FORENSIC SCIENCE ERROR MANAGEMENT, INTERNATIONAL FORENSICS SYMPOSIUM, JULY 20-24, 2015 WASHINGTON DC

ERROR RATES IN PROBABILISTIC GENOTYPING SOFTWARE FOR DNA MIXTURES IN HUMAN IDENTIFICATION – HOW TO COMPARE?

HEATHER MILLER COYLE, PH.D.

ASSOCIATE PROFESSOR

UNIVERSITY OF NEW HAVEN

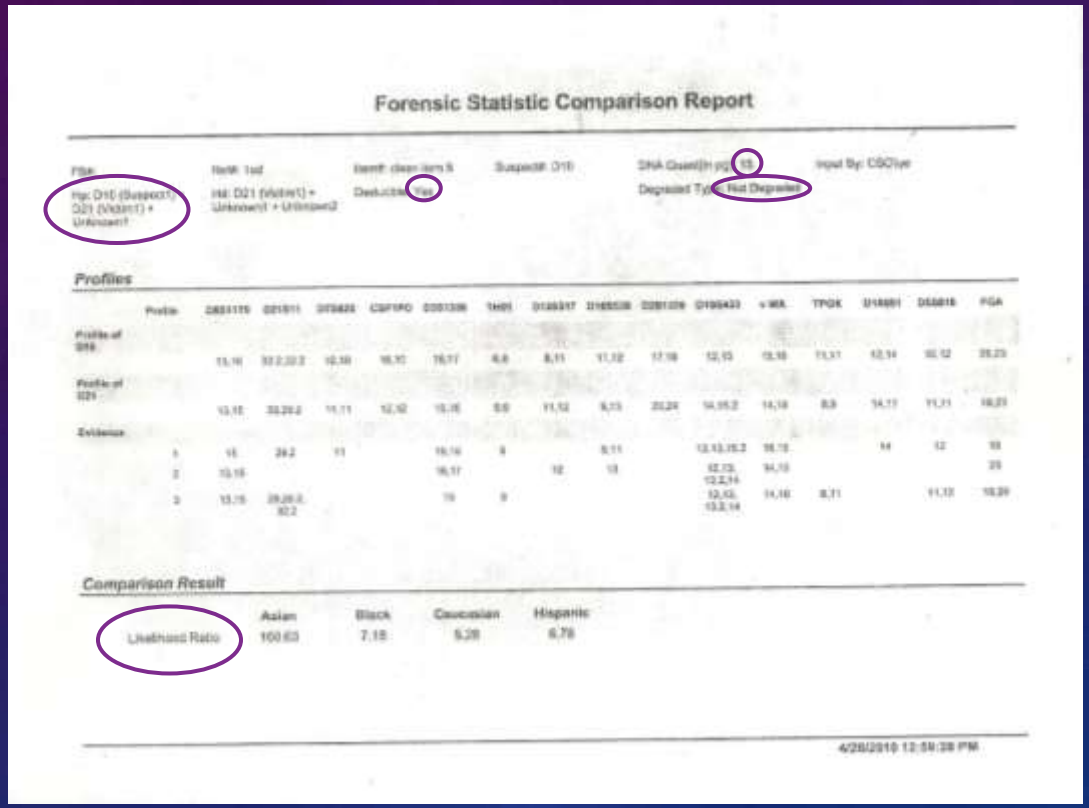
HCOYLE@NEWHAVEN.EDU

NEW YORK CITY FRYE HEARINGS FOR FST SOFTWARE

- FST – Forensic Statistical Tool
- Bronx – FST DOES meet Frye standards since it is based on data from standard PCR methodology by Justice Carruthers (*The People of the State of New York v. William Rodriguez, 2013*)
- Brooklyn – recent written decision by Justice Dwyer states FST DOES NOT yet meet Frye standards for general acceptance for numerous reasons including the manner in which the software was validated to assess drop-out values (*The People of the State of New York v. Andrew Peaks; The People of the State of New York v. Jaquan Collins, 2013*)
- Goal of probabilistic genotyping software (in general) – narrow the range of potential sources of DNA in mixtures by converting qualitative assessment by analyst to quantitative assessment by software for improved scientific accuracy in mixture analysis
- Emphasize use of likelihood ratios is not the Frye issue with FST; it is the software package (and use with LCN DNA in open populations) that is in question

ERROR RATES – HOW TO COMPARE?

