

Probabilistic Genotyping: The Use of the Forensic Statistical Tool (FST)

Craig O' Connor, PhD.
Criminalist III/ Assistant Technical Leader, Nuclear DNA Operations
Department of Forensic Biology
NYC Office of Chief Medical Examiner
coconnor@ocme.nyc.gov



Forensic Statistical Tool

Why Likelihood Ratio (LR)?

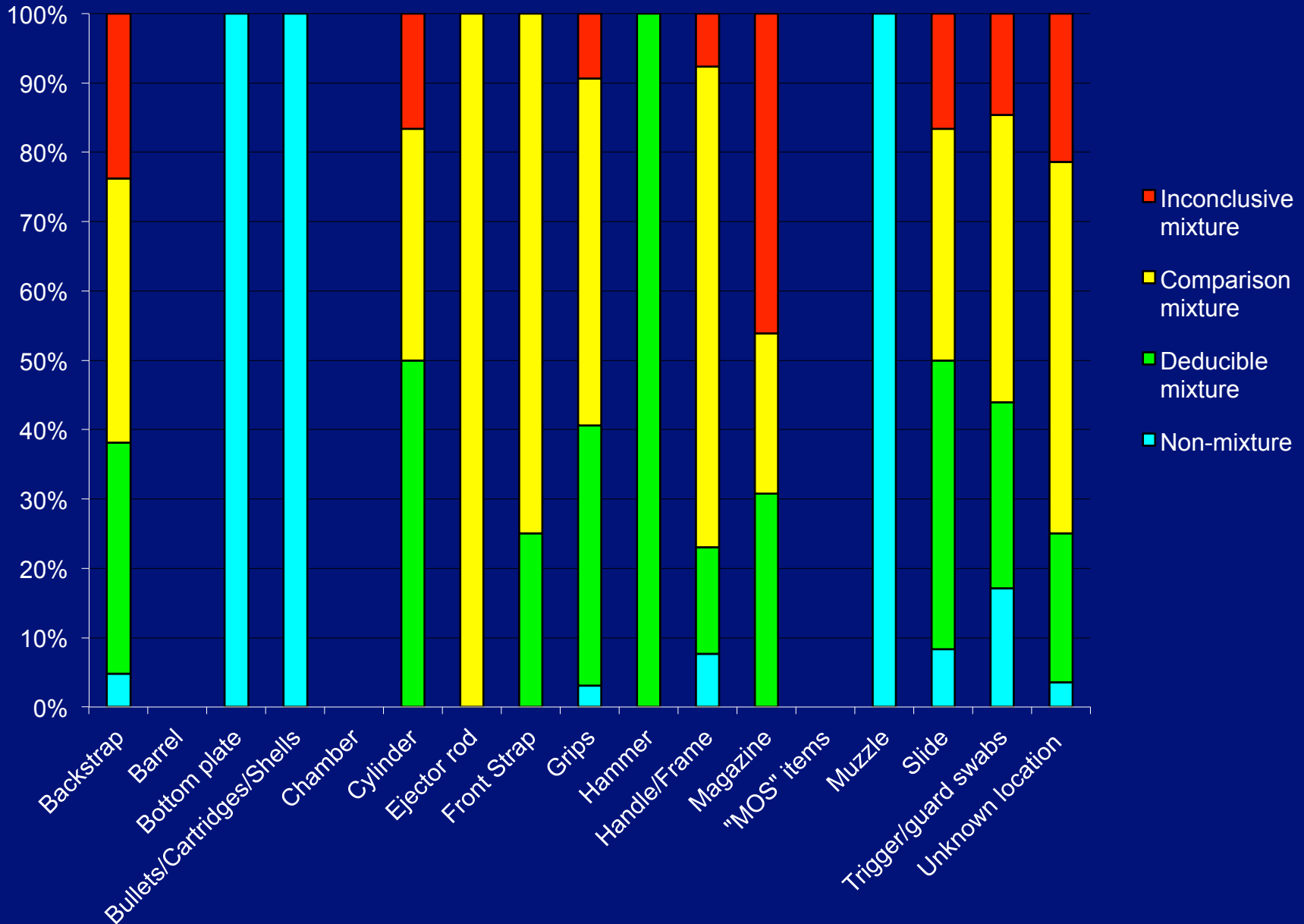
- National Academy of Science Report (*2009) / SWGDAM guidelines for interpretation of DNA mixtures 4.1 (2011)
 - Statistic should accompany all positive associations instead of qualitative conclusions
 - Can use RMNE (CPI) or LR
- The DNA commission of the ISFG recommends using the LR
 - More available data is utilized
 - Can incorporate DO/DI rates
- **Statistics for complex and Low Template DNA mixtures**

