A Picture is Worth 10,000 LRs

The Evaluation and Implementation of Tools to Improve Interpretation and Reporting of Mixtures Using Probabilistic Genotyping

Dr. Tim Kalafut[1]
Dr. Michael Coble[2,3]

[1]Sam Houston State University, Department of Forensic Science
[2]University of North Texas Health Science Center, Department of Microbiology, Immunology, and Genetics
[3]Center for Human Identification

Disclaimer

- The views and opinions expressed in this presentation are those of the presenter/author, and do not represent any official views or opinions of Sam Houston State University, University of North Texas Health Science Center, The National Institute of Justice, or the American Academy of Forensic Sciences.

- Certain commercial equipment, instruments, software, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the authors or the National Institute of Justice nor does it imply that any of the materials, instruments, software, or equipment identified are necessarily the best available for the purpose.

Why do Inter/Intra Lab Studies for PG?

- Ultimate Goal of Forensic Science(?)
  - Equal outcome for a given sample/case/Victim/Suspect
  - No matter who does the work
- Are we there yet?

Expect “different” results

- Different models
  - STRmix™
  - EuroForMix
  - TrueAllele
- Different parameters
  - Stutter variance
  - Analytical Thresholds
  - PCR cycles
  - F_{ST} values
- Different LRs
  - Conditioned vs Unconditioned
  - Point Source
  - Lowest Population
  - Normalized Population (Stratified)
  - Relatives (Unified)
  - An “interval” LR – (HPD)
- Reporting
  - The number no matter what
  - Categorized number
  - Verbal scale