- My goal - assess probabilistic genotyping FST software results and OCME DNA mixture validation studies to evaluate sources of experimental or computational error
- “Black Box”; software is not publically available for independent evaluation (ref. written decision, Justice Dwyer, 2015)
- FST published error rates are high (e.g. 3-person mixtures 1 per 1200 individuals in deducible mixtures) (ref. A. Mitchell et al. 2012. FSI: Genetics Supplement Series)
- TrueAllele (www.cybgen.com) published error rates are low (e.g. 1 per 20,000 individuals) (ref. M. Perlin et al. 2014. PLoS ONE)
- Why the difference?



Forensic Statistic Comparison Report

FSA:
hp: D10 (Suspect) + D21 (Victim1) + Unknown1

Item#: 1st
Hit: D21 (Victim1) + Unknown1 + Unknown2

Item#: clean item 5
Deduction: Fail

Suspect#: D10

DNA Quant(in pg): 18
Degraded Type: Not Degraded

Input By: CSCIye

Profiles

Profile	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	THO1	D13S317	D16S539	D2S1338	D19S433	v-WA	TPOX	D18S51	D5S818	FGA
Profile of D10	13,16	32.2,33.2	10,10	10,10	16,17	6,6	8,11	11,12	17,19	12,15	15,16	11,11	12,14	10,12	20,23
Profile of D21	13,15	30,29.2	11,11	12,12	15,16	9,9	11,12	9,13	20,24	14,15.2	14,16	8,8	14,17	11,11	19,21
Evidence															
1	15	29.2	11		16,19	9		9,11		12,13,15.2	16,18		14	12	19
2	13,15				16,17		12	13		12,13, 13.2,14	14,15				23
3	13,15	29,29.2, 32.2			19	9				12,13, 13.2,14	14,16	8,11		11,12	19,20

