JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2012
doi: 10.1111/1556-4029.12017
Available online at: onlinelibrary.wiley.com

CASE REPORT

CRIMINALITICS

Kirk E. Lohmueller,[1] *Ph.D. and Norah Rudin,*[2] *Ph.D.*

Calculating the Weight of Evidence in Low-Template Forensic DNA Casework*

ABSTRACT: Interpreting and assessing the weight of low-template DNA evidence presents a formidable challenge in forensic casework. This report describes a case in which a similar mixed DNA profile was obtained from four different bloodstains. The defense proposed that the low-level minor profile came from an alternate suspect, the defendant's mistress. The strength of the evidence was assessed using a probabilistic approach that employed likelihood ratios incorporating the probability of allelic drop-out. Logistic regression was used to model the probability of drop-out using empirical validation data from the government laboratory. The DNA profile obtained from the bloodstain described in this report is at least 47 billion times more likely if, in addition to the victim, the alternate suspect was the minor contributor, than if another unrelated individual was the minor contributor. This case illustrates the utility of the probabilistic approach for interpreting complex low-template DNA profiles.

KEYWORDS: forensic science, DNA typing, likelihood ratio, combined probability of inclusion, drop-out, low-template, stochastic threshold, statistics

This was a very complex case that involved physical evidence from several forensic disciplines, as well as a large amount of non-physical evidence. We do not intend to provide a comprehensive description of the evidence or the case, nor will we even address the complicated legal issues that arose. Our intention in this report is simply to outline a novel statistical approach that was used to clarify the weight of several ambiguous DNA profiles obtained from biological evidence items relevant to the case.

The victim in this case was found deceased in her bedroom. She had sustained multiple stab wounds to her body that were determined by the coroner to be the cause of death. Her husband, with whom she shared the residence, was suspected of killing her. In part because of the DNA results, the long-term mistress of the husband became an alternate suspect for perpetrating the murder of the victim.

A number of evidence items collected from the crime scene were analyzed and interpreted by the crime laboratory. These included three bloodstains around the kitchen sink that appeared as diluted vertical drips, and a swabbing of bloodstaining on the handle of a kitchen knife thought to be the murder weapon. No ambiguity existed as to the DNA profile of the major contributor to these bloodstains; this profile was concordant with that of the victim. Nor did any dispute exist as to the random match probability (RMP) reported for this profile. The difficulty arose in interpreting and assigning a statistical weight to the minor profiles detected in these samples. The government laboratory issued several reports over time in which they offered evolving interpretations of the evidence, assigning differing statistical weights to the minor profiles each time. Their calculations were based on variants of either a combined probability of inclusion (CPI) or a modified RMP adjusted for loci at which only a single peak was detected above a particular relative fluorescence units (RFU) threshold. The reason for the difficulty in interpreting the minor profiles was that they exhibited low peak heights, indicating the possibility of allelic drop-out. Allelic drop-out is the situation in which alleles from true contributors are missing from the evidence profile. At some stage of the typing process, the DNA present in the original sample may be nonrepresentatively sampled (stochastic effects), resulting in the failure to detect some alleles in the final profile. This in turn may lead to ambiguity in interpreting the DNA typing results. The likelihood of drop-out is higher when analyzing low-template profiles because the stochastic effects related to sampling DNA molecules during the typing process are greater when analyzing smaller quantities of template (1). This leads to the concept of a stochastic threshold, typically understood as the RFU value below which one of a pair of heterozygous alleles from a true contributor might fall below the detection threshold set by a laboratory's interpretation guidelines (2).

One approach to interpreting forensic DNA evidence appropriate for profiles exhibiting possible allelic drop-out emerges from a logical framework based on probabilistic reasoning. The likelihood ratio (LR) is the vehicle by which this approach can be used to provide evidential weight to many different types of complex DNA profiles frequently encountered in forensic cases, such as mixed samples or incomplete DNA profiles (3,4). Also, interpretational approaches based on probabilistic reasoning require the alternate hypotheses to be explicitly defined and stated. This is a strength of the approach

[1]Department of Integrative Biology, University of California Berkeley, Berkeley, CA.
[2]Forensic DNA Consultant, Mountain View, CA.
*Supported in part by a Ruth Kirschstein National Research Service Award from the National Human Genome Research Institute (F32HG005308) and by the Miller Research Institute at UC Berkeley.

Received 8 July 2011; and in revised form 6 Jan. 2012; accepted 8 Jan. 2012.

because it forces one to overtly consider various relevant explanations that could have generated the evidence. The probabilistic approach also enjoys substantial support in the scientific literature, whereas other methods (e.g., CPI, modified RMP, 2p) receive, at most, a tepid endorsement (e.g., 3–10). Therefore, using the case data from the laboratory, we calculated LRs that incorporated an empirically derived probability of allelic drop-out for the minor profiles.

