International Research Journal of Computer Science (IRJCS)
Issue 08, Volume 6 (September 2019)

ISSN: 2393-9842
www.irjcs.com

DNA MIXTURE INTERPRETATION: EFFECT OF THE HYPOTHESIS ON THE LIKELIHOOD RATIO

Heather Miller Coyle, Ph.D.*
Forensic Science Department, Henry C. Lee College of Criminal Justice & Forensic Sciences,
University of New Haven, 300 Boston Post Road, West Haven, CT 06516 USA
Hcoyle@newhaven.edu

Manuscript History
Number: **IRJCS/RS/Vol.06/Issue08/SPCS10081**
Received: 02, September 2019
Final Correction: 12, September 2019
Final Accepted: 20, September 2019
Published: **September 2019**
Citation: Heather (2019). DNA Mixture Interpretation: Effect of the Hypothesis on the Likelihood Ratio. IRJCS:: International Research Journal of Computer Science, Volume VI, 672-675. **doi://10.26562/IRJCS.2019.SPCS10081**
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

Abstract: Although nuclear forensic DNA tests are standard practice in most forensic science laboratories, complex DNA mixture analysis remains a challenge. Although new to many laboratories, the concept of probabilistic genotyping has been presented for over a decade as a tool to aid in mixture analysis. Probabilistic genotyping can be defined as a mathematical approach using the likelihood ratio (LR) to estimate if an individual is likely to be included or excluded in a DNA mixture based on statistical inference. Mathematical modelling of biological data has been shown to be less biased than using analyst discretion in determining an inclusion of a DNA donor to a complex mixture. Still, there are caveats to using probabilistic genotyping software that become evident when applied to forensic casework. The effect of allele sharing and the uncertainty of the number of contributors to the likelihood ratio hypothesis are discussed.

Keywords: forensic science; likelihood ratio (LR); probabilistic genotyping (PG); DNA mixtures; allele sharing;

I. INTRODUCTION

The likelihood ratio is a mathematical ratio between two potential hypotheses: the Prosecutor's Hypothesis (Hp) includes the defendant as a potential contributor to the mixture; the Defense Hypothesis (Hd) hypothesizes that another unknown person is the contributor to the DNA mixture. The resulting numeric ratio (LR) has a probability to include the defendant if a positive value; the defendant is excluded if the LR value is a negative. Although probabilistic genotyping software is a calculator that compares the two hypotheses as the either/or scenario, there does remain a small range of false positive values equivalent to an inconclusive or non-informative conclusion [1 - 4]. False negatives are also possible but not addressed here as that result would be conservatively reported as an exclusion. The point of the likelihood ratio is that it gives a quantitative value rather than a qualitative statement that allows one to express the degree of certainty in establishing the probability that the evidentiary DNA originated from the alleged reference contributor. The likelihood ratio, however, is only as good as the hypothesis being tested and the relative numeric range of values needs to be expressed for meaningful comparison.

Likelihood ratios provide a numerical measure of the effect of a result on probability. When there is a comparison between the evidence and the known reference sample, it is necessary to quantify the evidentiary value of the match. In cases where genetic relatives are present, it is not always possible to distinguish between individuals due to allele sharing (e.g. parent-child, 50%; full-sibling, 50%; half-sibling, 25%' cousin, 12.5%). In addition, some unrelated individuals in a given population by sheer random chance will share a percentage of alleles (i.e. coincidental match). The percentage of shared alleles will vary randomly on a case by case basis.

There is a formula, ($p^2 + 2pq + q^2 = 1$) the Hardy Weinberg Equilibrium (HWE) principle formula, used to represent allele frequencies in a population and to calculate match probability statistics in forensic DNA analysis. The letter "p" is used to designate the allele frequency of the first numeric allele value (allele 1); the letter "q" is used to designate the allele frequency of the second numeric allele value (allele 2). If the alleles are the same at a genetic locus (test position on the chromosome), the genetics indicate homozygosity and the part of the formula to be used is p^2. If the alleles are different at a genetic locus, the genetics indicate heterozygosity and the part of the formula to be used is 2pq. The HWE formula is the underlying science for establishing the likelihood ratio used in probabilistic genotyping software. The software calculates Hp/Hd as a ratio where:

- A Hp (LR) = 1 or greater if the hypothesis assumes 100% probability that the defendant is present in the DNA mixture.
- A Hd (LR) = 0 or less favors the hypothesis that it is another individual in the population in the DNA mixture.

Example:

The D13S317 locus with alleles 11 (p) and 14 (q) where p = 0.3394 and q = 0.04801 for allele frequencies in a given population. The LR = Hp/Hd = 1/2pq = 1/2 x (0.3394)(0.04801) = 1/0.03259 = 30.7; meaning it is 30.7 times more likely that the defendant left the DNA at the scene than an unknown unrelated individual at this locus. This is a simplified calculation using a single locus but this type of calculation is made for each genetic locus and the values multiplied together for the overall estimated LR.