Comparison Result

	Asian	Black	Caucasian	Hispanic
Likelihood Ratio	160.63	7.18	5.28	6.78

4/28/2010 12:59:39 PM

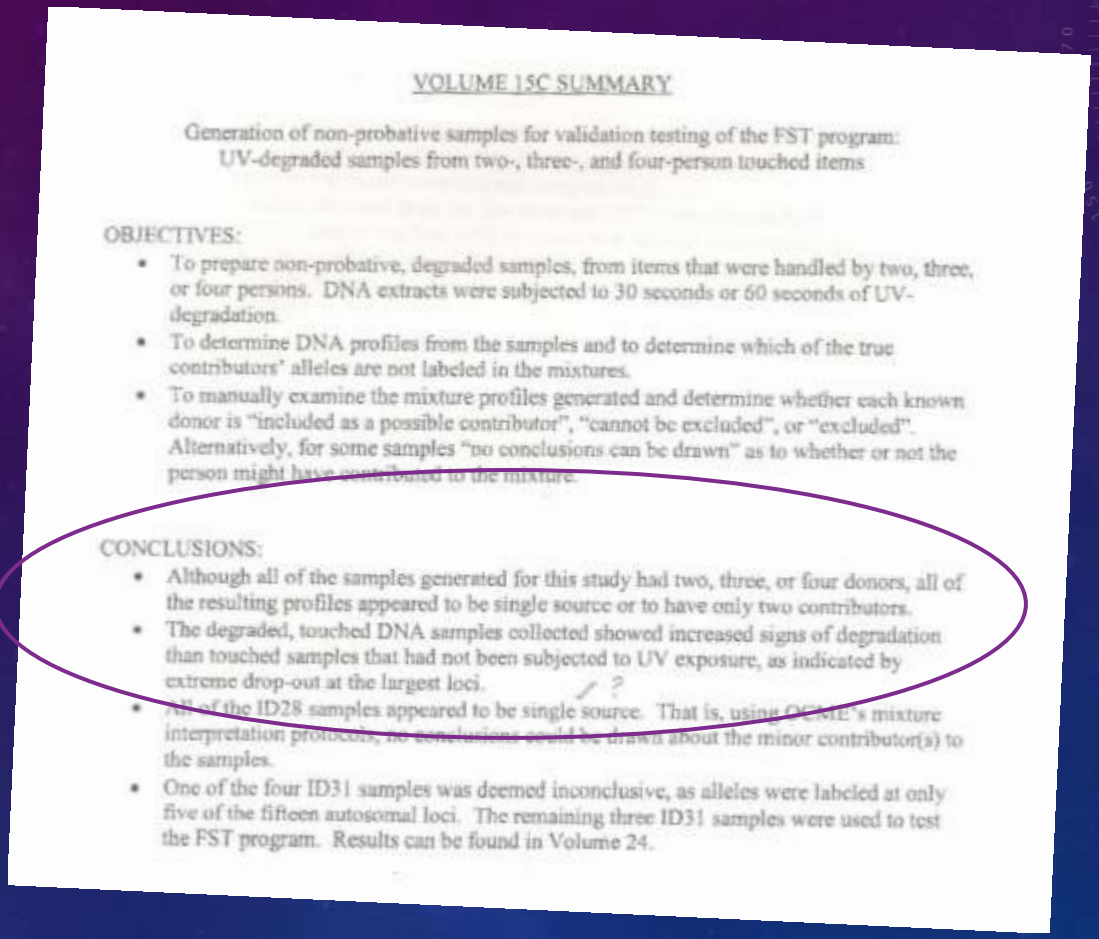
FST:

- Five assumptions that need to be correct:
  - Number of contributors
  - Degraded v. nondegraded
  - Deducible mixture v. nondeducible mixture (duplicate concordance)
  - Allele frequency database is appropriate
  - Quantity – drop out rates are linked to this value

FST:

- Software is claimed to be better since it uses empirically derived allelic drop-out rates but uses pristine DNA samples to derive the rates as well as the quantity value [ref. The Office of the Chief Medical Examiner of New York City (Department of Forensic Biology), “Forensic Statistical Tool Validation Summary Report, “ Volume 15C Summary]

VOL. 15C



VOLUME 15C SUMMARY

Generation of non-probative samples for validation testing of the FST program:
UV-degraded samples from two-, three-, and four-person touched items

OBJECTIVES:

- To prepare non-probative, degraded samples, from items that were handled by two, three, or four persons. DNA extracts were subjected to 30 seconds or 60 seconds of UV-degradation.
- To determine DNA profiles from the samples and to determine which of the true contributors' alleles are not labeled in the mixtures.
- To manually examine the mixture profiles generated and determine whether each known donor is "included as a possible contributor", "cannot be excluded", or "excluded". Alternatively, for some samples "no conclusions can be drawn" as to whether or not the person might have contributed to the mixture.

CONCLUSIONS:

- Although all of the samples generated for this study had two, three, or four donors, all of the resulting profiles appeared to be single source or to have only two contributors.
- The degraded, touched DNA samples collected showed increased signs of degradation than touched samples that had not been subjected to UV exposure, as indicated by extreme drop-out at the largest loci.
- All of the ID28 samples appeared to be single source. That is, using [illegible]'s mixture interpretation protocols, no conclusions could be drawn about the minor contributor(s) to the samples.
- One of the four ID31 samples was deemed inconclusive, as alleles were labeled at only five of the fifteen autosomal loci. The remaining three ID31 samples were used to test the FST program. Results can be found in Volume 24.

