



Forensic DNA Statistics: Still Controversial In Some Cases

Although forensic DNA testing is well established, experts sometimes disagree about the interpretation and statistical characterization of test results. This article will describe the key controversies and will explain what lawyers need to know to recognize and deal with controversial types of DNA evidence.

When labs try to “type” samples that contain too little DNA, or DNA that is too degraded, the results of the DNA test can be unreliable. The test may fail to detect certain genetic characteristics (called alleles) of people who contributed DNA to the sample — a phenomenon called allelic drop out; the tests may falsely detect characteristics that did not come from contributors — a phenomenon called allelic drop in; and the test results may be distorted in other ways that complicate interpretation.

Labs try to make allowances for these distortions when they deal with limited and degraded samples. Because two samples from the same person may (under these conditions) produce slightly different DNA profiles, labs must use lenient standards for declaring that a suspect “matches” or is “included as a possible contributor” to an evidentiary sample. They do not require that the suspect’s profile correspond exactly to the evi-

dentiary profile — some differences are allowed in order to account for drop out, drop in, and other types of distortion. But this leniency in matching increases the probability that an innocent suspect, who was not a contributor, will be incriminated — and therefore requires an adjustment to the statistics that describe the value of the DNA match. The controversy we will examine concerns how big this adjustment should be, and how labs should make it. Before delving into that debate, however, we will provide some background information on DNA evidence.

I. Admissibility of DNA Statistics

DNA evidence is typically accompanied by impressive statistics that purport to show the meaning (and power) of a DNA match. The most common statistic is the random match probability (RMP), which is an estimate of the probability that a randomly selected person from some reference population (e.g., Caucasians, African-Americans, Hispanics) would “match” or be “included as a potential contributor” to an evidentiary sample. These statistics can be extremely impressive — random match probabilities of one in billions, trillions, or even quadrillions and quintillions are typical, although more modest estimates of one in thousands, hundreds, or even less, are sometimes offered in cases where limited DNA profile information was detected in the evidentiary sample. In cases involving mixed DNA samples (samples with more than one contributor) labs sometimes compute another statistic called a combined probability of inclusion or exclusion (CPI or CPE), which reflects the relative probability of obtaining the observed results if the defendant was (and was not) a

BY WILLIAM C. THOMPSON, LAURENCE D. MUELLER, AND DAN E. KRANE

contributor.

The scientific community has long recognized the need for statistics in connection with DNA evidence. In 1992, the National Research Council, in its first report on DNA evidence, declared:

To say that two patterns match, without providing any scientifically valid estimate (or at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless. ... DNA "inclusions" cannot be interpreted without knowledge of how often a match might be expected to occur in the general population.¹

The Scientific Working Group on DNA Analysis Methods (SWGDAM), a group of forensic scientists chosen by the FBI to propose guidelines for DNA testing, has also declared that statistics are an essential part of forensic DNA analysis. Its most recent guidelines (approved in January 2010) include the following: "The laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of the case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis." (Interpretation Guideline 4.1).²

Courts in most jurisdictions will not admit DNA evidence unless it is accompanied by statistics to give meaning to the finding of a match. As one court explained, "[w]ithout the probability assessment, the

jury does not know what to make of the fact that the patterns match: the jury does not know whether the patterns are as common as pictures with two eyes, or as unique as the Mona Lisa."³ In many jurisdictions the law also requires, as a condition of admissibility, that the statistics presented in connection with the DNA evidence be accurate and reliable. In other words, the validity of the statistics is an issue going to the admissibility of DNA evidence, not just its weight.⁴

Since the mid-1990s, proponents of DNA evidence have generally had little difficulty establishing the admissibility of DNA statistics. As we will show, however, most laboratories are using statistical methods that are valid only in cases where the test results are reliable and accurate — which means cases where it can safely be assumed that all alleles (genetic characteristics) of the individuals who contributed DNA to the evidentiary samples have been detected. A growing number of laboratories are expanding their use of DNA testing — using it to examine limited and degraded samples where it is not possible to obtain complete results. In those cases, the standard statistical methods, which have long been accepted for the typical case, can be highly misleading.

II. Computing DNA Statistics in Cases With Ample Quantities of DNA and Clear Results

Forensic DNA testing in the United States is generally done using commercial

test kits that examine at least 12 locations (loci) on the human genome where there are STRs (short tandem repeats). At each STR a person will have two "alleles" (genetic markers), one inherited from each parent. If two DNA samples are from the same person, they will have the same alleles at each locus examined; if two samples are from different people, they will almost always have different alleles at some of the loci. The goal of forensic DNA testing is to detect the alleles present at the tested loci in evidentiary samples so that they can be compared with the alleles detected at the same loci in reference samples from possible contributors.⁵

DNA tests produce computer-generated charts (called electropherograms) in which "peaks" represent the alleles. Figure 1 shows results for five different samples — the top chart shows the alleles found in blood from a crime scene; the lower four charts show the alleles found in reference samples from four suspects. The charts show results for just three of the loci normally examined by the test — these three loci are called D3S1358, vWA, and FGA. At each locus the computer assigns a number to each peak (based on its position which in turn is determined by its size) to indicate which of the possible alleles it represents.

A quick look at the charts will show that only Suspect 3 has the same alleles as the blood from the crime scene — hence, based on these results, Suspect 3 would be "included" as a possible source of the blood, while the other three suspects would be excluded. To know how much weight to assign to such a "match," labs typically compute the random match