For complicated DNA mixtures, probabilistic genotyping is a convenient tool for performing these calculations [5]. If the D13S317 locus contains more than two alleles, it is a DNA mixture. For example, the alleles 14 (p), 15 (q), 16 (r) would be calculated for all possible pairwise combinations and the respective allele frequencies summed for the Hd denominator value.

II. MATERIALS AND METHODS

A review of criminal casework using probabilistic genotyping methods was made from multiple states. A few case examples were selected to illustrate some of the issues with LR calculations with (a) allele sharing and (b) establishing the number of contributors to minor components of a mixture in relationship to Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines for the validation of probabilistic genotyping systems.

III. RESULTS

The LR values will vary slightly based on a few factors using probabilistic genotyping software for DNA mixture analysis including: initial seed value, number of contributors (based on differences in analytical thresholds and allele counts per locus), choice of reference population database (racial groups), theta correction factor (inbreeding coefficient), probability of contamination or drop-in frequency, allele drop-out calculations and choice of hypothesis [6 – 10]. Examples of the last factor, choice of hypothesis, are described below.

Case Example #1 examines the effect of allele sharing in random, unrelated individuals. A swab was collected near a sink area that was in debate as to whether or not it contained the victim and defendant's DNA after a homicide. This was a two person mixture sample with 25pg of DNA and processed as a low copy number (LCN) sample in triplicate. The sample was genotyped for fifteen loci with forty percent allele sharing between the victim and the defendant. Multiple possible hypotheses were tested using Forensic Statistical Tool (FST) probabilistic genotyping software, a semi-continuous system. The report conclusions were as follows:

Possibility 1: It is 12.5 times more probable to have originated from two unknown, unrelated individuals than the defendant and one unknown, unrelated individual.

Possibility 2: It is 575 times more probable if the sample originated from the victim and one unknown, unrelated individual than if from victim and defendant.

Possibility 3: It is 71.2 million times more probable if the sample originated from the victim and one unknown, unrelated individual than from two unknown, unrelated individuals.

In the first two possibilities, the defendant is clearly excluded as a source of the DNA. However, conditioning the Defense hypothesis on the presence or absence of the victim substantially changes the LR value. By unrestricting the Defense Hypothesis (Hd) and allowing any two individuals to be part of the mixture (Possibility 3), the LR value greatly increases the confidence in the Prosecutor's Hypothesis (Hp) being correct.

This case example is used to illustrate the importance of testing a variety of hypotheses for Hp and Hd to find the best fit to the biological model for the DNA mixture.

In this case, the best fit of the data is the victim and one unknown individual in the DNA mixture at the sink compared to any other two unknown, unrelated individuals in a population. The effect of forty percent random allele sharing between the defendant and the victim is one explanation for the extreme reduction in the LR value for Possibility 2.

Coincidental matching due to allele sharing has been investigated in other probabilistic genotyping software systems (e.g. STRmix, a fully continuous system) in validation studies [3]. In one reported study, the highest false inclusion had a calculated LR value of 187, 504 for a two person mixture where a non-contributor coincidentally matched to a DNA mixture. This non-contributor shared six coincidental obligate minor alleles within the profile. The second highest LR value was 84, 283 and shared twenty-five alleles by coincidence with the true contributor to the mixture. Although a relatively rare event, it is cautionary as it is possible to be falsely included in a DNA mixture with a compelling likelihood ratio due to coincidental allele sharing. Many forensic laboratory validation studies perform non-contributor studies but few examine purposeful combinations of samples with varying percentages of allele sharing to evaluate the effect on the LR calculation as recommended (SWGDAM guideline 4.1.6.5) [11]. The SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems state "The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses. For example, if there are two persons of interest, they may be evaluated as co-contributors and, alternatively, as each contributing with an unknown individual. The hypotheses used for evaluation of casework profiles can have a significant impact on the results obtained (SWGDAM guideline 4.1.2.1)."

Case Example #2 examines the effect of uncertainty in partial contributors to the minor component of the mixture and the difficulty in formulating a hypothesis as to true number of contributors. In one case, three swabs were collected from a firearm and genotyped with standard forensic DNA testing. The swab of the "trigger/trigger guard" and the swab of the "backstrap/grips/frontstrap" each had insufficient DNA for further testing. However, the swab of the "slide grooves/release" generated a DNA mixture of at least two individuals, including one major donor. The forensic report stated "the DNA profile(s) of the minor contributor(s) to the mixture(s) could not be determined; however the results were suitable for comparison." In this case scenario, a LR calculation should be made for hypotheses with multiple numbers of contributors in the numerator and denominator since it is unclear how many individuals are present in the minor contributor fraction of the mixture; possibly one, two, three or more donors as the report states. SWGDAM guideline 4.1.6.4 states "if the number of contributors is input by the analyst, both correct and incorrect values (i.e., over- and under-estimating) should be tested" [11].