FST:

- Software requires a correct assumption of contributors to the DNA mixture which is frequently incorrect since it uses the allele counting method (ref. J Perez, AA Mitchell, N Ducasse, J Tamariz, T Caragine, “Estimating the Number of Contributors to Two-, Three-, and Four-Person Mixtures Containing DNA in High Template and Low Template Amounts”, Croat Med J, 52(3): 314-326, 2011.)

Error Rate Four Person Mixture Study	Error Rate Four Person Mixture Study
> 100 pg	50 – 100 pg
86% detection/ 14% error rate	76% detection/ 24% error rate

CONCORDANCE IN DUPLICATES

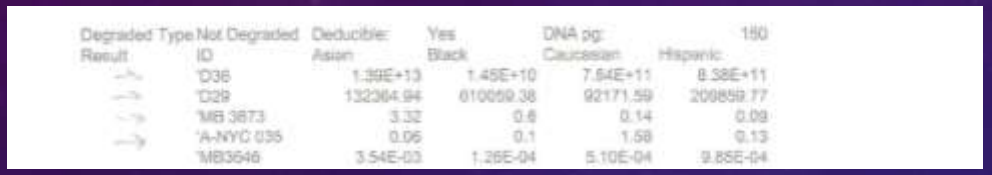
- PCR amplification efficiency (approximately 20 - 30% variance in peak heights tolerated and still considered from same source)
- New York State Inspector General's report (2013) – analyst debate between duplicates and a laboratory policy to include rather than exclude (bias)
- Contamination or addition of extraneous alleles – include or exclude in FST?

PCR Cycle No.	No. Contaminant Alleles
28	0
31	9

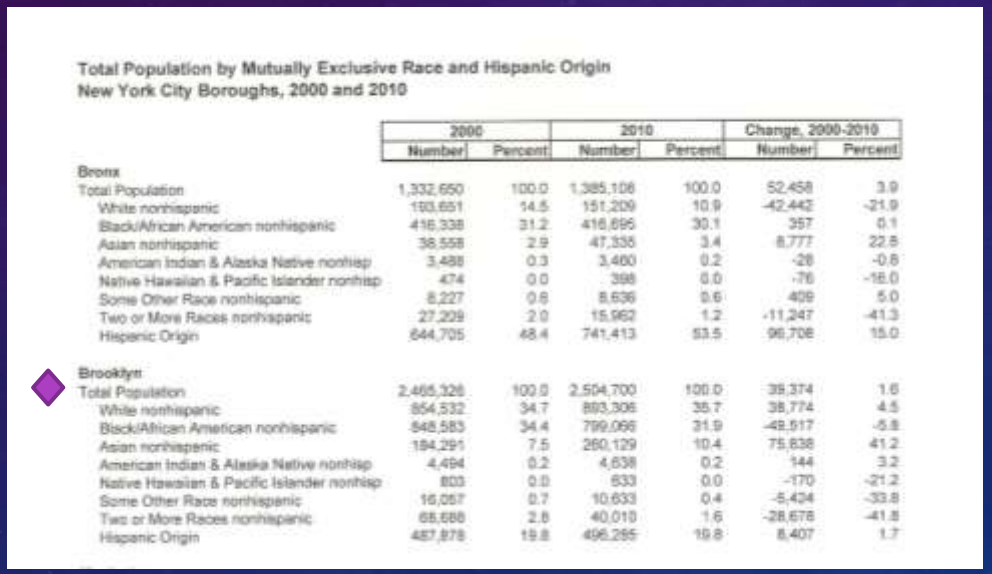
FST:

