

Suspect-centric Bias in DNA Mixture Interpretation

Mark W. Perlin, Ph.D., M.D.

Bias abounds in criminal justice. Predictive policing can bake bias into software, reflecting and reinforcing prior beliefs. Bail-risk computer programs may entrench pre-trial detention disparity. Human judgment pervades the process. Prosecutor and defender alike passionately argue their client's case, drawing opposite conclusions from identical facts.

Science is above the fray. Objective data suggest forensic match between crime scene and suspect. Statistical data analysis yields incontrovertible numbers for the strength of match. Cold DNA facts are presented as confirmed theories in court.

But what if DNA analysts could pick and choose their data? Or adjust software parameters to suit their theories? Changing data and parameters will alter forensic match results. Quantitatively, subjective manipulation can artificially inflate match strength. Qualitatively, some DNA evidence that excludes a suspect may be statistically twisted to include him.

Suspect-centric bias has long plagued forensic science. The mythic infallibility of fingerprint analysis was shattered when the FBI misidentified Brandon Mayfield in the Madrid bombing case. Confirmation bias just puts a number to a foregone match conclusion. Suspect-centric thought twists forensic facts to suit prosecution theories.

DNA evidence is not immune to suspect-centric bias. Most DNA evidence is a mixture of two or more people. Popular mixture protocols have crime laboratories first decide whether a suspect's DNA is in the evidence, and then pick and choose DNA data—all before calculating a match statistic. The FBI has abandoned their debunked Combined Probability of Inclusion (CPI) mixture statistic.¹ But the suspect-centric practice of “decide first, calculate later” continues, with experts selecting DNA data.

Fortunately, new “probabilistic genotyping” methods have been developed. Such software can objectively unmix DNA mixture data into component genotypes to deliver reliable match statistics. But should software let users pick their data? Or dial in chosen parameters? What if a software program lacks the math to use all the data, forcing users to make suspect-centric choices? This is not the unbiased science we expect, nor does it provide the impartial justice we require.

Here is a cautionary tale of bias in criminal justice,² of suspect-centric criminal investigation, prosecution and forensic science, and of how modern DNA software can be manipulated to falsely implicate an innocent man.

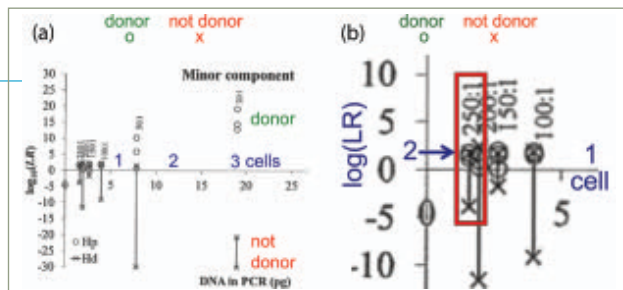


Figure 1. Validation match statistics for minor components. The x-axis gives the amount of DNA, represented in picograms (pg), number of cells (blue) and mixture ratios (major:minor). The y-axis gives the base 10 logarithm of the likelihood ratio (LR). The scatter plot shows true donors (Hp circles) and false non-donors (Hd crosses). (a) A zoomed out view to 25 pg. (b) A zoomed in view to 5 pg. One cell has 6 pg of DNA.

Death and DNA

Twelve-year old Garrett Phillips died from strangulation in his Potsdam, New York home on Oct. 24, 2011. Natural suspects were people close to the boy's widowed mother, including a local sheriff's deputy.

However, suspicion quickly centered on her ex-boyfriend Oral “Nick” Hillary, a soccer coach at Clarkson University. Jamaican-born Hillary was an obvious outsider in an otherwise close-knit upstate community.

The New York State Police (NYSP) crime laboratory in Albany developed 150 DNA samples from 35 evidence



items. Suspect-centric mixture interpretation failed to connect Hillary to the crime. The lab usually runs one short tandem repeat (STR) experiment per DNA item, but in this case they ran 26 Identifiler® tests on the boy's left hand fingernails. Unable to link the fingernails to Hillary via CPI, in 2013 the lab sent the mixture data to Cybergenetics for free TrueAllele® computer analysis.

The Pittsburgh company's DNA mixture system separates STR data into contributor genotypes (i.e., probabilistic profiles), and then compares them with suspect or evidence genotypes to determine match statistics. The process is objective—the mixture separation uses all the data and never sees a suspect, eliminating suspect-centric confirmation bias. The NYSP trusted the system, having published three peer-reviewed validation studies and conducted extensive internal validations.

The computer found a small foreign DNA fraction on the fingernails, statistically unconnected to Hillary. A year later, objective computer analysis of new NYSP data from the degraded DNA mixture again showed Hillary wasn't there. Ultimately, pro bono processing of all 150 samples would show no connection between Hillary and the crime scene. Evidence-centric science found no link to the soccer coach.