The Victoria Police Forensic Services Department (VPFSD) validation study for STRmix software with the Profiler Plus human identification kit addressed the issue of varying numbers of contributors to a DNA mixture [12]. The conclusion was the effect would vary and be dependent on the selected donor percentage contribution to the mixture and the number of contributor hypothesis. For example, in one DNA mixture a hypothesis of three contributors yielded an LR of 10; the same DNA mixture with a hypothesis of two donors yielded a likelihood ratio of 1.42, an approximate 7 fold difference in values. In addition, the Erie County Forensic Laboratory STRmix Validation for the Fusion kit reports that as the DNA concentration in the sample decreases, so does the LR calculation. This is intuitively correct as more alleles "drop-out" of the mixture, the statistical confidence decreases [13].

IV. CONCLUSIONS

Probabilistic genotyping software is beneficial in many circumstances for aiding in the interpretation of complex DNA mixtures. However, if the original hypothesis is incorrect, the validity of the LR calculation can be called into question. It would be best practice to establish if there is significant allele sharing between the victim and defendant as the LR calculation can change significantly based on the inclusion of the victim as a contributor as demonstrated in Case Example #1. In addition, establishing the true number of contributors to a complex DNA mixture is not always straight forward as profiles may be partial, with high levels of allele "drop-out" and evidence of DNA contributors below the analytical threshold indicating additional donors even if the alleles may not be reported by the forensic laboratory due to policy reasons. A change in the number of contributors, however, can have an effect on the statistical calculation based on the hypothesis as demonstrated in Case Example #2. The probabilistic genotyping software system is only as good as the original hypothesis selected by the software user and careful consideration of the hypothesis testing and all alternate theories is sound scientific practice for a likelihood ratio calculation to establish statistical probabilities.

ACKNOWLEDGMENT

Thank you to attorney Richard Potter for providing the review information for Case Example #1 (*New Jersey v. Daniel Rochat*) and attorney Jesse Hoberman-Kelly for Case Example #2 (*People of the State of New York v. Markey Linen*).

REFERENCES

1. K.B. Watters and H. Miller Coyle, "Forensic Statistical Tool (FST): a probabilistic genotyping software program for human identification,"Jurimetrics.,56(2), pp.183-195, 2016.
2. S. Noël, J. Noël, D. Granger, J. Lefebvre and D. Séguin, "STRmix(™) put to the test: 300 000 non-contributor profiles compared to four-contributor DNA mixtures and the impact of replicates," Forensic Sci. Int. Genet., vol. 41, pp. 24-31, 2019.
3. M. Coble and J. Bright, "Probabilistic genotyping software: An overview," Forensic Sci. Int. Genet., vol. 38, pp. 219-224, 2019.
4. J. Buckleton, J. Bright, S. Gittelson, T. Moretti, A. Onorato, F. Bieber, B. Budowle and D. Taylor, "The probabilistic genotyping software STRmix: utility and evidence for its validity," J. Forensic Sci., vol. 64, pp. 393-405, 2019.
5. M. Kruijver, J. Bright, H. Kelly and J. Buckleton, "Exploring the probative value of mixed DNA profiles," Forensic Sci. Int. Genet., vol. 41, pp. 1-10, 2019.
6. T. Moretti, R. Just, S. Kehl, L. Willis, J. Buckleton, J. Bright, D. Taylor and A. Onorato, "Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles," Forensic Sci. Int. Genet., vol. 29, pp. 126-144, 2017.
7. J. Bright, D. Taylor, C. McGovern, S. Cooper, L. Russell, D. Abarno and J. Buckleton, "Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles," Forensic Sci. Int. Genet., vol. 23, pp. 226-239, 2016.
8. H. Haned, P. Gill, K. Lohmueller, K. Inman and N. Rudin, "Validation of probabilistic genotyping software for use in forensic DNA casework: definitions and illustrations," Sci. Justice., vol. 56, pp. 104-108, 2016.
9. M. Perlin, J. Hornyak, G. Sugimoto and K. Miller, "TrueAllele(®) genotype identification on DNA mixtures containing up to five unknown contributors," J. Forensic Sci., vol. 60, pp. 857-68, 2015.
10. M. Perlin, J. Belrose and B. Duceman, "New York State TrueAllele ® casework validation study," J. Forensic Sci., vol. 58, pp. 1458-1466, 2013.
11. Scientific Working Group on DNA Analysis Methods (SWGDAM). (2015) Guidelines for the validation of probabilistic genotyping systems. [Online]. Available: https://www.swgdam.org/publications
12. L. Federie, A. Pollett, "Validation of STRmix v. 1.05 using Profiler Plus, Victoria Police Forensic Services Department (VPFSD) study, Victoria, Australia, pp. 1-36, 2013.
13. Erie County Central Police Services Forensic Laboratory. (2015) STRmix Implementation and Internal Validation. [Online]. Available: https://johnbuckleton.files.wordpress.com