Figure 1: Electropherograms showing DNA profiles of five samples at three loci

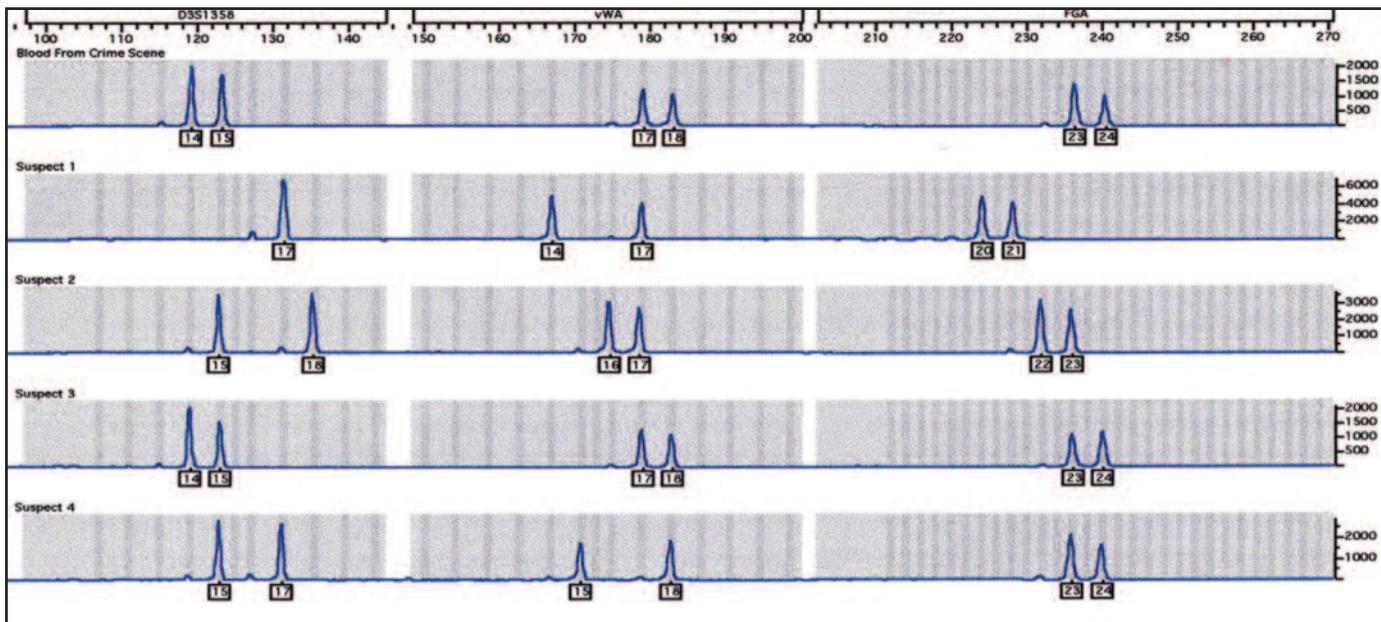


Table 1: Allele, genotype and profile frequencies (among U.S. Caucasians) for the blood from the crime scene

	Locus						Overall Profile Frequency (Random Match Probability)	
	D3S1358		vWA		FGA			
	Alleles	14	15	17	18	23	24	
Allele frequencies	.103	.262	.281	.200	.134	.136		
Genotype frequencies	0.054 or 1 in 18.5		0.112 or 1 in 8.9		0.036 or 1 in 27.8		0.0002177 or 1 in 4,592	

probability — that is, the probability that a random person would “match” the alleles found in the blood sample. Table 1 shows how this is done. A lab analyst consults a database to determine the frequency of each of the matching alleles in a reference population that reflects a pool of alternative suspects. (Table 1 shows the frequency of each allele among U.S. Caucasians). These frequencies are then multiplied together in a particular way. If there are two alleles at a locus, the frequency for the pair (called a genotype) can be determined with the simple formula: $2 \times p \times q$, where p and q are the frequencies of the individual alleles. For locus D3S1358, for example, the allele frequencies are 0.103 and 0.262 (which means that 10.3 percent of the alleles observed for this locus are “14” and 26.2 percent are “15”), so the frequency of the pair of alleles (genotype) is $0.103 \times 0.262 \times 2 = 0.054$, which means that among U.S. Caucasians approximately 1 person in 18.5 would have this genotype. If there is only one allele at a locus (which would occur if the person inherited the same allele from both parents), then the frequency for such a genotype is simply p^2 , where p is the frequency of that allele. The frequencies of the genotypes at each locus are then all multiplied together to produce a frequency estimate for the overall profile.⁶ The overall frequency (among U.S. Caucasians) of the three-locus DNA profile seen in the blood from the crime scene (and also in Suspect 3) is approximately 0.000212, which means that approximately one person in 4,592 would be expected to have this particular three-locus profile.

The profile frequency is often called the random match probability (RMP) because it delivers the answer to a very specific question: the probability that a randomly chosen, unrelated person from the reference population would happen to have the same DNA profile found in an evidence sample. The DNA test places a matching suspect, like Suspect 3, in a cate-

gory of individuals who might have been the donor of this sample; the profile frequency (RMP) tells us something about the size of this group (in our example, it includes 1 in 4,592 Caucasians, which is a lot of people given that more than 100 million Caucasians live in the United States). The profile frequency does *not* tell us the probability that Suspect 3 was (or was not) the source of the blood at the crime scene because the DNA evidence cannot tell us whether Suspect 3 was more or less likely to be the donor than any of the other people who would also match. But it does tell us how broad a net was cast by the test — it was broad enough to incriminate about one person in 4,592, just by chance. Hence, in a case like the one illustrated here, it is a useful statistic for characterizing the value of a DNA match and is widely accepted as valid and appropriate by scientific experts. Unfortunately, not all cases are as clear and easy to interpret as this one — for more complex cases, the answer to the question asked by the random match probability (RMP) can be highly misleading.

III. Computing Statistics in Cases With Incomplete Or Unreliable Data

In 2001, Bruce Budowle, then a senior scientist at the FBI Crime Laboratory, and some of his colleagues, issued a warning to the DNA testing community about the way they were beginning to use (and abuse) the new STR technology:

Because of the successes encountered with STR typing, it was inevitable that some individuals would endeavor to type samples containing very minute amounts of DNA. ... When few copies of DNA template are present, stochastic amplification may occur, resulting in either a substantial imbalance of two alleles at a given heterozygous

locus or allelic drop out.⁷

Budowle's concern, in short, was that labs were beginning to use STR technology in ways that were unreliable — examining samples so limited or so degraded that the tests could not be counted on to produce a complete and accurate profile.

It is now well accepted in the scientific community that STR tests become unreliable when used to type samples containing too little DNA. Such limited samples are sometimes called low-copy number (LCN) samples because they contain a low number of copies of the targeted DNA. When such samples are examined the STR test may pick up more copies of one allele than another at a particular locus simply as a result of sampling error. This phenomenon, known as a stochastic effect, can distort the resulting DNA profiles in various ways. For example, some alleles simply may not be sampled (allelic drop out); it can also cause two alleles from the same person to have widely different peak-heights, which falsely suggests that they came from two different people. Another problem with LCN samples is the detection of spurious alleles due to a phenomenon called “allelic drop in” as well as increased incidence of a complicating artifact known as “stutter.” As Budowle et al. explained in 2001, “[m]ixture analyses and confirmation of a mixture are not reliable with LCN typing, because ... imbalance of heterozygote alleles, increased production of stutter products, and allele drop in can occur.”

Although these underlying problems are widely recognized, some forensic laboratories are pushing the limits of STR testing and ignoring Budowle's warning. Increasingly STR testing is used on aged and extremely limited samples, such as “touch DNA” samples, that are likely to yield quantities of DNA in the range where stochastic effects occur. Analysts try to take these effects into account by being more lenient about their standards

for declaring a match. They do not require that a suspect's DNA profile correspond exactly to an evidentiary sample because stochastic effects, or other phenomena associated with low copy number testing, might cause discrepancies. This leniency in matching is necessary to avoid false exclusions, but it makes it difficult if not impossible to estimate the statistical meaning of a "match."

Consider, for example, the DNA profiles that are compared in Figure 2. The chart on top shows the profile of an evidentiary sample from a sexual assault case in which very little DNA was recovered. The chart on the bottom shows the profile of the defendant. If this were a conventional case, the defendant would be excluded as a possible contributor because his profile obviously differs. One of his alleles at locus D13S317 (allele "14") was not found in the evidentiary sample. But he was not excluded — in fact, he is now in prison on the strength of this evidence. The lab attributed the discrepancy to "allelic drop out" and concluded that he was a possible contributor notwithstanding the discrepancy.⁸

This lenient approach to declaring an "inclusion" or "match" is problematic in several ways. First, it depends in part on circular reasoning. There is no way to verify that a discrepancy like the one illustrated in Figure 2 was, in fact, caused by "allelic drop out." An alternative possibility is that the true contributor of the sample is a homozygote (a person who inher-

ited allele 8 from both of his parents and has no other allele to contribute). The conclusion that allelic drop out occurred here is inferred, in part, from the very fact that it purports to explain — the discrepancy between the evidentiary sample and the defendant's profile. At best this type of analysis creates uncertainty about the value of DNA evidence; at worst it undermines that value entirely.

