Elevating forensics
— FOR 30 YEARS —

ISHI
INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION
PALM SPRINGS, CA | SEPT. 23-26, 2019

DNA Mixture Interpretation Principles and Best Practices

John M. Butler, PhD

Hari K. Iyer, PhD

Sheila Willis, PhD

Acknowledgments and Disclaimer

Points of view are the presenters and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

Certain commercial entities 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 National Institute of Standards and Technology, nor does it imply that any of the entities identified are necessarily the best available for the purpose.

Workshop Description

In the process of working on a NIST Scientific Foundation Review on DNA mixture interpretation, our team gathered and studied the scientific literature on the topic and carefully considered principles and best practices. The report generated from our study will be discussed along with specific thoughts regarding case context and measurement and interpretation issues involving binary approaches and probabilistic genotyping.

Learning Outcomes

1. The literature and scientific foundations of DNA mixture interpretation

2. Case context considerations in examining complex, low-level DNA mixtures

3. Measurement uncertainty and reliability considerations for forensic mixture DNA interpretation involving binary approaches and probabilistic genotyping

Intended Audience: DNA analysts and technical leaders

ISHI 2019 Workshop Attendees
28 states and 11 other countries

96 registered + 3 presenters
(as of 8/6/2019)

Alaska

Hawaii

Puerto Rico

Australia
Bahamas
Bahrain
Brazil
Canada
Germany
India
Iraq
Qatar
Russia
Singapore

In addition to state and local forensic laboratory analysts, we have representatives from:

Bode Technology
DFSC (USACIL)
FBI Laboratory
Federal Public Defender (CA)
Florida International University
Hennepin Co. Public Defender
Identilab
LevitasBio
MN Board of Public Defense
National Institute of Justice
NYC Legal Aid Society
Promega Corporation
ThermoFisher Scientific

Green = participants
Gray = no attendees

Planned Workshop Schedule

Time	Topic	Presenter(s)
1:00pm (10 minutes)	Introductions and Goals for Workshop	John
1:10pm (20 minutes)	NIST Scientific Foundation Review and Mixture Principles	John
1:30pm (60 minutes)	Reliability Considerations	Hari
2:30pm (15 minutes)	BREAK	
2:45pm (60 minutes)	Relevance Considerations	Sheila
3:45pm (15 minutes)	Some Key Takeaways & Best Practices, Q&A	John, All
4:00pm	Conclude Workshop	

Background and Qualification of Presenters

all are members of the NIST team conducting a scientific foundation review on DNA mixture interpretation

- **John M. Butler**:
  - Author of five textbooks (2001, 2005, 2010, 2012, 2015) and >170 research articles
  - Conducted dozens of workshops on DNA mixture interpretation

- **Hari K. Iyer**:
  - NIST Statistical Engineering Division (2014-present), statistics professor at Colorado State University for over 30 years
  - Author with Steve Lund of *Likelihood Ratio as Weight of Forensic Evidence: A Closer Look* (https://nvlpubs.nist.gov/nistpubs/jres/122/jres.122.027.pdf)

- **Sheila Willis**:
  - Retired director of Forensic Science Ireland (Dublin, Ireland)
  - Chaired effort to write 2015 ENFSI Guidelines for Evaluative Reporting in Forensic Science
  - NIST guest researcher (2017-2019)

From Yesterday…   Los Angeles Times

'Rapid DNA' promises breakthroughs in solving crimes.
So why does it face a backlash?
By Maura Dolan (September 25, 2019)

There is no question the devices work well on "single-source DNA," genetic material that comes from a single person. The problem comes when there is a mixture of DNA from multiple individuals, said Vincent A. Figarelli, superintendent of Arizona's Crime Laboratory System. In those situations, a trained forensic scientist is needed to interpret it.

"**Mixture interpretation is the most difficult thing that crime laboratory analysts have to do by far**," Figarelli said. "There's no way you want a Rapid DNA operator doing a mixture analysis."

https://www.latimes.com/california/story/2019-09-24/rapid-dna-forensics-crime-police

Lots of Change in the Past Few Years for DNA Mixture Interpretation…

- Growth of **probabilistic genotyping software (PGS)** use throughout the U.S. forensic DNA community
    - >50 U.S. laboratories now using STRmix, TrueAllele, or Lab Retriever

- Many new publications on theory and data behind probabilistic genotyping models (primarily those used in STRmix)

- Widespread adoption of new STR megaplex kits and in some cases new CE instrumentation that has required additional validation studies

- New guidelines and standards released and in development (e.g., SWGDAM 2017, FBI QAS 2020)

Historical Overview and Timeline

PGS development and application to complex mixtures

1985

Gill *et al.* (1985) differential extraction

1990

Evett *et al.* (1991) LR

NRC I (1992) CPI described

1995

QAS (1998)

NRC II (1996) LR preferred

SWGDAM STR Interpretation (2000) *4 pages*

FSS CAI & hierarchy of propositions (1998)

Clayton *et al.* (1998) mixture steps

2000

DAB Stats (2000)

German Stain Commission *Mixture Categories* (2006)

ENFSI DNA WG Mixture Principles Consensus (2007)

Technical UK WG on mixtures (2008)

Gill *et al.* (2000) FSS LCN & PG theory

ISFG *Mixtures* (2006)

2005

NIST MIX05 (2005)

O.J. Simpson Case (1995)

Amanda Knox Case (2007, 2011)

ISFG meeting (2009)

Rome meeting (2012)

Melbourne Lab Closure (2010)

EuroForGen-NOE (2012-2017)

STRmix created

ISFG *Dropout* (2012)

2010

QAS (2009)

QAS (2011)

SWGDAM Autosomal STR Interpretation (2010) *28 pages*

NAS Report (2009)

2015

SWGDAM PGS Validation (2015)

SWGDAM STR EDM (2014)

DC Lab Closure (2015)

NIST MIX13 & TL Summit (2013)

ENFSI BPM PGS Validation (2017)

ISFG *Software* (2016)

ISFG *Propositions* (2018)

UK FSR Mixture Interpretation & PGS Validation (2018)

2020

QAS (2020)

SWGDAM Autosomal STR Interpretation (2017) *90 pages*

SWGDAM Validation (2016)

U.S. labs begin to adopt PGS (~2014-present)

TX FSC Review (2015-)

PCAST Report (2016)

ASB 020 Mixture Method Verification (2018)

SWGDAM LR Verbal Equivalents (2018)

Butler *et al.* (2018) MIX05 & MIX13 publication

Bright *et al.* (2018) STRmix PCAST response

Bright *et al.* (2019) STRmix interlab

PGS = probabilistic genotyping software

DNA Mixture Interpretation Approaches

- **Binary methods with simple** (mostly 2-person) **mixtures**
  - Statistical approaches: **LR** (Evett et al. 1991, NRC 1996), **CPI** (NRC 1992, Budowle et al. 2009, Bieber et al. 2016)
  - Deconvolution/interpretation: (Clayton et al. 1998, Evett et al. 1998, Bill et al. 2005)

- **Probabilistic genotyping software (PGS)**
  - Theory: probability of drop-out and drop-in (Gill et al. 2000, Balding & Buckleton 2009)
    - Early implementation: LoComationN (Gill et al. 2007), gamma model (Cowell 2007)
  - Current probabilistic genotyping software:
  - **Discrete models: FST** (Mitchell et al. 2012), **LRmix** (Gill et al. 2013), **likeLTD** (Balding 2013), **LiRa** (Puch-Solis & Clayton 2014), **Lab Retriever** (Inman et al. 2015)
  - **Continuous models: TrueAllele** (Perlin et al. 2011), **STRmix** (Taylor et al. 2013), **DNAmixtures** (Cowell 2015), **EuroForMix** (Bleka et al. 2016), **CEESIt** (Swaminathan et al. 2016), *Kongoh* (Manabe et al. 2017), **GenoProof Mixture 3** (Götz et al. 2017), **DNA Mixture Solution** (Brenner 2015), **MaSTR** (Adamowicz et al. 2018)

Models Used in Mixture Interpretation

Binary	Discrete (semi-continuous)	Continuous (fully-continuous)
Binary	**Discrete**	**Continuous**
Binary models require that all peaks are designated as alleles or ignored completely. No allowance for drop-in or drop-out.	Discrete models use the presence/absence of peaks but do not take their heights into consideration. Drop-in and drop-out are accommodated.	Continuous models take into account the peak heights of each allele as well as their presence or absence. Again, drop-in and drop-out are accommodated.

https://cdnmedia.eurofins.com/european-west/media/1418957/lgc_lira_fact_sheet_en_0815_90.pdf

Probabilistic Genotyping Software (PGS)

Yes (probability = 1)
or
No (probability = 0)

Peak probabilities may be >0 and <1

A model is a lie that helps you see the truth.

- Howard Skipper (a cancer researcher at Southern Research Institute in Alabama)

Science Magazine News (Sept 23, 2019)

Representative Mark Takano (D–CA) says, "Intellectual property rights should not be able to trump due process." AP PHOTO/J. SCOTT APPLEWHITE

This U.S. lawmaker wants greater scrutiny of algorithms used in criminal trials

By Catherine Matacic | Sep. 23, 2019 , 4:10 PM

Since it was introduced in the 1980s, DNA evidence has become a "gold standard" of U.S. courtrooms, leading to the convictions—and exonerations—of thousands of accused criminals. But experts struggle to analyze degraded or contaminated samples, and many have started to use sophisticated **probabilistic genotyping software** to estimate the likelihood that a suspect's DNA matches DNA at the crime scene. Such so-called **forensic algorithms** are far from rare: Increasingly, they're used to estimate matches for everything from fingerprints to gun barrels to faces in security camera footage.

Last week, Representative Mark Takano (D–CA) introduced legislation that would make it easier for defendants facing federal criminal charges to gain access to forensic algorithms, and further **require the makers of computational forensic software to meet minimum standards set by the National Institute of Standards and Technology (NIST).**

https://www.sciencemag.org/news/2019/09/us-lawmaker-wants-greater-scrutiny-algorithms-used-criminal-trials

Recent Government Accountability Office (GAO) Publication

Since January 2019, first four Science & Tech Spotlights cover:

Hypersonic Weapons
Probabilistic Genotyping Software
Opioid Vaccines
Blockchain & Distributed Ledger Technologies

https://www.gao.gov/assets/710/701363.pdf

SEPTEMBER 2019

WHY THIS MATTERS

New developments in software to analyze contaminated or partly degraded DNA could greatly facilitate criminal investigations. However, the validity of the analysis and the implications for constitutional due process protections remain unsettled.

NIST Scientific Foundation Review on DNA Mixture Interpretation

John

NIST Forensic Science Activities

Conduct Research and Collaborate

Intramural Research

DNA
Digital
Fingerprints
Firearms
Footmarks
Statistics
Toxins
Trace
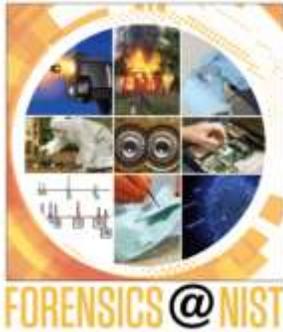
FORENSICS@NIST

Extramural Research
funding a NIST Center of Excellence in Forensic Science (CSAFE: since 2014)

1920s - present

Partner with Community to Strengthen Policies and Practices

National Commission on Forensic Science (NCFS) with DOJ

2013 - 2017

Organization of **OSAC** Scientific Area Committees for FORENSIC SCIENCE

2013 - present

Convene Meetings to Examine Issues

Improving Biometric and Forensic Technology: The Future of Research Datasets
January 26-27, 2015

TECHNICAL COLLOQUIUM · WEIGHT OF EVIDENCE

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

Research Innovation to Implementation in Forensic Science Symposium
RI2I
NIST / Gaithersburg, MD / June 19-20, 2019/ NIST.GOV/RI2I

Human Factors Working Groups (with NIJ)

2009 - present

Explore Scientific Foundations

Initial efforts with DNA mixture interpretation and bitemark analysis

NISTIR 8225 DRAFT
NIST Scientific Foundation Reviews

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Associate Director of Laboratory Programs

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs
Director's Office

September 2018

U.S. Department of Commerce

National Institute of Standards and Technology

2017 - present

FORENSIC SCIENCES

https://www.nist.gov/topics/forensic-science

Purpose of our DNA Mixture Interpretation Review

Primary Goals:

1. **Develop a bibliography of relevant literature**
2. Define underlying principles, characterize capabilities and limitations of methods for mixture analysis
3. Identify knowledge gaps for future research
4. Inform the forensic community and non-specialists of findings (judges, attorneys,& general public)
5. **Create a framework for potential future NIST foundational reviews in forensic science (bitemarks already started)**

AAFS workshop conducted in February 2019
Working to complete a draft report for release soon

AAFS 2019 Workshop W10

DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review

Chair: John M. Butler (NIST),
Co-Chair: Sheila Willis (NIST Guest Researcher)

8 hours, 17 presenters, 19 talks, 406 slides

https://strbase.nist.gov/AAFS2019-W10.htm

Speakers (*left-to-right*):
NIST team & Resource Group

Joel Sutton (DFSC)
Jack Ballantyne (UCF)
Keith Inman (Cal State East Bay)
John Butler (NIST)
Lisa Schiermeier-Wood (VA DFS)
Peter Vallone (NIST)
Melissa Taylor (NIST)
Ray Wickenheiser (NYSP)
Robin Cotton (BU)
Bruce Heidebrecht (MSP)
Hari Iyer (NIST)
Eugene Lien (NYC OCME)
Sheila Willis (NIST associate)
Jennifer Breaux (MoCo, MD)
Charlotte Word (consultant)
Roger Frappier (CFS-Toronto)
Rich Press (NIST)

Who Is Involved in the NIST DNA Study?

- **NIST Review Team**
  - Role: conducting review & writing report
  - 6 people who **have met weekly for the past two years**
    - John Butler, Hari Iyer, Rich Press, Melissa Taylor, Pete Vallone, Sheila Willis (guest researcher)
  - Expertise: research, DNA literature, statistics, human factors, casework management, communications

- **Resource Group**
  - Role: providing input & sounding board
  - 13 practitioners & academics/consultants (Federal, state, local, and international) who provide periodic input & feedback **across 12 meetings with the NIST team**
  - Expertise: DNA casework
  - *Reviewed initial draft report (in June 2019) but are not being asked to endorse report conclusions or key takeaways*

NIST DNA Mixtures Explainer

Credit: N. Hanacek/NIST
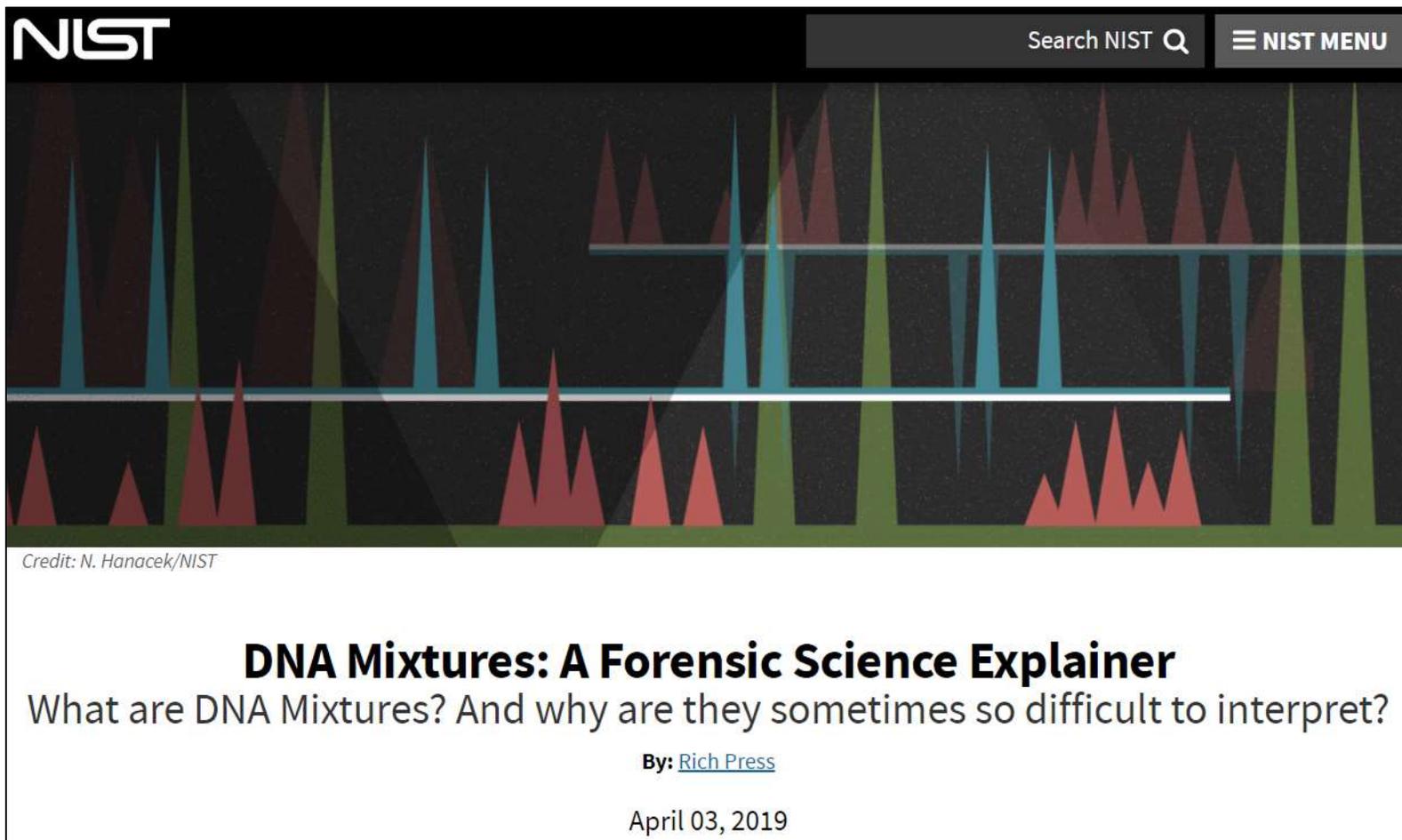
DNA Mixtures: A Forensic Science Explainer
What are DNA Mixtures? And why are they sometimes so difficult to interpret?

By: Rich Press

April 03, 2019

Topics Covered

- **Why have DNA mixtures and trace DNA become so prevalent?**
- **Are all DNA mixtures difficult to interpret?**
- **Why are complex DNA mixtures difficult to interpret?**
- **UNCERTAINTY #1: When is a peak a peak?**
- **UNCERTAINTY #2: Whose peak is it anyway?**
- **What is probabilistic genotyping software, and how does it help?**
- **How confident can one be that the DNA is related to the crime?**
- **Should labs just stop analyzing complex DNA mixtures altogether?**

https://www.nist.gov/featured-stories/dna-mixtures-forensic-science-explainer

Initial Draft Report (~350 pages; *too long - being revised*)

- Front Material: Acknowledgments and Disclaimer
- **Chapter 1**: Introduction to the Review
- **Chapter 2**: DNA Mixture Approaches, Principles, and History
- **Chapter 3**: Data Sources and Study Input
- **Chapter 4**: Reliability (validation and LR discrimination & calibration)
- **Chapter 5**: Relevance (DNA transfer & activity)
- **Chapter 6**: New Technologies (potential & limitations)
- **Chapter 7**: Training and Continuing Education
- **Chapter 8**: Key Takeaways Summarized
- **Appendix 1**: Foundational Bibliography **(605 references)**
- **Appendix 2**: DNA Basics & Glossary **(122 terms defined)**
- **Appendix 3**: Comments on PCAST Requirements for Scientific Validity

Personal Reflections on the NIST Scientific Foundation Review of DNA Mixture Interpretation

- **Valuable input from our Resource Group feedback** and discussions have illustrated common challenges across laboratories

- We do not always use terminology the same and as a community we can **benefit from having a more uniform language and terminology** (standardized definitions that are used and understood)

- In some cases, we need to **consider what questions we are really addressing** when we are working with small amounts of material that can be transferred

- Looking more towards **performance based testing** (what do my validation data actually demonstrate?) instead of task-driven efforts (did I meet a set of required studies?)

