

D.H. Kaye, Commentary, "False, But Highly Persuasive": How Wrong Were the Probability Estimates in *McDaniel v. Brown*?, 108 Mich. L. Rev. First Impressions 1 (2009), <http://www.michiganlawreview.org/firstimpressions/vol108/kaye.pdf>.

## APPENDIX

### A. URMP Calculations

Romero testified that she arrived at 1/3,000,000 using the “interim ceiling principle” described in a 1992 Report of a committee the National Research Council. The committee itself did not agree that it was necessary to be this conservative in estimating random-match probabilities (RMPs), and a subsequent committee soon recommended alternatives that produce much smaller estimates for RMPs. Because neither Romero’s report nor her testimony specified the VNTR allele lengths, however, it is impossible to use the materials before the Supreme Court to state just how much smaller a reasonable estimate would be. The Scholars’ Brief gives the allele designations and their frequencies among Caucasians. Table 1 uses these data with the population-genetics model recommended in the 1996 NRC report (p. 102) to obtain an RMP of less than 1/181,000,000 (for a co-ancestry coefficient  $\theta = 0.03$ ) or 1/155,000,000 (without the population-structure correction).

Table 1. Estimated Five-locus VNTR Frequencies in Caucasians

VNTR locus	Allele		$2p_1p_2$	$2p_1p_2(1-\theta)$	
	length	p		$\theta = 0.01$	$\theta = 0.03$
D2S44	3297	0.089	0.0251	0.0248	0.0243
	1640	0.141			
D4S139	14577	0.124	0.0308	0.0304	0.0298
	10101	0.124			
D10S28	4386	0.065	0.0012	0.0012	0.0011
	653	0.009			
D14S13	1846	0.166	0.0693	0.0687	0.0673
	1517	0.209			
D17S79	1757	0.186	0.102672	0.1017	0.0996
	1525	0.276			
Five-locus genotype frequency			$6.433 \times 10^{-9}$	$6.118 \times 10^{-9}$	$5.53 \times 10^{-9}$
1 per			155,400,000	163,500,000	181,000,000

The frequency at all 11 loci (the five VNTR loci plus the six PCR-based DQ $\alpha$  and

Polymarker loci) is smaller. If the VNTRs are statistically independent of the genes (which is likely to be approximately true), a rough estimate is  $1/3,000,000 \times 1/4,800 = 1/14,400,000,000 \approx 1/15,000,000,000$ . This calculation uses the artificially high 1/3,000,000 ceiling VNTR RMP and the largest PCR-based RMP given by the Nevada laboratory for the major population groups. Geneticists have been willing to testify to uniqueness based on matches for the two genetic systems combined.

## B. SRMP Calculations

The sibling random match probability SRMP at a single locus can be found from equations 4.9a and 4.9b on page 113 of the 1996 NRC Report. Table 2 uses these formulas together with the VNTR allele frequencies for Caucasians given in Table 1 to estimate the SRMP for the five VNTR loci. The same calculations can be found in the Scholars' Brief.

Table 2. Estimated Sibling Random-Match Probabilities in Caucasians (VNTRs)

VNTR locus	Allele length	p	$(1+p_1+p_2+2 p_1p_2)/4$
D2S44	3297	0.089	0.314
	1640	0.141	
D4S139	14577	0.124	0.320
	10101	0.124	
D10S28	4386	0.065	0.269
	653	0.009	
D14S13	1846	0.166	0.361
	1517	0.209	
D17S79	1757	0.186	0.391
	1525	0.276	
Five-locus SRMP			0.00381 (1/263)

Table 3 uses the 1996 NRC formulas together with allele frequencies for the PCR-based genetic systems reported in Bruce Budowle et al., Validation and Population Studies of the Loci LDLR, GYPA, HBGG, D7S8, and Gc (PM Loci), and HLA-DQ $\alpha$  Using a Multiplex Amplification and Typing Procedure, 40 Am. J. Hum. Genetics 45, 50, 52 (1995), to estimate the SRMP for these loci.