The lenient approach to interpretation also creates a second problem — it undermines the objectivity of DNA evidence. The results of a comparison like that shown in Figure 2 depend on an analyst's subjective judgment about whether drop out did or did not occur, which creates room for expert disagreement. A recent study by Itiel Dror and Greg Hampikian illustrates the degree of subjectivity (and disagreement) that is possible in DNA interpretation.⁹ They asked 17 qualified DNA analysts from accredited laboratories to evaluate independently the DNA evidence that had been used to prove that a Georgia man participated in a gang rape. The analysts were given the DNA profile of the Georgia man, and were given the DNA test results obtained from a sample collected from the rape victim, but were not told anything about the underlying facts of the case (other than scientific details needed to interpret the test results). The analysts were asked to judge, based on the scientific results alone, whether the Georgia man should be included or excluded as a possible con-

tributor to the mixed DNA sample from the victim. Twelve of the analysts said the Georgia man should be excluded, four judged the evidence to be inconclusive, and only one agreed with the interpretation that had caused the Georgia man to be convicted and sent to prison — i.e., that he was included as a possible contributor to the DNA mixture. The authors found it "interesting that even using the 'gold standard' DNA, different examiners reach conflicting conclusions based on identical evidentiary data."¹⁰

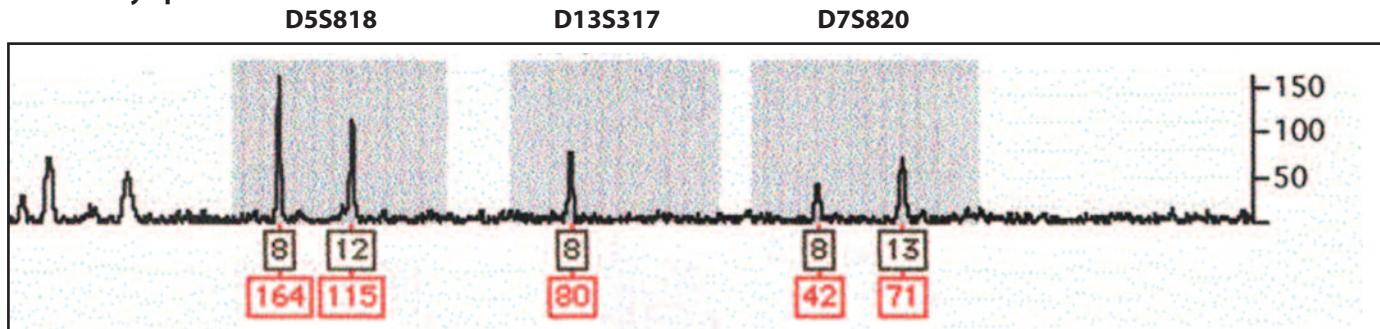
Dror and Hampikian also highlighted a third problem with the lenient approach to interpretation required when typing low copy number samples — it creates a potential for bias in the interpretation of DNA tests. Noting that the analyst who testified in the Georgia case had been exposed to investigative facts suggesting that the Georgia man was guilty, Dror and Hampikian suggested that this "domain irrelevant information may have biased" the analyst's conclusions.¹¹ Other commentators have also documented the potential for bias in the interpretation of DNA test results.¹²

We will focus here, however, on a fourth problem with the lenient approach to DNA interpretation illustrated in Figure 2 — it makes it difficult, if not impossible, to assess the statistical meaning of the resulting DNA "match." The standard approach of computing the frequency (in various reference populations) of people who have the same profile as the

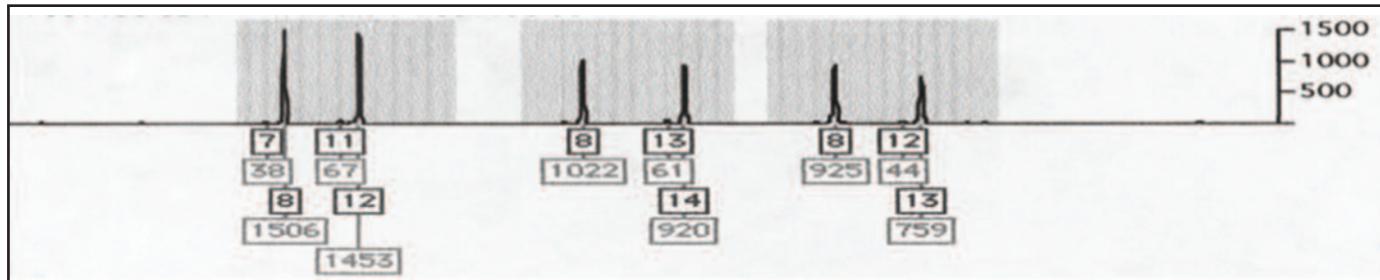
Figure 2: DNA profile of an evidentiary sample in a sexual assault case and DNA profile of a suspect

(The height of each peak in relative fluorescent units — RFU — is shown in the boxes below the labels for the peaks.)

Evidentiary Specimen



Defendant's Reference Profile



Thank You to our Cocktail Reception Sponsor at the White Collar Seminar in New York



The reception was held in the Fordham Law Center Atrium.

evidentiary sample will not work here. The frequency of the evidentiary profile is irrelevant when analysts are willing to “include” as possible contributors people (like the defendant) who do not have that profile. When analysts widen the net to include people whose profiles do not match the evidentiary profile perfectly, the frequency of the evidentiary profile no longer reflects and will necessarily underestimate the likelihood that an innocent person will be included by chance.

A few labs have tried (incorrectly) to deal with this problem by computing the frequency, in a reference population, of people who have the alleles that the defendant *shares* with the evidentiary sample. In other words, they try to compute the frequency of people who would “match” the evidentiary sample *the way that the defendant matches*. For the case illustrated in Figure 2, they would compute the frequency of people who have the defendant’s genotype at locus D5S818 and locus D7S820, and who have the “8” allele (in combination with any other allele) at locus D13S317. But the net cast by this test has the potential to “include” a far broader group of people than those defined by this method. Suppose, for

example, that the defendant had genotype 13,13 rather than 8,13 at locus D7S820. Would he have been excluded? We strongly doubt it. We think that in that case the analyst might simply have concluded that the “8” allele observed at that locus was caused by “allelic drop in” or by DNA of a second contributor. Analysts are often reluctant to “exclude” a suspect unless there is no plausible way to account for discrepancies between his profile and the evidentiary sample.¹³ When the analyst can invoke allelic drop out, drop in, secondary contributors, and various other explanatory mechanisms to account for discrepancies, however, the number of profiles that might plausibly be consistent with the evidentiary sample can expand rapidly. This means the net cast by the test is far broader than the category of individuals who match *the way the defendant matches*.