Although the government laboratory provided for the use of LRs incorporating a probability of drop-out in their DNA procedures manual, they did not have a convenient tool at their disposal to perform the calculations. We took advantage of computer code written by Professor David Balding, based on equations published by Balding and Buckleton (5), to calculate LRs incorporating a probability of allelic drop-out. We followed Tvedebrink et al. (11) to empirically estimate the probability of allelic drop-out.

We applied this approach to four different low-template samples in this case. For simplicity, we present here the details for only one sample. Item 24 was a bloodstain near the knife block next to the kitchen sink. The stain appeared to have resulted from a vertical drip, and also appeared substantially diluted, as if it had been mixed with water. We chose this sample because it illustrates the most extreme dichotomy between the laboratory's various calculations and our LR calculations. It is the details of these calculations that we present here.

Materials and Methods

Calculation of LRs for Mixed Samples When Accounting for Allelic Drop-out

Several workers have offered statistical approaches to calculate LRs for mixed DNA samples that also take into account the possibility of allelic drop-out (3,12–14). Professor Balding has recently written scripts using the R statistical package (15) that perform the Balding and Buckleton version of these calculations (5). The program is available from the corresponding author.

The overall approach of Balding and Buckleton (5) follows the same general logic as the standard methods to calculate LRs. Typically:

$$\mathrm{LR} = \frac{\Pr(\text{DNA profile}|H_1)}{\Pr(\text{DNA profile}|H_2)}$$

where H_1 and H_2 are two different competing explanations for the DNA profile. For example, H_1 may be that the person in question contributed the evidence, while H_2 may be that some other person left the evidence. The novelty of the Balding and Buckleton approach (5) is that it includes terms that allow for the incorporation of a probability of allelic drop-out (denoted $P(D)$). Specifically, under H_2, the program sums over all possible genotypes at each locus for the unknown contributor, weighted by the probability of sampling that genotype from the population (i.e., the genotype frequency) and, conditional on that genotype, the probability of finding the detected alleles in the evidence sample. We will discuss approaches to estimate $P(D)$ later in this paper.

For the calculations in this case, we started with the R code provided on Professor Balding's website and made some minor adjustments. To account for possible shared ancestry between the victim and the unknown donor(s) of the evidence, Professor Balding's method uses the allele frequency correction that is analogous to NRC equation 4.10d (9,16,17). We set θ equal to 0.01, which is the customary value used in the United States (9). We did not use the database sample size correction available in the R code because all of the alleles observed in the alternate suspect's profile were present in each population database 9 or more times (18), indicating that an adjustment for the presence of rare alleles was not critical for this case. Professor Balding also performed calculations and provided an initial report in this case. While we ultimately chose slightly different parameters, which were reflected in the R code that we used, our results were concordant with those of Professor Balding.

Interpretation of the Electropherograms

We assigned each peak in the electropherogram for stain 24 to one of three categories: major peaks, minor peaks, and minor peaks at stutter positions. Table 1 illustrates the categorization of alleles for this sample, and also lists the reference profiles for the victim, suspect, and alternate suspect. The first category contains the major peaks that can be accounted for by known individuals who are not disputed to be contributors. In this case, the alleles from the victim fall into this category. For example, at D2S1338, the 17 and the 21 alleles fall into this category (Fig. 1, Table 1). The second category consists of the minor peaks that are of lower intensity and cannot be explained by the victim's contribution. These alleles must be from a minor contributor. The 19 allele at D2S1338 falls into this category (Fig. 1, Table 1). The third category contains observable peaks in stutter positions that could either represent only stutter or could also be masking an allele from the minor contributor. The 16 and the 20 peaks at D2S1338 fall into this category (Fig. 1, Table 1). When some of the minor peaks in the electropherogram are of the same intensity as the stutter peaks, it can be difficult to distinguish between these possibilities. The R code implementing the method of Balding and Buckleton (5) allows for both possibilities. We also considered several drop-out and drop-in (the sporadic appearance of an allele) probabilities and performed each calculation using allele frequencies from three different major population groups (18).

Empirical Estimation of Allelic Drop-out Probabilities

The approach of Balding and Buckleton (5) to calculate LRs incorporating a probability of drop-out requires the user to input some value for the drop-out probability ($P(D)$). While the forensic DNA community has cited a lack of empirical data as one reason for resisting the adoption of an LR approach, we argue that a vast repository of data exists in the years, if not decades, worth of validation studies; the data just need to be parsed. Further, while previous papers have either estimated drop-out or used a range of drop-out probabilities for demonstrating the value of the approach, empirical estimates are preferred for any specific case. Thus, we used the data provided by the laboratory to derive estimates of drop-out probabilities for application to this case.