- High false positive associations in DNA mixtures to non-contributor DNA databases, sometimes with high likelihood ratio (LR) values (ref. A. A. Mitchell, J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz, T. Caragine, "Validation of a DNA Mixture Statistics Tool Incorporating Allelic Drop-Out and Drop-In, " Forensic Science International: Genetics, vol. 6(6), pp. 749-761, 2012.)
- It could be argued that the Bayes Factor (BF) with odds calculation would be an effective manner in which to accurately testify to a LR ratio for evidence without ignoring false positive rate
- Courtroom testimony is being monitored for accurate reporting of error rate on case by case basis

FALSE POSITIVE MATCHES TO NONCONTRIBUTORS AND INCREASE WITH ASSUMED NUMBER OF CONTRIBUTORS



Degraded Type Result	Not Degraded ID	Deducible: Asian	Yes Black	DNA pg: Caucasian	150 Hispanic
	'D3S	1.39E+13	1.45E+10	7.64E+11	8.38E+11
	'D2S	132364.94	610659.38	92171.59	209859.77
	'MB 3873	3.32	0.6	0.14	0.09
	'A-NYC 035	0.06	0.1	1.58	0.13
	'MB3646	3.54E-03	1.26E-04	5.10E-04	9.85E-04



Total Population by Mutually Exclusive Race and Hispanic Origin
New York City Boroughs, 2000 and 2010

	2000 Number	2000 Percent	2010 Number	2010 Percent	Change, 2000-2010 Number	Change, 2000-2010 Percent
Bronx						
Total Population	1,332,650	100.0	1,385,108	100.0	52,458	3.9
White nonhispanic	193,651	14.5	151,209	10.9	-42,442	-21.9
Black/African American nonhispanic	416,338	31.2	416,695	30.1	357	0.1
Asian nonhispanic	38,558	2.9	47,335	3.4	8,777	22.8
American Indian & Alaska Native nonhisp	3,488	0.3	3,460	0.2	-28	-0.8
Native Hawaiian & Pacific Islander nonhisp	474	0.0	398	0.0	-76	-16.0
Some Other Race nonhispanic	8,227	0.6	8,636	0.6	409	5.0
Two or More Races nonhispanic	27,209	2.0	15,962	1.2	-11,247	-41.3
Hispanic Origin	644,705	48.4	741,413	53.5	96,708	15.0
Brooklyn						
Total Population	2,465,326	100.0	2,504,700	100.0	39,374	1.6
White nonhispanic	854,532	34.7	893,306	35.7	38,774	4.5
Black/African American nonhispanic	848,583	34.4	799,066	31.9	-49,517	-5.8
Asian nonhispanic	184,291	7.5	260,129	10.4	75,838	41.2
American Indian & Alaska Native nonhisp	4,494	0.2	4,638	0.2	144	3.2
Native Hawaiian & Pacific Islander nonhisp	803	0.0	633	0.0	-170	-21.2
Some Other Race nonhispanic	16,057	0.7	10,633	0.4	-5,424	-33.8
Two or More Races nonhispanic	68,688	2.8	40,010	1.6	-28,678	-41.8
Hispanic Origin	487,878	19.8	496,285	19.8	8,407	1.7

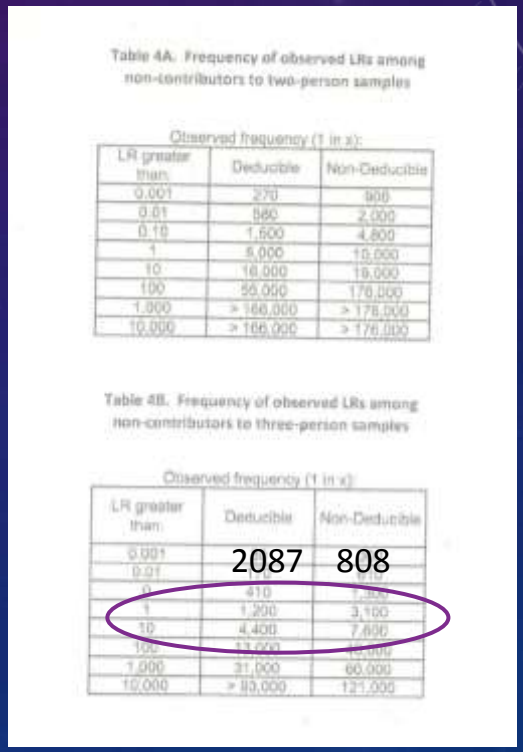


Table 4A. Frequency of observed LRs among non-contributors to two-person samples

Observed frequency (1 in x):