Another approach labs have taken to address this problem is to base statistical computations only on loci where all of the suspect’s alleles were detected in the evidence, ignoring (for statistical purposes) any locus where the suspect does not match. But two prominent experts of forensic statistics have recently published

an article calling this approach unsupportable.¹⁴ Using statistical modeling, they showed that “this approach may produce apparently strong evidence against a surprisingly large fraction of noncontributors.”¹⁵ It allows labs to ignore potentially exculpatory results (the loci where the suspect does not perfectly match) while counting against him those where he does match, and this cherry-picking of data greatly expands the net cast by the test in ways that are not adequately reflected in inclusion statistics. Such an approach answers a substantially different question than the one addressed by the random match probability. It essentially asks: “What is the chance that a randomly chosen individual from a given population would match the defendant if we focus only on evidence consistent with a match and ignore evidence to the contrary?” When analysts present such statistics to the jury, however, they typically (and misleadingly) describe them as “random match probabilities.”

The fundamental problem facing those who try to design statistical procedures for such cases is that no one knows how broad the net cast by the test really is. Estimating the percentage of the pop-

ulation who would be “included” as a possible contributor is like estimating the length of a rubber band. Just as a rubber band may be longer or shorter, depending on how far one is willing to stretch it, the size of the “included” population may be larger or smaller, depending on how leniently or strictly the analyst defines the criteria for an “inclusion.” Because it is unclear just how far the laboratory might stretch to “include” a suspect, the true size of the “included” population cannot be determined.

In some cases, forensic laboratories try to deal with this problem by using statistics known as likelihood ratios rather than frequencies and RMPs. To compute a likelihood ratio one estimates how probable the *observed results* (in the evidentiary sample) would be if the defendant was (and was not) a contributor. Because likelihood ratios focus on the probability of obtaining the exact results observed (under different hypotheses about how they arose), they avoid the difficulty of estimating the size of the “included” group. But they run smack into a related problem. To compute an accurate likelihood ratio, in a case like that shown in Figure 2, one must know the probability that allelic drop out (and drop in) occurred. Without knowledge of the drop out probability, one cannot know the probability of obtaining the observed results if the defendant was a contributor, which means one cannot compute a likelihood ratio (at least not accurately). It is well understood that as the quantity of DNA in a sample decreases, the probability of drop out and drop in increases. But, as the quantity of DNA decreases, the ability to reliably estimate its quantity also decreases. So estimates of the probability of drop out or drop in are often little more than guesses, which we find unacceptable given that the results of the DNA test — whether it is reported as a powerful incrimination or a definitive exclusion — may depend on what guess the expert happens to make.

In 2006, a DNA Commission of the International Society of Forensic Genetics proposed a statistical method that takes drop out probabilities into account when computing a likelihood ratio. Unfortunately, this approach is difficult to apply even in simple cases due to uncertainty about the drop out probability.¹⁶ Matters quickly become even more complicated when formulae incorporating drop out probabilities are applied to mixtures. The Commission acknowledged that “[e]xpansion of these concepts to mixtures is complex and that is why they are not generally used.” But that, of

course, leaves forensic DNA testing in the unsatisfactory position of having no generally accepted method for computing statistics in such cases.

A more defensible way to deal with problems arising from stochastic effects is for the lab to ignore for statistical purposes any locus where it is suspected that stochastic effects (leading to drop out) may have occurred, whether or not the suspect “matches” at that locus. One clue to whether the sample in question is subject to stochastic effects is the height of the “peaks” seen in the electropherogram. Peak heights are measured in “relative fluorescent units” (RFU) and their height can be determined by reference to the vertical indices on an electropherogram. As a general policy, some labs ignore for statistical purposes any results from a locus where there are peaks below a “stochastic threshold,” which is often set somewhere between 50 and 150 RFU. That means that (for statistical purposes) they rely only on loci where they are willing to assume that all of the contributors’ alleles have been detected.

“Stochastic thresholds” are not a perfect solution to the problem posed by unreliable DNA data because it means that the lab may ignore potentially exculpatory data such as the mismatch seen at locus D13S317 in Figure 2. And experts disagree about what the “stochastic threshold” should be. But this approach is less likely than other approaches to produce results that are unfairly biased against a suspect. Unfortunately, the conservatism of this approach is the very reason that many laboratories reject it; this approach reduces the impressiveness of the statistics that can be presented in connection with a DNA match.

When there is no valid way to compute a statistic, or when the only valid method produces statistics that are modest and unimpressive, labs sometimes elect to present DNA evidence without statistics. The jury is told that the defendant’s profile “matches” the evidentiary profile, or that the defendant “cannot be excluded,” but is given no statistics that would allow them to assess the strength of this evidence. This approach violates the SWGDAM guidelines as well as the legal requirement in many jurisdictions that DNA evidence is inadmissible without statistics. Defense lawyers sometimes raise no objection, however, thinking that without statistics the DNA evidence will do less harm, and fearing that an objection may prompt the government to generate a statistic that will prove more incriminating. The danger of this approach is that even in the absence of

Glossary

Allele (peak): One of two or more alternative forms of a gene, a peak appears on an electropherogram for each allele that is detected.

Bayes’ Rule: A mathematical equation that describes how subjective estimates of probability should be revised in light of new evidence.

Degradation: The chemical or physical breaking down of DNA.

Domain irrelevant information: Information that should have no bearing on an expert’s interpretation of scientific data; generally from outside of an expert’s area of expertise.

Drop in: Detection of an allele that is not from a contributor to an evidence sample, usually due to low levels of contamination.

Drop out: Failure to detect an allele that is actually present in a sample, usually due to small amounts of starting material.

Electropherogram: The output of a genetic analyzer, typically displayed as a graph where individual peaks correspond to the presence of alleles detected in a tested sample.

Likelihood ratio: A statistic reflecting the relative probability of a particular finding under alternative theories about its origin.

Locus (pl. loci): The physical location of a gene on a chromosome.

Low-copy number (LCN)/low-template (LT) DNA: DNA test results at or below the stochastic threshold.

Monte Carlo-Markov Chain (MCMC) modeling: A computer-intensive statistical method that proposes millions of possible scenarios that might have produced the observed results, computes the probability of the observed results under each scenario, and uses the resulting distributions (and Bayes’ Rule) to determine which scenarios best explain the observed results.

Random match probability (RMP): The probability that a randomly chosen unrelated individual would have a DNA profile that cannot be distinguished from that observed in an evidence sample.

Scientific Working Group on DNA Analysis Methods (SWG-DAM): A group of forensic scientists from Canada and the U.S. crime laboratories appointed by the director of the FBI to provide guidance on crime laboratory policies and practices.

Stochastic effects: Random fluctuations in testing results that can adversely influence DNA profile interpretation (e.g., exaggerated peak height imbalance, exaggerated stutter, allelic drop-out, and allelic drop-in).

STR (short tandem repeat) testing: A locus where alleles differ in the number of times that a string of four nucleotides are tandemly repeated.

Stutter: A spurious peak that is typically one repeat unit less (or more) in size than a true allele. Stutter arises during DNA amplification because of strand slippage.

statistics the jury will assume that the DNA evidence is highly probative, which could easily cause the jury to over-value the problematic kind of DNA evidence we are discussing here.