One complication to estimating drop-out probabilities is that drop-out varies under different conditions encompassing both the quantity and quality of DNA (3,4,12,19–22), and hence no one drop-out probability applies to all situations. The inverse relationship of drop-out probability to the amount of DNA amplified forms the basis for the approach used here to estimate drop-out probabilities. One way to estimate the amount of DNA that was amplified is by inference from the peak heights in the electropherogram. Although other factors such as degradation or inhibition can also moderate peak height, as a simple guideline, higher peak

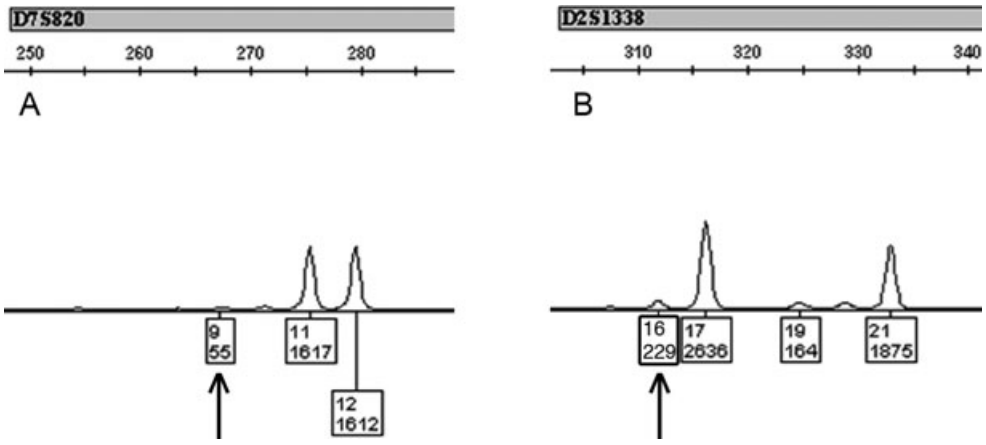

FIG. 1—*The two loci from the electropherogram of Item 24 upon which the laboratory based their initial exclusion of the alternate suspect from the minor profile. The arrows indicate the problematic alleles. (A) At D7S820, the 9 allele was detected at 55 RFU in the 5-sec injection, shown here. (B) At D2S1338, a peak in the 16 position, the stutter position of the 17 allele, did not rise above the stutter threshold in the 5-sec injection because it was only 8.7% the height of the parent peak. The stutter filter in the software treats peaks in stutter positions at this locus as true stutter if the percentage of the parent peak is 11.1% or less. The allele call and peak height at the 16 position were manually added to this figure. Note that the peak in the D2S1338 20 position, although not labeled here, was also included in our likelihood ratio calculations (Table 1).*

TABLE 1—*Designation of peaks for item number 24.*

Locus	Major Peaks	Minor Peaks	Minor Peaks at Stutter Positions	Victim	Suspect	Alternate Suspect
D8S1179	13, 14	15	12	13, 14	13, 15	13, 15
D21S11	30, 31.2	32.2	29, 30.2	30, 31.2	30, 31.2	30, 32.2
D7S820	11, 12	9	10	11, 12	8, 11	9, 11
CSF1PO	10, 13	11, 12*	9	10, 13	12, 13	11, 12
D3S1358	15, 17	16†	14	15, 17	15	16
THO1	7, 8	9	6	7, 8	8, 9.3	8, 9
D13S317	11	8, 12	10	11	10, 11	8, 12
D16S539	11, 13	9	10, 12	11, 13	11, 14	9, 11
D2S1338	17, 21	19	16, 20	17, 21	17, 26	16, 19
D19S433	13	14	12	13	13	13, 14
vWA	17	15, 18	16‡	17	17	15, 18
TPOX	8, 9	10		8, 9	9, 11	9, 10
D18S51	13, 17		12, 16	13, 17	13, 17	17
D5S818	11, 12		10	11, 12	11	11, 12
FGA	20, 21	24	19	20, 21	21, 24	19, 24

*Note, although allele 12 falls in a stutter position of a major allele, it is too high (157 RFU) to be solely attributable to stutter. The stutter percentage (the RFU of the 12 allele divided by the RFU of the 13 allele multiplied by 100) is 10.9%, which exceeds the AB stutter value of 9.2%.

†Note, although allele 16 falls in a stutter position of a major allele, it is too high (1321 RFU) to be solely attributable to stutter. The stutter percentage (the RFU of the 16 allele divided by the RFU of the 17 allele multiplied by 100) is 22.1%, which exceeds the AB stutter value of 10.7%.