Politics and prosecution

St. Lawrence County District Attorney Nicole Duve respected the DNA science, and did not charge Hillary. But the 2013 election featured a new DA candidate—former police officer and public defender Mary Rain. Running on a “Justice for Garrett” campaign, Rain defeated Duve with a mandate to pursue the case. Rain had Hillary arrested and jailed. Her indictment was later dismissed for lack of evidence.

Rain assembled her DNA prosecution team. She partnered with a prominent New York forensic prosecutor. Dissatisfied with the exculpatory American computer mixture analysis, they reached half-way around the world to a New Zealand company, recruiting a DNA expert who had developed his own mixture software.

This foreign software could perform probabilistic genotyping, analyzing mixtures to calculate a match statistic for a suspect. But its mathematical model

was incomplete. Laboratory calibration helped fill in parameters it couldn't derive from evidence. Human software operators could pick and choose their data, dialing in parameters the computer couldn't get from data. The process was susceptible to confirmation bias.

On Dec. 7, 2015, the new DA had her New Zealand answer. The foreign expert reported a match statistic of 10 million, connecting suspect Hillary to the fingernails. With DNA evidence in hand, the case against Hillary could now proceed.

Subjective computer analysis

Calibration

Modern Bayesian computing derives parameter probabilities directly from the data. However, limited DNA modeling may lack that math capability, and instead substitute historical data for case evidence. Crime labs usually develop calibration data to tune the foreign analysis software. However, no lab-specific calibration was done in the Hillary case. The foreign expert had to pick and choose calibration parameters in order to run his software on the NYSP lab data.

Validation

No lab-specific software validation was available. Instead, the prosecution presented studies from another crime lab and the software manufacturer. But these results only demonstrated why their software was unreliable on the DNA evidence in this case.

The prosecution expert reported a 1:220, or 0.45 percent, foreign unknown fraction in the deceased's fingernail mixture. This minuscule DNA portion amounted to less than half a human cell.

The defense expressed concern about lack of validation at extreme mixture ratios. The manufacturer then reassessed software performance down to 1:250 (Figure 1a). A threshold of 30 RFU was used.

Zooming in shows the software cannot distinguish between true (circle) or false (cross) inclusions at very low DNA levels (Figure 1b). For 1:250 to 1:100 amounts (arrow), the software's match statistics hovered around a hundred (two powers of 10) *whether or not* a suspect contributed his DNA to the mixture. Moreover, at these low levels the validation did not yield higher match numbers like the 10 million (seven powers of 10) reported in this case.

RFU	Data choices	LR	All stutters
80		0	9
70		30	51
60		250	1,660
50	15,500,000		69,200
40		0	0
30		0	0

Table 1. Match statistic LR values on the same data are shown (blue) for computer runs using different threshold (rows) and stutter (columns) data assumptions.

Case data

STR peaks have heights on a scale from zero to 10,000 relative fluorescence units (RFU). Taller peaks reflect more allelic DNA; shorter peaks less. Relevant peaks for the foreign 1:220 fraction were in the low 30 to 70 RFU range. Objective analysis requires considering all of those data peaks.

Robust Bayesian modeling of baseline, stutter and other artifacts can use all the data. Without that machine intelligence, DNA software users must apply a threshold, discarding low-level data below a fixed peak height. For the foreign software, this level was usually set and studied at 30 RFU. Indeed, the company had objected when others used a higher level of 50 RFU in independent studies.

The prosecution expert chose a threshold of 50 RFU, picking only taller foreign data as software input. Potentially exculpatory peaks between 30 and 50 RFU were ignored. This human data choice impacted the software's calculated statistics.

Independent analysis

The defense scientists ran the foreign software, showing how a chosen threshold affected match statistics (Table 1). The six rows list different peak height thresholds: lower RFU considers more data, higher RFU less data. The two columns give the reported expert data choices, with and without discarding potential stutter peaks.

The prosecution presented only the 50 RFU threshold results that gave high match statistics. But at the software's usual 30 RFU threshold, using the same prosecution operator parameters, Hillary wasn't there. And at 40 RFU, the New Zealand software again excluded him from the fingernail DNA. Exceeding 50 RFU, the match statistic faded away. Only at the chosen 50 RFU sweet spot was there an apparent connection to the defendant.

Different human choices of input data gave diametrically opposed answers. The software contradicted itself. It both included and excluded Hillary.

Locus notes

At the D8 STR locus, choosing a 50 RFU threshold retained three alleles: two victim and one foreign. But that choice ignored five other peaks pointing away from Hillary.