IV. Computers to the Rescue? The TrueAllele® Casework System

A Pittsburgh company called Cybergeneitics has been marketing an automated system for interpreting DNA evidence that relies on high-powered computers, and a form of Bayesian analysis called Monte Carlo-Markov Chain (MCMC) modeling, to draw conclusions about the profiles of possible contributors to evidentiary DNA samples. Promoters of this system claim that it is an objective and scientifically valid method for assessing the statistical value of DNA evidence. They claim it can be used in all types of cases, including problematic cases in which sample limitations render the test results less than perfectly reliable. Better yet, the system often produces statistics that are even more impressively incriminating than the statistics produced by conventional methods. This sales pitch appears to be working. The company has sold its system (which consists of software and associated computer hardware) to several forensic laboratories for use in routine casework. It has also helped forensic laborato-

ries come up with statistical estimates in a number of specific cases. In a few instances it has helped defense lawyers by reanalyzing evidence in order to see whether TrueAllele® agreed with the interpretation offered by a human forensic analyst. Because evidence generated by this system is increasingly appearing in courtrooms, it is important that lawyers understand its strengths and limitations.

The system relies on a form of statistical modeling (called MCMC) that has been widely used in the field of statistics to model complex situations. Although the application of this technique to forensic DNA testing is novel, the underlying approach has been used successfully elsewhere. The computer is programmed to make certain assumptions about how forensic DNA tests work, as well as how they fail to work. The assumptions allow the computer to predict how an electropherogram will look when the sample tested has DNA from an individual (or individuals) with specific profiles, how the pattern of peaks should change with variations in the quantity of DNA from each contributor, with variations in the degree to which the DNA is degraded, etc. The assumptions cover such issues as when allelic drop out and drop in would be expected to occur, and how probable these phenomena are under various conditions. Based on the assumptions that are programmed into the system, it can predict the “output” of a forensic DNA test for any given set of “inputs.” In other words, it can predict the probability that a forensic DNA test will produce electropherograms showing a particular pattern of peaks, given that the sample tested contained DNA of an individual (or individuals) who have specific DNA profiles, and given various other assumptions about the quantity and quality of the samples tested. Whether the predictions it makes are accurate is a matter we will consider in a moment—but there is no doubt that the computer can make such predictions in a consistent manner.

In order to analyze an evidentiary sample, such as the one shown in Figure 2, the computer is programmed to propose millions of possible scenarios that might have produced the observed results and then to compute the probability of the observed results under each scenario. The scenarios include all possible combinations of explanatory variables (i.e., genotypes of contributors, mixture proportion, degradation, stutter, etc.). Each scenario is, effectively, a hypothesis about a set of factors that might explain the results; it is evaluated relative to other hypotheses in terms of how well it fits the observed data.

Most hypotheses are effectively ruled out because they cannot explain the observed results. As more and more hypotheses are tested, however, the system identifies a limited number of hypotheses that might possibly explain the data. If all of those hypotheses require that the contributor have the same genotype, then the system assigns a probability of one (certainty) to that genotype. To give a simple example, if such a system assumed a single contributor to the evidentiary sample shown in Figure 2, and tried out various hypotheses about how that single contributor could have produced the observed results at locus D5S818, it would conclude that the contributor must have genotype 8, 12 — any other hypothesis would not fit the data. Hence, it would assign that genotype a probability of one.

For a locus like D13S317, the analysis would be more complicated. The hypothesis that there was a single contributor with genotype 8, 8 would fit the data, but to the extent allelic drop out is possible, the data could also be explained by a single contributor with genotype 8, x where “x” is any other allele at this locus. In that case, the system would compute the probability of each of the possible genotypes by using a formula known as Bayes’ Rule to combine data on the frequency of the genotype in the population with the system’s estimates of the probability of obtaining the observed results if the contributor had that genotype. For example, the system would compute the probability that the contributor had genotype 8, 14 (and therefore matched the defendant) by considering both the underlying frequency of this genotype in the population and the probability of obtaining the observed results (an 8 allele with a peak height of 80) if the single contributor had genotype 8, 14. To do this, the computer would obviously need to estimate the probability that drop out (and/or drop in) occurred in this case. It would make this estimate based on pre-programmed assumptions about such matters as the relationship between peak height and drop out (and drop in: how sure can we be that the contributor at the D5S818 locus was 8, 12 and not really 8, 8 or 12, 12?). As suggested above, the accuracy of these assumptions will be a key issue when evaluating TrueAllele® and similar systems.

Greater complications arise if the system is instructed to assume there could have been more than one contributor. If there are two contributors, then many additional hypotheses become plausible candidates for explaining the data. The results at locus D5S818, for example, could be explained if one con-

THANK YOU TO OUR COCKTAIL RECEPTION SPONSORS AT THE DRUG SEMINAR IN LAS VEGAS



Arrascada & Aramini, Ltd

THE LAW OFFICES OF
Chesnoff and Schonfeld, PC

GORDON  **SILVER**

NACJ

Nevada Attorneys for Criminal Justice

tributor has genotype 8, 8 and the other contributor has genotype 12, 12; but might also be explained by two contributors who both have genotype 8, 12. Or perhaps one contributor has genotype 8, 12 and the other has genotype 8, x, where x dropped out, and so on. To compute the probability of the genotypes, the computer would need to consider every possible combination of explanatory variables and compute the probability of the observed results under each possible scenario for every possible combination of all of the other explanatory variables. This requires the computer to make enormous numbers of calculations, which is why the method requires high-powered computers and analysis of a case often requires many hours (or even days) of computer time.

Once the system has computed the probabilities of the various possible genotypes at each locus for the evidentiary sample, those results are compared with the suspect's genotypes. If any of the suspect's genotypes has been assigned a probability of zero, then the suspect is excluded. If all of the suspect's genotypes have a probability greater than zero, then the suspect is included as a possible contributor. The statistic used to describe the

value of this "inclusion" is a likelihood ratio that indicates how much more probable the observed results would be (based on TrueAllele®'s calculations) if the contributor was the suspect than if the contributor was a random person.¹⁷

Is TrueAllele® a Valid Method?

As with any type of computer modeling, the accuracy of TrueAllele® depends, in part, on the accuracy of the underlying assumptions. The system is designed to take into account the possibility of phenomena like allelic drop out, drop in, and "stutter," but does it really know the probability that those phenomena occurred in a particular case? One expert who commented on automated systems of this type noted that a key problem is that knowledge of such matters is limited.¹⁸ So a central issue, when considering the admissibility of this method in court, will be the scientific foundation for the assumptions on which the model relies. Before accepting evidence produced by such a system, courts should demand to see a careful program of validation that demonstrates the system can accurately classify mixtures of known samples under conditions comparable to those that arise in actual forensic

cases. The fact that an automated system can produce answers to the questions one puts to it is no assurance that the answers are correct. While automated systems appear promising, their ability to handle "hard cases" like those discussed in this article remains to be fully evaluated.

Those promoting TrueAllele® have published a number of articles that explain the theoretical background of the method and describe the validation performed to date.¹⁹ The validation research focuses on two aspects of the system's performance: "efficacy" and reproducibility. "Efficacy" is measured by comparing the likelihood ratios computed by TrueAllele® to likelihood ratios derived from conventional methods like the CPI and CPE. These studies show that TrueAllele® generally produces much larger likelihood ratios than conventional methods. Its "efficacy advantage" stems from the fact that it considers more information when making calculations than the conventional methods do. The conventional methods generally consider only whether an allele is present or absent in a sample; TrueAllele® also considers the height of the underlying peak and the presence or absence of technical artifacts that often accompany actual alleles.