- The community will benefit from developing a **comprehensive, curated reference list** of foundational publications

- Spelling out **key principles** that we need to understand will help with training more consistently across laboratories and analysts

Data Resources Sought for Examination in Our Review
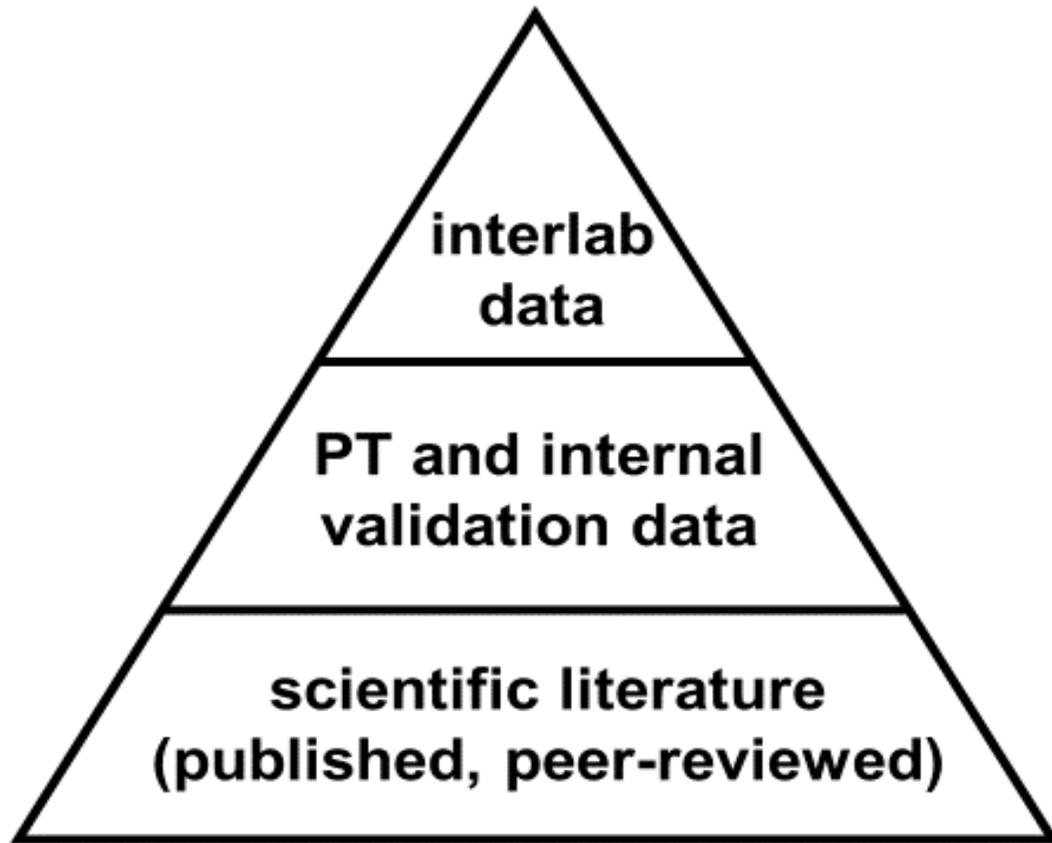
An illustration of general relationships for information in support of a method and its use

Interlaboratory data reveal the degree of reproducibility with a method across multiple laboratories.

Proficiency test (PT) and internal validation data demonstrate the ability to obtain reliable results under specific laboratory conditions in a single laboratory.

Published articles in peer-reviewed scientific journals typically establish the broad base of what is possible.

"Factor Space" in DNA Mixture Studies

- **Total DNA amount** (e.g., 1 ng or 100 pg)
  - Lowest amount of DNA in a minor contributor

- **Contributor component ratios** (e.g., 10:1 or 1:1:1)

- **Degree of allele overlap** across mixture components
  - Minor contributor alleles in stutter positions of major contributor alleles
  - Almost never discussed in publications

- **Number of contributors**

What are the range of casework samples typically seen in your laboratory?

Examination of Factor Space in a PGS Publication

Greenspoon et al. 2015 "Establishing the limits of **True Allele Casework**: a validation study" (*J. Forensic Sci.* 60(5): 1263-1276)

- **17-1p, 18-2p, 15-3p, 7-4p mixtures** explored with PowerPlex 16, ABI 3130xl; some single-source DNA examined down to 10pg; all mixtures examined at 1 ng total DNA with varying ratios (2p - 97:3, 90:10, 80:20, 70:30, 60:40, 50:50; 3p – 90:5:5, 80:10:10, 70:20:10, 60:30:10, 40:30:30; 4p – 60:20:10:10, 50:20:20:10, 40:30:20:10, 40:40:10:10, 85:5:5:5); some general discussion of the degree of allele sharing; seven synthetic "sons" and "brothers" were created to examine specificity for differentiating relatives in the 2p, 3p, and 4p mixtures

Impact of More Contributors

Number of Alleles (L)

N \ L	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total Genotype Combinations
1	1	1															2
2	1	4	6	3													14
3	1	8	29	52	45	15											150
4	1	13	84	297	600	690	420	105									2,120
5	1	19	192	1,116	3,933	8,661	11,970	10,080	4,725	945							41,642
6	1	26	381	3,321	18,080	63,919	150,332	236,978	247,275	163,800	62,370	10,395					956,878
7	1	34	687	8,484	66,645	346,644	1,231,857	3,052,008	5,316,885	6,483,330	5,415,795	2,952,180	945,945	135,135			25,955,630
8	1	43	1,155	19,428	210,645	1,529,064	7,687,512	27,472,653	71,004,690	133,874,415	184,033,080	182,338,695	126,756,630	58,648,590	16,216,200	2,027,025	811,819,826

Number of Contributors (N)

Principles

- A principle is a fundamental truth from which others are derived. An understanding of foundational principles can provide the basis for why something is important and can assist in deciding what should be done in specific situations.

Biology

1. Our DNA generally remains unchanged across time and cell type

2. DNA transfers and persists and can be collected and analyzed

3. A forensic DNA profile comes from analysis of DNA at specific sites in the genome

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.
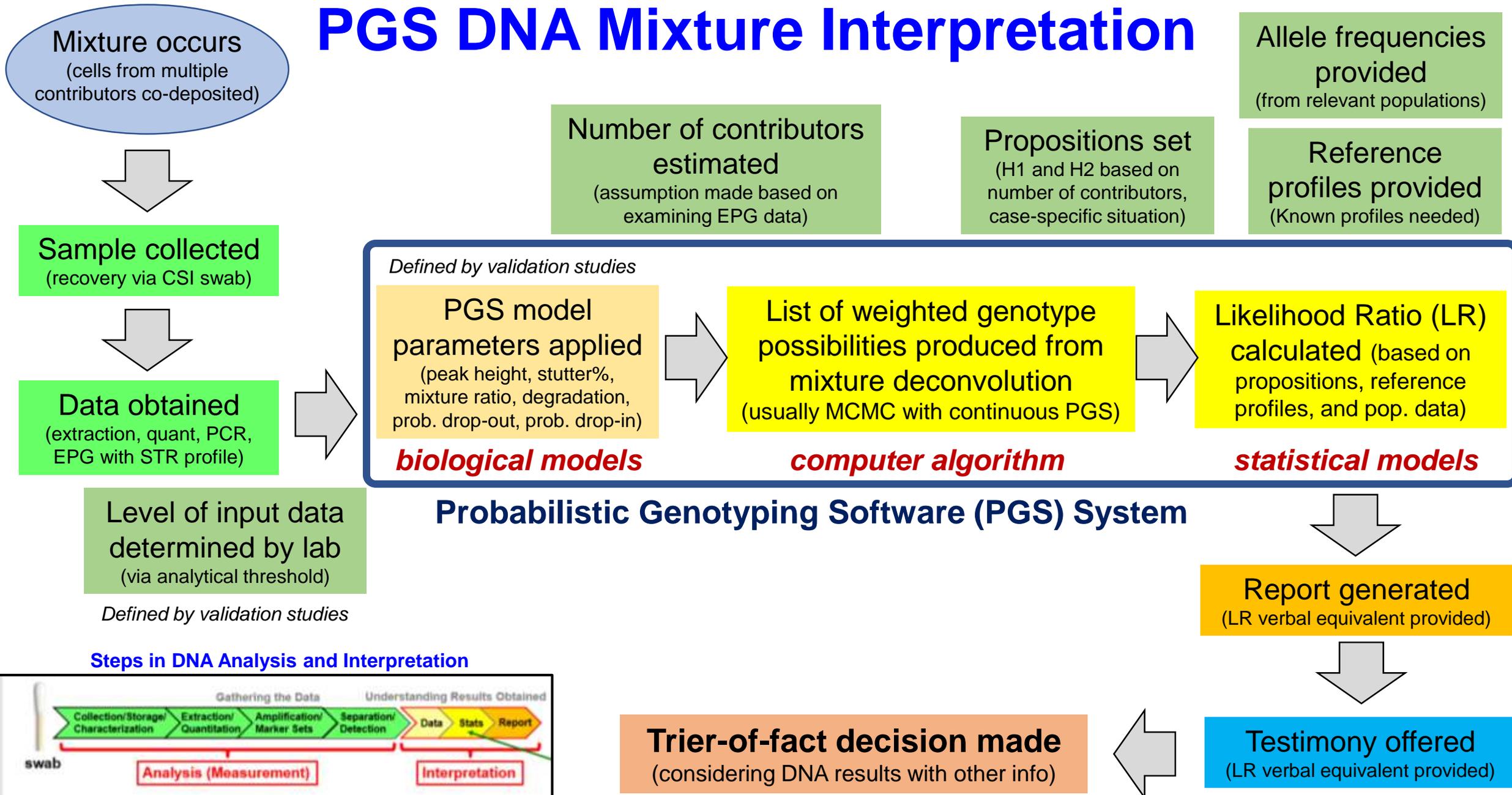
Genetics

4. DNA passes from parent to offspring according to genetic inheritance patterns

5. Genetic inheritance patterns enable weight-of-evidence statistical calculations

6. DNA profiles from close relatives are more similar than DNA from unrelated people

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.

Analysis

7. With PCR amplification, a copy of a sample's DNA template is attempted, but it may not be exactly representative of the original sample

8. PCR enables sensitive detection but may introduce artifacts (e.g., STR stutter products)

9. When copying low amounts of DNA, the chance of allele drop-out increases

10. Stochastic (random) effects influence repeatability and relative amounts of detected alleles

11. Length-based separations (e.g., capillary electrophoresis) may not fully resolve different STR alleles

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.

Statistics

12. Ability to differentiate DNA profiles generally increases as more DNA sites are tested

13. Probability is used to account for uncertainties when interpreting complex DNA data

14. Validation and calibration with known samples can be used to assess reliability of probability assignments

15. Different statistical approaches (e.g., LR, RMP, CPI) can produce different results

16. Weight-of-evidence assessments require at least two (mutually exclusive or different) points of view

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.

Mixture Interpretation

17. Mathematical models can digitally deconvolute mixture components that cannot be physically separated

18. Continuous models use more information than discrete or binary approaches

19. Ability to deconvolute mixture components depends on the degree of allele sharing and stochastic effects, which can limit reliable pairing of alleles into potential contributor genotypes

20. Probabilistic genotyping software (PGS) can assist in complex DNA mixture interpretation but should not replace the judgment of DNA analysts

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.

Propositions

21. Use of the hierarchy of propositions helps address different questions (e.g., who? vs. how?)

22. Information from DNA transfer and persistence studies should inform activity level evaluations

23. Sub-source likelihood ratios, if taken in isolation, can potentially be misleading

These principles and concepts have been distilled out of various publications and aspects of DNA mixture interpretation. They have been grouped by theme and ordered arbitrarily.

What Question Are We Answering?

- Depends on the propositions (hypotheses made)
  - Need to consider **the hierarchy of propositions** (Cook et al. 1998)

- The likelihood ratio (LR) is **the probability of getting the evidence** *if* the defendant is a contributor compared to the probability of getting the evidence *if* that defendant is not a contributor.
  - Involves assumptions as to the number of contributors in the mixture
  - LR is influenced by a number of inputs

Factors Influencing LR Values Determined by PGS Systems

Input	By Who	Impact/Example
Modeling choices	PGS system architect(s)	Peak height ratio variance allowed, how potential degradation is modeled, etc.
Data input choices	DNA analyst	Defining alleles (setting analytical threshold), categorizing artifacts from alleles (e.g., stutter)
Proposition choices and assumptions	DNA analyst	Use of unrelated individuals vs. relatives or conditioning on a victim's profile with an intimate sample
Population database choices	DNA analyst/ laboratory policy	Different allele frequency values will influence LR values
Reporting statistic choices	DNA analyst/ laboratory policy	Handling sampling variation (e.g., HPD*)

*HPD=highest posterior density-defines interval most likely to contain the true value

PGS DNA Mixture Interpretation

Mixture occurs
(cells from multiple contributors co-deposited)

↓

Sample collected
(recovery via CSI swab)

↓

Data obtained
(extraction, quant, PCR, EPG with STR profile)

→

Level of input data determined by lab
(via analytical threshold)

Defined by validation studies

Steps in DNA Analysis and Interpretation

Number of contributors estimated
(assumption made based on examining EPG data)

Propositions set
(H1 and H2 based on number of contributors, case-specific situation)

Allele frequencies provided
(from relevant populations)

Reference profiles provided
(Known profiles needed)

Defined by validation studies

PGS model parameters applied
(peak height, stutter%, mixture ratio, degradation, prob. drop-out, prob. drop-in)

biological models

→

List of weighted genotype possibilities produced from mixture deconvolution
(usually MCMC with continuous PGS)

computer algorithm

→

Likelihood Ratio (LR) calculated (based on propositions, reference profiles, and pop. data)

statistical models

Probabilistic Genotyping Software (PGS) System

↓

Report generated
(LR verbal equivalent provided)

↓

Testimony offered
(LR verbal equivalent provided)

←

Trier-of-fact decision made
(considering DNA results with other info)

Improved Sensitivity is a Two-Edged Sword

"As sensitivity of DNA typing improves, laboratories' abilities to examine smaller samples increases. This improved sensitivity is a two-edged sword. **With greater capabilities comes greater responsibilities to report meaningful results.** Given the possibility of DNA contamination and secondary or even tertiary transfer in some instances, **does the presence of a single cell (or even a few cells) in an evidentiary sample truly have meaning?...**"

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation* (Elsevier Academic Press: San Diego), p. 458

Reliability Considerations

Hari

DISCLAIMER

Viewpoints expressed are those of the presenter and are not intended to reflect the viewpoints of anyone else at NIST (except perhaps Steve Lund and a few friends 🙂)

The discussion is based on what makes sense to us and no claim is made that these are new perspectives

ACKNOWLEDGEMENTS

- Sheila Willis

- John M. Butler

- Melissa K. Taylor

- Rich Press

- Peter M. Vallone

- Steve Lund

- Sarah Riman

- All the members of the NIST DNA Scientific Foundation Review Resource Group

Reliability

Reliability
/re-ly-a-bi-li-ti/

1. To be able to produce good results time after time. 2. How much a person can be depended on.

Colourbox.com

Reliability

The Cambridge Dictionary describes "Reliability" as "how accurate or able to be trusted someone or something is considered to be."

reliability
noun [U] · UK 🔊 /rɪˌlaɪəˈbɪləti/ US 🔊

[https://dictionary.cambridge.org/us/dictionary/english/reliability]

⭐ **how accurate or able to be trusted someone or something is considered to be:**

Competence and reliability are prerequisites for any job.

Public confidence in **the reliability of** a company's financial statements depends on investors perceiving the company's auditors as being independent from the company.

reliability of information/data We have made significant improvements in the reliability of the information held on our database.

Reliability

The plain English meaning of the word 'reliability' is 'trustworthiness'. This is the sense in which we use this term in the report.

In the fields of psychology and sociology the term RELIABILITY is used to describe Consistency. This has led to much confusion.

Reliability implies consistency but consistency alone does not imply reliability.

Reliability

- **A Method is RELIABLE if it produces 'good' results time after time.**

- What is meant by 'good'? Rather than give binary answers (reliable or not reliable) or personal assessments (method has a high degree of reliability) what we require are FACTS and DATA.

- Personal Assessment: "this surgical procedure has an <u>excellent track record</u> of being successful".

- Facts&Data: "90 out of 100 patients who underwent this type of surgery survived and lived for at least 5 more years. The other 10 died on the operating table."

  Judgements of reliable/unreliable are personal.
  But facts and data are not personal.

Reliability

- **Accuracy**
- Precision
- Repeatability
- Reproducibility
- Uncertainty
- Error

Accuracy: 'how close is the result to the true value?' or 'how often does this procedure lead to correct decisions or conclusions?'

Inaccuracy: 'how far is the result from the true value?'

Note: True value can be an elusive quantity. Usually substituted with 'highly trusted reference value' [Standard Reference Materials (SRMs): values from NIST ☺]

Or a 'consensus value' based on various authoritative national metrology labs.

Reliability

- Accuracy

- **Precision**

- **Repeatability**

- **Reproducibility**

- Uncertainty

- Error

Precision: 'To what extent do repeated measurements of the 'same' quantity agree with one another?

Imprecision: 'To what extent do repeated measurements of the 'same' quantity disagree with one another?

When repeated measurements give different values (there is measurement variability) we can all see that the process does not produce perfectly accurate results. The variability among the results gives us a lower bound on the inaccuracy.

Repeatability and **Reproducibility** explore the extent to which measurements of the 'same' quantity differ under varying conditions.

Reliability

- Accuracy
- Precision
- Repeatability
- Reproducibility
- **Uncertainty**
- Error

Uncertainty is the <u>doubt regarding the underlying truth</u> that remains before or after observing the measurements.

Often this may be considered in the context of the difference between a measured value and some concept of the most appropriate value.

Reliability

- Accuracy
- Precision
- Repeatability
- Reproducibility
- Uncertainty
- **Error**

Error:

<u>Conventional meaning</u>: **Mistake**

<u>Statistical usage</u>: Difference between offered result and 'truth'.

Reliability and Science

Science has two key components:

1. Logic: Mathematics, Probability Theory

    These fields examine the (undeniable) logical consequences that follow from an initial set of assumptions (they do not have anything to say about the correctness of the assumptions). Reliability of these methods is generally not questioned.

2. Experimental (Empirical) Measurements

    These help us examine the correctness of
    (a) the initial set of assumptions, and/or
    (b) the logically predicted consequences

- When a scientific theory is proposed, logical arguments help us judge the plausibility of the theory but the claim may be considered scientifically supported only when empirical studies support the claim.

- Generally speaking, multiple independent empirical studies are needed before a claim is accepted as reliable.

If you toss this quarter twice, what is the probability that both tosses will give 'HEADS' ?

Reliability and Science

- IF the probability of 'HEADS' in a single toss of a coin is ½ and
- IF the result of one toss does not affect the result of any other toss (independence),

THEN

- The probability of getting both heads in two tosses of the coin must be ¼.

EXPERIMENT: A coin is tossed two times and the number of 'heads' is recorded (0 or 1 or 2). The experiment is repeated 1000 times. The following data are obtained.

	Both "Tails"	One 'Head' one 'Tail'	'Both Heads'	TOTAL
EXPECTED	250	500	250	1000
OBSERVED	400	500	100	1000

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong."

Richard P. Feynman
Nobel Laureate, 1965
Quantum Electrodynamics & Physics of Elementary Particles

TRUST

Trust can arise in several ways:

(1) Belief (in a higher authority, e.g. expert)
(2) Empirical demonstrations of a claim in ground truth known situations
   - Making predictions and verifying if the predictions come true

Absolute truth is difficult or impossible to establish but one can be 'convinced' that something is true based on a combination of the above modes of forming trust.

Each individual has his/her own thought processes involving combination of empirical knowledge with intuition and belief that lead him/her to form a degree of acceptance of a claim.

Science attempts to provide a common ground for all by providing empirical foundations for theories and methods that are developed.

To what extent one is convinced of the truth of a scientific claim is a personal matter.
Science attempts to minimize the level of belief one needs to accept a claim by replacing belief with empirical demonstrations of the claim's correctness."

Federal Rules of Evidence 702 (FRE 702)

<u>Rule 702</u>. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on **sufficient facts** or **data**;

(c) the testimony is the product of **reliable** principles and methods; and

(d) the expert has **reliably** applied the principles and methods to the facts of the case.

https://www.law.cornell.edu/rules/fre/rule_702

Daubert

Rule 702 has been amended in response to *Daubert v. Merrell Dow Pharmaceuticals, Inc*., 1993

- In *Daubert* the Court charged trial judges with the responsibility of acting as gatekeepers to **exclude unreliable expert testimony**, and

- the Court in *Kumho* clarified that this gatekeeper function applies to all expert testimony, not just testimony based in science (*Kumho Tire Co. v. Carmichael, 1999)*

https://www.law.cornell.edu/rules/fre/rule_702

Daubert

Daubert set forth a non-exclusive [*non-exhaustive?*] checklist for trial courts to use in assessing the **reliability of scientific expert testimony**.

The specific factors explicated by the *Daubert* Court are

1) whether the expert's technique or theory can be or has been tested—that is, whether the expert's theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability;

2) whether the technique or theory has been subject to peer review and publication;

3) the known or potential rate of error of the technique or theory when applied;

4) the existence and maintenance of standards and controls; and

5) whether the technique or theory has been generally accepted in the scientific community.

https://www.law.cornell.edu/rules/fre/rule_702

Federal Rules of Evidence 403 (FRE 403) New Slide

Rule 403. Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons

The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following:

(a) unfair prejudice,
(b) confusing the issues,
(c) misleading the jury,
(d) undue delay,
(e) wasting time, or
(f) needlessly presenting cumulative evidence.