‡Note, although allele 16 falls in a stutter position of a major allele, it is too high (663 RFU) to be solely attributable to stutter. The stutter percentage (the RFU of the 16 allele divided by the RFU of the 17 allele multiplied by 100) is 14.6%, which exceeds the AB stutter value of 12.6%. However, this peak is not counted as minor allele because it has an unusual shape and its signal appears to be increased by pull-up from the 7 allele at THO1.

heights indicate that more DNA has been amplified. Thus, the higher the peak heights, the lower the probability of drop-out.

The method we used to estimate drop-out probabilities is similar to that described in Tvedebrink et al. (11). Importantly, these researchers allow the drop-out probability to differ from locus to locus, while in this analysis we assume the same drop-out probability for all loci. While their approach is likely to be more biologically realistic, with the limited validation data that were available to us for this case (four or five profiles for each set of experimental parameters), we did not think it wise to use more complicated models.

We received validation studies from the government laboratory through discovery in this case. These data included electropherograms from the laboratory's sensitivity study. The laboratory typed 4–5 different dilutions for each set of experimental conditions. The various parameters included injection time (5, 10 sec) and injection volume (1.5, 3 μL). They performed these experiments for both single-source samples and mixtures of two individuals. We used these data to find the relationship between the average peak heights over all peaks within a profile and the fraction of alleles that dropped out of the entire multilocus profile under different experimental conditions. Then, applying this relationship to the average RFU values of the evidentiary profiles in this case, we estimated the drop-out probabilities for those samples.

The following describes how the statistics were calculated from the single-source validation profiles. First, define d_{ij} to be the number of alleles that dropped out from the jth locus in the ith replicate profile for a given set of experimental conditions. Then, d_i or the total number of alleles that dropped out of profile i is simply

$$d_i = \sum_{j=1}^{H} d_{ij}$$

where H is the total number of loci where the individual typed was heterozygous. Then,

$$P_i = \frac{d_i}{2H}$$

where P_i is the proportion of alleles that dropped out of the ith validation profile. Only peaks at loci where the typed individual was heterozygous were included in this analysis. Second, we tabulated the average heights (in RFU) of the peaks present in the electropherogram. Let X_i be the average height of all detected peaks at loci where the typed individual was heterozygous in the ith profile. We chose to use average peak heights across all peaks in the profile, rather than calculate individual statistics for each locus because less sampling variation exists in estimating an average compared with a single data point.

The laboratory also typed profiles from mixtures of two individuals. The same statistics (X_i and P_i) were tabulated from these profiles. However, only alleles from the minor profile were considered. Further, only those loci where the minor contributor was heterozygous were included. Specifically, here d_{ij} was defined to be the number of alleles that dropped out from the minor profile of the ith replicate at locus j for a given set of experimental conditions. P_i was then defined as

$$P_i = \frac{d_i}{\sum_{j=1}^{H} A_j}$$

where A_j was the total number of alleles at locus j from the minor contributor that do not overlap with alleles from the major contributor. A_j was zero if the two alleles from the minor contributor overlapped with two alleles from the major contributor. A_j was one if one of the two alleles from the minor contributor overlapped with an allele from the major contributor. A_j was two if neither of the alleles from the minor contributor overlapped with alleles from the major contributor. Here, H refers to the total number of loci where the minor contributor was heterozygous. Similarly, for the mixtures, X_i only included those peaks at loci where the minor contributor was heterozygous and that did not overlap with an allele from the major contributor.

Finally, we determined the relationship between the fraction of alleles that had dropped out (the drop-out probability) and the average RFU value of each single-source or minor profile from the validation data. To do this, we used logistic regression (as suggested by Tvedebrink et al. [11]) using the *glm* function implemented in the R statistical package (15). The logistic model relates the proportion of alleles that dropped out from a particular profile as a function of the average height of the peaks in the profile using the following equation,

$$\ln\left[\frac{P_i}{1-P_i}\right] = \alpha + \beta X_i$$

In this model, α and β are the intercept and slope, respectively, of the logistic function. The logistic model was fit separately to the various profiles analyzed under different experimental conditions. As an example, Fig. 2 shows the relationship between the proportion of alleles that dropped out and the average RFU value derived from the minor component of a mixture typed using validation samples with 5-sec injections of 3 μL. Each point on the graph represents a particular 15 locus Identifiler® (Life Technologies, Foster City, CA) profile. As expected, as the average RFU value decreases, the drop-out probabilities increase. The curved line shown on the plot is the best-fit curve from the logistic regression model. The model appears to fit the data quite well as the points fall on, or fairly close to, the line.