Greater efficacy is not, however, the same thing as greater accuracy. A gasoline gauge that tells you there are 100 gallons in your tank would, by this definition, have more “efficacy” than a gauge that tells you there are only 10 gallons. Before deciding which gauge to rely on, however, you would want to know which one produced an accurate reading of the amount of gas actually in the tank — in other words, you would need to know which gauge is valid and accurate. Although studies have shown that TrueAllele® produces more impressive numbers than other methods, these studies do not address the question of whether those numbers are valid and accurate.

A second line of validation research focuses on whether the results of TrueAllele® are reproducible. In other words, does the system produce the same likelihood ratio each time it is run on a particular sample? The answer is “no.” Because there are random elements in the way the system does its modeling, such as the random choice of which hypotheses to consider, in which order, no two computer runs will be exactly the same. Proponents argue that the variation in the answers produced by the system is small enough not to matter although, as discussed below, that assertion has been challenged in recent litigation.

One limitation of all the published TrueAllele® validation studies is that the number of samples tested was relatively small. Some studies included forensic samples from laboratory casework, which have the advantage of being realistic, but are somewhat problematic for validation because the exact number of contributors and their profiles cannot be known with certainty. We would like to see additional studies that create and test samples from known contributors where the samples have sufficiently low levels of DNA that allelic drop out is likely to occur. Although the existing validation demonstrates that TrueAllele® can, in some cases, make inferences about the genotypes of contributors that human analysts would have difficulty making, we think the existing validation is insufficient to prove that TrueAllele® can consistently make correct genotype inferences in challenging, problematic cases such as mixture cases with unequal contributions from the contributors, limited quantities of DNA, degradation due to environmental insult, etc. And the published validation studies have not yet answered the more difficult question of whether the likelihood ratio estimates produced by the system are appropriate. Consequently, grounds may well exist to challenge the

admissibility of results from TrueAllele® under either the *Frye* or *Daubert* standards, although (to our knowledge) no such challenges have yet been mounted.

Issues have also arisen concerning the way the TrueAllele® system has been used in particular cases and the way in which its results have been reported. For example, in a case in Belfast, Northern Ireland,²⁰ testimony revealed that the company ran the software four separate times on a particular electropherogram and produced four different likelihood ratios for incriminating the defendant: 389 million, 1.9 billion, 6.03 billion, and 17.8 billion.²¹ These varying results illustrate the issue of reproducibility discussed above — they show that there is an element of uncertainty (a margin of error) in the likelihood ratios generated by the system. How this uncertainty is handled, when reporting the results to the jury, is an issue on which experts may well differ. Cybergeneitics reported that the likelihood ratio for the electropherogram was six billion, a number that the company's president defended on grounds that it was the center of the range of values produced in the four runs. A defense expert (Laurence Mueller, co-author of this article) contend that the company should have taken account of the margin of error by computing a 95 percent confidence interval based on the four estimates and reporting its lower bound — which would have yielded an estimate of 214 million (which is lower than the reported value by a factor of 28). While all the numbers for this particular evidence sample were highly incriminating, a difference of this size might well be consequential in a cold-hit case in which there is little other evidence to incriminate the defendant.

More important, perhaps, was another issue that arose in the same case. A critical evidentiary sample in the case, a swab from a cell phone, was amplified three times in an effort to produce complete results. Each time the result was imperfect — each electropherogram showed a slightly different set of alleles, suggesting that drop out and perhaps drop in had occurred. When TrueAllele® was run on the results of all three amplifications (treating them as samples of the same underlying set of alleles), TrueAllele® produced a likelihood ratio for incriminating the defendant of 24,000. Cybergeneitics elected not to report this likelihood ratio; it instead reported the likelihood ratio of six billion, which was based on one of the three amplifications. When questioned about this decision, the company president stated that he thought that the particular amplification he had chosen to rely upon

when reporting statistics was the most informative of the three, and that relying on the three amplifications together was less informative than relying on the one he deemed the best. This decision was criticized by defense experts, who suggested that the company had cherry-picked its data, deciding that one of the amplifications was more “informative” because it was more incriminating for the defendant. This dispute suggests that even the use of a purportedly objective computer-based method for interpretation and statistical computation will not end questions about the potential for analytic bias in DNA testing. It is unclear, however, whether this criticism had any impact on the judge who was trying the case. He convicted defendant Shiver but acquitted Duffy after a bench trial.²²

To reduce the potential for analytical bias, a defense lawyer in one case asked Cybergeneitics to analyze the evidentiary samples and derive the DNA profile (or profiles) of possible contributors without knowing the profiles of the defendant or any other suspects. To its credit, the company agreed to perform the testing in this “blind” manner. The company was told the profiles of the defendant and other possible contributors (so that the company could draw conclusions about the statistical value of “inclusions”) only after it had provided an initial report regarding the genotypes of possible contributors. The dispute in the Belfast case about whether the company had cherry-picked its data might well have been eliminated had the company chosen which evidentiary samples were most “informative” without having access to information about the profiles of any suspects. In our opinion, lawyers who choose to have samples analyzed by an automated system (or, indeed, by any independent form of review) would be well advised to ask that it be done in a manner that keeps the analyst “blind” to the profiles of suspects until after the evidentiary samples are fully interpreted.

V. Identifying Problematic Cases

One of the goals in writing this article was to alert lawyers to the special problems that arise in cases in which sample limitations affect the reliability of DNA profiles. In such cases there is a special need for expert assistance to help lawyers evaluate the statistical significance of the evidence, and grounds may well exist for challenging the admissibility of some findings. But how can lawyers tell whether the DNA evidence in a case they are handling falls into the problematic

Presented in Partnership by The National Association of Criminal Defense Lawyers
and the California Attorneys for Criminal Justice

MAKING SENSE OF SCIENCE VI: FORENSIC SCIENCE & THE LAW

April 5-6, 2013 * The Cosmopolitan Hotel * Las Vegas, NV



NACDL & CACJ's 6th Annual Forensic Science Seminar will be a two-day event in the City of Lights — Las Vegas! In the modern world, you need to know and understand the forensic sciences in order to effectively represent your clients. Attend this one-of-a-kind CLE seminar and leave with a better understanding of forensic evidence and technology to use in the arsenal of tools to win your next case. If it involves forensic evidence or technology in a criminal case, it will be covered at this one-of-a-kind seminar with an unprecedented faculty.

Topics will include:

- ◆ Designer Drugs: The Trend
- ◆ DNA: Be More Skeptical
- ◆ Determining Cause of Death
- ◆ Audio/Video Enhancement in the George Zimmerman Case
- ◆ Ballistics and Toolmarks 101
- ◆ Postmortem Interval: Calculating Time of Death
- ◆ Arson Investigation & The Souliotes Case
- ◆ The Science Behind Substance Abuse and its Impact on the Brain
- ◆ Memory and Eyewitness Identification
- ◆ An Update on the NAS Report
- ◆ DNA Statistics
- ◆ The Science Behind False Confessions
- ◆ Shaken Baby Syndrome

Registration Information

Circle your membership category fee and note the Grand Total due. Confirmation letters will be sent to all registrants in early March. Registrants must stop by the on-site registration desk before attending any events to receive their badge and materials. **Cancellations must be received in writing by 5:30 pm ET on April 1, 2013 to receive a refund, less a \$75 processing fee.** A \$15 processing fee will be applied for returned checks. To register online, visit www.nacdl.org/meetings; or fax this form with credit card information to 202-872-8690; or mail with full payment to: **2013 Forensic Seminar, 1660 L St., NW, 12th Floor, Washington, DC 20036. Questions? Contact Viviana Sejas at 202-872-8600 x632.**