Table 3. Estimated Sibling Random-Match Probabilities in Caucasians (DQ $\alpha$  and PM Loci)

Locus	Allele		(1+2p+p <sup>2</sup> )/4 (if homozygous)
	Name	p	(1+p <sub>1</sub> +p <sub>2</sub> +2 p <sub>1</sub> p <sub>2</sub> )/4 (if heterozygous)
DQ $\alpha$	3	0.216	0.538
	3	0.216	
LDLR	A	0.453	0.624
	B	0.547	
GYPA	A	0.584	0.725
	A	0.584	
HBGG	B	0.524	0.361
	B	0.524	
D7S8	A	0.615	0.618
	B	0.385	
Gc	A	0.257	0.555
	A	0.257	
Six-locus SRMP			0.0575 (1/17)

If the VNTR loci are independent of the gene markers, the 11-locus sibling random-match probability is the product of the overall SRMP from each table—about 0.000219, or 1/4563.

### C. Transposition of the Conditional

The text asserted that the prosecution's equation

$$P(Troy | Match) = 1 - P(Match | Unrelated) \quad (A.1)$$

is true if only if

$$P(Match | Unrelated) = P(Unrelated | Match). \quad (A.2)$$

This follows directly from the axioms of probability and propositional calculus. To begin with,

$$P(M \wedge S) = P(M), \quad (A.3)$$

where  $S$  is the disjunction of all possible states of nature and  $M$  is *Match*. I posited that the only two possible states of nature were that the true source is Troy ( $T$ ) and that the true source is someone else unrelated to him:

$$S = T \vee U. \quad (A.4)$$

Hence,

$$P[(T \vee U) \wedge M] = P(M) \quad (A.5)$$

$$P[(T \wedge M) \vee (U \wedge M)] = P(M) \quad (A.6)$$

$$P(T \wedge M) + P(U \wedge M) = P(M) \quad (\text{A.7})$$

$$P(T \wedge M) / P(M) + P(U \wedge M) / P(M) = 1 \quad (\text{A.8})$$

$$P(T | M) + P(U | M) = 1 \quad (\text{A.9})$$

$$P(T | M) = 1 - P(U | M) \quad (\text{A.10})$$

Substituting (A.10) into (A.1) yields (A.2). Reasoning in the opposite direction establishes the asserted “if and only if” relationship.

The text also stated that transposition can favor the defendant rather than the prosecution. The remainder of this appendix describes the conditions under which the dubiously named “prosecutor’s fallacy” favors the defendant. This situation occurs when the probability of a match given that the suspect is not the source,  $P(M | \sim S)$ , that is mistakenly transposed exceeds the probability  $P(\sim S | M)$  computed according to Bayes’ rule:

$$P(M | \sim S) > P(M | \sim S) = P(\sim S) P(M | \sim S) [P(\sim S) P(M | \sim S) + P(S) P(M | S)]^{-1} \quad (\text{A.11})$$

$$P(\sim S) P(M | \sim S) + P(S) P(M | S) > P(\sim S) \quad (\text{A.12})$$

$$P(M | \sim S) + P(M | S) [P(S) / P(\sim S)] > 1 \quad (\text{A.13})$$

$$\text{Odds}(S) > [1 - P(M | \sim S)] / P(M | S). \quad (\text{A.14})$$

In the absence of laboratory error,  $P(M | S) = 1$  and  $P(M | \sim S) = p$ , where  $p$  is the random match probability. Hence, the condition (A.11) is satisfied when the prior odds exceed  $1 - p$ .

$$\text{Odds}(S) > 1 - p. \quad (\text{A.15})$$

Thus, for  $p \ll 1$  and prior odds  $> 1$ , as in *Brown*, the transposed probability is too large (favoring the defendant) when compared to the correctly computed posterior probability.