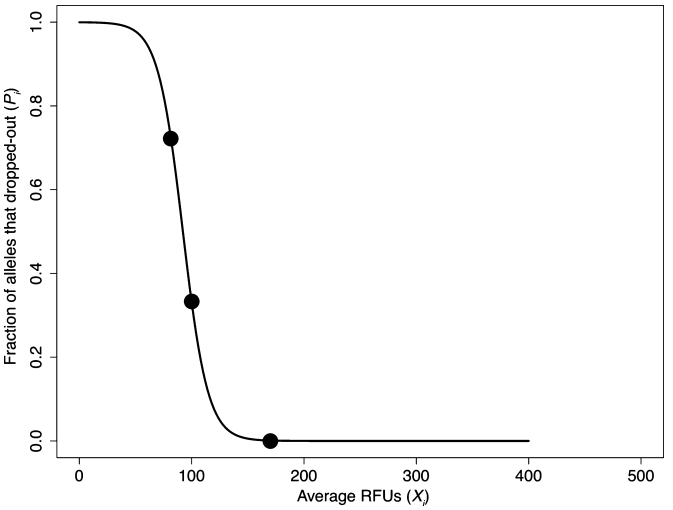

FIG. 2—*Relationship between the drop-out probability (vertical axis) and the average RFU in the laboratory validation data (horizontal axis). Each dot represents a particular 15-locus Identifiler® profile. Note, the fourth profile (at roughly 1100 RFU) from this set was omitted for clarity. The black line is the best-fit line from the logistic regression analysis. The equation from this line is used to estimate drop-out probabilities from the evidence.*

From the logistic regression analysis, we obtained an equation for the best-fit curve.

We then used this equation to estimate the drop-out probability for each evidentiary profile in the case. Specifically, for item 24, the average RFU value for the minor profile was calculated from the "minor peaks" column in Table 1. Entering this variable into the logistic regression equation fit to the validation data allowed us to determine a drop-out probability specific to the evidentiary profile.

Results

We performed LR calculations for four samples in this case, all typed using the Identifiler® kit. In all samples, the major profile is concordant with that of the victim; minor profiles that are similar, but not identical, to each other are observed in all four samples. Because of the low level of the minor contributor peaks (generally <300 RFU at most loci) in the samples, we must consider the possibility that the minor profiles are incomplete, and may not represent the totality of the minor component(s). Because of this, assigning a statistical strength to the minor profiles proved a challenging exercise. Depending on various analysis parameters, assumptions, and choice of statistical tool, the apparent weight of the evidence varied radically, in some cases by many orders of magnitude. In an extreme demonstration, these variables even determined whether a particular reference sample could be included or should be excluded.

Laboratory Interpretation of the Mixture

A mixed profile was obtained from item 24 in which it was possible to differentiate a major profile and a minor profile. At 13 of the 15 loci, alleles present in the alternate suspect reference profile were either observed in the minor profile or could have been masked by alleles from the major profile. Nevertheless, the laboratory, in their initial report, excluded the alternate suspect, as well as the suspect, as a possible contributor of the minor profile to this sample (Table 2). According to testimony,

the exclusion was based on two problematic loci. First, although a peak is clearly present at the D7S820 9 position in the 3-sec injection, it does not rise above the laboratory's 50 RFU analytical threshold and thus was not labeled by the GeneMapper® (Life Technologies) software (data not shown); however, the D7S820 9 allele is clearly labeled in the 5-sec injection (Fig. 1). The laboratory chose to report only the data from the 3-sec injection because some of the larger loci were slightly overloaded in the 5-sec injection; they apparently ignored all data from the 5-sec injection. Second, at D2S1338, the peak in the stutter position for the 17 allele did not rise above the manufacturer's suggested stutter threshold (Fig. 1). Thus, the laboratory designated the peak as originating *only* from stutter and did not consider the possibility that a 16 allele from the minor contributor could also have contributed to the peak in the stutter position.

The laboratory subsequently amended their conclusion to include the alternate suspect. To assess the weight of this evidence, the laboratory calculated a CPI, using only those three loci for which all detected peaks at the locus rose above their 200 RFU stochastic threshold. All other loci were disregarded in assessing the weight of this evidence. Using only these three loci, the laboratory offered a CPI of 16% (Table 2). In a later report, the laboratory chose to use a different strategy to assess the weight of this item of evidence. Rather than employ their 200 RFU stochastic threshold or use a CPI, they calculated a modified RMP assuming a single minor contributor to the mixture. For loci in which a single minor peak was detected above 100 RFUs, the laboratory multiplied the frequency of this allele by 2 (commonly known as the 2p calculation). Loci where no minor alleles were detected above 100 RFU were not included in the calculation; for loci where both minor alleles rose above 100 RFU, they used a standard RMP calculation. This statistical approach resulted in an estimate of one in every 10,000 individuals as possible contributors to the minor profile of this sample. This calculation also excluded a peak that was, in the interim, determined to be an artifact (Table 2).