Name _____

Badge name (if different) _____

CLE State(s) _____ Bar # (s) _____

Address _____

City _____ State _____ Zip _____

Office phone _____ Home phone _____

E-mail _____ Fax _____

Guest(s)/children _____

Check enclosed (payable to NACDL)
 AMEX VISA MASTERCARD DISCOVER

Card Number _____

Name on Card _____ Exp. Date _____

Billing Address _____

City _____ State _____ Zip _____

Authorized Signature _____

Fax to: (202) 872-8690

Seminar Registration Fees

NACDL Members: Regular, Life and Sustaining	\$389
CACJ Members	\$389
All NEVADA Lawyers (membership not required)	\$229
All Public Defenders	\$285
<u>Non-Members</u> of NACDL or CACJ	\$559
1-Day Seminar	\$205
Groups of (4) or more lawyers (Must register at the same time with one payment; no changes, replacements, or cancellations allowed)	\$249

Discounted NACDL Membership & Seminar

Regular Membership & Seminar (First Time Members Only)	\$535
PD Membership & Seminar (First Time Members Only)	\$385

Seminar Total

Select your seminar materials preference:

Pre-Seminar Download (FREE) Hard copy (add \$30)
Emailed to you the week before the seminar.

Materials Fees

Seminar Attendee Audio CD	\$150
Non-Attendee Audio CD	\$200
Non-Attendee Written Materials & Audio CD	\$250
DVD Video (Pre-Sale)	\$400

Materials Total

Subtotal

GA CLE fee add \$60.00*
NC CLE fee add \$40.00*
PA CLE fee add \$21.00*
IL, NE, UT CLE fee add \$15.00*
TX CLE fee add \$10.00*

*State fees must be paid in advance to receive CLE credit.
If left blank, you will not receive credit.

Grand Total

category being discussed here? There are several important clues:

❖ *LCN cases.* If the laboratory identifies the case as one involving “low copy number” (LCN) testing (sometimes also called “low template,” “trace,” or “touch” DNA testing), then the lawyer can be certain that it raises the statistical issues discussed in this article. It is important to understand, however, that not all cases raising these issues are identified as LCN cases by forensic laboratories. Laboratories sometimes apply the label LCN only when they employ special testing procedures designed to increase the sensitivity of DNA testing (in order to find results with samples that are otherwise too limited for STR testing). But laboratories frequently encounter problems like allelic drop out even when they do not employ these procedures, so one should not assume that the problems discussed here are limited to LCN cases.

❖ *The samples tested contain small quantities of DNA.* Forensic STR tests are designed to work best when they test samples that yield 0.5 to 1 ng (nanogram) of DNA for analysis. When preparing samples for STR testing, labs typically perform an operation known as quantitation in order to estimate the amount of DNA available. They then try to concentrate or dilute the samples to the proper level for STR testing. Sometimes, however, the estimated quantity of DNA in the samples is too low to achieve the ideal concentration. In such cases labs often go ahead with the test anyway and hope for the best. It is well known, however, that stochastic effects become more likely as the concentration of DNA decreases. If it is below 250 pg (picograms), and certainly if it is below 100 pg, which is one-tenth of a nanogram, then most experts agree that stochastic effects are almost certain to be in play. The moral of this story for lawyers is that it is important to know the lab’s estimates of the quantities of DNA in the samples tested, which can be found in the laboratory case file or bench notes. But these estimates are not always accurate (especially as they get into the 100 pg range and below), so it is important to attend to the other clues listed here as well.

❖ *Defendant is identified as a possible “minor contributor” to a mixed sample.* In mixed samples there sometimes is enough DNA to produce a reliable profile for the major contributor, but not enough to produce reliable profiles for minor contribu-

tors. Suppose, for example, that the sample containing 1 ng is a 10:1 mixture of DNA from a major and minor contributor. The quantity of DNA from the minor contributor will be less than 100 pg, which means that it will almost certainly be plagued by stochastic effects.

❖ *One or more alleles of alleged contributors were not detected in the evidentiary sample.* If the laboratory must invoke the theory of allelic drop out (and/or drop in) in order to avoid excluding putative contributors, then the lawyer can be certain that the case falls in the problematic category discussed here.

❖ *Low peak heights.* As noted above, the height of the peaks in electropherograms usually corresponds to the quantity of DNA present from a particular contributor. As the peaks associated with that contributor get shorter, the chance of stochastic effects increases. They are most common when peak heights are below 150 RFU.

Lawyers who see any of these clues in a case they are handling should take special care in evaluating the DNA evidence, and particularly the associated statistics. Expert assistance may well be required to make an adequate evaluation, and consideration should be given to challenging the admissibility of problematic testimony as well as attacking it at trial. Although DNA evidence has long been regarded as the gold standard of forensic science, lawyers should not assume that the tests were performed or interpreted properly in every case.

VI. Admissibility Challenges And Regulatory Reform

In 2009 the National Research Council issued a scathing report about the state of forensic science in the United States. It found that entire disciplines rest on deficient scientific foundations, that procedures routinely used for interpretation are lacking in rigor, that analysts take inadequate measures to avoid error and bias, and that they testify with unwarranted certainty. The report noted that the legal system has been “utterly ineffective” in dealing with these problems. It recommended the creation of a new federal regulatory agency — the National Institute of Forensic Science (NIFS) — to establish best practice standards and oversee the field.

The problems with forensic DNA statistics discussed in this article provide an excellent example of why NIFS, or some similar body, is needed. The legal

system has not been effective, thus far, in dealing with these problems. DNA analysts are increasingly presenting problematic DNA evidence of the kind discussed here and are characterizing it in a variety of different ways, which are often clearly wrong and misleading. The existing regulatory structure, which consists of accrediting agencies like ASCLD-LAB and advisory bodies like SWGDAM, has not addressed these problems in an adequate manner. In an ideal system, a body like NIFS would step in and provide authoritative guidance about how (and when) analysts should present evidence in problematic cases, and what statistics should be used to characterize such evidence. But efforts to create a national regulatory body like NIFS have foundered in Congress, due, in large part, to opposition from enforcement groups, particularly prosecutors. It seems that they prefer the status quo in which labs can present problematic forensic evidence in ways that exaggerate its value to a regulatory regime with the power to address such problems.

In the absence of effective self-regulation by forensic scientists, the only check on misleading forensic science testimony is provided by criminal defense lawyers. Their task is extremely challenging. Judges and jurors perceive DNA evidence to be virtually infallible — the gold standard of forensic science. At this point in history it will not be easy to persuade judges even to hear challenges to the admissibility of DNA evidence, let alone exclude it. But legal grounds for successful challenge surely exist in some cases, and a great deal is riding on the ability of defense lawyers to bring such challenges successfully where they are warranted. Successful challenges not only protect the rights of accused individuals, they may well persuade forensic scientists to be more cautious about their interpretations in other cases; they might even persuade policy-makers to rethink the need for effective forensic science regulation.²³ It is vitally important that defense lawyers advocate effectively on these issues.