Notes of Advisory Committee on Proposed Rules states:
…**"Unfair prejudice" within its context means an** *undue tendency to suggest decision on an improper basis*, **commonly, though not necessarily, an emotional one.   …**

(see https://www.law.cornell.edu/rules/fre/rule_702)

DNA Mixture Interpretation

Reliability Considerations

DNA: MEASUREMENT & INTERPRETATION
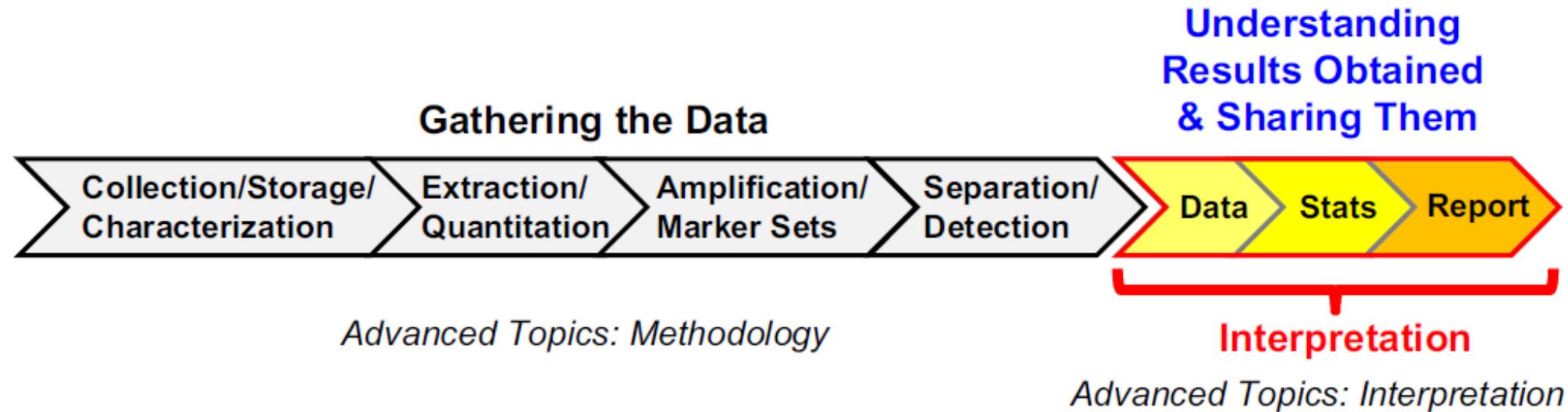
FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

JOHN M. BUTLER

National Institute of Standards and Technology
Gaithersburg, Maryland, USA

Advanced Topics in Forensic DNA Typing: Interpretation
http://dx.doi.org/10.1016/B978-0-12-405213-0.00001-4

DNA: MEASUREMENT & INTERPRETATION

SAMPLE

Gathering the Data

Understanding Results Obtained & Sharing Them

Collection/Storage/Characterization → Extraction/Quantitation → Amplification/Marker Sets → Separation/Detection → Data → Stats → Report

Weight of Evidence

End-to-End SYSTEM

DNA: MEASUREMENT & INTERPRETATION

FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

JOHN M. BUTLER

National Institute of Standards and Technology
Gaithersburg, Maryland, USA

Advanced Topics in Forensic DNA Typing: Interpretation
http://dx.doi.org/10.1016/B978-0-12-405213-0.00001-4

DNA: MEASUREMENT & INTERPRETATION

FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

JOHN M. BUTLER

National Institute of Standards and Technology
Gaithersburg, Maryland, USA

DNA: MEASUREMENT & INTERPRETATION

Gathering the Data

LR

Collection/Storage/Characterization → Extraction/Quantitation → Amplification/Marker Sets → Separation/Detection → Data → Stats → Report

FIGURE 1.1   Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

JOHN M. BUTLER

National Institute of Standards and Technology
Gaithersburg, Maryland, USA

DNA: MEASUREMENT & INTERPRETATION

PROPOSITIONS

H_p: DNA from POI is in the sample

H_d: DNA from POI is not in the sample
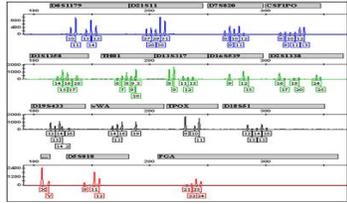
$$LR = \frac{Pr(E|H_p, I)}{Pr(E|H_d, I)}$$
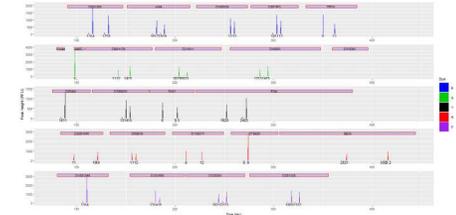
H_p: DNA from POI is in the sample

H_d: DNA from POI is not in the sample

E = Expert Findings

Expert's findings vs EPG

I = Background Information prior to examining crime sample

What is I ?

Any information that is taken to be "true" while assessing probability of interest.

- Background crime-related information prior to examining crime sample

- Statistical models used

- Number of contributors (if known)

- Etc.

$$LR = \frac{Pr(E|H_p, I)}{Pr(E|H_d, I)}$$

What is I ?

$$LR_{Needed} = \frac{Pr(E|H_p, I)}{Pr(E|H_d, I)}$$

$$LR_{Provided} = \frac{Pr(E|H_p, Ip, Iu)}{Pr(E|H_d, Ip, Iu)}$$

I_p consists of additional assumptions by the LR system

I_u consists of background (ideally, uncontestable) information

$$LR_{Provided} \overset{?}{=} LR_{Needed}$$

System Reliability vs Component Reliability
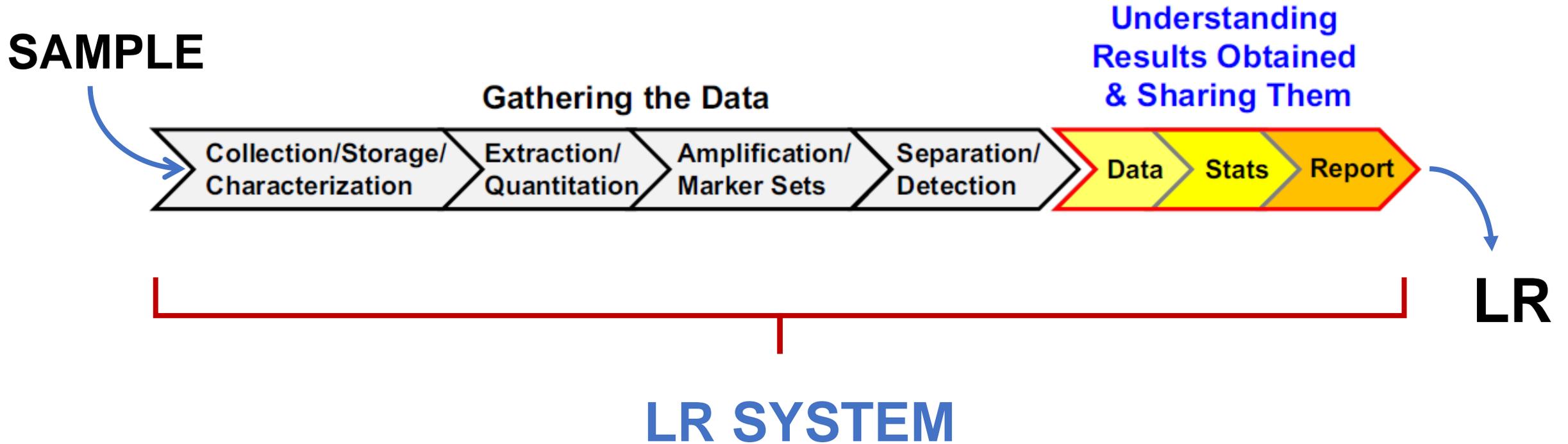
System Reliability vs Component Reliability
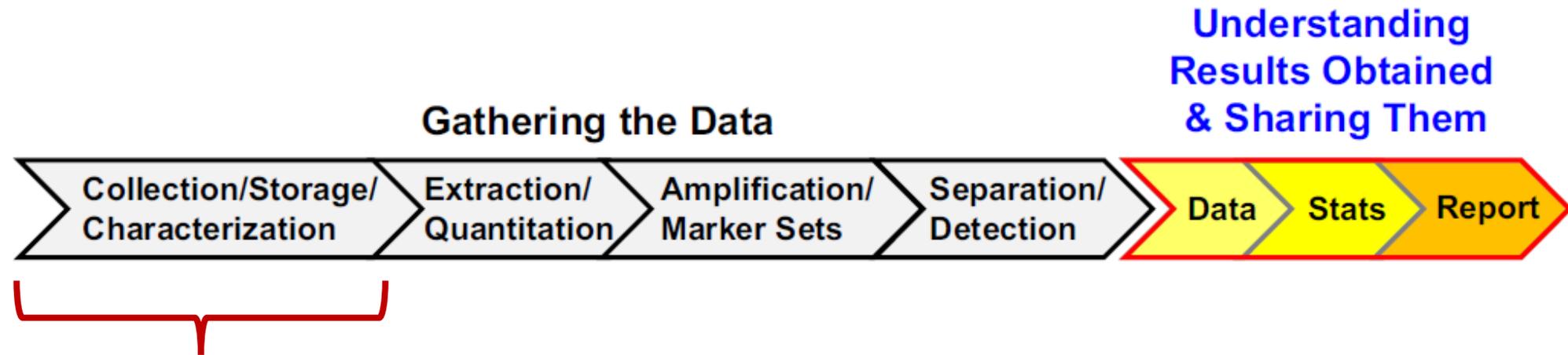
System Reliability vs Component Reliability
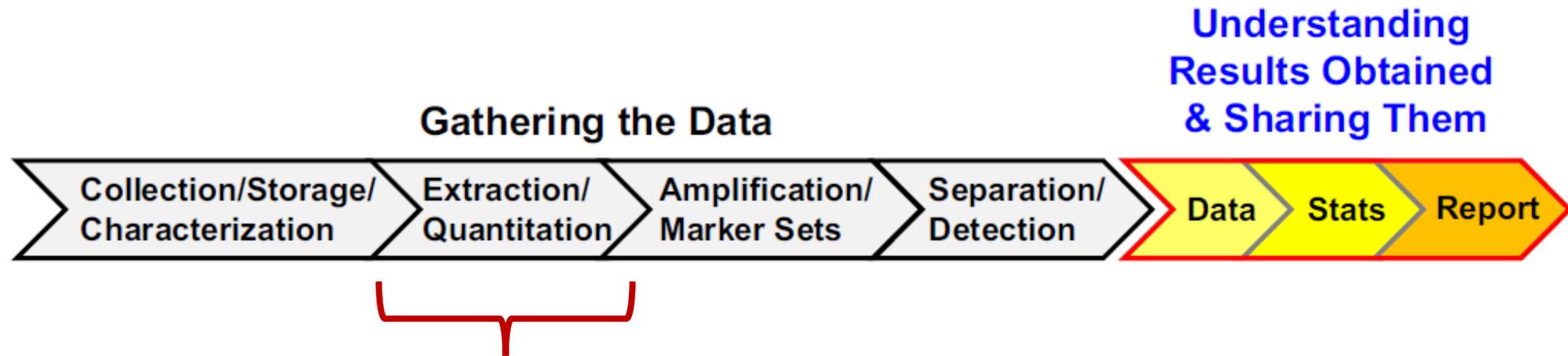
System Reliability vs Component Reliability
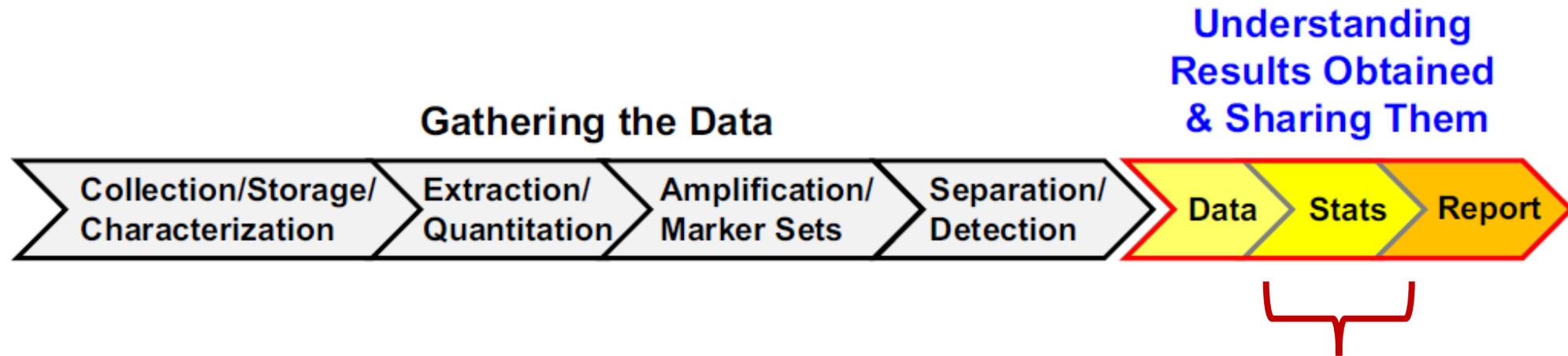
Some Factors Affecting Reliability of an LR System

1. Sample
   a) Sample amount (contributor template amounts)
   b) Sample quality (degradation level)
2. Labs
   a) Kits used
   b) Equipment Used
   c) Number of PCR cycles
   d) Analyst
   e) Choice of Analytical Threshold (AT)
3. Probabilistic Genotyping (PG) Model
   a) Choice of model
   b) Choice of laboratory specific parameters for use in the PG model
   c) **Propositions Chosen (H_p and H_d)**
4. Software Implementing the PG Model
   a) Choice of numerical methods for computing LR (MCMC, Numerical Integration)
   b) Choice of number of iterations OR numerical integration parameters (such as grid size)