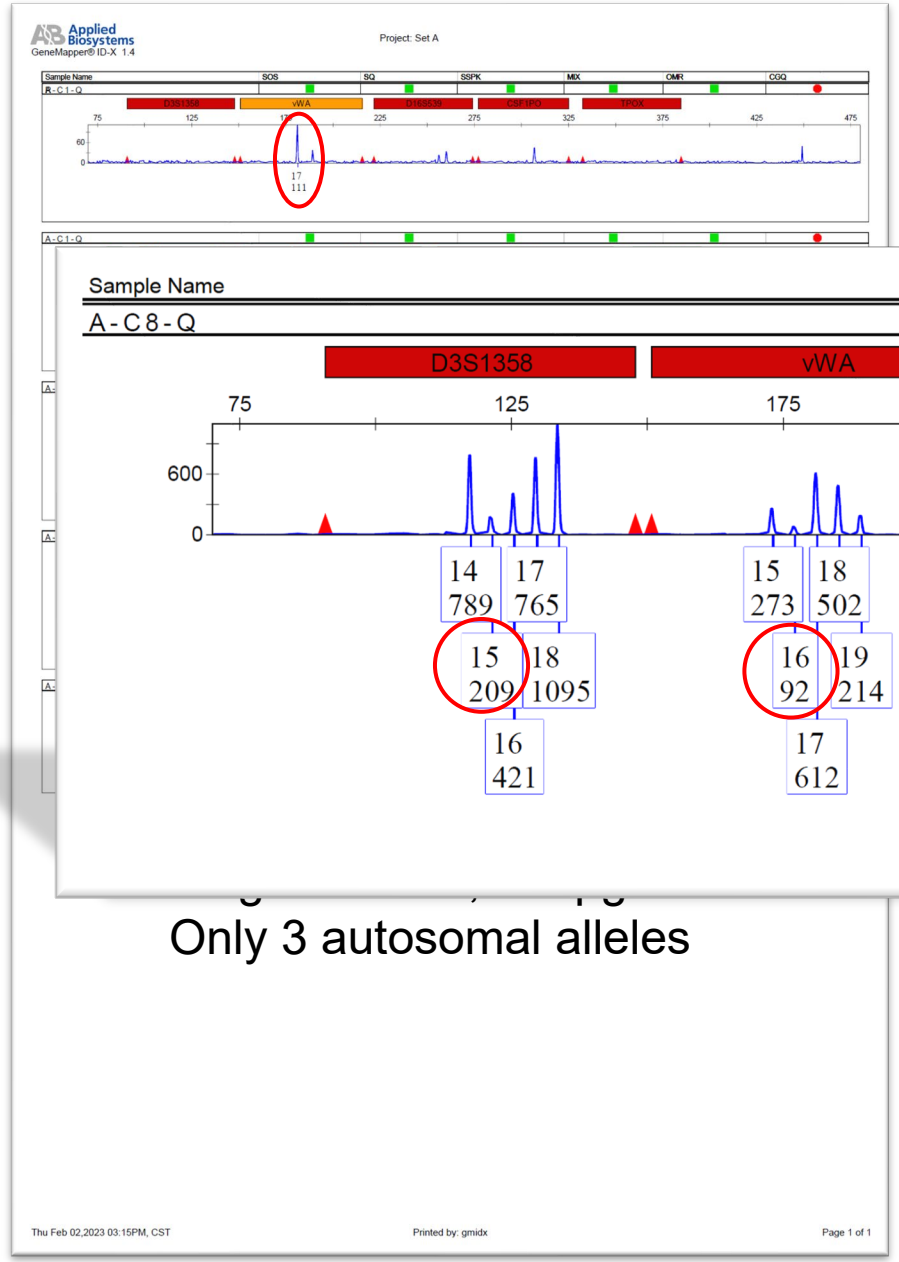
This study

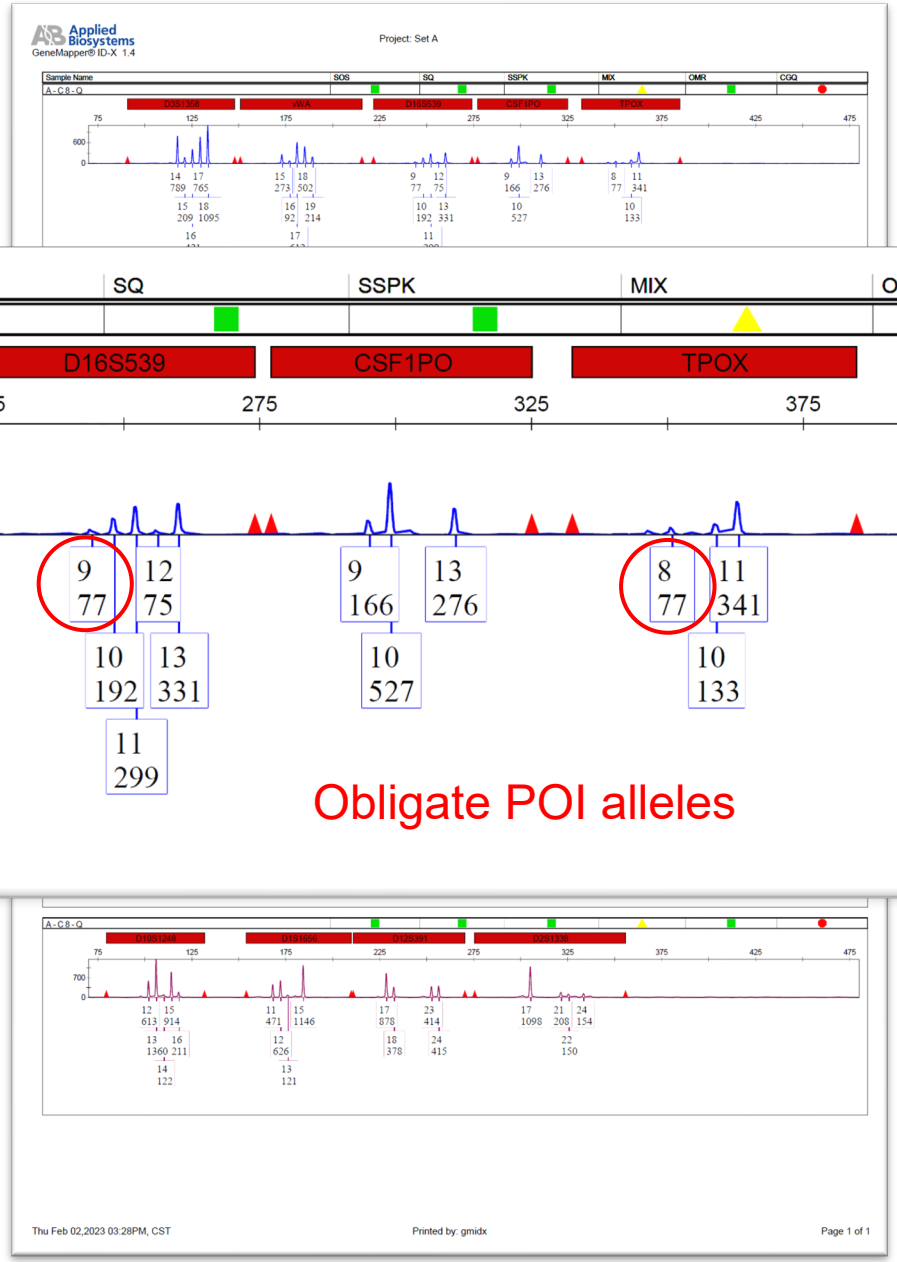
- An attempt to look at “lousy” samples
- Very low-level donors; volunteers were told 1 – 4 persons
- Expect low LRs
- Compared ground truth donors to the mixtures