Probabilistic Interpretation of the Mixture

We interpreted the electropherogram from the 5-sec injection. The nominal peak designations were all the same as those that were labeled in the 3-sec injection. Additionally, re-analysis by an independent laboratory using a different typing system (not reported here) further confirmed the D7S820 9 allele. We assigned each peak in the electropherogram to one of the three categories described above. Table 1 lists all of the peaks in the electropherogram that fall into each category, using a 50 RFU detection threshold. We determined an average RFU value of 262 RFU over all the minor peaks in the sample. Using the logistic regression equation derived from the validation data, we estimated that the drop-out probability was <10% (Table 3) across all of the experimental conditions considered. To allow for imprecision in estimating the drop-out probability, we considered a range of values from 0.2 to 10%.

To calculate the LR, we considered the two propositions:

H_1: The mixture contains DNA from the victim and the alternate suspect, and

H_2: The mixture contains DNA from the victim and one unknown individual.

The LR is calculated as:

$$\mathrm{LR} = \frac{\Pr(\text{DNA profile}|H_1)}{\Pr(\text{DNA profile}|H_2)} = \text{at least 47 billion}$$

This calculation results from using a drop-out probability of 10%, which was chosen to deliberately underestimate the final strength of the LR (Table 4). The DNA profile detected in the sink stain is substantially more likely if the alternate suspect, rather than an unrelated individual, is a minor contributor, even when using what is likely to be an over-estimate of the allelic drop-out probability (Table 4). This figure contrasts with the three different calculations provided by the laboratory for this sample, variously, exclusion of the alternate suspect, a CPI of 16%, and a modified RMP of 1/10,000 (Table 2).

Discussion

We initially became interested in item 24, as well as one other sample, because we were concerned that the laboratory had falsely excluded the alternate suspect. Over time, additional samples became of interest as the laboratory amended their conclusions and statistical estimates several times in an attempt to provide a "conservative" interpretation of evidence. This buzzword is usually invoked as a positive attribute and is meant to avoid overweighting the evidence against a suspect. Unfortunately, that is only true when it is the inclusion of the suspect in an evidence profile relevant to the case that is at issue. Interestingly, this attempt to be overly "conservative," usually understood as aggressively avoiding a false inclusion, in this case actually resulted in a false exclusion. This did no favor for the suspect in this case.

This case highlights two major issues in the interpretation of low-template DNA profiles: (i) The current requirement to *a*

TABLE 2—*Summary of statistical conclusions with regard to the alternate suspect.*

Report Number	24 BS Near Sink	Statistical Approach	Conditions
Laboratory report 1	Excluded		
Laboratory report 2	Excluded		
Laboratory report 3	16%	CPI	200 RFU stochastic threshold, 50 RFU analytical threshold
Laboratory report 4	1/10,000	RMP assuming a single minor contributor	2p for single peaks above 100 RFU, 50 RFU analytical threshold, peaks between 50 and 100 RFU omitted; dropped D19 9.2 "allele" that was determined to be an artifactual peak
KL/NR report	47 billion times more likely if AS + V than UK + V	LR incorporating drop-out	50 RFU analytical threshold; DO probability of 10%

BS, bloodstains; CPI, combined probability of inclusion; LR, likelihood ratio; RFU, relative fluorescence units; DO, drop-out; RMP, random match probability; AS, alternate suspect; V, victim; UK, unknown individual.

TABLE 3—*Summary of the estimates of the drop-out probability for item 24 made using the laboratory's validation data.*

Type	Injection Volume (μL)	Injection Time (Sec)	α	β	Estimated $P(D)$*
Single source	1.5	5	4.1	−0.040	1.80×10^{-3}
		10	7.7	−0.045	1.70×10^{-2}
	3	5	4.3	−0.041	1.50×10^{-3}
		10	6.3	−0.038	2.27×10^{-2}
Mixture	1.5	5	8.6	−0.091	2.63×10^{-7}
		10	7.2	−0.060	1.90×10^{-4}
	3	5	8.3	−0.090	2.20×10^{-7}
		10	10.5	−0.083	1.26×10^{-5}

*This is the estimated drop-out probability for the minor peaks in item 24, using an average peak height of 262 RFU.

TABLE 4—*Likelihood ratios (LRs) for item 24 using different drop-out probabilities and allele frequencies from different population groups.*

Drop-out Probability	LR African American	LR Hispanic	LR European
0.2%	6.9×10^{11}	1.0×10^{11}	2.2×10^{11}
1%	6.3×10^{11}	9.6×10^{10}	2.1×10^{11}
5%	4.2×10^{11}	7.0×10^{10}	1.5×10^{11}
10%	2.6×10^{11}	4.7×10^{10}	1.0×10^{11}

priori include or exclude a possible contributor, and (ii) subsequently assessing the strength of the evidence. Importantly, no extra polymerase chain reaction cycles were used in this case. This emphasizes the fact that low-template profiles may be encountered in routine forensic casework, whether or not deliberate enhancement procedures are used, and whether or not the laboratory is intentionally typing a sample that is thought, *a priori*, to contain a low quantity of DNA. Even when the total amount of DNA in a sample appears sufficient, minor components may be present at much lower quantities, presenting a low-template situation.