FACTOR SPACE

RELIABILITY vs COMPLEXITY

RELIABILITY vs COMPLEXITY of the Mixture
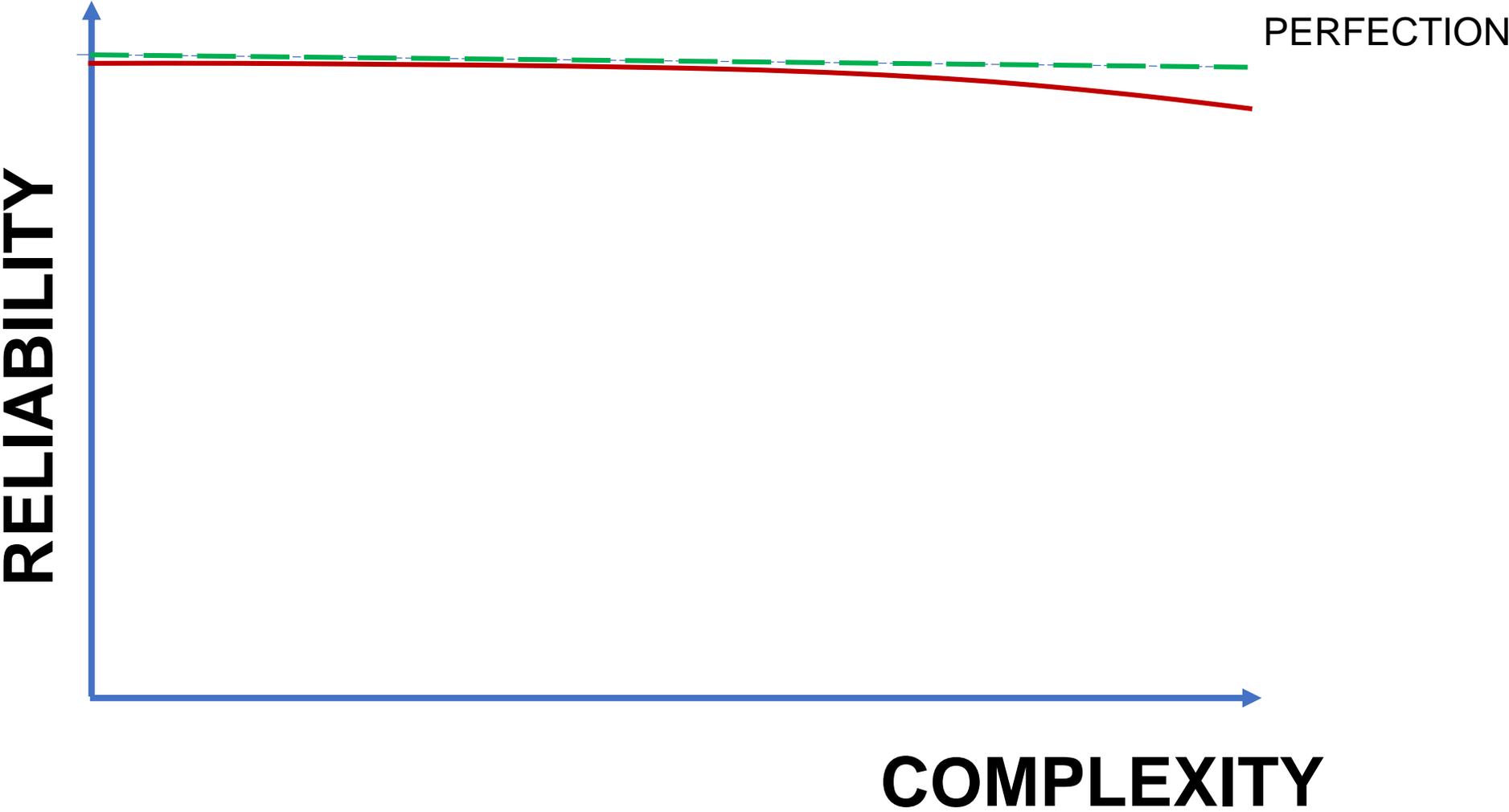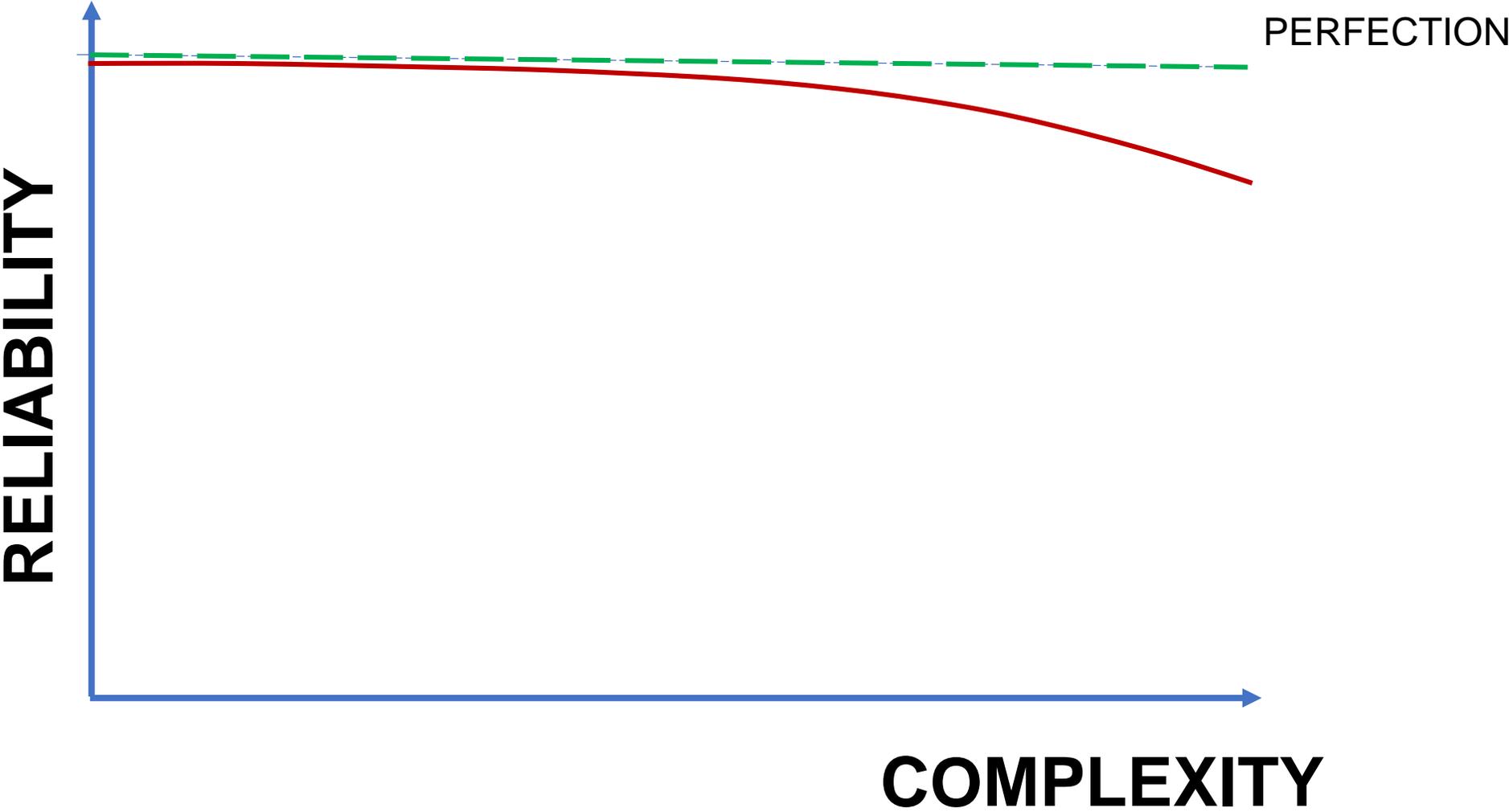
RELIABILITY vs COMPLEXITY of the Mixture
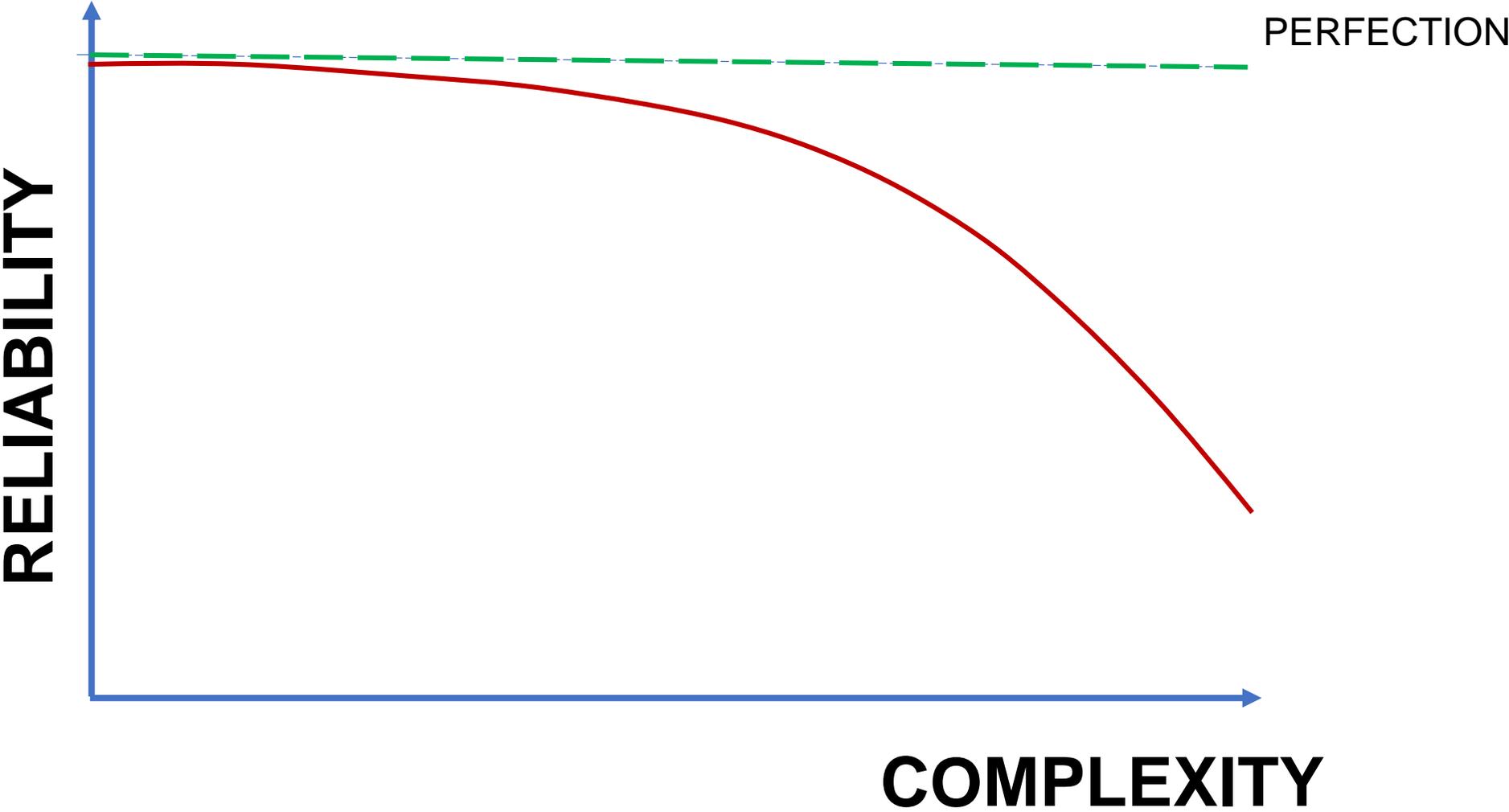
RELIABILITY vs COMPLEXITY of the Mixture
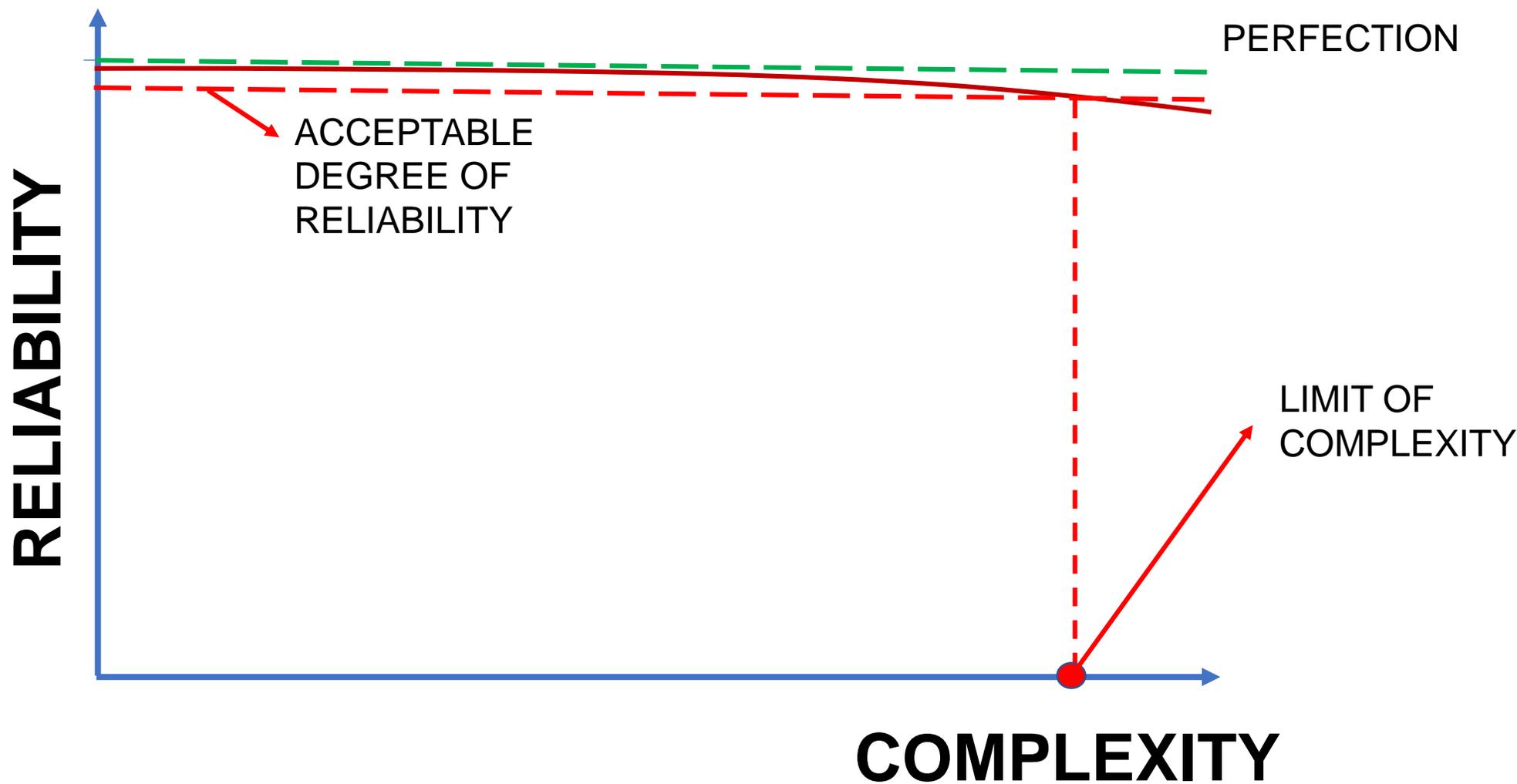
RELIABILITY vs COMPLEXITY of the Mixture
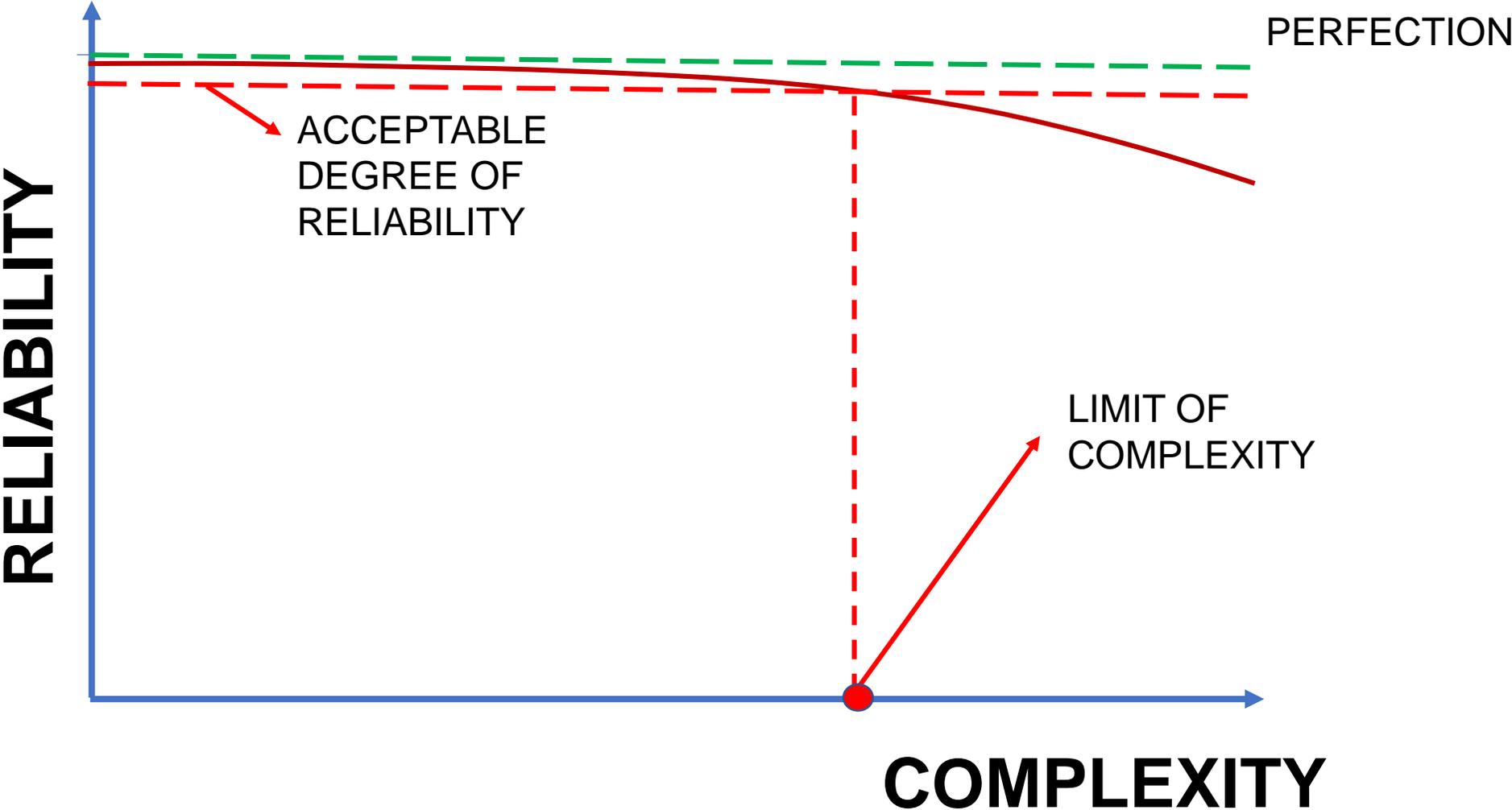
RELIABILITY vs COMPLEXITY of the Mixture
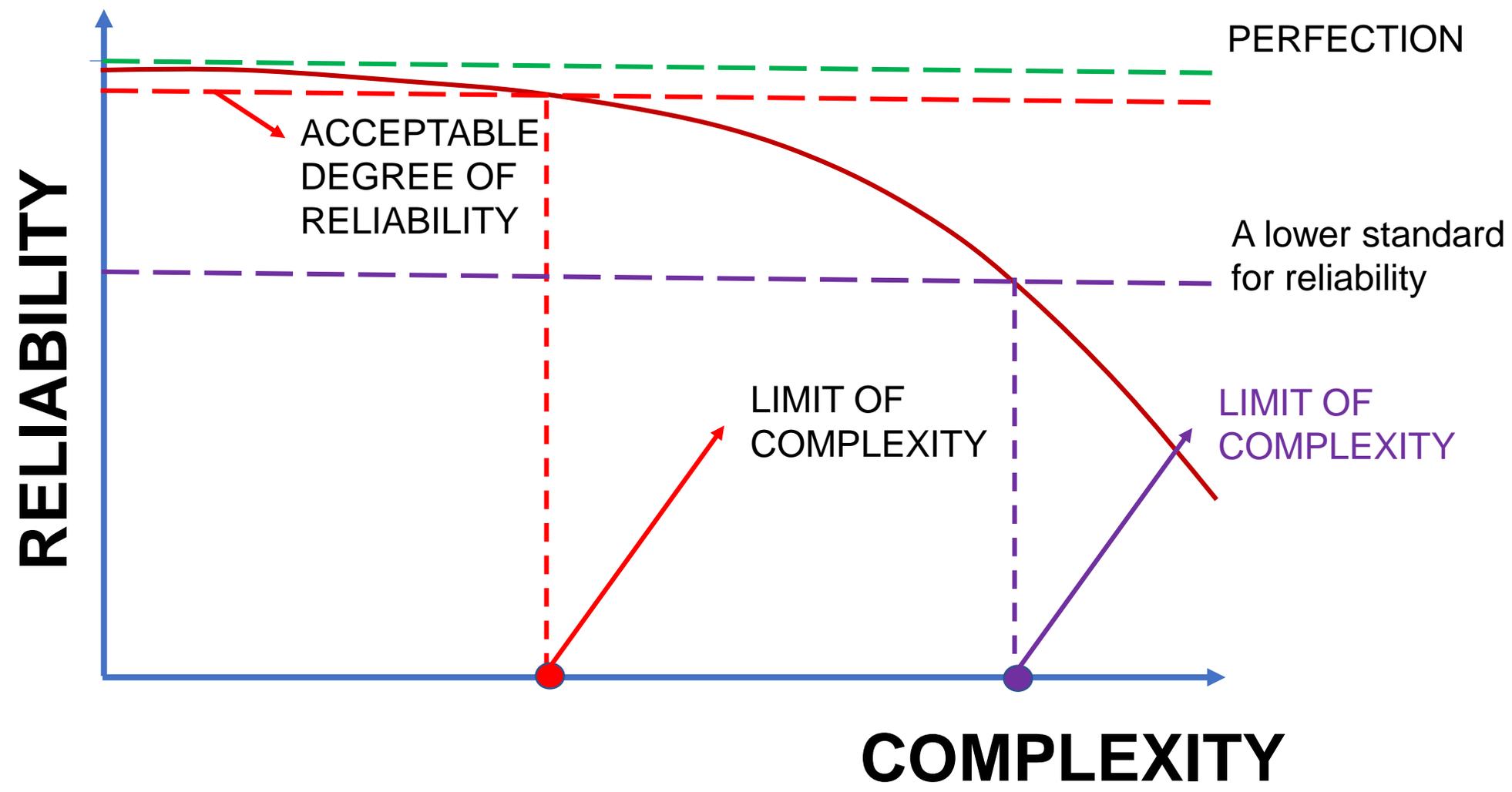
RELIABILITY in PRACTICE

Suppose we send portions of a DNA mixture sample to different DNA labs along with the profile of the defendant in this case.

Will the lab results be close enough to one another that differences between them would be inconsequential to the outcome of this case?

Can we trust that the result from this lab will be close enough to the 'true value' that difference between the reported value and the true value would be inconsequential to the outcome of this case?

RELIABILITY in PRACTICE

Will the lab results be close enough to one another that differences between them would be inconsequential to the outcome of this case?

Inter-lab Trials can help assess degree of agreement among labs

Can we trust that the result from this lab, in this case, will be close enough to the 'true value' that difference between the reported value and the true value would be inconsequential to the outcome of this case?

Trials using ground truth known (and casework like) samples can provide useful information to make this assessment.

EXAMPLE

Cholesterol measurement of 5 aliquots of a sample of blood from 5 labs

148 mg/dL

180 mg/dL

215 mg/dL

375 mg/dL

450 mg/dL

EXAMPLE

Cholesterol measurement of 5 aliquots of a sample of blood from 5 labs

346 mg/dL

351 mg/dL

362 mg/dL

355 mg/dL

366 mg/dL

WHAT IS THE TRUE VALUE?

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 911c

Cholesterol

Details	
Description:	Cholesterol
Lot:	N/A
Expiration Date:	12/31/2024
Unit Price *:	$814.00
Unit of Issue:	2 g

Certified Cholesterol Mass Faction: 99.2 % ± 0.4 %

This Standard Reference Material (SRM) is certified as a chemical of known purity. It is intended primarily for use in the calibration and standardization of procedures for the determination of cholesterol in research samples and for routine evaluations of daily working standards used in these procedures. A unit of SRM 911c consists of 2 g of material.

Certified Value: A NIST certified value is a value for which NIST has the highest confidence in its accuracy in that all known or suspected sources of bias have been investigated or taken into account [1]. This certified value is the equally weighted mean of results obtained from the analytical methods. The expanded uncertainty in the certified concentration is calculated as $U = ku_c$. The quantity u_c is the combined standard uncertainty calculated based on a Bayesian approach in reference 1 and the ISO/JCGM Guide [2]. The coverage factor, $k = 2$, represents an approximate 95 % level of confidence. The measurand is the total mass fraction of cholesterol. Metrological traceability to the SI derived unit for mass fraction (expressed as percent).

Cholesterol measurement of 5 aliquots of a sample of blood (NIST SRM 911c) from 5 labs

148 mg/dL

180 mg/dL

215 mg/dL

375 mg/dL

450 mg/dL

REFERENCE VALUE = 180 mg/dL !!

(test sample traceable to the SRM)

WHY DISCUSS RELIABILITY?

Cholesterol measurement of 5 aliquots of a sample of blood from 5 labs

346 mg/dL
351 mg/dL
362 mg/dL
355 mg/dL
366 mg/dL

REFERENCE VALUE = 180 mg/dL !!

WHY DISCUSS RELIABILITY?

Cholesterol measurement of 5 aliquots of a sample of blood from 5 labs

346 mg/dL
351 mg/dL
362 mg/dL
355 mg/dL
366 mg/dL

REFERENCE VALUE = 180 mg/dL

DEVIATIONS (ERRORS):   166, 171, 182, 175, 186

RELIABILITY

Degree of agreement among a group of labs by itself does not characterize degree of reliability

but

Degree of agreement with respect to a true value or a highly trusted reference value, on a consistent basis, is what characterizes reliability

WHY DISCUSS RELIABILITY?

ELSEVIER

FSI GENETICS

Research paper

GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: Results and evaluation

Check for updates

P.A. Barrio[a,b,c], M. Crespillo[a,c,*], J.A. Luque[a,c], M. Aler[d], C. Baeza-Richer[e], L. Baldassarri[f], E. Carnevali[g], P. Coufalova[h], I. Flores[i], O. García[j], M.A. García[k], R. González[l], A. Hernández[m], V. Inglés[n], G.M. Luque[b], A. Mosquera-Miguel[o], S. Pedrosa[p], M.L. Pontes[q], M.J. Porto[r], Y. Posada[s], M.I. Ramella[t], T. Ribeiro[u], E. Riego[v], A. Sala[w], V.G. Saragoni[x], A. Serrano[c], S. Vannelli[y]

Participants were provided with the thresholds values used/employed: analytical threshold of 50 RFUs, stochastic threshold of 150 RFUs, and stutter threshold for each of the markers/kits according to the manufacturer's specifications.

GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: Results and evaluation

Table 1

Hypothesis and LR values obtained by each of the participating laboratories. All laboratories used the *LRmixStudio* software, except those marked as * (*EuroForMix*) and ** (*DNAMIX*). Legend: V (Victim), S (Suspect), P (Regular partner), U (Unknown).

Labs	LR value	Hypothesis	Other evaluations	
			LR value	Hypothesis
GHEPMIX_08*	1.7200E + 02	V + S + P/V + U + P		
GHEPMIX_23	2.6000E + 03	V + S + P/V + U + P		
GHEPMIX_26	6.1640E + 03	V + S + P/V + U + P		
GHEPMIX_17	6.5565E + 04	V + S + P/V + U + P		
GHEPMIX_07	6.8487E + 04	V + S + P/V + U + P		
GHEPMIX_05	1.4800E + 05	V + S + P/V + U + P		
GHEPMIX_22	2.8776E + 05	V + S + P/V + U + P		
GHEPMIX_06	3.2224E + 05	V + S + P/V + U + P		
GHEPMIX_16	4.3423E + 05	V + S + P/V + U + P		
GHEPMIX_18	1.3900E + 06	V + S + P/V + U + P		
GHEPMIX_03	1.8200E + 06	V + S + P/V + U + P		
GHEPMIX_02	2.7323E + 06	V + S + P/V + U + P		
GHEPMIX_20	5.5183E + 06	V + S + P/V + U + P		
GHEPMIX_15	1.9820E + 07	V + S + P/V + U + P		
GHEPMIX_27	1.3587E + 08	V + S + P/V + U + P	7.4048E + 19	P/U
GHEPMIX_13**	2.7300E + 10	V + S + P/V + U + P		
GHEPMIX_10	3.2032E + 14	V + S + P/V + U + P	1.1551E + 07	V + S + P/V + U1 + U2
GHEPMIX_24			1.3400E + 19	V + P/V + U

WHY DISCUSS RELIABILITY?

Contents lists available at ScienceDirect

Forensic Science International: Genetics

ELSEVIER

journal homepage: www.elsevier.com/locate/fsigen

GENETICS

DNA mixtures interpretation – A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples

E. Alladio[a,b,*], M. Omedei[b], S. Cisana[b], G. D'Amico[b], D. Caneparo[b], M. Vincenti[a,b], P. Garofano[b,c]

[a] Dipartimento di Chimica, Università degli Studi di Torino, Via P. Giuria 7, 10125, Torino, Italy
[b] Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Regione Gonzole 10/1, 10043, Orbassano, Torino, Italy
[c] Accademia Italiana di Scienze Forensi, Viale Regina Margherita 9/D, 42124, Reggio Emilia, Italy

**Lab Retriever
LRmix Studio**

**DNA•VIEW®,
EuroForMix and
STRmix**

Page 145

Furthermore, log(LR) results provided by fully-continuous models proved similar and convergent to one another, with slightly higher within-software differences (i.e. approximatively 3–4 degrees of magnitude).

A factor of 1000 to 10000 ?

WHY DISCUSS RELIABILITY?