No defendant allele was present in the D16 locus data from three replicated lab tests. The suspect was excluded from all of that input data. Yet the software's statistical result still included him. The software didn't need inclusionary evidence to implicate Hillary.

Science on trial

Admissibility hearings were held in 2016. Judge Felix Catena, the presiding trial court judge, found the prosecution software to be reliable. But the defense further asked whether this software had been *reliably applied* to the half-cell fingernail data.

On cross-examination, the DNA defenders reviewed the software validation studies, as applied to the case data. They highlighted human choices of STR data and software parameters. When asked if he had "curiosity about what the [match] value would be ... at 30 RFU," the prosecution expert replied, "Yes, indeed, I am curious. In fact, I'd like to go to 10." But it was left for the defense team to discover that the software actually excluded Hillary at the usual data threshold of 30 RFU.

Advanced methods contain enough mathematics to fully examine DNA evidence data and eliminate human bias. Others require laboratory calibration, and user selection of data and parameters to compensate for



missing math. With math-deficient models, internal validation is needed to guide human operators on how to reliably pick their data and choose parameters.

On Aug. 26, 2016, Catena precluded the foreign expert and his software results. “No internal validation studies were performed,” wrote the judge. “As a result, the expert was forced to *pick and choose data ... and input parameters* into the program.” The prosecution’s DNA evidence would not be heard at trial.

Concerned about potential racial bias with an all-white jury, Hillary chose a non-jury bench trial. Lacking physical evidence, the prosecution’s case collapsed. On Sept. 28, Catena found Hillary not guilty.³ A civil rights claim is pending.

Suspect-centric bias issues continue. Some news from June of 2018:

- Texas Judge David Wahlberg dismissed the foreign software’s DNA test results in an Austin homicide case. His unfavorable Daubert decision said a forensic scientist had used flawed techniques when analyzing DNA data with the software.
- The New York Appellate Court suspended former District Attorney Mary Rain from practicing law for two years. Her misconduct as district attorney included attempting to withhold potentially exculpatory evidence in the Hillary case.
- The NYSP crime lab was again accused of suspect-centric bias. After initially reporting inconclusive DNA, a revised match report implicated a suspect. Choosing some loci, and omitting others, had produced an inclusionary CPI statistic.

The path forward

Objective scientists do not “pick and choose” their data. Published behavioral mixture studies show that suspect-centric bias can lead DNA experts to give conflicting answers. When cued with a hopeful hypothesis—like “we know it’s him”—forensic analysts can succumb to bias, picking data that falsely implicates the innocent. Even scientific experts are prone to the confirmation bias of normal human reasoning.

Forensic scientists have an ethical duty to not “materially misrepresent data or scientific principles” underlying their conclusions, according to

the American Academy of Forensic Sciences. All choices affecting an outcome should be disclosed, along with conflicting results. When software permits human choices of data, parameters or hypotheses, the DNA analyst should calculate and give results for alternative scenarios.

Lawyers have their own ethical obligations. Prosecutors must ensure full disclosure of data, methods and assumptions for DNA match results. Such potentially exculpatory Brady evidence “is material to either guilt or to punishment.” A vigilant defender will request all electronic DNA data files, laboratory reports and computer results. DNA defense experts need prosecution software to replicate or challenge match statistics; such access was essential in Hillary’s defense.

Some crime labs already embrace forensic software that objectively analyzes evidence to accurately connect or eliminate a suspect. Other labs select data to suit their theory, a capital mistake in light of potential suspect-centric bias. As artificial intelligence permeates society, judges and juries will demand impartial objectivity from forensic science. In time, all DNA experts will report results that suit all the facts. ●

References

1. M.W. Perlin, “When DNA is not a gold standard: failing to interpret mixture evidence,” *The Champion*, 42(4):50-56, May, 2018. <https://www.nacdl.org/champion.aspx>
2. M.W. Perlin, W.P. Allan and C. Hughes, “On the threshold of injustice: manipulating DNA evidence,” *American Academy of Forensic Sciences 69th Annual Meeting*, New Orleans, LA, February 17, 2017. On-line presentation at: <https://www.cybgen.com/information/presentations/2017/AAFS/Perlin-Hughes-On-the-threshold-of-injustice-manipulating-DNA-evidence/page.shtml>
3. J. McKinley, “Oral Nicholas Hillary acquitted in Potsdam boy’s killing,” *The New York Times*, September 28, 2016. <https://www.nytimes.com/2016/09/29/nyregion/oral-nicholas-hillary-potsdam-murder-trial-garrett-phillips.html>

Dr. Mark Perlin is chief scientist and executive at *Cybergenetics*, a Pittsburgh-based bioinformatics company that provides patented TrueAllele® automation for accurate and objective interpretation of complex DNA evidence.