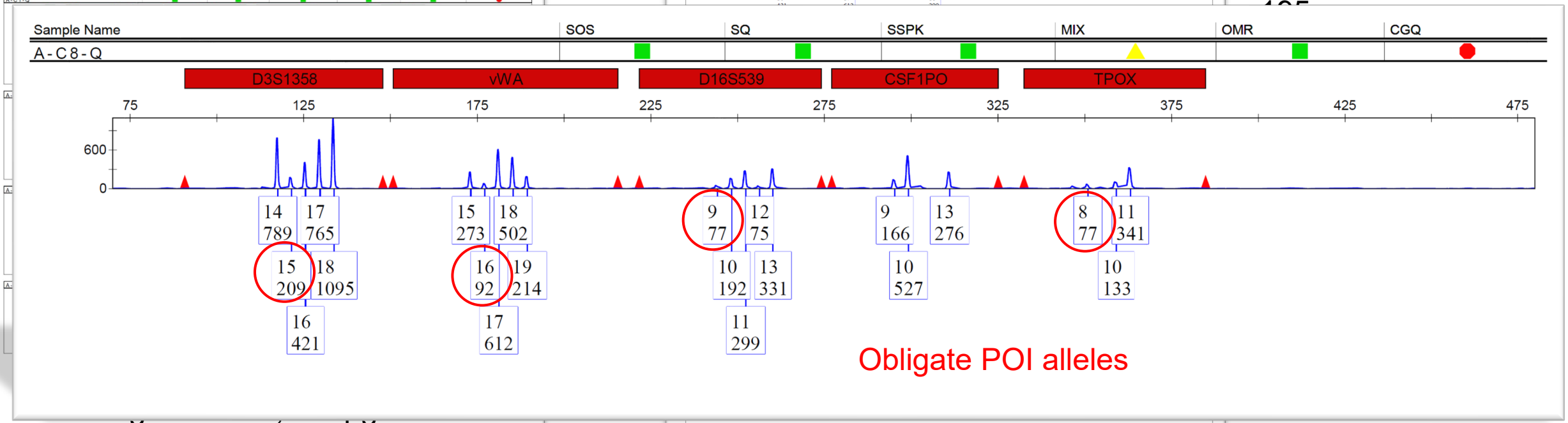
- PROVEDIt mixtures used; validated parameters (Kelly, et al)
- “Same” GMID analysis parameters
- Same STRmix™ parameters
- 10 laboratories, 25 volunteers

- Goal was to look at the “bottom line” – including report wording

How lousy is a “lousy” profile?







Obligate POI alleles

4 person; 1:4:4:4

Only 3 autosomal alleles

Most likely by realizing the other samples were 1, 2, or 3 persons

Samples examined twice

- Set A and Set B
  - 8 mixtures each
  - Assigned to different people
  - We asked for no peaking at the other set

- Round 1
  - Interpret exactly as your protocol says (Caveat: AT's provided by us)
  - Reporting statement, including LR (if any), verbal scale (if used), etc.

- Round 2 and 3
  - Swap sample sets (still no peaking/"TR")
  - Use two different tools to evaluate the LR
  - Hybrid reporting statement (normal language, but based on tools)

Round 1 Results (Lab Protocols)

- We recorded the reporting category (bottom line)

- Lots of different terms used
- Verbal scale used by some but not all
  - Verbal scale with “context”
  - Verbal scale wording – “This _____ provides…”
    - “analysis”
    - “LR”
    - Just “this”
    - “evidence” is best term?
- Suitability threshold (no interpretation/STRmix™)
- Report the number no matter how big/small
- Some have inconclusive zones

Round 1 converted to common terms

- Inclusionary = "included as a possible contributor"
- Exclusionary = LR<1
- Inc = inconclusive
- Uninformative = LR rounds to 1
- Unsuitable = Not used
- We counted them up