Effect of 3 to 4 orders of magnitude:

Suppose prior odds = 1: 1000000   = (1/1,000,000)
(Crime occurred in the city of New York, say)

LR1 = 50000   (Strong evidence)

LR2 = 50000000 (Very Strong Evidence)   [a factor of 1000 higher than LR1]

Posterior Probability 1 = 0.048 = 4.8%

Posterior Probability 2 = 0.98   = 98%

Posterior Odds = Prior Odds x LR

$$\text{Posterior Probability} = \frac{(\text{LR x prior odds})}{1 + (\text{LR x prior odds})}$$

Then, is any value an acceptable value for LR?

No.

Some LR Systems may be considered sufficiently reliable for use in case work and others may be considered unreliable.

VALIDATION

It is often claimed that a model or a process has been 'validated' (implying reliability)

VALIDATION is not a binary concept. Nor is it universal.

The purpose of an LR SYSTEM VALIDATION STUDY is to gather sufficient information regarding the process used to make weight of evidence assessments and to provide empirical support for judging how reliable the process might be when applied to a given casework situation.

For any given casework application an LR system may be highly reliable, somewhat reliable, or unreliable, etc. A good validation study would collect data that will allow the receiver of the weight of evidence information to make these reliability judgements.

A Validation study cannot give a PASS/FAIL verdict unless the 'limits' of applicability and error rate thresholds are explicitly stated.

EMPIRICAL ASSESSMENT OF LR SYSTEMS

There are two aspects to judging the reliability of an LR system for assessing value of forensic DNA evidence

1. Accuracy of Claim or **Calibration**

2. Discrimination power

H_p: DNA from POI is in the sample

H_d: DNA from POI is not in the sample

ACCURACY/CALIBRATION

- LR value of 1 occurs as often under H_p as it does under H_d

- LR value of 10 occurs 10 times more often under H_p than it does under H_d.

- LR value of 100 occurs 100 times more often under H_p than it does under H_d.

- LR value of 0.1 occurs 10 times more often under H_d than it does under H_p.

LR value of x is x times more likely to occur under H_p than under H_d.

LR value of x is x times more likely to occur under H$_p$ than under H$_d$.
(LR of LR is LR)

> ..... the likelihood ratio of the likelihood ratio is the likelihood ratio. That is
>
> $$l[l(e_k)] = \frac{P_1[l(e_k) \mid h_1]}{P_2[l(e_k) \mid h_2]} = l(e_k) \qquad (1.32)$$
>
> for all events e_k.
>
> Green and Swets, 1966, page 26, section 1.8, equation (1.32)

In principle, this property can be empirically tested

ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT

- Suppose we have a large collection of ground truth known DNA samples representing different scenarios (degradation, number of contributors, template amounts) we expect to encounter in case work

- For each sample, select a known contributor profile or a known noncontributor profile (say by coin toss) and send them through the LR pipeline, from analysis to interpretation. (blinded)

- Record the value of LR obtained along with whether it is for an H_p true case or for an H_d true case.

- At the end of this exercise we will have a pool of H_p true LR values and a pool of H_d true LR values.

ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT

Known Noncontributors LR	Known Contributors LR
0.00E+00	3.68E+08
6.69E-03	2.10E+07
1.48E-03	7.34E+10
1.60E-03	1.26E+09
1.04E+00	1.45E+08
0.00E+00	3.87E+10
1.32E-01	3.12E+07
3.98E-03	1.71E+06
1.12E-02	6.56E+10
1.85E-06	1.95E+08
1.56E-01	1.61E+06
5.48E-09	4.13E+10
3.97E-04	1.87E+08
0.00E+00	1.11E+06
6.07E-13	5.18E+09
5.03E-04	2.99E+07
7.10E-03	1.87E+05
0.00E+00	1.86E+09
0.00E+00	8.08E+08
5.81E-01	7.17E+17
8.81E-08	5.81E+13
1.32E-01	2.76E+09
2.26E-14	3.18E+17
2.12E-01	4.66E+13
2.78E-01	4.78E+07

ETC

1.21E+00	1.01E+17
1.09E-03	1.16E+12
2.09E-13	1.41E+06
0.00E+00	9.87E+16
0.00E+00	2.61E+10
2.60E-01	2.34E+03
1.37E-04	1.05E+16

Noncontributors LR

Contributors LR
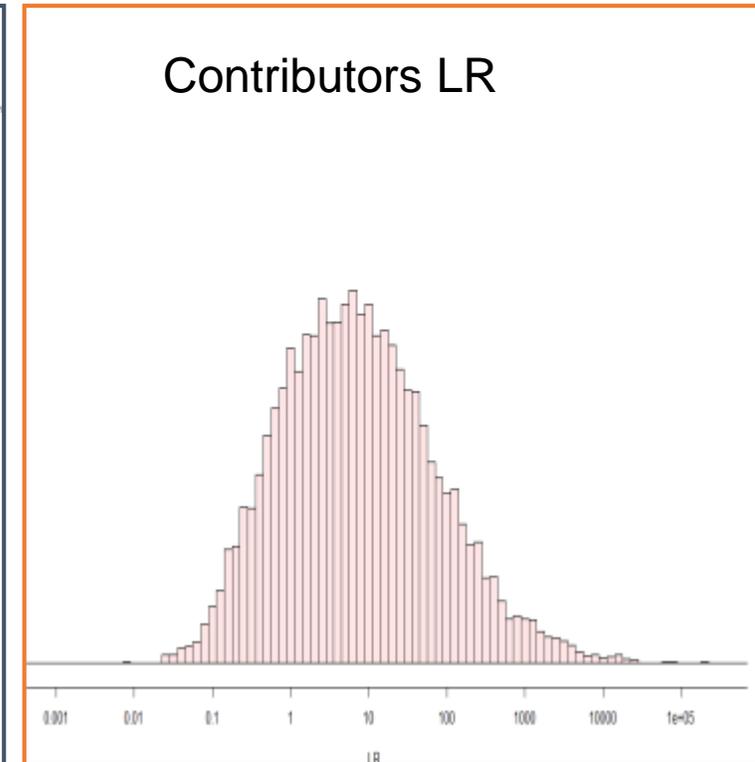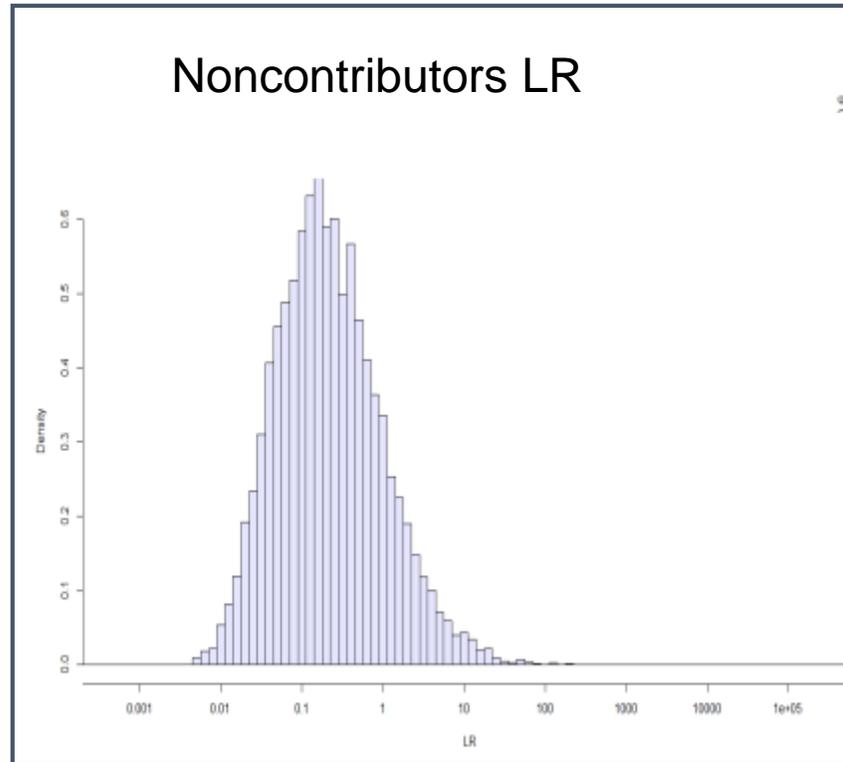
Disclaimer: This is only a thought experiment.
Actual assessment will require a well thought out experimental design.

ACCURACY/CALIBRATION

Conceptually:

- check whether an LR value of 1 occurs (roughly) equally often under H_p and under H_d

- check whether an LR value of 10 occurs (roughly) 10 times more often under H_p than it does under H_d.

- Check whether an LR value of 100 occurs (roughly) 100 times more often under H_p than it does under H_d.

- Check whether an LR value of 0.1 occurs (roughly) 10 times more often under H_d than it does under H_p.

- And, **in general, whether an LR value of x occurs (roughly) x times more often under H_p than under H_d.**
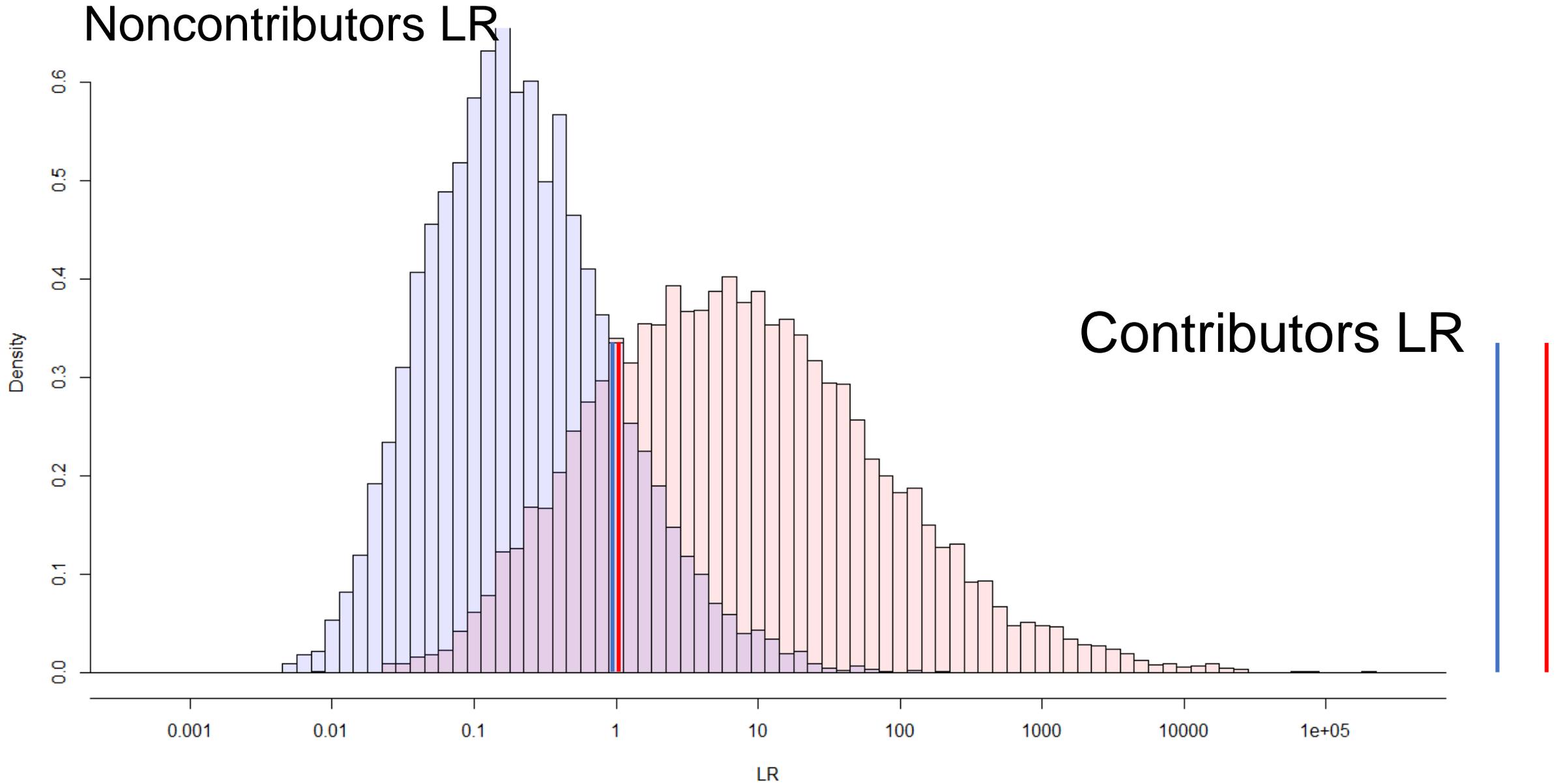
Noncontributors LR

Contributors LR

ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT

Noncontributors LR

Contributors LR
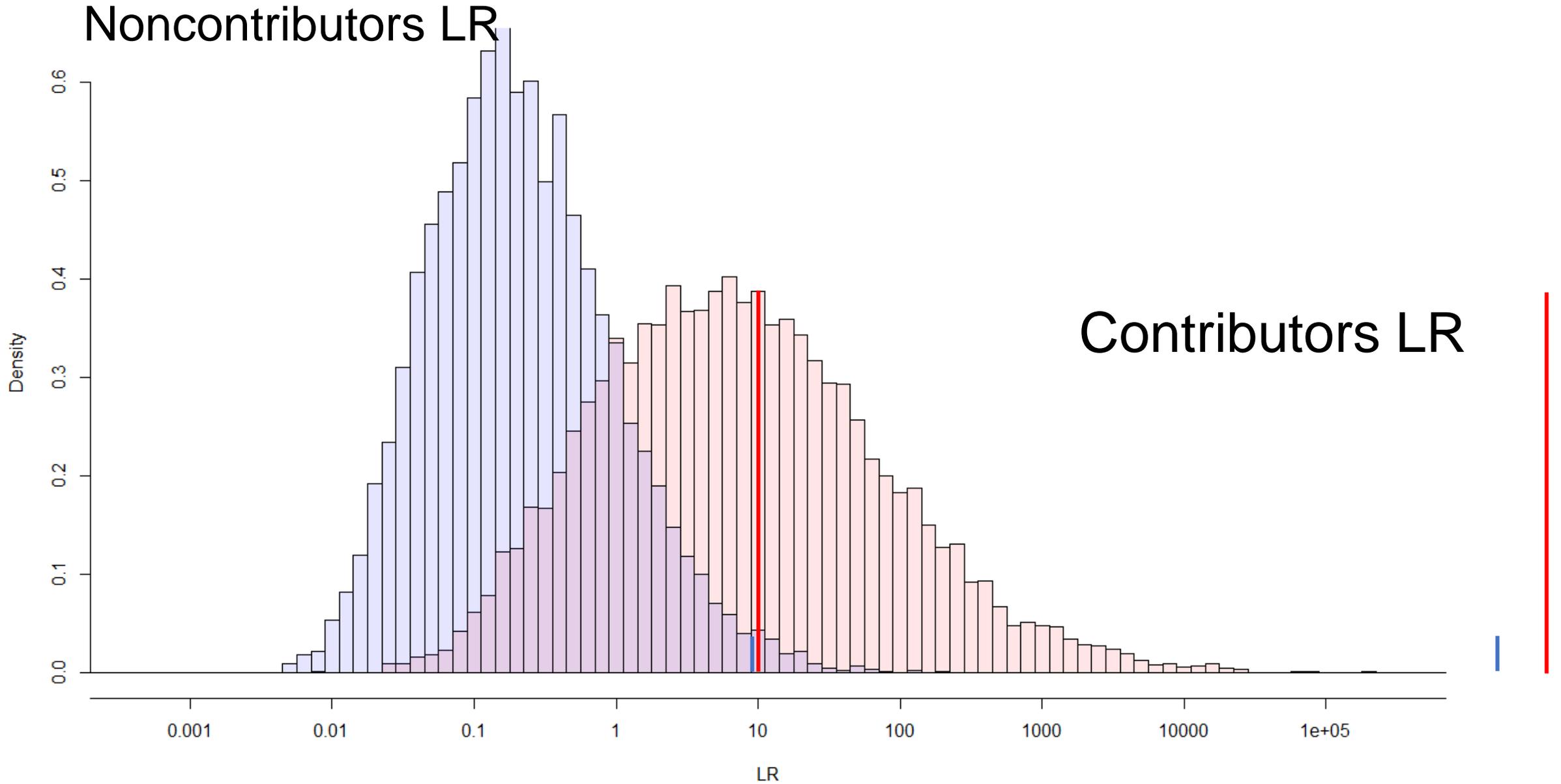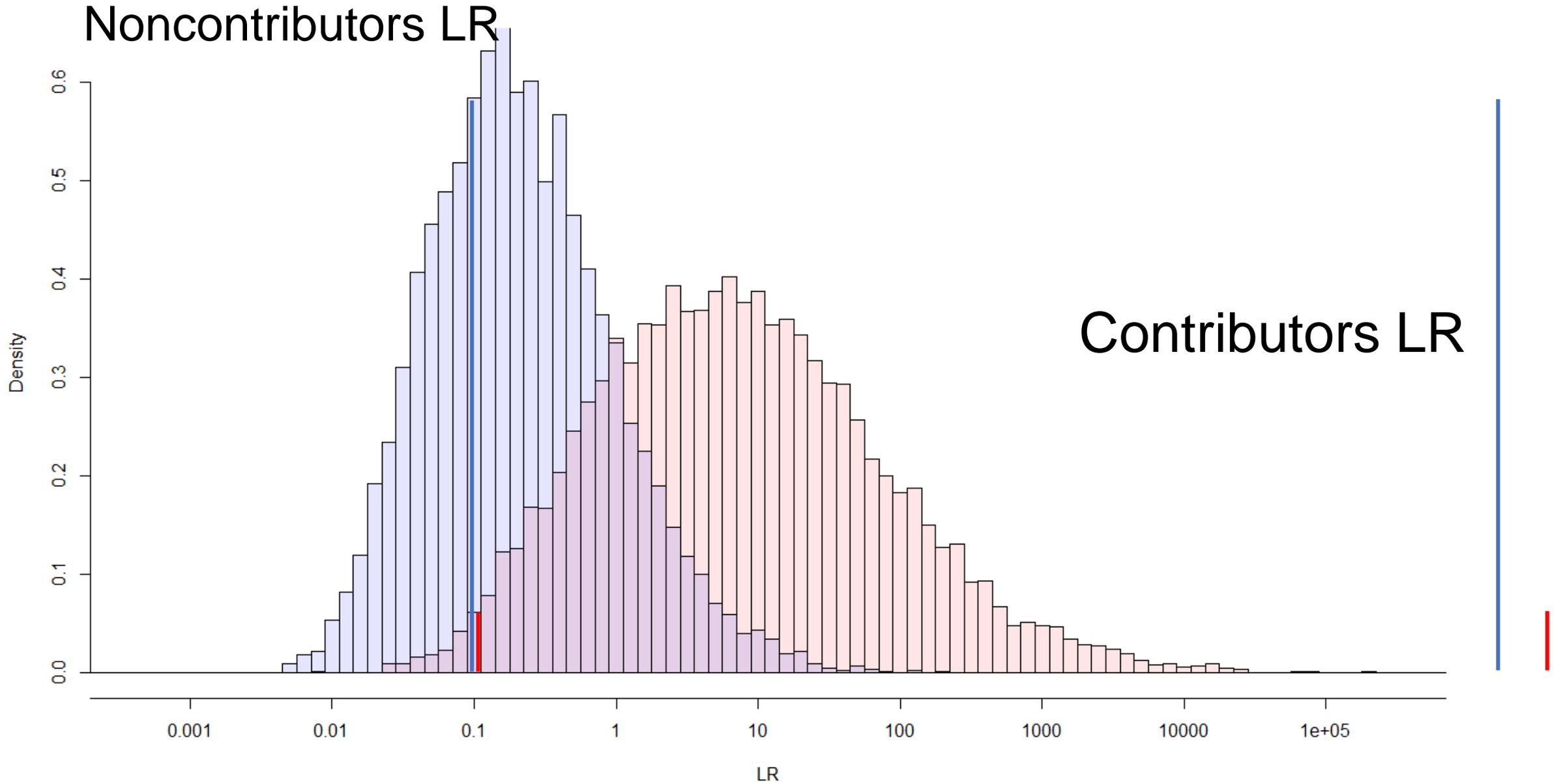
Example Where Calibration Check Fails