* Strengthening Forensic Science in the United States: A Path Forward (2009) The National Academies Press, Washington, DC



Forensic Statistical Tool

Types of Profiles Generated



LRs are Ubiquitous

Pub Med Searches

- Over 5,500 published articles on LRs in medicine and/or genetics
- **243** articles with Likelihood ratio directly in their title



Forensic Applications of LR already in use

- Random match probability (RMP) is a LR
 - RMP: 1 in a trillion
 - LR: sample is one trillion times more likely if suspect is the source than if a random person is the source
- LR is used for kinship calculations
 - DNA View
 - Software used for WTC identifications
- CODIS Popstats software (FBI) performs mixture analysis using LR



TWO COMPETING SCENARIOS IN FORENSICS

The suspect is a contributor to a mixture

OR

The suspect is not a contributor, and an
unknown person is the contributor



Forensic Statistical Tool

How did you learn about the software?

Comparisons to other LR software

- *TrueAllele*
 - Perlin et al., J Forensic Sci, 2001, 46:1372-1378; Perlin et al., PLoS One, 2009, 4:e8327; Perlin et al., J Forensic Sci, 2011, 56:1430-1447
- *LoComatioN*
 - Curran et al. Forensic Science International, 2005, 148:47-53; Gill et al. Forensic Science International, 2007, 166:128-138
- *Forensim*
 - Haned, Forensic Science International Genetics, 2011, 5:265-268; Haned et al., Forensic Science International: Genetics, 2012, 6:762-774
- *LikeLTD*
 - Balding et al., Forensic Science International: Genetics, 2009, 4:1-10
- *Lab Retriever*
 - Lohmueller & Rudin, J Forensic Sci, 2013, 58: S243-249



Forensic Statistical Tool (FST)

- Similar LR framework as other programs
- “Semi-continuous” approach to LR calculations which incorporates drop-out and drop-in
- Differences in methods used to estimate probability of drop-out and drop-in
 - Other programs:
 - User-specified rates
 - Probabilistic models
 - Empirical estimates using peak heights
 - FST uses empirical estimates based on the quantity of DNA in the sample – *in house quant assay*



FST uses in casework

- LR statistic applied primarily to 2- and 3- person mixtures
 - Comparison sample is positively associated
 - Deconvoluted mixtures where the major/minor donor is not consistent with your suspect
 - Non-deconvoluted mixtures
- Can be used on older cases if needed
- All interpreting analysts trained



Challenges in implementing FST

- Program FST with empirically derived drop-out/
drop-in rates
- Training/testimony
 - Staff
 - Police detectives and Attorneys
 - Judges and juries



Challenges in implementing FST

- Challenge in incorporating drop-out and drop-in rates
 - Dynamic vs static? Simulated vs empirical
- OCME developed and validated FST using empirical drop-out and drop-in rate estimates
 - Drop-out
 - Locus-specific
 - Homozygote or heterozygote genotype
 - Depends on input DNA and approximate mixture ratio
 - Drop-in
 - Most drop-in occurred in -4 stutter position of a true allele
 - Not dependent on locus



Development of FST

- Modeled after LoComatioN
- Can be used with:
 - Single source samples and mixtures
 - High and low template samples
- Simultaneously consider data from one, two or three amplifications of evidence sample
- Not an expert system
- Approved for use with criminal casework samples by the New York State Forensic Science Commission, December 2010



Forensic Statistical Tool

FST Reliably Assigns a Quantitative Value to a Comparison of Forensic Samples

- 400+ samples tested over a range of template amounts and mixture ratios
 - Purposeful mixtures from blood and buccal swabs
 - Touched items
 - 2 and 3 person mixtures
- Many samples were purposefully or naturally degraded
- 500,000+ comparisons with non-contributors performed



Forensic Statistical Tool

Validation Conclusions

- LR for true contributors support qualitative assessments
- In some cases, LR is more conservative than qualitative assessment
- LRs showed a good separation between true and non-contributors* .

* Due to allele sharing, for some mixtures, chance positive associations were noted, but FST assigned an appropriate weight.



Forensic Statistical Tool

Example 1: Single Source LR with drop-out / drop-in

- Single source sample
 - S_p : Suspect; S_d : Unknown, unrelated person
- The LR is constructed with two additional factors considered
 - “If this person is a contributor to the mixture, did any of their alleles drop out?”
 - “Are there any alleles that are not explained by this set of contributors (i.e., drop-in)?”