Sample Set	Case #	Internal Protocol				
		Unsuitable	Exclusionary	Inc	Uninformative	Inclusionary
A	C1	8		1		3
A	C2	6				6
A	C3	1		1		10
A	C4		1		1	10
A	C5	1		1		10
A	C6	2	1	1		8
A	C7	4	1	1		6
A	C8	3	2	1		6
B	C1	8				5
B	C2	8				5
B	C3	5				8
B	C4	1	1			11
B	C5	1		2		10
B	C6	2	2	1	1	7
B	C7	5	1	1		6
B	C8	4	1	1	2	5

Use of Advent*LR* and DBLR™

- Two tools that allow for non-contributor testing
- “Calibration” of the LR for POI compared to non-donors

- Advent*LR* (Schuerman *et. al.* 2020, *JFS*)
  - Graphical presentation of a STRmix™ database search
  - 10,000 random profiles based on NIST 1036 allele frequencies (static)
  - H_d or H_2 or H_A only

- DBLR™ (Kelly *et. al.* 2021, *FSI: Reports*)
  - STRmix™ companion tool
  - 10,000 to 10 million non-donors base on allele freqs (dynamic)
  - Both H_p/H_1 and $H_d/H_2/H_A$

Proposed use of non-contributor/calibration

- Schuerman et al (based on Gill, etc) - 99.9%ile LR
  - 99.9%ile LR as a “utility threshold” to aid the expert and the jury

  - LR_{POI} > 99.9%ile LR; Inclusionary support
  - LR_{POI} < 99.9%ile LR; Inconclusive due to non-contributor testing

- Initial testing of Advent*LR* and DBLR™ were both suitable
- High level of agreement (10,000 static profiles vs 100K)

- ***Volunteers were trained in use of both tools and theory***

Non-contributor Testing

- Used the full NIST 1036 combined allele frequencies for “test LR”
  - For POI in the case
  - For all profiles in the database

- Compared point source (Factor of N! or Sub-source) LR for POI
- To LRs generated from profile database (10K or 100K)

- If LR_{POI} “passes” then asked labs to report their “normal” LR
  - HPD from smallest of 3 or 4 populations
  - Unified LR was used also (4 populations normalized per census)

A picture is worth 1000 words - Advent*LR*

- Picture that shows how LR_{POI} compares to population

Set-A-Q1_K

CASE NUMBER	Set-A-Q1
SAMPLE ID	K1
NOC	1
H_p Profiles (K)	
H_d Profiles (K)	
AdventLR: V	2.9.0
STRmix™: V	2.5.11

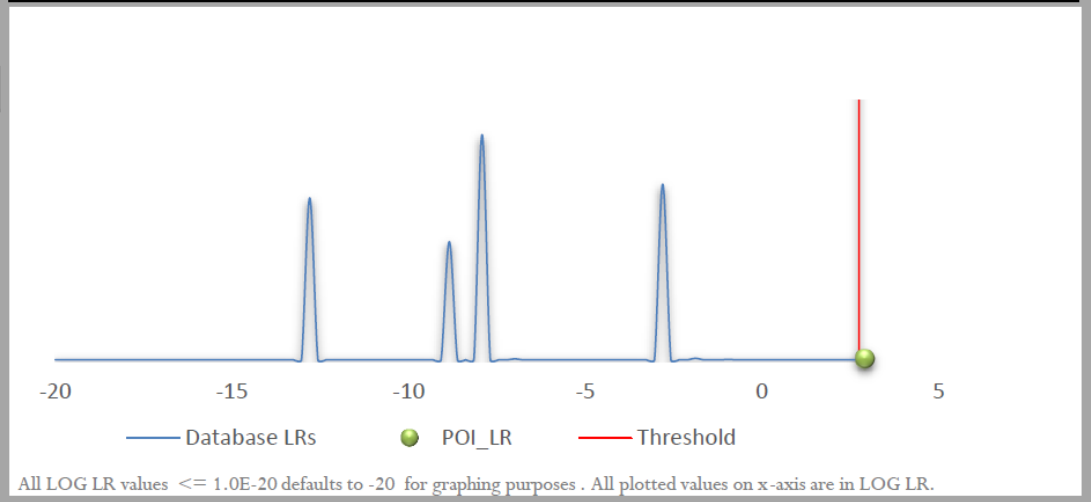
CASE NOTES	

SUMMARY OF PROFILES IN DATABASE

	Total Profiles	Average LR
Entire DB	10000	2
LR = 0	0	
LR > 0	10000	2
LR >= 1	21	714