Noncontributors LR

A Different LR system
Fails Calibration Check

Contributors LR
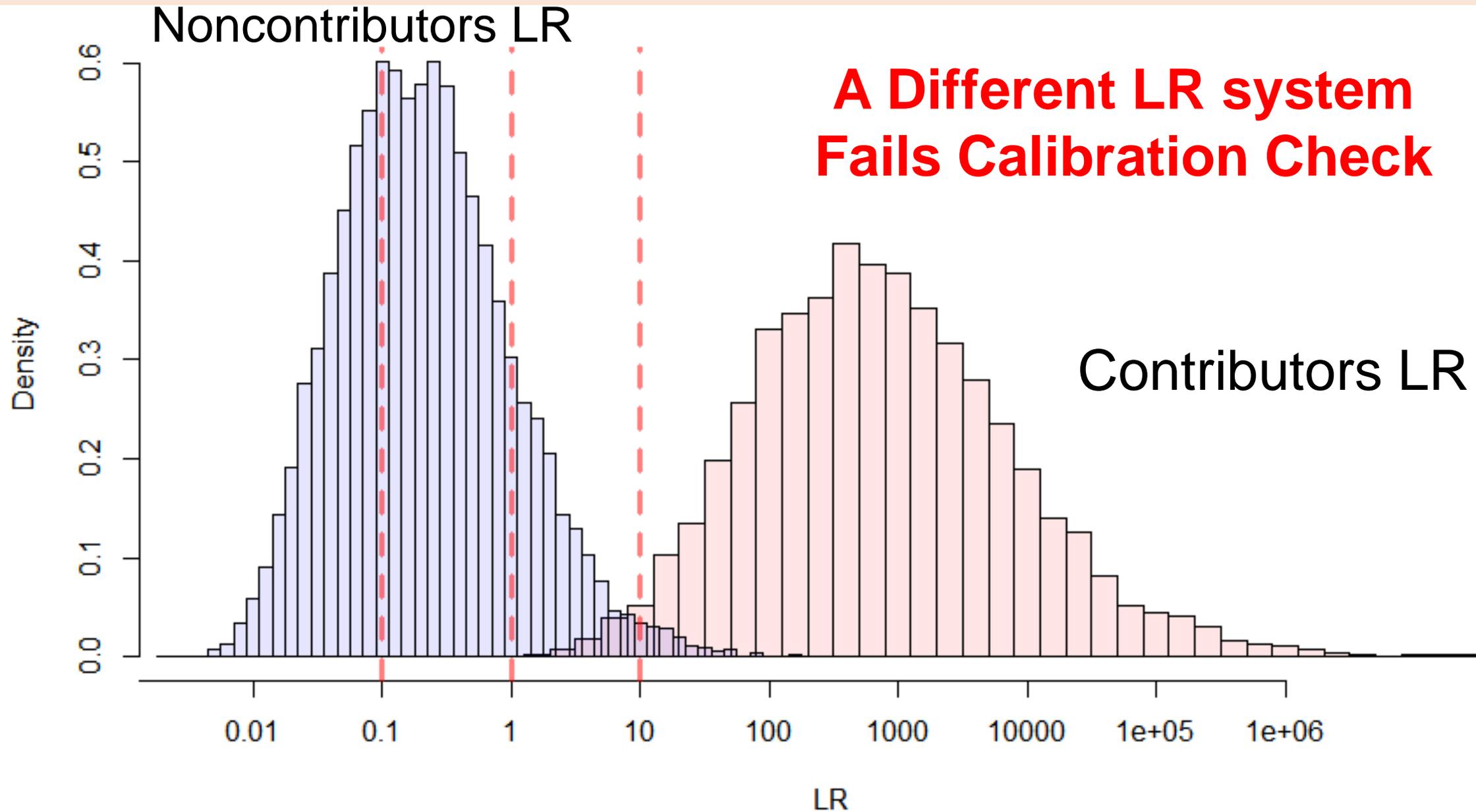
DISCRIMINATING POWER

The ability of an LR system to discriminate between H_p and H_d depends on

1. How much of the discriminating information in the sample is measured or extracted?

2. Does the interpretation make effective use of such information?

DISCRIMINATION POWER
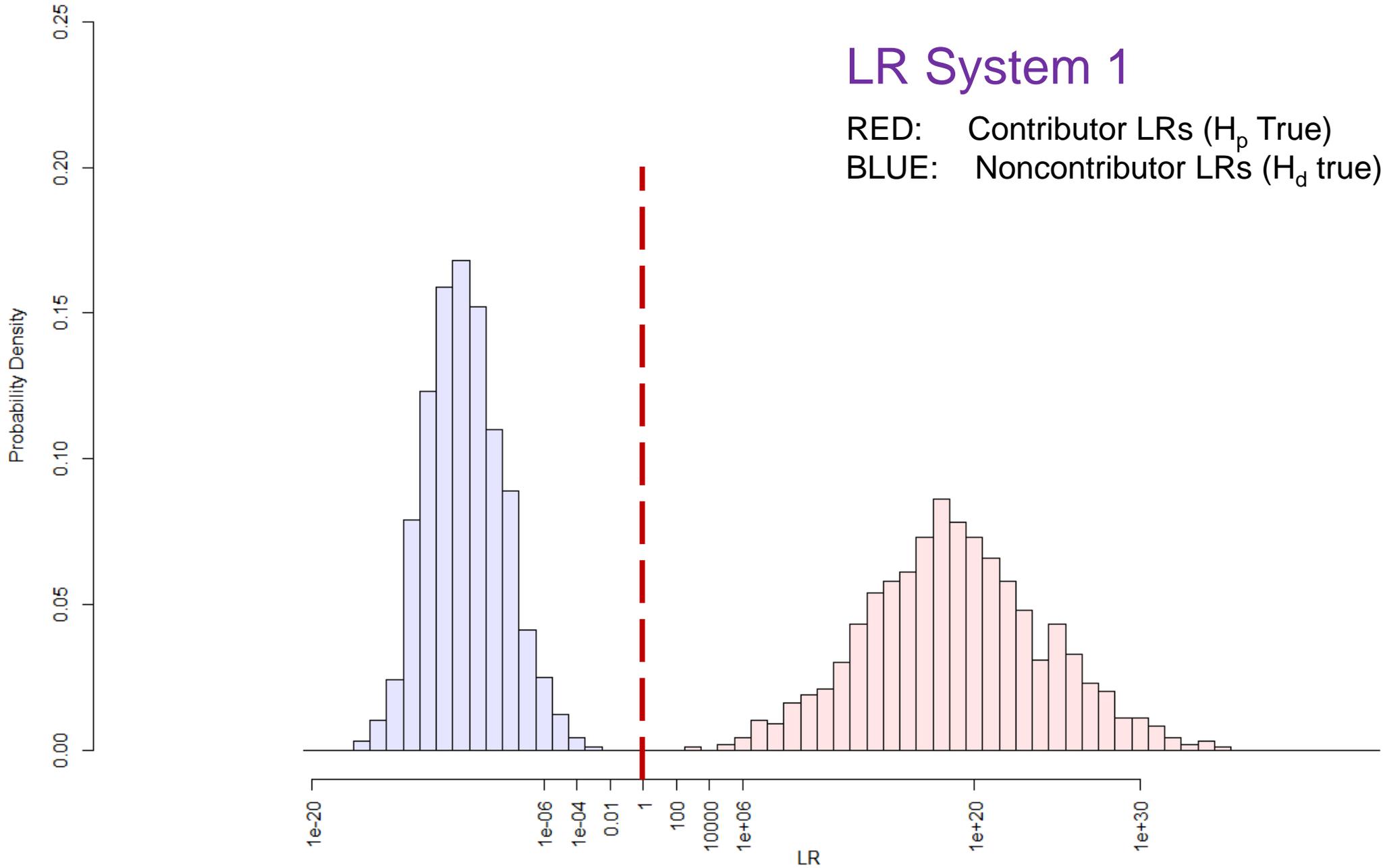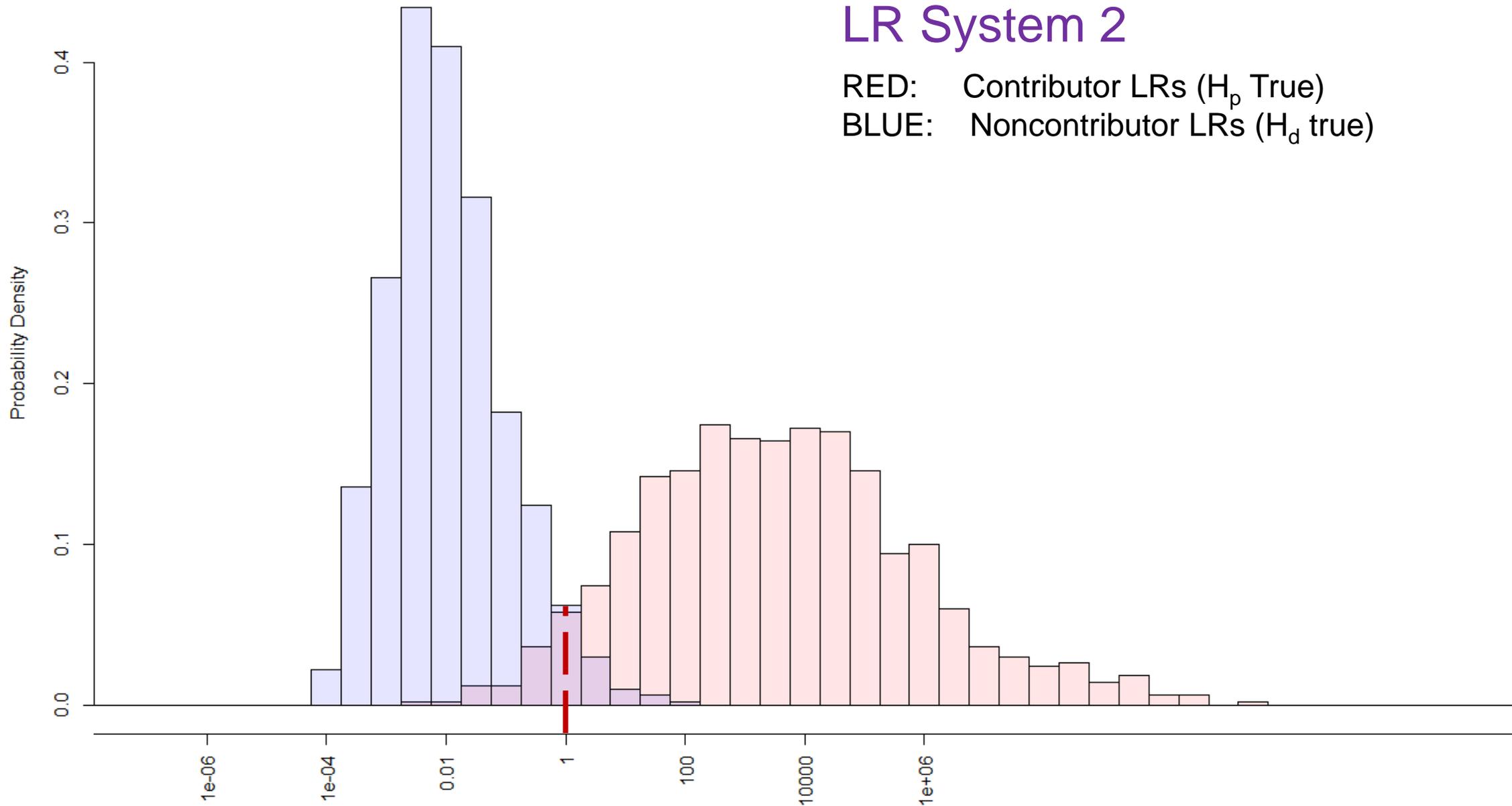
LR System 1

RED: Contributor LRs (H_p True)
BLUE: Noncontributor LRs (H_d true)

DISCRIMINATION POWER

LR System 2

RED: Contributor LRs (H_p True)
BLUE: Noncontributor LRs (H_d true)

DISCRIMINATION POWER

LR System 2

RED: Contributor LRs (H_p True)
BLUE: Noncontributor LRs (H_d true)
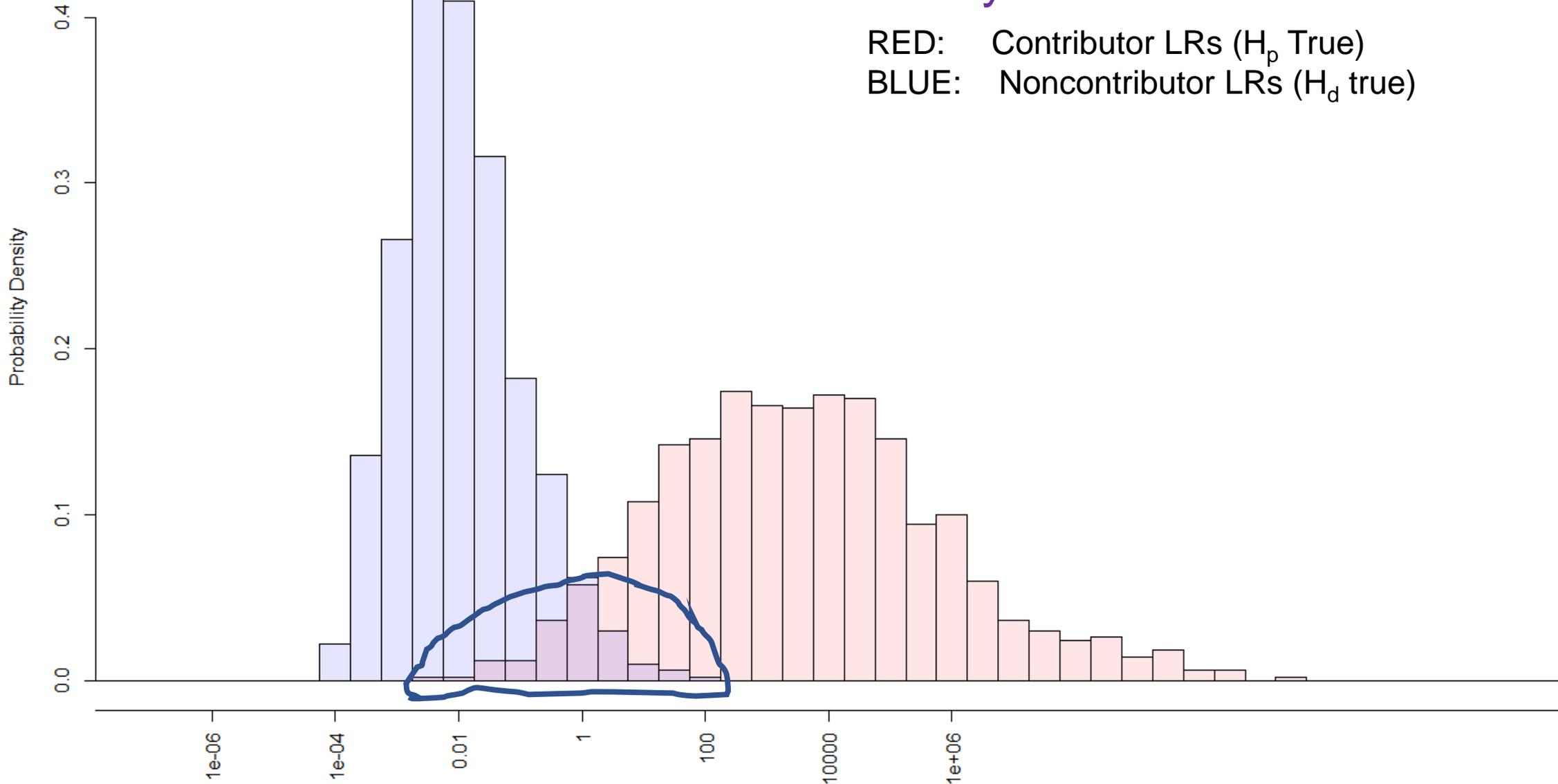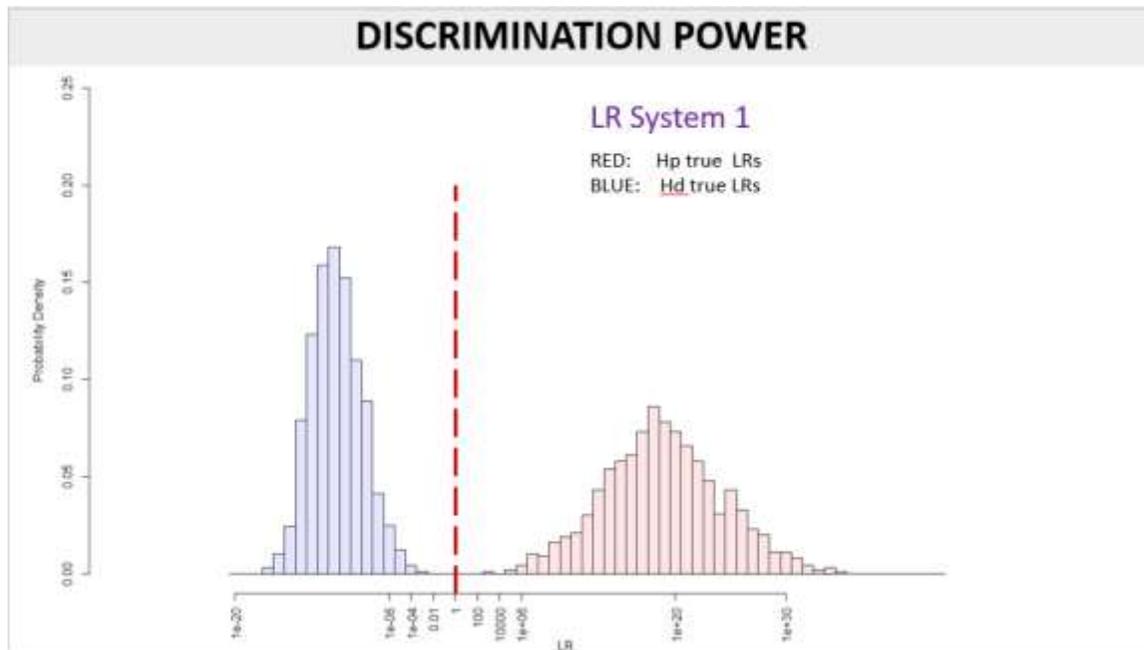
LR System 1
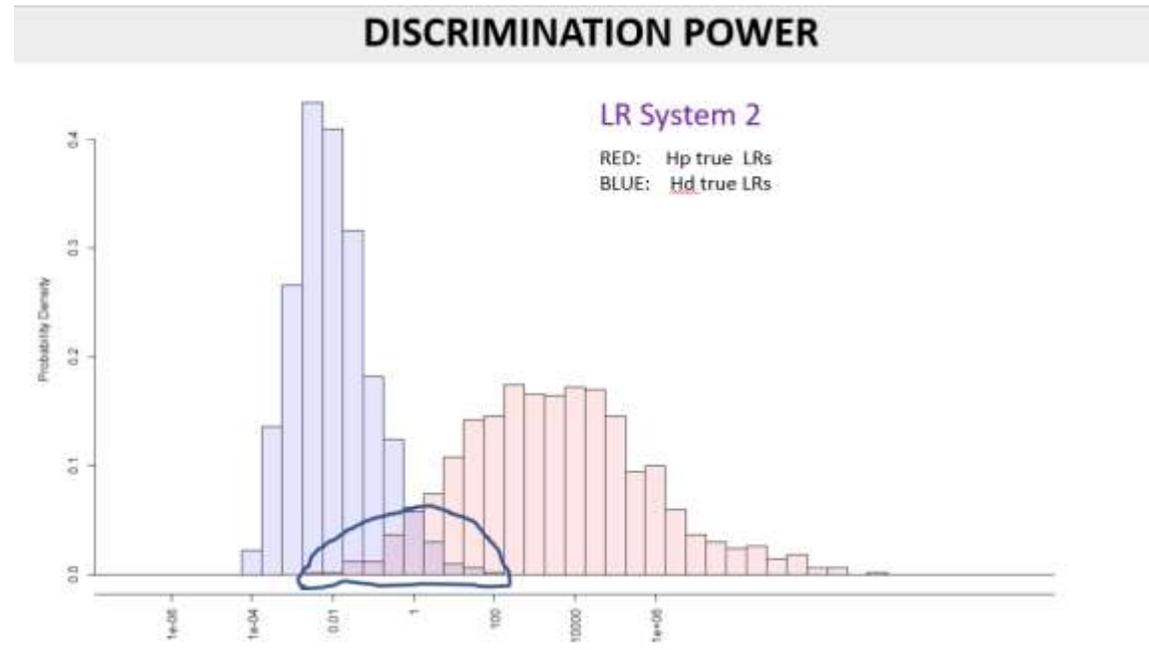
LR System 2

