13
	5	-1.21	6.63	36	71	13	24
	10	-0.34	10.35	61	118	22	39

tion time. This suggests that ATs should be determined for each method to assess the point at which amplification noise is expected to be observed. This ensures the AT chosen will be one that maximizes the labeling of alleles, while minimizing the labeling of noise. At the AT_{M6}, 2 ng samples did not have a significant number of noise peaks labeled, suggesting that at these thresholds, amplification noise is more likely to be distinguished from true alleles. It should be noted that although AT_{M5} resulted in thresholds that would negate the labeling of all non-allelic peaks, the AT's calculated using Method 5 may be unnecessarily large for the purposes of detecting alleles.

Sensitivity

Another “figure of merit” associated with an analytical method is that of sensitivity. In general, the sensitivity is often erroneously used to describe the AT or detection limit. The AT is not synonymous with sensitivity because sensitivity is defined as the change in signal with analyte concentration, whereas the AT is the smallest signal that can be distinguished from baseline. Therefore, it may be the case that a method may be very sensitive (i.e., large change in signal with increasing concentration), but have a high AT (i.e., baseline is very noisy). The two most common definitions of sensitivity are calibration sensitivity (Eq. [6]) and analytical sensitivity (Eq. [7]).

$$m = \left(\frac{dS}{dc} \right)_c \quad (6)$$

$$\gamma = \frac{m}{SD_c} \quad (7)$$

where m is the slope of the tangent at concentration/input c of an analytical curve—and is constant if the curve is linear—and SD_c is the standard deviation of the signal at a given concentration/input. The first definition, calibration sensitivity, is the most common and is the one accepted by the IUPAC. It is the slope of the calibration curve and does not take into consideration the reproducibility (or lack thereof) of the signal at a given concentration. In contrast, Eq. (7) is termed the “analytical sensitivity” (13) and takes into account the change in signal with respect to concentration and the precision of the measurement at a given

concentration. The analytical sensitivity therefore reflects the ability to discern small differences in input amounts at a given concentration/mass. Table 5 summarizes both types of sensitivity for the positive samples and shows that although the calibration sensitivity increased with injection time, the analytical sensitivity did not. This is a direct result of the fact that although the RFU increased with injection time, so did its standard deviation. Therefore, the ability to discern between small changes in input DNA mass (within a color) is low and constant between injections. This is qualitatively seen in Fig. 6, where the error bars between the RFU values of 1 and 0.5 ng overlap. This is noteworthy as it implies that although it may be possible to detect small amounts of DNA, it may not be possible to readily distinguish between small changes in DNA input.

Discussion

Six methods for determining ATs were discussed and compared, where the first four methods were based on determining ATs through the analysis of blank signals after signal processing. Although it has been suggested that threshold determination considers noise prior to peak/software processing such that the noise measurements take into account the raw signal of the instrument (1,7), others have suggested that determination of the AT takes into account all laboratory procedures (10). Because interpretation of peaks is performed post-software analysis, this study endeavors to determine the AT for the entire laboratory process. As such, all measurements were exported from the peak detection software (GeneMapper ID v 3.1) subsequent to processing. To ensure the formulae are applicable to the blank signal data obtained, an analysis of the distribution was performed. A least squares fit was executed and showed that the data were better fit by a log-normal distribution than a Gaussian one. Comparisons between ATs from negatives and run blanks showed that ATs derived from negatives are slightly larger, and as a result, ATs derived from negatives are recommended as they are more representative of the method employed. To calculate ATs using negatives, the laboratory can use data already present in the laboratory, that is, analysis of 20–30 negatives (which have no indication of allelic “drop-in”) run with case-work samples may be utilized for this purpose. This eliminates the need for laboratories to design a new validation experiment and use additional reagents. Additionally, the methods proposed herein do not require purchase of costly and/or complicated statistical software.