For item 24 (as well as one other sample), the laboratory initially excluded the alternate suspect as a possible minor contributor to these samples. For stain 24, this exclusion was based on ignoring a peak that was clearly genuine, although falling below 50 RFU in the injection the laboratory picked to interpret, as well as a failure to consider that a peak in the stutter position of a large peak in the major profile might also harbor an allele from the minor contributor. The issues were slightly different for each sample, but included dogmatic adherence to an analytical threshold and an inability to probabilistically address the possibility of allelic drop-out.

Both of these issues trace back to the historical decision of forensic laboratories to assign strict threshold limits to the determination of whether a peak either constitutes a true signal or represents noise or stutter. In our experience, this approach is proving more problematic than helpful in accurately interpreting and assigning statistical weight to evidence samples, especially for low-level, complex mixtures; this case is but one example. A better approach, which will be explored more fully in forthcoming work, enjoys a long history in analytical chemistry, that of setting an analytical threshold based on the quality of validation data, or, better yet, the quality of a particular run. This is often accomplished by setting a limit of detection and a limit of quantitation (2,23–26). At least one paper has been published applying that approach to forensic DNA work (27), and we are aware of laboratories that have implemented it. This approach allows the analyst to use much more of the data, increasing the reliability of the interpretation. In our experience, much interpretational strife results from gratuitously discarding data based on inflexible laboratory guidelines based as much on policy as science.

Stutter (the *in vitro* loss or addition of an STR repeat unit) was acknowledged as a limitation of STR markers from the beginning of their forensic implementation; however, the contribution of stutter to the ambiguity of low-level complex mixtures has vastly increased the problem of interpreting such samples. Because no molecular biological solution seems imminent, or even theoretically possible, all we can do is attempt to increase our understanding of the phenomena that affect the proportion of stutter, and also how minor components contribute to peak height in the stutter position of major donors. In the meantime, analysts should remember that, for low-level minor profiles, even stutter peaks that do not rise above a proportion threshold determined for high-quality single-source samples could be masking a true minor allele.

The statistics provided by the laboratory to assess the weight of the DNA evidence substantially underestimated the strength of all four evidence items that we addressed in the case. Indeed, for two items, including item 24, the focus of this paper, they patently excluded the alternate suspect as a possible contributor to the minor profile. Later, the laboratory revised their statistics to account for the possibility of drop-out by excluding (i.e., set the CPI equal to 1) all loci with any peaks between their detection threshold of 50 RFU and their stochastic threshold of 200 RFU. While this is not an uncommon approach, this case illustrates why discarding large amounts of data wholesale is not necessarily a good, useful, or even "conservative," procedure. Rather, this approach substantially understated the strength of the evidence against the alternate suspect.

Our calculations assessed the strength of the evidence using a probabilistic approach, specifically an LR that contained an explicit model of allelic drop-out (calculations available upon request). As such, we used all loci and alleles that rose above the allelic detection threshold of 50 RFU. Importantly, no need exists to apply a stochastic threshold using LRs with drop-out; thus, we were able to incorporate much more of the information present in the evidence profiles. As a result, our statistics are more likely to reflect the true weight of the evidence. While one challenge to the probabilistic approach used here is that it requires the user to input a value for the drop-out probability, we have also shown how laboratory validation data can be used to generate empirical estimates of drop-out probabilities. Any laboratory can apply this approach using their own validation data, adding LRs incorporating allelic drop-out to their forensic casework toolbox.

Conclusions

The conclusions and statistical weighting by the laboratory regarding the four samples we addressed were either patently incorrect or greatly underweighted the evidence with respect to the inclusion of the alternate suspect. For sample 24, discussed in this paper, as well as one other, the alternate suspect was initially falsely excluded as a possible contributor of the minor profile. The numerous factors discussed above culminated in an inability of the laboratory to fully incorporate all of the data present in the sample using appropriate statistical methods. The method of Balding and Buckleton (5) allowed us to incorporate empirically derived drop-out estimates into a more relevant and reliable estimate of the weight of each item of evidence. This

turned out to be critical to understanding the evidence as it relates to the alternate suspect.

For example, the DNA results obtained from stain 24 are substantially more likely if, in addition to the victim, the alternate suspect is a contributor to the evidence samples than if a random unrelated individual is the minor contributor.