Largest LR from search	1547
Known non-contributors > POI	6
*Point source LR for POI	876
Uninformative threshold	560

* Point source LR calculated using the NIST1036_Combined allele frequencies



REPORTED LR=93

Point source LR = 876

Green Dot

99.9%ile LR = 560

Red Line

Single Source sample (previous epg)

A picture is worth 1000 words - Advent*LR*

- Picture that shows how LR_{POI} compares to population

A-Q8_K

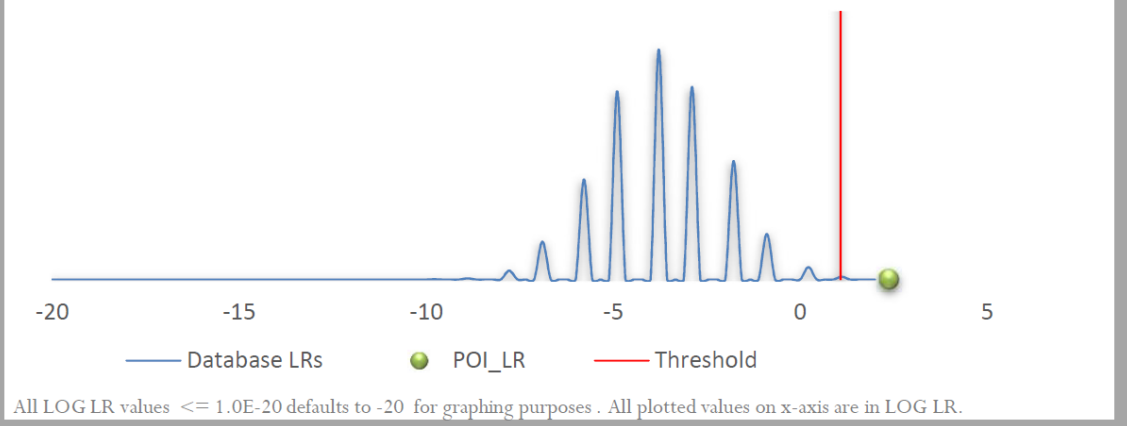
CASE NUMBER	
SAMPLE ID	
NOC	
H_p Profiles (K)	
H_d Profiles (K)	
AdventLR: V	2.9.0
STRmix™: V	2.5.11

CASE NOTES	

SUMMARY OF PROFILES IN DATABASE

	Total Profiles	Average LR
Entire DB	10000	1.02E-01
LR = 0	0	
LR > 0	10000	1.02E-01
LR >= 1	80	11

Largest LR from search	246
Known non-contributors > POI	0
*Point source LR for POI	247
Uninformative threshold	12



* Point source LR calculated using the NIST1036_Combined allele frequencies

REPORTED LR=10

Point source LR = 247

99.9%ile LR = 12

4p sample (previous epg)