TABLE 5—The calibration and analytical sensitivities calculated using Eqs (6) and (7) using a calibration plot consisting of five points of 16 replicates each and weighted least squares regression. Analytical sensitivities are calculated using the standard deviation of signal at 1 ng. The y-intercepts and standard error of the weighted regressions are depicted in Table 3.

Color Channel	Injection Time (s)	Slope of Weighted Regression—Calibration Sensitivity	Standard Deviation of RFU at 1 ng Target	Analytical Sensitivity Calculated Using Eq. (7) at 1 ng
Blue	2	338	76	4.4
	5	722	159	4.5
	10	1330	316	4.2
Green	2	318	67	4.7
	5	671	158	4.2
	10	1228	328	3.7
Yellow	2	309	71	4.3
	5	647	135	4.8
	10	1171	233	5.0
Red	2	231	71	4.0
	5	463	98	4.7
	10	793	176	4.5

Signal-to-noise comparisons showed that with amplification of DNA, comes a substantial rise in baseline noise. Although manufacturers have recommended amplification targets that typically range between 0.5 and 1.5 ng, and quantitation of evidentiary samples to prevent high levels of DNA input into the PCR is regularly practiced, the addition of a specified amount of DNA into the PCR reaction is directly affected by the ability of the laboratory to accurately quantitate the amount of DNA in an extract. This has proven difficult as qPCR precision has been shown to significantly vary between assays (23). This not only emphasizes the need for laboratories to validate their qPCR procedures to ensure stable quantitation over time, but it also emphasizes the need to establish ATs that are representative of samples and signals typically observed in casework. Therefore, the use of ATs derived from negatives is not appropriate for samples exhibiting significant signal because the baselines of negatives are not representative of the baselines of samples with significant levels of DNA. However, because the baselines of negatives are representative of baselines of low-level DNA amplifications, thresholds derived from negatives may be used for these types of samples. It is also noted that Methods 1–4 may be used to calculate ATs from positives as well, that is, the baseline signal not associated with alleles or artifacts can be used to calculate the ATs of positive samples. However, the intensity and number of noise peaks are dependent upon input mass. Therefore, calculating an AT using a number of positives at varying inputs needs to take into consideration the resulting distribution and may result in an AT, which is too large for low-input samples.

The question of what k value to use is also of concern when calculating ATs. If Method 1 is used to calculate the AT of blanks or low-level (i.e., “near blank”) DNA samples, a k value of 4 or 3 is warranted. To determine an appropriate value of k , a laboratory must consider the AT derived via Method 1 reflects the value of the “true” signal, which is significantly different from the negative signal value and is expressed in terms of population statistics (i.e., the population mean and the population standard deviation), that is, this method only considers a type I error (i.e., α -error or false positive) and does not consider that there is still a 50% probability the “true” signal might not be distinguishable from baseline (i.e., false negative). This error is known as type II or β error (9). Other k values have also been suggested and range from 6 to 10, and the various threshold values have been referred to as limit of purity, limit of quantita-

tion, and identification limit (6–8,24). Generally, signals that are greater than $\bar{Y}_{bl} + 9s_{Y,bl}$ are described as being in the quantitation range and samples with signal between $\bar{Y}_{bl} + 3s_{Y,bl}$ and $\bar{Y}_{bl} + 9s_{Y,bl}$ are in the limit of detection range (9). Arbitrarily choosing a large k as high as 10 is not recommended particularly if the laboratory is interested in determining whether the peak can be distinguished from baseline and is therefore determining “detection.” Although STR peak height/area information can be used to calculate ratios (i.e., peak height ratios, mixture ratios, etc.), it is not recommended that the methods described herein be used to establish a mixture interpretation or stochastic threshold. Mixture interpretation needs to encompass analyses of the probability of allelic dropout and peak height ratio discrepancies at small input levels (25–27). Therefore, for purposes of detection, a lower value of k (i.e., 3 or 4) is recommended, while a percentile rank of at least 99% is recommended for Method 2.