Notes

1. NATIONAL RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE 74-75 (1992).
2. <http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines>.
3. *United States v. Yee*, 134 F.R.D. 161, 181 (N.D. Ohio 1991), *aff’d. sub nom. United States v. Bonds*, 12 F.3d 540 (1993).
4. See *People v. Venegas*, 18 Cal.4th 47 (1998) (FBI DNA test inadmissible due to failure of lab to follow correct statistical

procedures when computing match probabilities); W. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons From the 'DNA War'*, 84 J. CRIM. L. & CRIMINOLOGY 22, 84-86 (1993).

5. See generally J. BUTLER, *FORENSIC DNA TYPING: BIOLOGY, TECHNOLOGY AND GENETICS OF STR MARKERS* (2005); W.C. THOMPSON, S. FORD, T. DOOM, M. RAYMER & D. KRANE, *Evaluating Forensic DNA Evidence: Essential Elements of a Competent Defense Review: Part 1*, 27(3) THE CHAMPION 16-25 (2003); W.C. THOMPSON, S. FORD, T. DOOM, M. RAYMER & D. KRANE, *Evaluating Forensic DNA Evidence: Essential Elements of a Competent Defense Review: Part 2*, 27(4) THE CHAMPION 24-28 (2003).

6. In some instances the labs also introduce a correction factor called "theta" into the computation to account for subgroup variation in population frequencies, but this correction has a relative minor effect on the resulting numbers. For the sake of simplicity, we will ignore it here.

7. B. BUDOWLE ET AL., *Low Copy Number — Consideration and Caution*, Proc. 12th International Symposium on Human Identification (2001) (<http://www.promega.com/~/media/files/resources/conference%20proceedings/ishi%202012/oral%20presentations/budowle.pdf?la=en>).

8. For purposes of illustration, we present here the results for only three of the nine loci examined by the laboratory in this case. There were discrepancies between the defendant's profile and the evidentiary profile at some of the other loci as well, but all of the discrepancies were attributed to the unreliability of the STR test for typing a sample with the limited quantity of DNA found in the evidentiary sample. The short peaks that appear in the defendant's reference sample — for example, peaks "7" and "11" at locus D5S818 — are examples of the phenomenon known as "stutter."

9. Itiel E. Dror & Greg Hampikian, *Subjectivity and Bias in Forensic DNA Mixture Interpretation*, 51 SCIENCE AND JUSTICE 204-208 (2011).

10. *Id.*

11. *Id.*

12. W.C. THOMPSON, *Painting the Target Around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation*, 8 LAW, PROBABILITY AND RISK 257-276 (2009); available at <http://lpr.oxfordjournals.org/cgi/reprint/mgp013?ijk=ey=ehrDeU74Gi6ERv1&keytype=ref>.

13. *Id.*

14. J.M. CURRAN & J. BUCKLETON, *Inclusion Probabilities and Dropout*, 55 J. FORENSIC SCI. 1171-73 (2010).

15. *Id.* at 1171.

16. P. GILL, C.H. BRENNER, J.S. BUCKLETON, ET AL., *DNA Commission of the International Society of Forensic Genetics: Recommendations*

for the Interpretation of Mixtures

160 FORENSIC SCI. INT'L 90-101 (2006). The International Commission noted that many laboratories have carried out experiments that provide information relevant to estimating drop out probabilities. In general, the probability of drop out of one allele in a genotype increases as the peak height of the other allele decreases — in other words, allelic drop out is more common when peak heights are low due to limited quantities of DNA. Empirical studies may allow determination of a threshold peak height above which drop out is unlikely to occur but estimation of drop out probabilities when the observed peaks are below this threshold inevitably requires a considerable element of guesswork, which means the ultimate statistics will always depend, in part, on the subjective judgment of experts on questions about which experts are likely to differ.

17. To compute the likelihood ratio, the system simply divides its overall estimate of the probability the contributor had the defendant's profile by the probability a random person would have this profile. The likelihood ratio thus corresponds to the degree to which the DNA analysis increases certainty that the suspect was a contributor.

18. J.M. CURRAN, *A MCMC Method for Resolving Two Person Mixtures*, 48 SCIENCE & JUSTICE 168-177 (2008).

19. See M.W. PERLIN & B. SZABADY, *Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples*, 46 J. FORENSIC SCI. 1372-1378 (2001); M.W. PERLIN & A. SINELNIKOV, *An Information Gap in DNA Evidence Interpretation*, 4 PLoS ONE e8327 (2009); M.W. PERLIN, J.B. KADANE & R.W. COTTON, *Match Likelihood Ratio for Uncertain Genotypes*, 8(3) LAW, PROBABILITY AND RISK 289-302 (2009); M.W. PERLIN, *Simple Reporting of Complex DNA Evidence: Automated Computer Interpretation*, Promega's Fourteenth International Symposium on Human Identification (2003); M.W. PERLIN, M.M. LEGLER, C.E. SPENCER, J.L. SMITH, W.P. ALLAN, J.L. BELROSE & B.W. DUCEMAN, *Validating TrueAllele® DNA Mixture Interpretation*, 56 J. FORENSIC SCI. 1430-1447 (2011).

20. *Regina v. Duffy and Shivers*, 2012 NICC (available at [http://www.courtsni.gov.uk/en-GB/Judicial%20Decisions/PublishedByYear/Documents/2012/\[2012\]20NICC%201/j_j_HAR8394Final2.htm](http://www.courtsni.gov.uk/en-GB/Judicial%20Decisions/PublishedByYear/Documents/2012/[2012]20NICC%201/j_j_HAR8394Final2.htm)).

21. Laurence Mueller and Dan Krane, who are co-authors of this article, served as experts for the defendant.

22. *Regina v. Duffy and Shivers*, *supra*.

23. See W.C. THOMPSON, *The National Research Council's Plan to Strengthen Forensic Science: Does the Path Forward Run Through the Courts?* 50 JURIMETRICS JOURNAL 35-51 (2009) (arguing that policy-makers

are unlikely to take seriously the need for forensic science reform unless, or until, courts begin excluding problematic forensic science evidence). ■

About the Authors

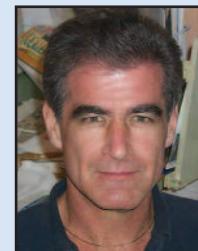
William C. Thompson is a Professor in the Department of Criminology, Law & Society and at the School of Law, University of California, Irvine. He is a former co-chair of NACDL's Forensic Evidence Committee.



Professor William C. Thompson

Department of Criminology,
Law & Society
School of Social Ecology
University of California, Irvine
Irvine, CA 92697
949-824-6156
E-MAIL william.thompson@uci.edu

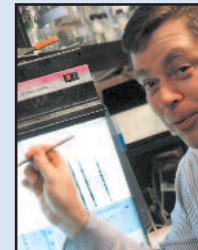
Laurence D. Mueller is a Professor and Chair of the Department of Ecology and Evolutionary Biology at the University of California, Irvine.



Professor Laurence D. Mueller

Department of Ecology
& Evolutionary Biology
University of California, Irvine
Irvine, CA 92697
949-824-4744
E-MAIL ldmuelle@uci.edu

Dan E. Krane is a Professor in the Department of Biology at Wright State University in Dayton, Ohio.



Professor Dan E. Krane

Wright State University
3640 Colonel Glenn Hwy.
Dayton, OH 45435
937-775-3333
E-MAIL dan.krane@wright.edu