B29 Mathematics Takes a Holiday — The Scientific Working Group on DNA Analysis Methods' (SWGDAM's) Imaginative Y Haplotype Guidelines

Charles H. Brenner, PhD*, 6801 Thornhill Drive, Oakland, CA 94611-1336

After attending this presentation, attendees will understand that the formulas recommended by SWGDAM for Y haplotype evidence are not based on mathematics or on population genetic models, but rather represent confused analogies and cookbook statistics.[1]

This presentation will impact the forensic science community by raising the level of scepticism about trusting expert or committee authority rather than insisting on logic and data.

Mathematical treatment of forensic DNA evidence has some weak spots, the worst being Y haplotypes. SWGDAM's recommendations for computation rest on several confused ideas about mathematics and genetics.

Start with genetics. SWGDAM offers the claim, "It is recognized that population substructure exists for Y-STR haplotypes" to justify an "Eq 3," $Pr(A \mid A) = \theta + (1 - \theta) \, p_A$, adapted from a similar formula in autosomal practice. In the autosomal world θ comes into play to cater to the idea of population substructure from preferential mating — sexual reproduction. But Y haplotypes reproduce clonally, not sexually. There cannot be Y haplotype "population substructure," not in anything like the sense of the autosomal forerunner to Eq 3. The conclusion is that Eq 3 comes about not from clear thinking, but by a careless false analogy that confuses two different senses of "population substructure" — one based on sexual mating, the other on geography or on tribal identity.

What does the guideline mean by θ? The guidelines are vague on that, but reverse-engineering from tables of numbers in an appendix in the guideline reveals that θ is quite close to the average matching chance of two randomly selected haplotypes. The left-hand side, $Pr(A \mid A)$, of Eq 3 must also be on average close to the average matching chance. It would follow that p_A, which purports to be a probability, is negative about half the time!

The guidelines avoid confronting the embarrassment of negative probabilities by bad mathematics: though p_A is meant to be a probability — which is an inference from *data* ("the degree of expectation of its occurrence, which we are warranted in entertaining by our present *evidence*" as the 19th-century thinker JS Mill put it) — the guidelines insist on confusing it with population frequency and on estimating that frequency by a clumsy cookbook statistical procedure which is intended — both misguidedly and unreliably — to be generous.[1]

Hence, the entire process is founded on confusion. Nobody can possibly understand it because there is nothing there to understand. But the above-average analyst, who mistakenly trusts the expertise of the committee, is likely to imagine that the recommendations are fair, useful, and meaningful. The analyst thus gulled testifies in court, unfairly in a Native American context, that the involvement of θ means that the formula takes into account tribal structure. In another case, the pointless generosity means a big break to an offender for no principled reason. Confusion is awkward and expensive — and careless reasoning is bad precedent.

Reference(s):
1. SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories, accessed August 9, 2017, http://media.wix.com/ugd/4344b0_da25419ba2dd4363bc4e5e8fe7025882.pdf.

Y Haplotypes, SWGDAM Guideline, Forensic Mathematics