Because the baselines of DNA samples are not necessarily similar to the baselines of negatives, consideration regarding the use of negatives as the basis for AT determination was tested. Thresholds that stem from analysis of positive samples take into account the uncertainty in the signal to input relationship, and therefore, the y -intercept and its upper boundary determine the signal that is expected at 0 ng. For all injection times and colors, the \bar{Y}_{bl} calculated from the analysis of 31 negatives is contained within the y -intercept $\pm 2S_y$, and is a reasonable indication that Method 6 can be used in place of Method 5. However, it should be noted that the ATs derived in this manner need to be calculated for each color and for all preparations (i.e., different injections, sample preparation volumes, post-PCR cleanup, etc.).

The ATs derived from negatives are an alternative for samples with low input DNA. An example of such a methodology may be by analysis of the largest peaks. For example, if the largest peaks of the profile are <1000 RFU (2-sec injection), ATs derived from negatives may be implemented. However, it should be stated that for samples with low input DNA from a minor contributor, where the minor contributor exhibits peak heights <1000 RFU, but the major contributor’s peak heights are in excess of 1000 RFU, the ATs derived from negatives may not be representative of these baselines, and the ATs derived from samples containing DNA is suggested. Regardless of the method, it should be noted the ATs would need to be re-evaluated when there is a quality control failure and/or significant instrument modification (i.e., lamp, detector replacement). Otherwise, the uncertainty of the ATs can be accounted for during

validation where the determination of ATs can include the analysis of multiple samples, analyzed several times, using different kit lots and pipettes.

Ideally, the development of ATs should be fairly simple and easily applied in casework. In typical forensic analysis software, the RFU thresholds are input for each color channel. Therefore, using different ATs for each color is easily employed and is not expected to significantly impact DNA analysis and the time required to do such analysis. One AT for all colors is not necessarily easier to apply, and because the ATs are significantly different between color channels, it is recommended that color specific thresholds be utilized.

Additionally, clarification of the proper terminology is needed. A method with a low AT is not synonymous with a method that is sensitive. Analytical sensitivity, which is the calibration sensitivity divided by the standard deviation at a given concentration, ranged from 3.7 to 5.0. Because analytical sensitivity is a metric which describes the ability to discern small differences in the analytical concentration, large standard deviations and a small slope suggest that small differences in DNA input cannot be easily distinguished using peak heights.

Conclusion

Determination of ATs needs to be fairly simple if it is to be widely implemented. Careful consideration with regard to the method by which laboratories choose to calculate the ATs needs to be representative of the baselines typically observed. The utilization of Method 1 or 2 or 4 (with a k of 3 or 4 or rank of at least 99%) is recommended if ATs for “near negative” samples are desired. ATs based on these calculations are not recommended for medium to high signal samples as they do not take into consideration amplification noise that might occur with significant amounts of input DNA. Medium to high target forensic DNA samples require an AT based on samples containing DNA. Further studies that focus on the application of ATs to mixture sets would be required to assess the implications of using such methods and are currently under way.

Acknowledgments

The authors thank John Butler, Charlotte Word, and Mike Coble for helpful discussions throughout this work. We also thank the anonymous reviewers for their comments and critique.