This work was all proffered at an admissibility hearing prior to the start of the trial. Although we knew that this was the first time that LRs incorporating drop-in and drop-out would be offered in a U.S. Court of Law, we were unaware that no appellate decision existed with regard to even standard LRs used for mixtures where drop-out was not required to explain a conclusion. In this case, the judge admitted these calculations to be used at trial.

Acknowledgments

We are indebted to David Balding for writing and posting the original R code, and also for performing the first set of calculations for this case and issuing a report. His generous contribution of time and assistance throughout was much appreciated. Berch Henry, Thomas Wahl, and Megan Palmer of the North Dakota State University Forensic DNA Facility performed retesting of some of the items and DNA extracts that helped to confirm some of the alleles and clarify the profiles. John Buckleton, Keith Inman, and Rori Rohlfs provided insightful comments that helped to clarify and improve the manuscript. We also appreciated the thoughtful comments and suggestions of our anonymous reviewers. We wish to acknowledge the cooperation of the Government Laboratory in providing their validation data as part of the discovery in this case. Finally, we are grateful to the defense attorneys in this case, Lisa Maguire and Renee Berenson, for supporting our efforts to introduce this novel statistical evidence at trial.

References

1. Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, et al. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res 1996;24(16):3189–94.
2. SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories [Internet]. 2010 [updated 1/14/2010]; http://www.fbi.gov/about-us/lab/codis/swgdam.pdf (accessed May 1, 2010).
3. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the international society of forensic genetics: recommendations on the interpretation of mixtures. Forensic Sci Int 2006;3:90–101.
4. Gill P, Buckleton J. A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. Forensic Sci Int Genet 2010;4(4):221–7.
5. Balding DJ, Buckleton J. Interpreting low template DNA profiles. Forensic Sci Int Genet 2009;4(1):1–10.
6. Buckleton J, Triggs C. Is the 2p rule always conservative? Forensic Sci Int 2006;3:206–9.
7. Buckleton J, Curran J. A discussion of the merits of random man not excluded and likelihood ratios. Forensic Sci Int Genet 2008;2(4):343–8.
8. Curran JM, Buckleton J. Inclusion probabilities and drop-out. J Forensic Sci 2010;55(5):1171–3.
9. The National Research Council. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press, 1996.
10. Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KA, Buckleton J. Interpreting DNA mixtures. J Forensic Sci 1997;42(2):213–22.
11. Tvedebrink T, Eriksen PS, Mogensen HS, Morling N. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Sci Int Genet 2009;3(4):222–6.
12. Buckleton J, Gill P. Low copy number. In: Buckleton J, Triggs CM, Walsh SJ, editors. Forensic DNA evidence interpretation. Boca Raton, FL: CRC Press, 2005;275–97.
13. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 2000;112(1):17–40.
14. Gill P, Kirkham A, Curran J. LoComatioN: a software tool for the analysis of low copy number DNA profiles. Forensic Sci Int 2007;3:128–38.
15. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2009;ISBN 3-00051-07-0 [updated 1/14/2010], http://www.R-project.org (accessed February 22, 2010).
16. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;3:125–40.
17. Balding DJ, Nichols RA. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 1995;2:3–12.
18. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48(4):908–11.
19. Buckleton J. Validation issues around DNA typing of low level DNA. Forensic Sci Int Genet 2009;3(4):255–60.
20. Budowle B, Eisenberg AJ, van Daal A. Validity of low copy number typing and applications to forensic science. Croat Med J 2009;50(3):207–17.
21. Gill P. Application of low copy number DNA profiling. Croat Med J 2001;42(3):229–32.
22. Gill P, Buckleton J. Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810–21. J Forensic Sci 2010;55(1):265–8.
23. Anderson DJ. Determination of the lower limit of detection. Clin Chem 1989;35(10):2152–3.
24. Armbruster DA, Tillman MD, Hubbs LM. Limit of detection (LQD)/limit of quantitation (LOQ): comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. Clin Chem 1994;1:1233–8.
25. Rubinson KA, Rubinson JF. Sample size and major, minor, trace, and ultratrace components. In: Rubinson KA, Rubinson JF, editors. Contemporary instrumental analysis. Upper Saddle River, NJ: Prentice Hall, 2000;150–8.
26. Thomsen V, Schatzlein D, Mercuro D. Limits of detection in spectroscopy. Spectroscopy 2003;18(12):112–4.
27. 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(1):97–101.

Additional information and reprint requests:
Kirk E. Lohmueller, Ph.D.
Department of Integrative Biology
University of California
1005 Valley Life Sciences, Building 3140
Berkeley, CA 94720-3140
E-mail: k.lohmueller@berkeley.edu