LR greater than	Deducible	Non-Deducible
0.001	270	900
0.01	860	2,000
0.10	1,500	4,800
1	5,000	10,000
10	16,000	19,000
100	55,000	176,000
1,000	> 166,000	> 176,000
10,000	> 166,000	> 176,000

Table 4B. Frequency of observed LRs among non-contributors to three-person samples

Observed frequency (1 in x):

LR greater than	Deducible	Non-Deducible
0.001	2087	808
0.01	[illegible]	[illegible]
0	410	[illegible]
1	1,200	3,100
10	4,400	7,600
100	13,000	40,000
1,000	31,000	60,000
10,000	> 93,000	121,000

FST - LACK OF GENERAL ACCEPTANCE (DWYER DECISION)

- Other forensic science laboratories do not routinely use the LCN process in the United States (drop-out rates are predicated on the validation study for LCN)
- Drop-out rates from pristine DNA studies make it difficult to establish number of contributors with accuracy since validation with UV treated DNA was not used for drop out rates
- Contamination rates with LCN are high (8-11%) so DNA in sample does not accurately reflect true contributors and not all true contributors are detected
- HID kits are not optimal at <100pg; stochastic effects and high stutter confound DNA mixture interpretation and there is nonconcordance between duplicates
- False positive rates for inclusion in DNA mixtures are exceptionally high when compared to other software with low error rate and need to be conveyed accurately in reporting

HOW TO COMPARE?

- Few validation studies are easily accessible for exhaustive review
- Error rates are published but it is difficult to equate software programs due to
  - Different assumptions for contributors, threshold values and missing data points
  - Different allele frequency databases for generating the likelihood ratio or LR values
  - FST uses a local city DNA database with some unusual construction features including hybrid profiles and lack of Hardy Weinberg Equilibrium (HWE)
  - LR values are rough approximations rather than exact scientific values; would like to see improved scientific accuracy – wide variation in error estimates between computational programs

RECENT ARTICLES

- H. MILLER COYLE. 2015. QUALITY CONTROL AND DUPLICATION FOR CONCORDANCE IN FORENSIC DNA SAMPLES: IMPLICATIONS FOR INTERPRETATION OF MIXTURES. INTERNATIONAL RESEARCH JOURNAL OF COMPUTER SCIENCE. 2(6): 16-18.
- H. MILLER COYLE. 2015. SOURCES OF COMPUTATIONAL ERROR IN PROBABILISTIC GENOTYPING SOFTWARE USED FOR DNA MIXTURE INTERPRETATION. INTERNATIONAL RESEARCH JOURNAL OF COMPUTER SCIENCE. 2(5): 12-16.
- M. HAZELL-SMITHEN, T. CALLAHAN, H. MILLER COYLE. 2014. TOUCH DNA AND THE ABILITY TO DETECT THE CORRECT SOURCE. INTERNATIONAL JOURNAL OF ADVANCED TECHNOLOGY AND SCIENCE. 1 (1): 45-51.

ACKNOWLEDGEMENTS

- The Legal Aid Society of New York City (www.legal-aid.org)
- Kyle B. Watters, Watters & Svetkey, 286 Madison Avenue, New York, NY 10017 (ref. *The People of the State of New York v. Carlos Marin, 2013)*
- University of New Haven (www.newhaven.edu) – touch DNA studies/DNA database sampling
  - Tim Callahan, Mesha Smithen, Stephanie Tedeschi