References

1. <http://www.fbi.gov/about-us/lab/codis/swgdam.pdf> (accessed November 22, 2010).
2. Brown RJ. The use and abuse of limits of detection in environmental analytical chemistry. *Sci World J* 2008;8:796–801.
3. Corley J. Best practices in establishing detection and quantification limits for pesticide residues in foods. In: Lee PW, Aizawa H, Barefoot AC, Murphy JJ, editors. *Handbook of residue analytical methods for agrochemicals*. Hoboken, NJ: John Wiley & Sons Ltd., 2002; 1–18.
4. Currie LA. Limits for qualitative detection and quantitative determination: application to radiochemistry. *Anal Chem* 1968;40:586–93.
5. Currie LA. The measurement of environmental levels of rare gas nuclides and the treatment of very low-level counting data. *IEEE Trans Nucl Sci* 1972;19:119–26.
6. Currie LA. Detection and quantification limits: origins and historical overview. *Anal Chim Acta* 1999;391:127–34.
7. Gilder JR, Doom TE, Inman K, Krane DE. Run-specific limits of detection and quantitation for STR-based DNA testing. *J Forensic Sci* 2007;52:97–101.
8. Mocak J, Bond AM, Mitchell S, Scollary G. A statistical overview of standard (IUPAC and ACS) and new procedures for determining the limits of detection and quantification: application to voltammetric and stripping techniques. *Pure Appl Chem* 1997;69:297–328.
9. Winefordner JD, Long GL. Limits of detection: a closer look at the IUPAC definition. *Anal Chem* 1983;55:712A–24A.
10. Kaiser H. Part II: quantitation in elemental analysis. *Anal Chem* 1970;42:26A–59A.
11. Skoog DA, Holler FJ, Nieman TA. *Principles of instrumental analysis*. South Melbourne, Victoria: Brooks/Cole, 1998.
12. Currie LA. Detection: international update, and some emerging dilemmas involving calibration, the blank and multiple detection decisions. *Chemometrics Intell Lab Systems* 1997;37:151–81.
13. Ingle JD. Sensitivity and limit of detection in quantitative spectrometric methods. *J Chem Educ* 1974;51:100–5.
14. Applied Biosystems. AmpFISTR(R) Identifier(R) PCR Amplification Kit user’s manual. Foster City, CA: Applied Biosystems, 2006.
15. <http://www.bumc.bu.edu/biomedforensic/faculty-and-staff/faculty/catherine-grgicak/tools/> (accessed November 22, 2010).
16. Sturges H. The choice of a class-interval. *J Am Stat Assoc* 1926;21: 65–6.
17. Wavemetrics I. IGOR Pro v 6.2 user’s manual. Lake Oswego, OR: WaveMetrics Inc, 2010.
18. Inczedy J, Lengyel T, Ure AM. *Compendium of analytical nomenclature*. Oxford, UK: Blackwell Science, 1997.
19. Hubaux A, Vos G. Decision and detection limits for linear calibration curves. *Anal Chem* 1970;42:849–55.
20. Glaser JA, Foerst DL, McKee GD, Quave SA, Budde WL. Trace analyses for wastewater. *Environ Sci Technol* 1981;15:1426–35.
21. Clayton A, Hines JW, Elkins PD. Detection limits with specified assurance probabilities. *Anal Chem* 1987;59:2506–14.
22. <http://www.bowdoin.edu/~rdelevie/excellaneous/> (accessed November 22, 2010).
23. Grgicak CM, Urban ZM, Cotton RW. Investigation of reproducibility and error associated with qPCR methods using Quantifiler duo DNA quantification kit. *J Forensic Sci* 2010;55:1331–9.
24. Thomsen V, Schatzlein D, Mercurio D. Limits of detection in spectroscopy. *Spectroscopy* 2003;18:112–4.
25. Gill P, Brenner C, Buckleton J, Carracedo A, Krawczak M, Mayr W, et al. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. *Forensic Sci Int* 2006;160:90–101.
26. Gill P, Puch-Solis R, Curran J. The low-template DNA (Stochastic) threshold—its determination relative to risk analysis for national DNA databases. *Forensic Sci Int Genet* 2009;3:104–11.
27. Lucy D, Curran J, Pirie A, Gill P. The probability of achieving full allelic representation of LCN-STR profiling of haploid cells. *Sci Justice* 2007;47:168–71.

Additional information and reprint requests:

Catherine Grgicak, Ph.D.
Biomedical Forensic Sciences
Boston University School of Medicine
72 E. Concord Street, Room R806B
Boston, MA 02118
E-mail: cgrgicak@bu.edu