Example 1:

Single Source LR with drop-out / drop-in

- Drop-out and drop-in terms included in numerator and denominator
- Drop-out of suspect's or unknown person's allele(s) may have occurred

	Locus 1	Locus 2	Locus 3	Locus 4
Evidence	14, 15	29	8	11, 12
Suspect	14, 15	29	8, 12	11, 12



Example 1: Single Source LR with drop-out / drop-in

- When drop-out is not modeled and profiles match
 - Numerator is 1.0
 - Denominator is RMP
- When drop-out is modeled
 - Numerator < 1.0
 - Unknown contributor's allele(s) may have dropped out



Example 1 – Locus 3 only

- **Numerator:**

- Suspect is 8, 12; evidence is 8
- $\Pr(\text{Data} | S_p) = \Pr(\text{one drop-out from heterozygote}) \times \Pr(\text{no drop-in})$



Example 1 – Locus 3 only

- **Denominator:**

- Unknown contributor may have genotype:
 - 8, 8
 - 8, w (w is any allele other than 8)
 - w, w
- Find expected population frequency of each genotype
- Multiply by probability of drop-out and drop-in required to obtain evidence profile

$$\begin{aligned} & p_8^2 \quad \times \quad \text{Pr (no homozygote drop-out)} \times \text{Pr (no drop-in)} \\ + & 2p_8p_w \quad \times \quad \text{Pr (one heterozygote drop-out)} \times \text{Pr (no drop-in)} \\ + & p_w^2 \quad \times \quad \text{Pr (homozygote drop-out)} \times \text{Pr (one drop-in)} \end{aligned}$$



LR with Drop-out / Drop-in

DO and DI rates counted and programmed in FST

- D_0 = no drop-out (heterozygote): $1 - D_1 - D_2$
- D_1 = one drop-out: **counted**
- D_2 = two drop-outs: **counted**

- D_{H0} = no drop-out (homozygote): $1 - D_{H1}$
- D_{H1} = drop-out: **counted**

- C_0 = no drop-in (per locus): $1 - C_1 - C_{2+}$
- C_1 = one drop-in allele: **counted**
- C_{2+} = two or more drop-in alleles: **counted**



Forensic Statistical Tool

Example 1

$$LR = \frac{D_1 C_0}{p_8^2 D_{H0} C_0 + 2p_8 p_w D_1 C_0 + p_w^2 D_{H1} C_1}$$



Example 2

- Two-person mixture
- S_p : Suspect and one unknown, unrelated person
- S_d : Two unknown, unrelated people
- Consider a single locus with three labeled alleles
- All possible genotypes for the three unknown contributors (one in numerator, two in denominator) must be generated



Example 2

- Evidence Sample: 11, 12, 14
- Suspect Profile: 11, 12
- Numerator includes one unknown person with possible genotypes:

11, 11	12, 12	14, 14	w, w
11, 12	12, 14	14, w	
11, 14	12, w		
11, w			



Example 2: Numerator

Evidence: 11, 12, 14; Suspect 11, 12

Drop-out? Drop-in?

Unknown Contributor Genotype	Drop-out Required? (Type?)	Drop-in Required? (Which allele(s)?)
11, 11		
11, 12		
11, 14		
11, w*		
12, 12		
12, 14		
12, w*		
14, 14		
14, w*		
w*, w*		

*w is any allele other than 11, 12, 14

Example 2: Numerator

Evidence: 11, 12, 14; Suspect 11, 12

Drop-out? Drop-in?

Unknown Contributor Genotype	Drop-out Required? (Type?)	Drop-in Required? (Which allele(s)?)
11, 11	No	Yes (14)
11, 12	No	Yes (14)
11, 14	No	No
11, w*	Yes (partial heterozygous)	Yes (14)
12, 12	No	Yes (14)
12, 14	No	No
12, w*	Yes (partial heterozygous)	Yes (14)
14, 14	No	No
14, w*	Yes (partial heterozygous)	No
w*, w*	Yes (total homozygous)	Yes (14)

*w is any allele other than 11, 12, 14

Example 2: Numerator

Evidence: 11, 12, 14; Suspect 11, 12

Unknown's genotype frequency

Unknown Contributor	Frequency	Drop-out?	Code	Drop-in?	Code
11, 11		No		Yes (14)	
11, 12		No		Yes (14)	
11, 14		No		No	
11, w*		Yes		Yes (14)	
12, 12		No		Yes (14)	
12, 14		No		No	
12, w*		Yes		Yes (14)	
14, 14		No		No	
14, w*		Yes		No	
w*, w*		Yes		Yes (14)	