A picture is worth 1000 words – DBLR™

SIMULATION RESULTS TABLE

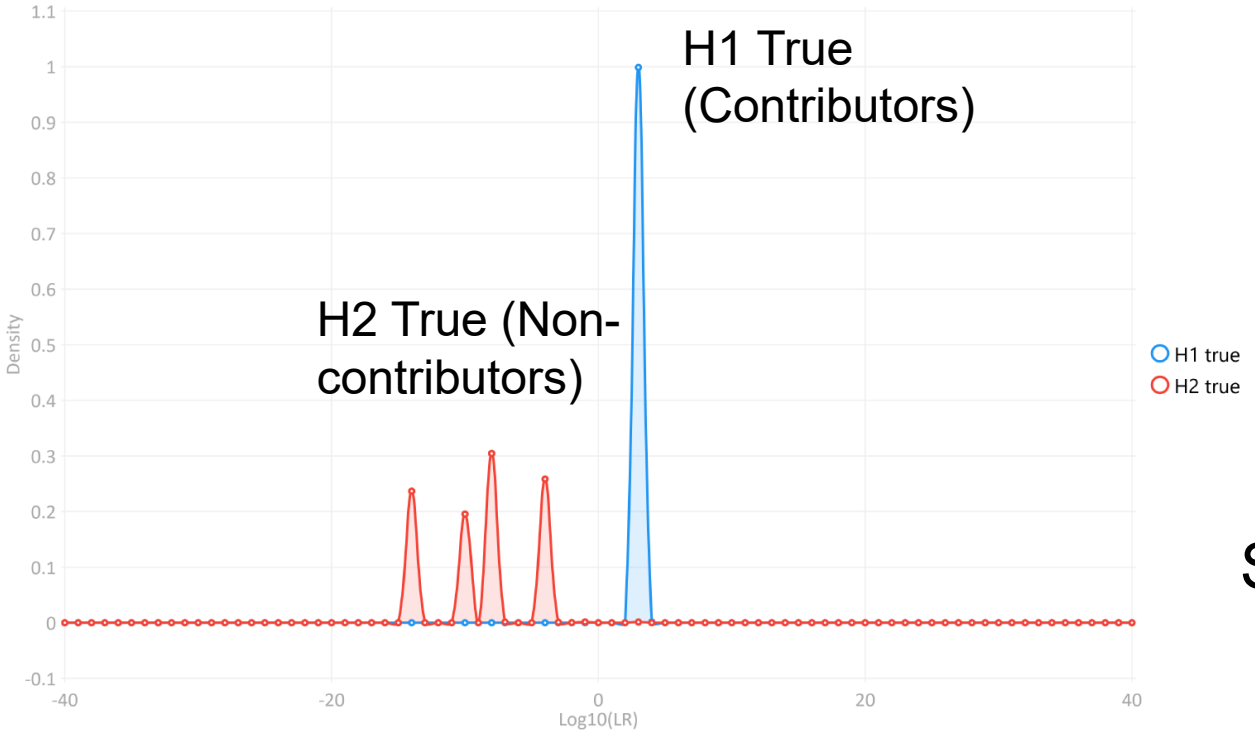
	ESTIMATE	
P(Log10LR >= 2.93752 | H1)	0.50831	
P(Log10LR >= 2.93752 | H2)	0.00048	
	Importance sampling from H1:	0.00040791

P(Log10LR>=2.93752|H2) = 0.00040791

Probability of LR > POI Given H2 is 0.040791%

(logLR used instead of LR)

SIMULATION RESULTS PLOT



1.0 – 0.00040791 = 0.99959

LR of POI is greater than 99.9%ile LR

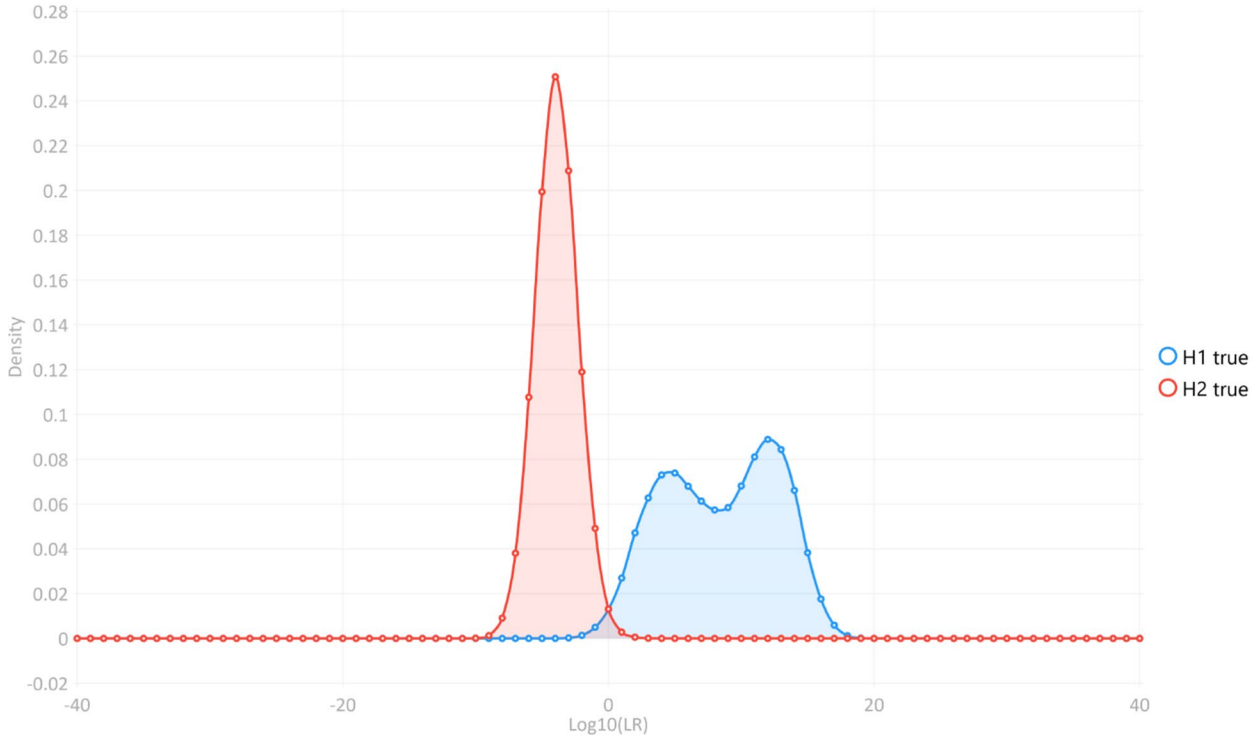
Single Source sample (previous epg)