LR System 1 is more discriminating between H_p and H_d than LR system 2

ROC = Receiver Operating Characteristic
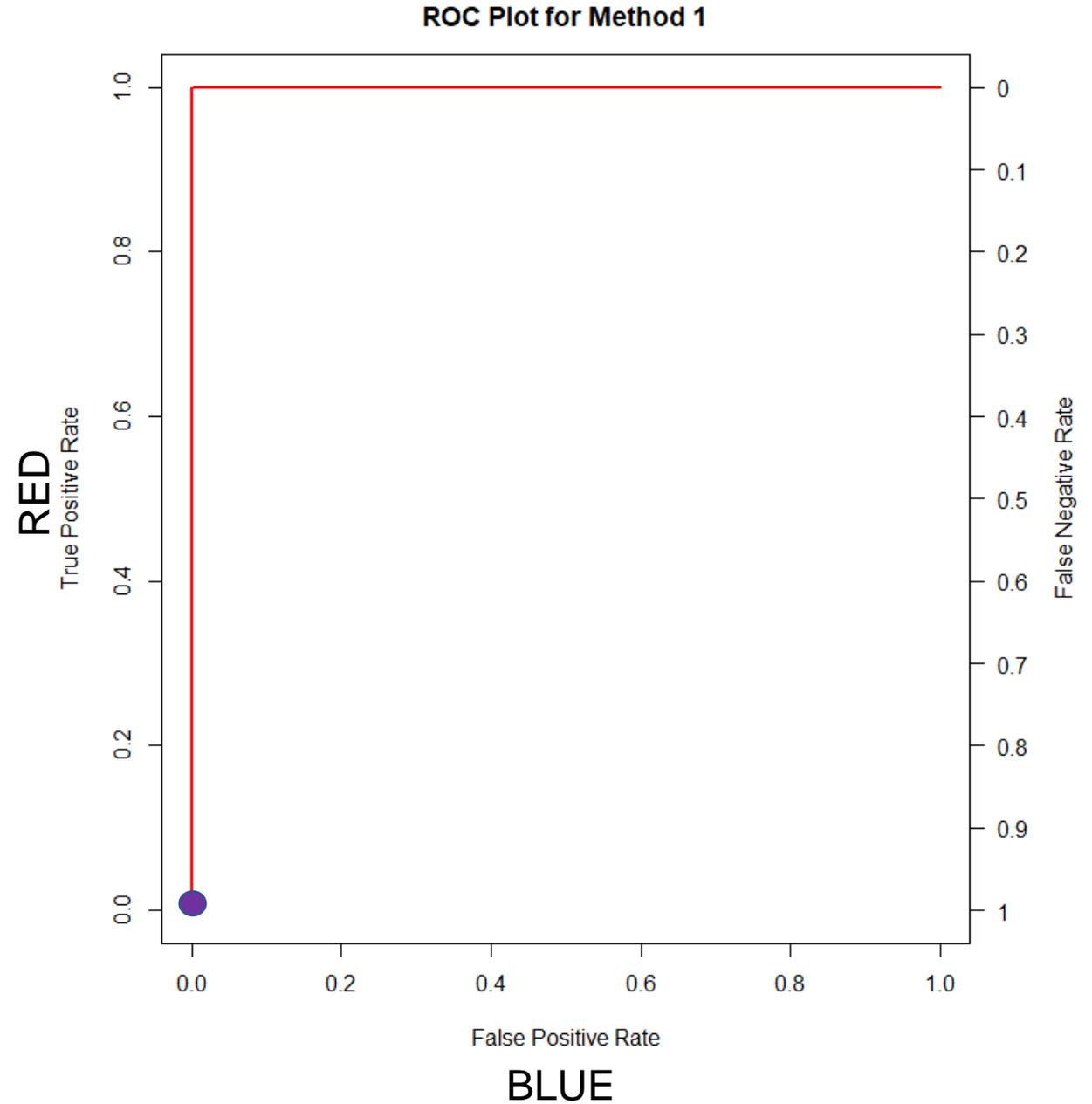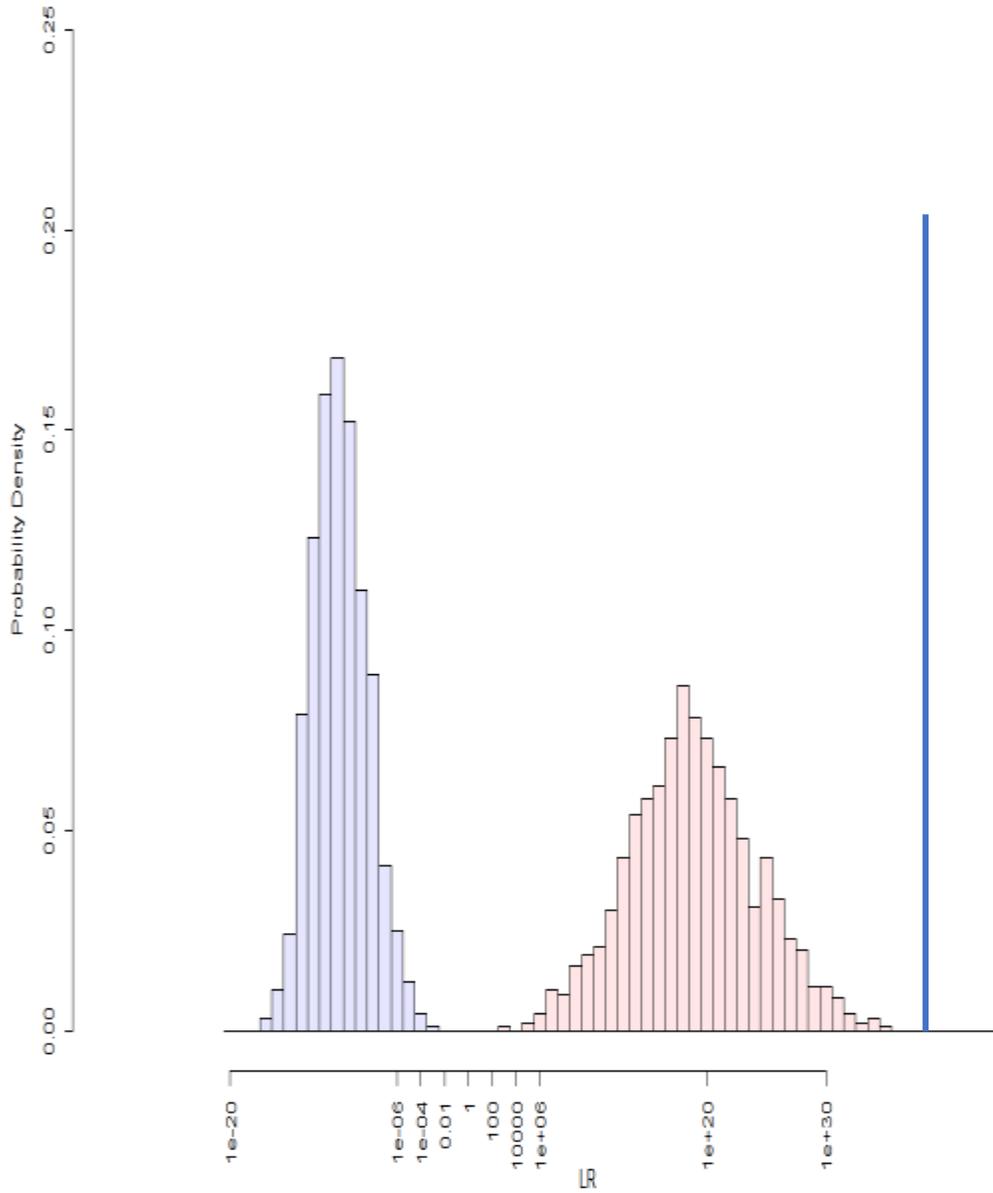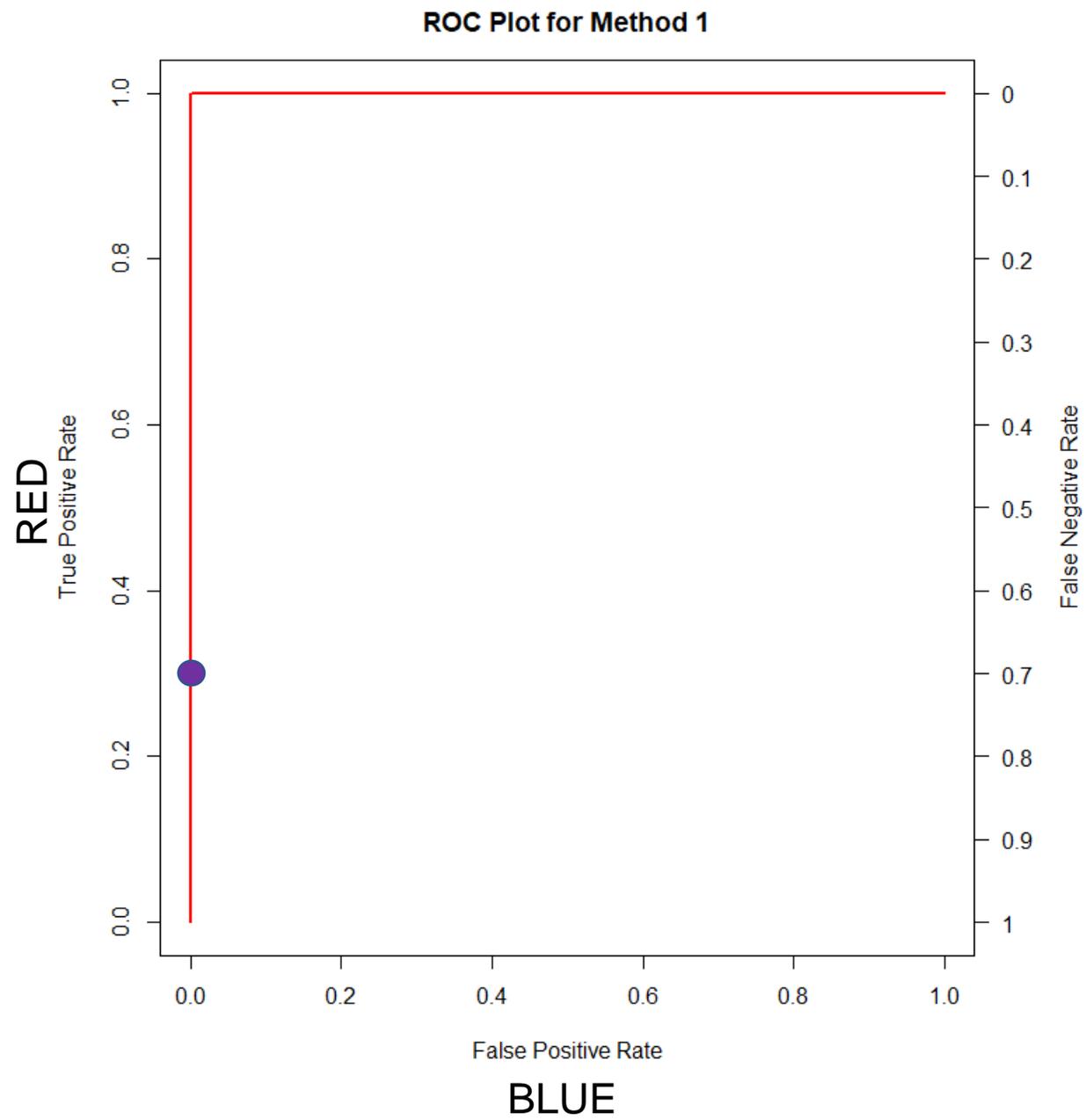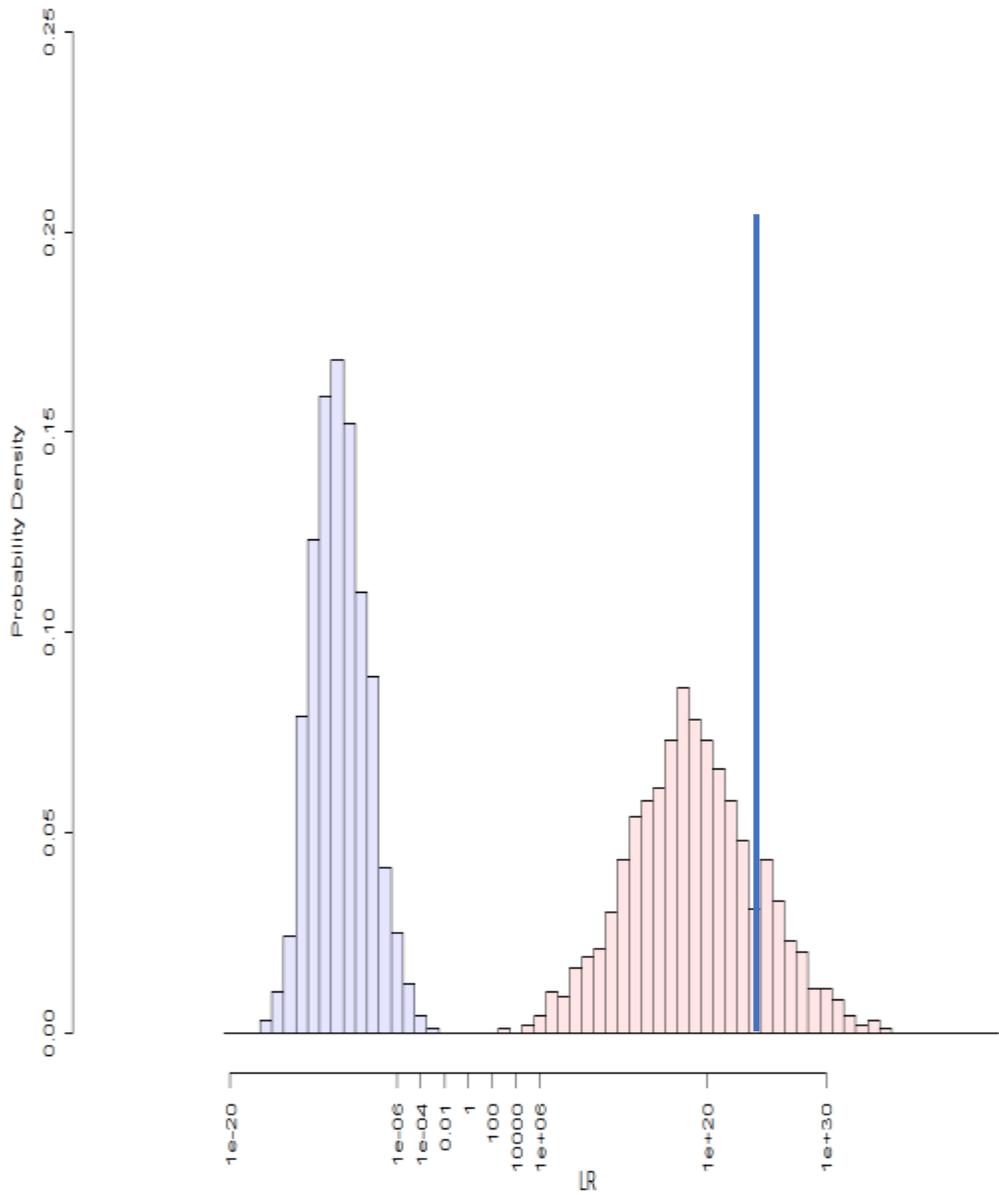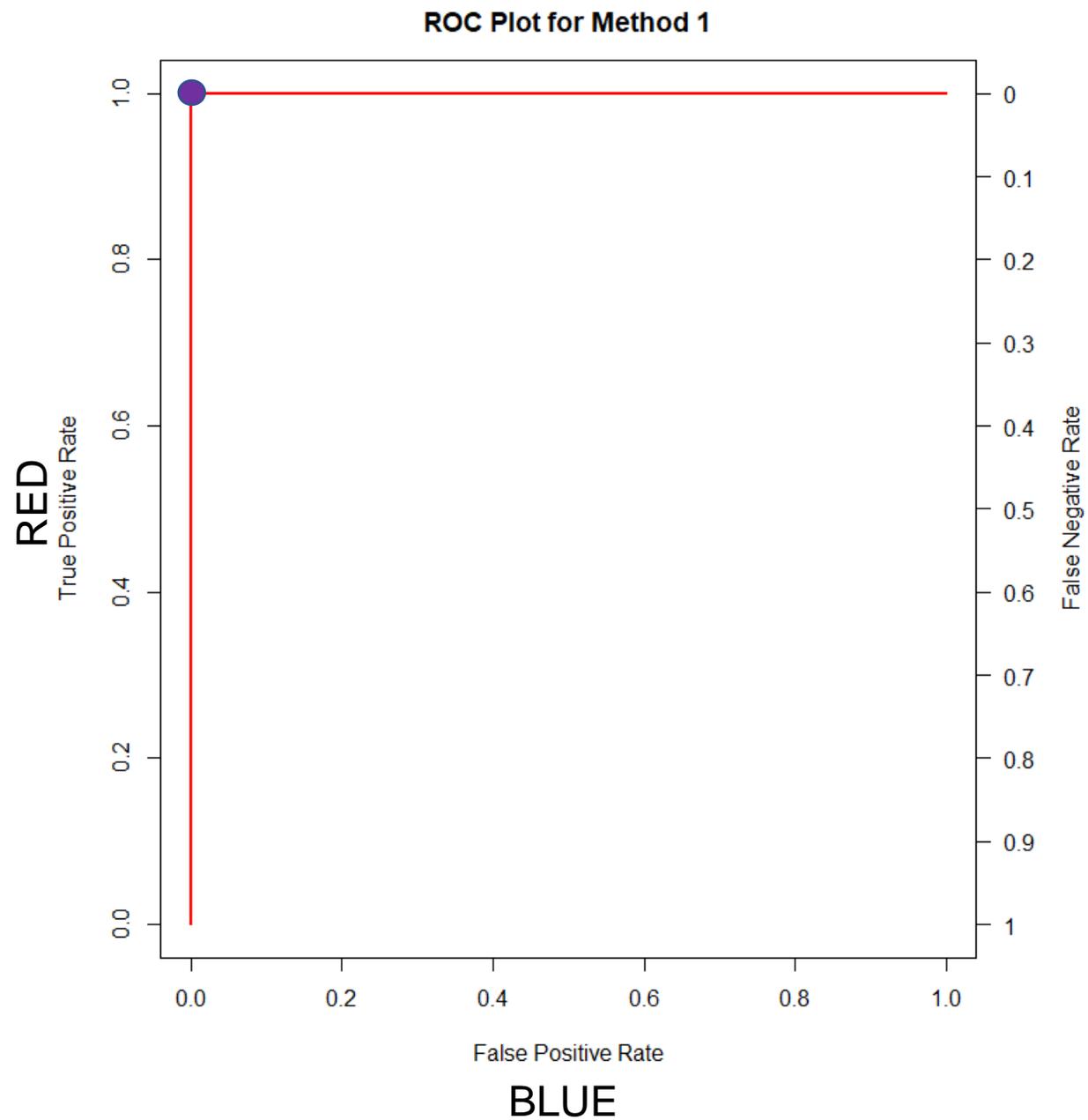