Example 2: Numerator

Evidence: 11, 12, 14; Suspect 11, 12

Unknown's genotype frequency

Unknown Contributor	Frequency	Drop-out?	Code	Drop-in?	Code
11, 11	P_{11}^2	No	D_{H0}	Yes (14)	C_1
11, 12	$2P_{11}P_{12}$	No	D_0	Yes (14)	C_1
11, 14	$2P_{11}P_{14}$	No	D_0	No	C_0
11, w^*	$2P_{11}P_w$	Yes	D_1	Yes (14)	C_1
12, 12	P_{12}^2	No	D_{H0}	Yes (14)	C_1
12, 14	$2P_{12}P_{14}$	No	D_0	No	C_0
12, w^*	$2P_{12}P_w$	Yes	D_1	Yes (14)	C_1
14, 14	P_{14}^2	No	D_{H0}	No	C_0
14, w^*	$2P_{14}P_w$	Yes	D_1	No	C_0
w^*, w^*	P_w^2	Yes	D_{H1}	Yes (14)	C_1

Example 2: Numerator

Evidence: 11, 12, 14; Suspect 11, 12

Unknown Contributor	Frequency	Drop-out?	Code	Drop-in?	Code
11, 11	P_{11}^2	No	D_{H0}	Yes (14)	C_1
11, 12	$2P_{11}P_{12}$	No	D_0	Yes (14)	C_1
11, 14	$2P_{11}P_{14}$	No	D_0	No	C_0
11, w	$2P_{11}P_w$	Yes	D_1	Yes (14)	C_1
etc...					

$$\begin{aligned}
 & p_{11}^2 \times D_{H0} \times C_1 + 2p_{11}p_{12} \times D_0 \times C_1 \\
 & + 2p_{11}p_{14} \times D_0 \times C_0 + 2p_{11}p_w \times D_1 \times C_1 \\
 & \text{etc...}
 \end{aligned}$$

Example 2: Denominator

Evidence: 11, 12, 14

U1	Frequency	Drop-out	U2	Frequency	Drop-out	Drop-in
11, 11	P_{11}^2	No (D_{H0})	11, 11	P_{11}^2	No (D_{H0})	Yes (C_{2+})
			11, 12	$2P_{11}P_{12}$	No (D_0)	Yes (C_1)
			etc...			
			w, w	P_w^2	Yes (D_{H1})	Yes (C_{2+})
11, 12	$2P_{11}P_{12}$	No (D_0)	11, 11	P_{11}^2	No (D_{H0})	Yes (C_1)
			11, 12	$2P_{11}P_{12}$	No (D_0)	Yes (C_1)
			etc...			
			w, w	P_w^2	Yes (D_{H1})	Yes (C_1)
etc...						

Example 2: Denominator

Evidence: 11, 12, 14

U1	Frequency	Drop-out	U2	Frequency	Drop-out	Drop-in
11,11	P_{11}^2	No (D_{H0})	11, 11	P_{11}^2	No (D_{H0})	Yes (C_{2+})
			11, 12	$2P_{11}P_{12}$	No (D_0)	Yes (C_1)
			etc...			
			w, w	P_w^2	Yes (D_{H1})	Yes (C_{2+})

$$\begin{aligned}
 & p_{11}^2 \times D_{H0} \left(p_{11}^2 \times D_{H0} \times C_{2+} + 2p_{11}p_{12} \times D_0 \times C_1 \right. \\
 & \left. + \dots + p_w^2 \times D_{H1} \times C_{2+} \right) \\
 & + 2p_{11}p_{12} \times D_0 \left(p_{11}^2 \times \dots \text{etc} \dots \right)
 \end{aligned}$$

Identifiler

Forensic Statistic Comparison Report

v2.5

FB#1: FB13-xxxx

FB#2: FBS13-xxxx

Item: 3P Pen C

Comparison: D5

DNA Template Amount (pg):
340

Input By: CSClaamitchell

Hp: D5 (Comparison) + Unknown

Hd: 2 Unknowns

Deducible: Yes

Profiles

Profile	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
D5 (Comparison)	13,14	29,32.2	8,10	9,13	15,16	6,9	8,13	11,11	20,25	12,14	16,16	8,8	12,16	11,11	21,23
Evidence															
1	13, 14, 15	29, 30.2	10, 11	11	14, 16, 17	6, 7, 9	11, 12	11, 13	20, 25	12, 13, 14, 15	17	8	12, 15	10, 11, 12	22, 23
2	13, 14, 15	29, 30.2, 31.2, 32.2	10	11	14, 15, 16, 17	6, 7, 9	8, 11, 12, 13	11, 13	20, 25	12, 13, 14, 15	16, 17	8	12, 15, 16	10, 11, 12	22
3															

Comparison Result

	Asian	Black	Caucasian	Hispanic
Likelihood Ratio	502.71	778.78	9.85	50.85

Report the lowest value



Forensic Statistical Tool

Thanks and Appreciations

- NYC Office of Chief Medical Examiner Forensic Biology Department
- National Institute of Standards and Technology



Forensic Statistical Tool