A picture is worth 1000 words – DBLR™

SIMULATION RESULTS TABLE

	ESTIMATE	
P(Log10LR >= 1.79239 | H1)	0.94167	
P(Log10LR >= 1.79239 | H2)	0.00041	
	Importance sampling from H1:	0.00035068

SIMULATION RESULTS PLOT



1 – 0.00035068 = 0.99964

LR of POI is greater than 99.9%ile LR

4p sample (previous epg)

Round 2/3 reporting in common terms

- Dropped unsuitable
- Added “Inc – NCT”

Sample Set	Case #	AdventLR Non-contributor testing (10,000)				DBLR Non-contributor testing (100,000)			
		Exclusionary	Inc-NCT	Uninformative	Inclusionary	Exclusionary	Inc-NCT	Uninformative	Inclusionary
A	C1		1		11		9		4
A	C2				12				13
A	C3				12				13
A	C4				11			1	12
A	C5				12				13
A	C6				12				13
A	C7	1			11				13
A	C8	1			11	1			12
B	C1		1		11		8		4
B	C2				12				12
B	C3				12				12
B	C4				12				12
B	C5				12				12
B	C6	1		1	10	1		1	10
B	C7				12				12
B	C8	1		3	8	1		3	8

Overall comparison

- Much higher agreement between laboratories
- “Classification” based on empirical data

Sample Set	Case #	Internal Protocol					AdventLR testing (10,000)				DBLR (100,000)			
		Unsuitable	Exclusion	Inc	Uninformative	Inclusion	Exclusion	Inc-NCT	Uninformative	Inclusion	Exclusion	Inc-NCT	Uninformative	Inclusion
A	C1	8		1		3		1		11		9		4
A	C2	6				6				12				13
A	C3	1		1		10				12				13
A	C4		1		1	10				11			1	12
A	C5	1		1		10				12				13
A	C6	2	1	1		8				12				13
A	C7	4	1	1		6	1			11				13
A	C8	3	2	1		6	1			11	1			12
B	C1	8				5		1		11		8		4
B	C2	8				5				12				12
B	C3	5				8				12				12
B	C4	1	1			11				12				12
B	C5	1		2		10				12				12
B	C6	2	2	1	1	7	1		1	10	1		1	10
B	C7	5	1	1		6				12				12
B	C8	4	1	1	2	5	1		3	8	1		3	8

Bonus information

- Controlling variables in comparison studies is hard!
- GMID settings affected number of called peaks
  - No trend for difference in 99.9%ile
  - Reported LR in billions compared to 10,000s
- STRmix version less influence than GMID settings
- (Just initial observations; digging into this is taking time)

Looking forward

- If the goal is to get same “inclusion/exclusion” – no matter who does the work
- Non-contributor testing may be useful
  - Empirical data approach
  - Considers the quirks of the actual mixture
- But there are still unexpected things that play a role

Acknowledgments

- Participating laboratories (esp. Management)
- The volunteers
- Curt Schuerman for help with Advent*LR*
- ESR for DBLR™ trial licenses

- Todd Bille – helped with inspiration to apply to NIJ

- NIJ Grant #15PNIJ-21-GG-02710-SLFO

Dr. Tim Kalafut: tim.kalafut@shsu.edu

Dr. Michael Coble: michael.coble@unthsc.edu