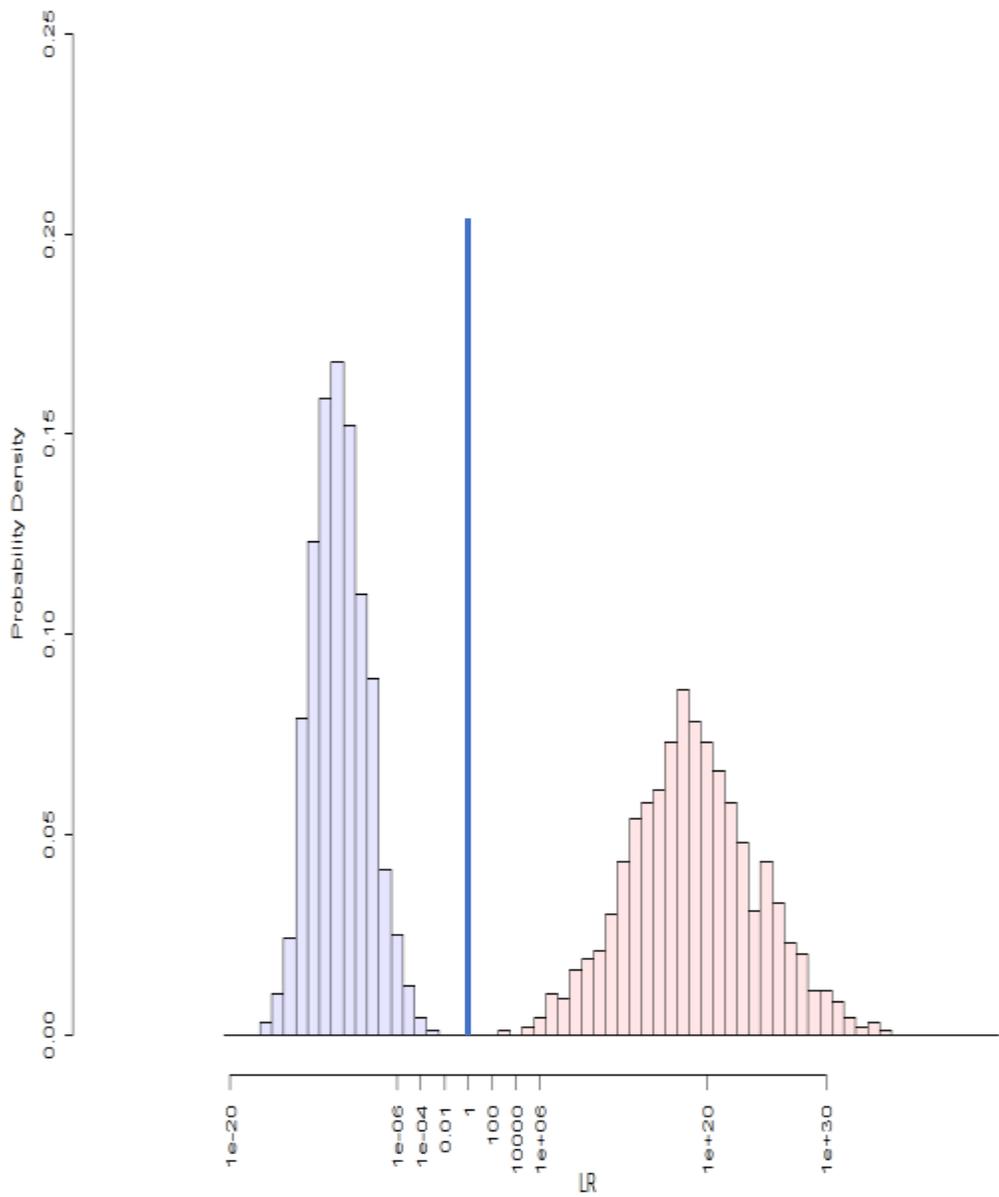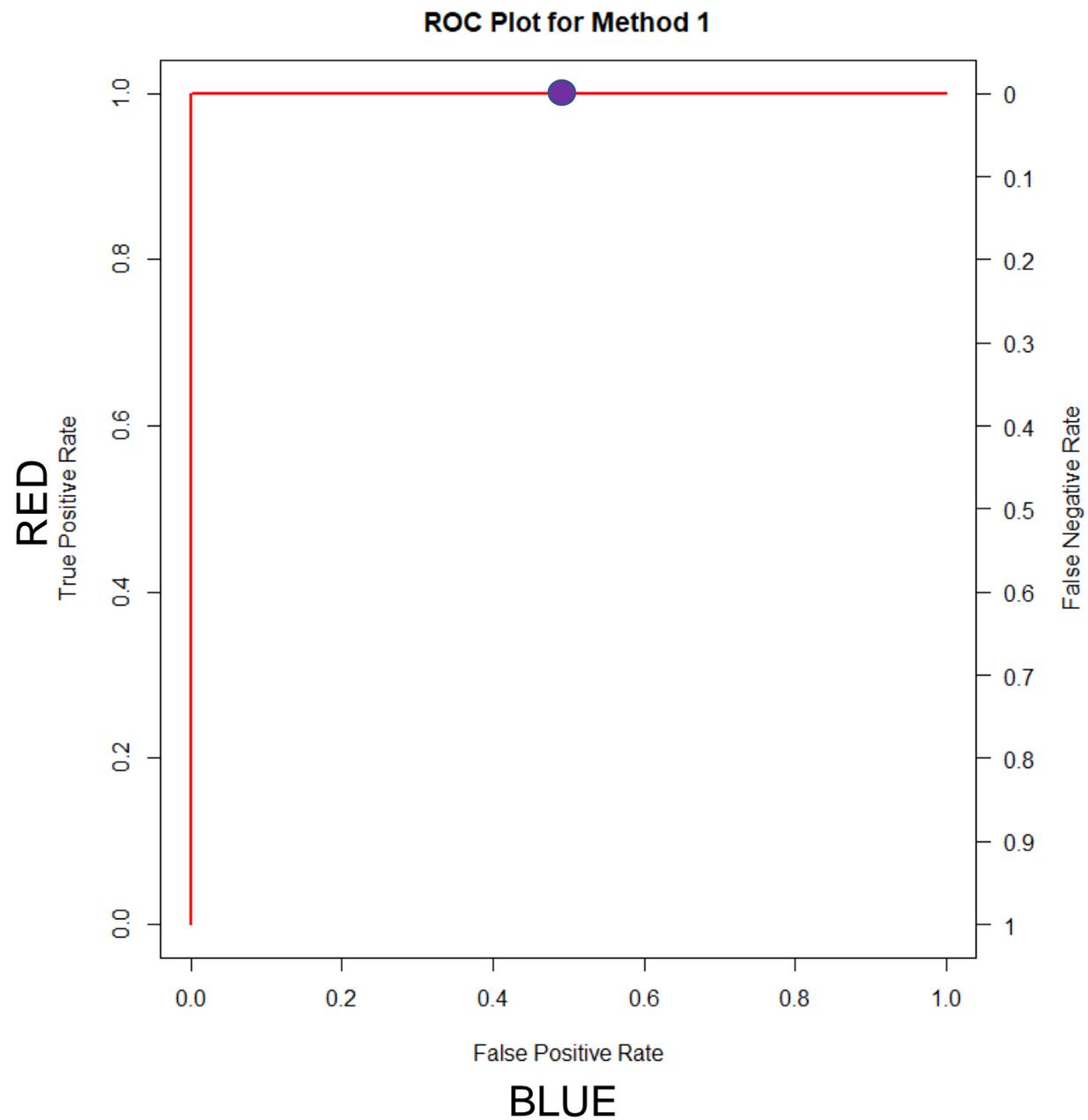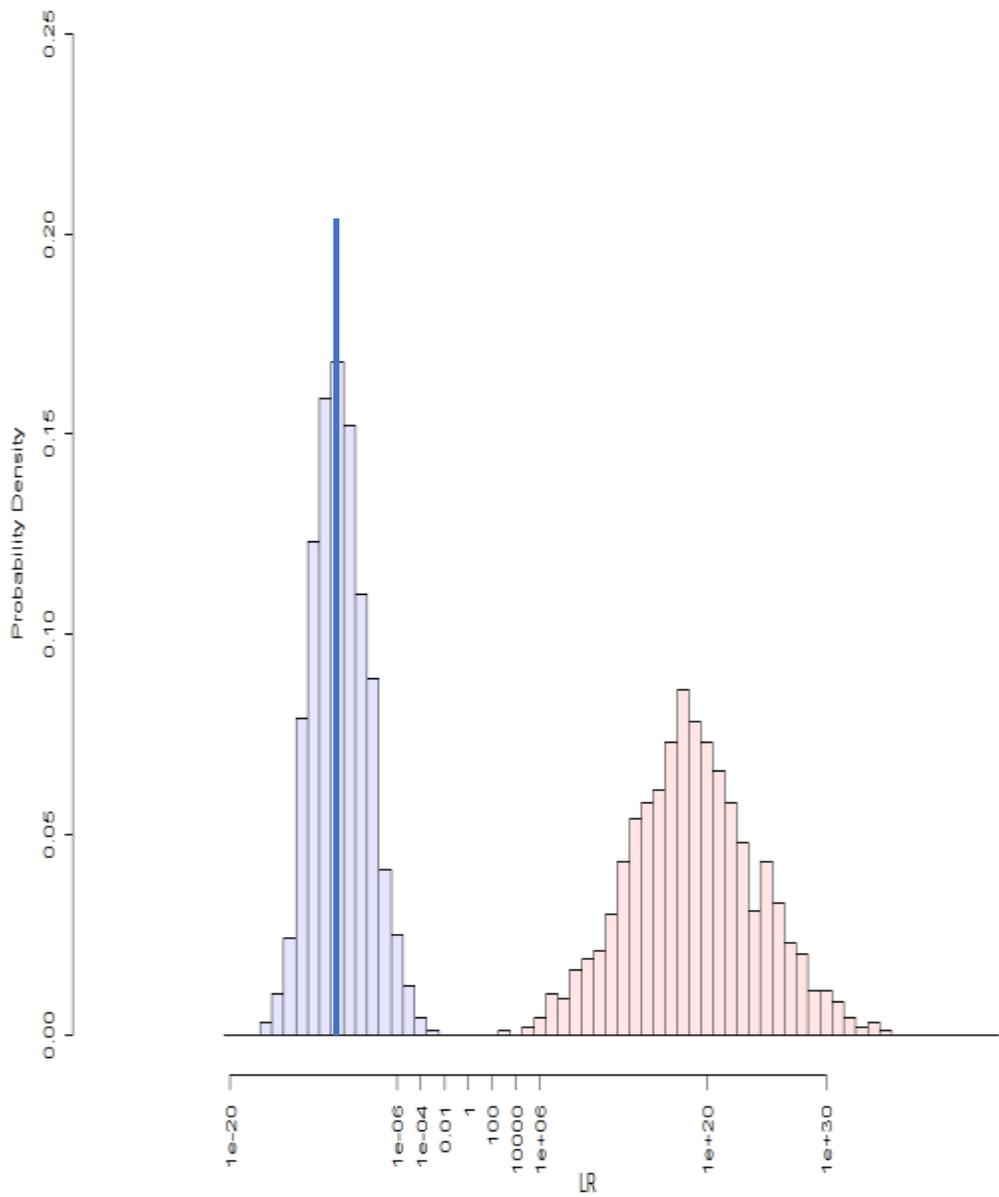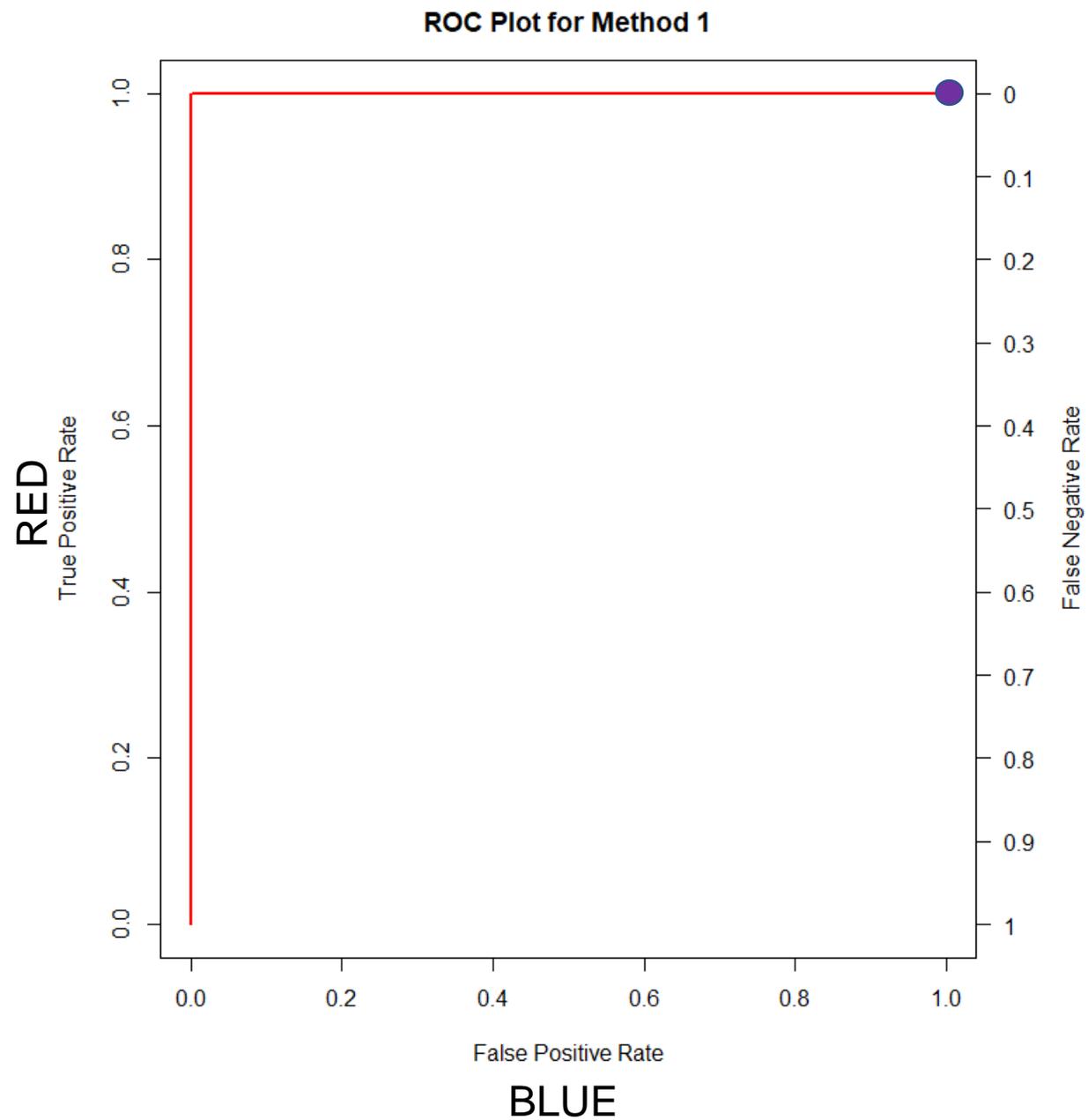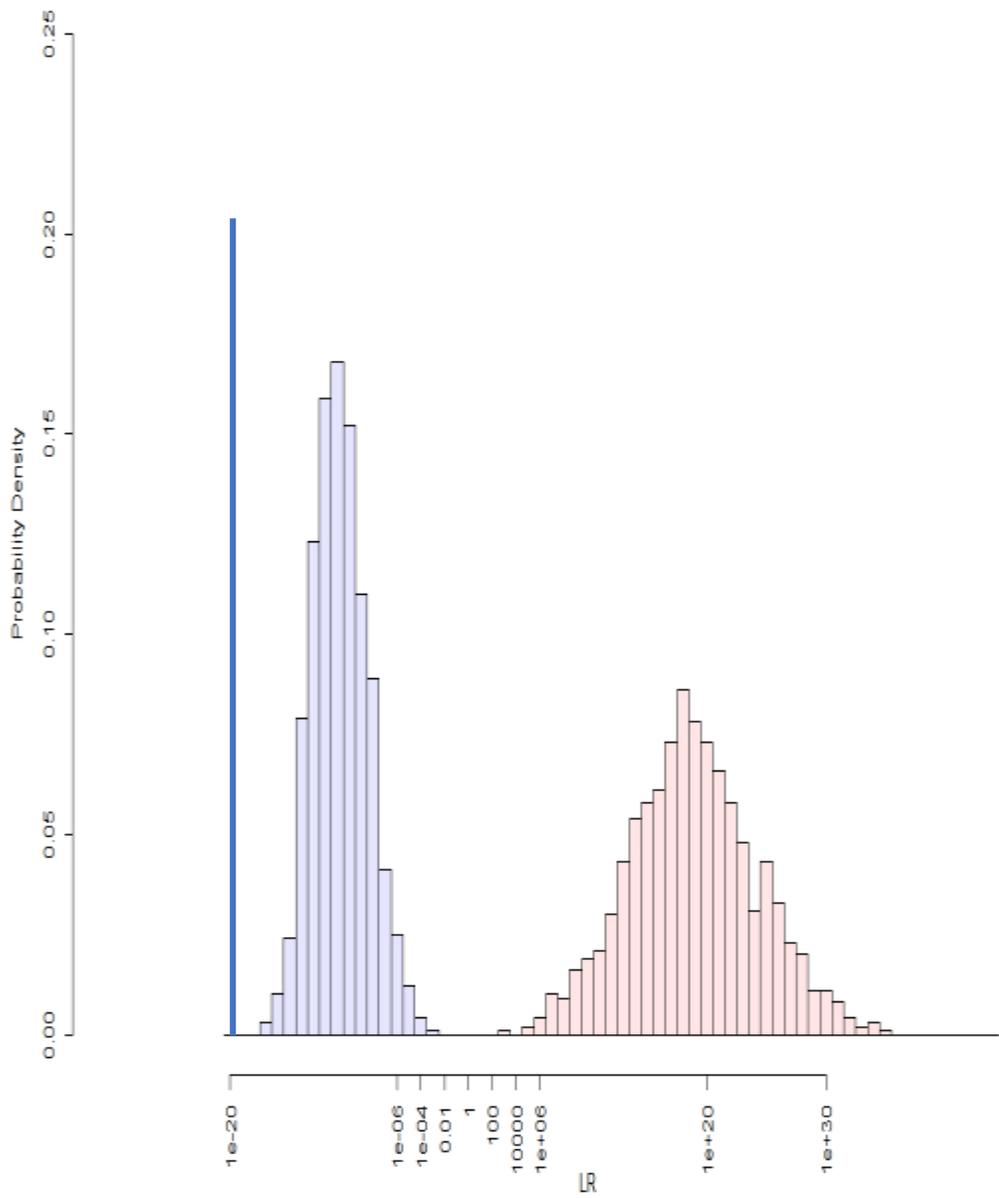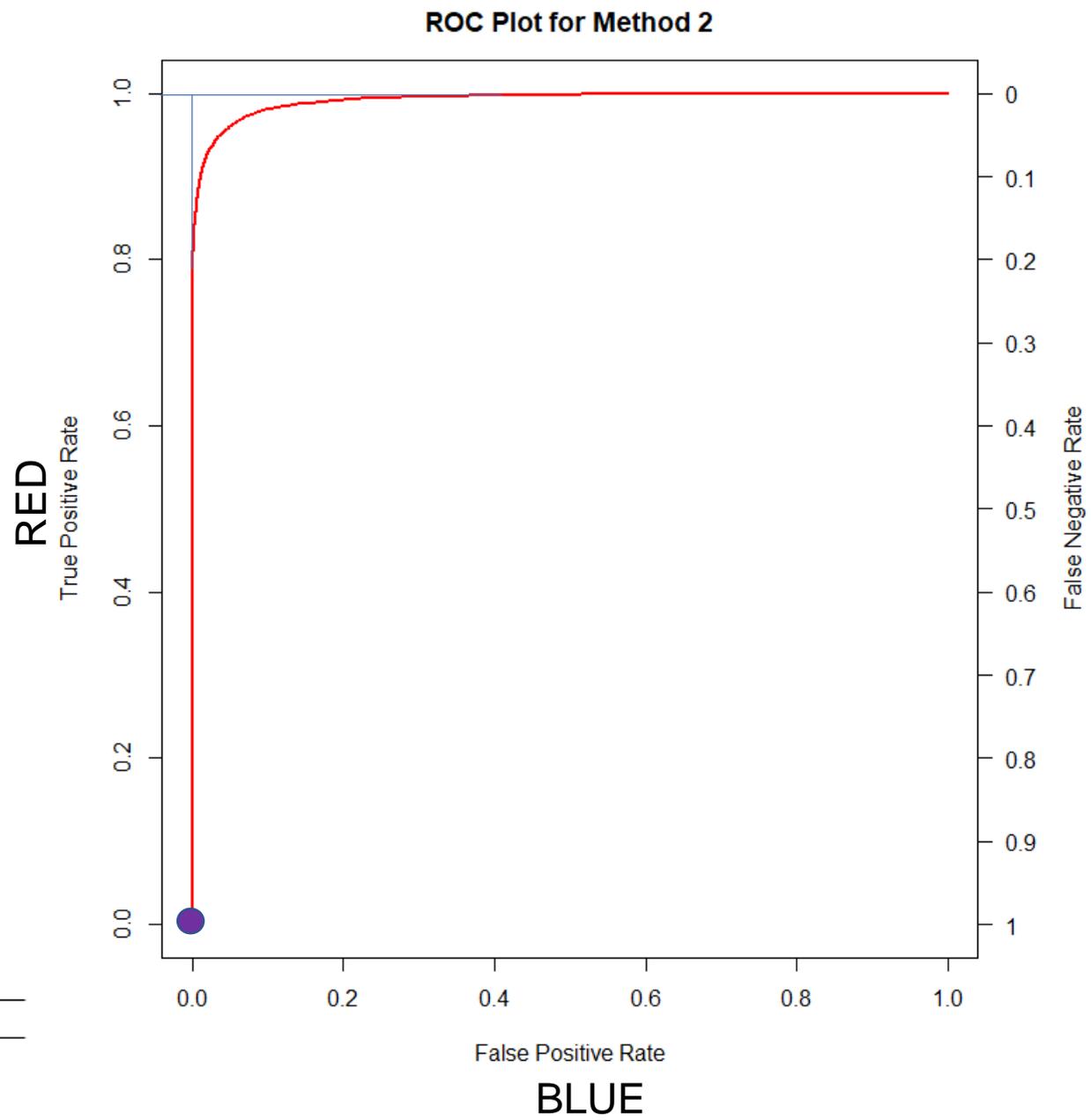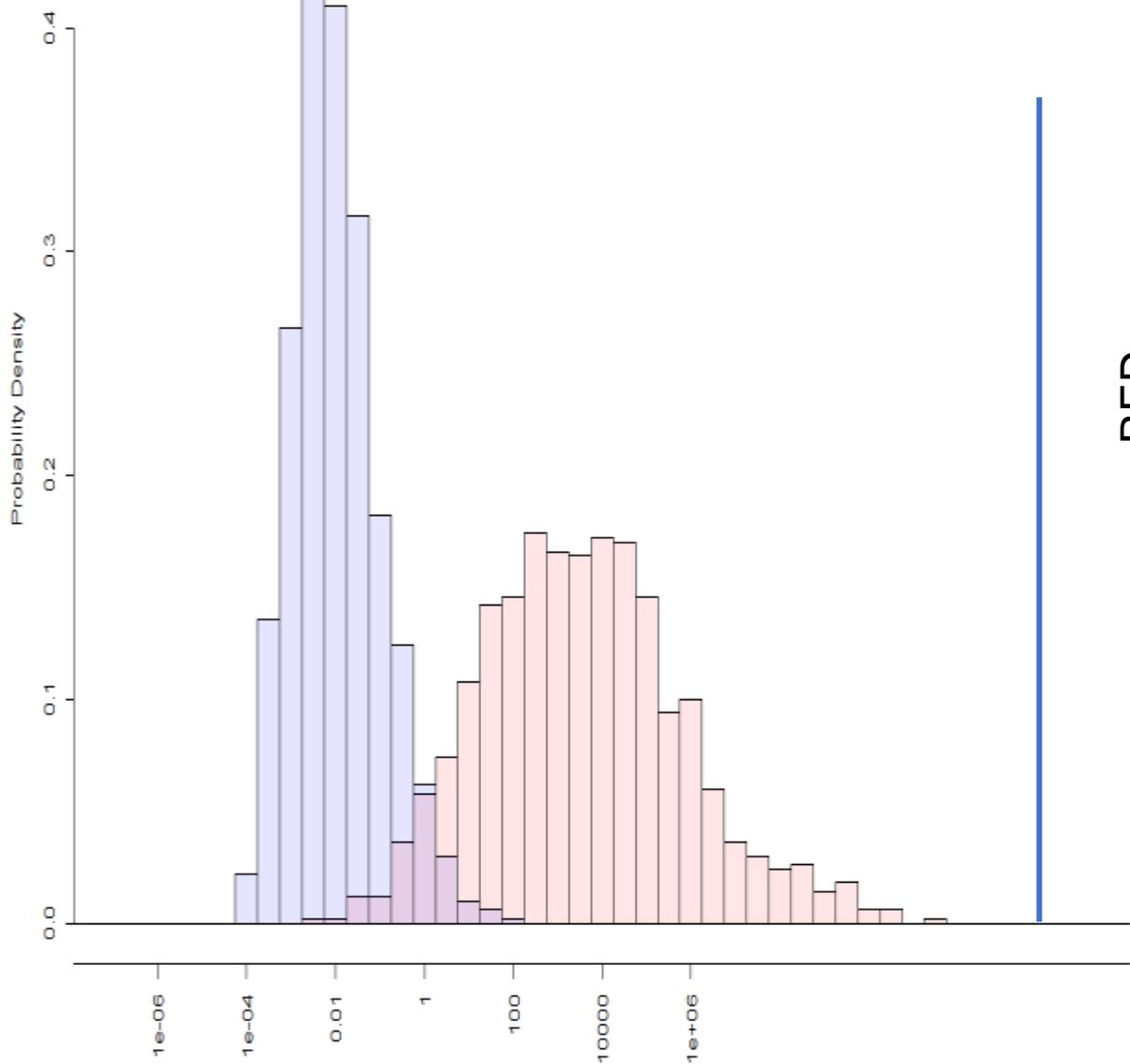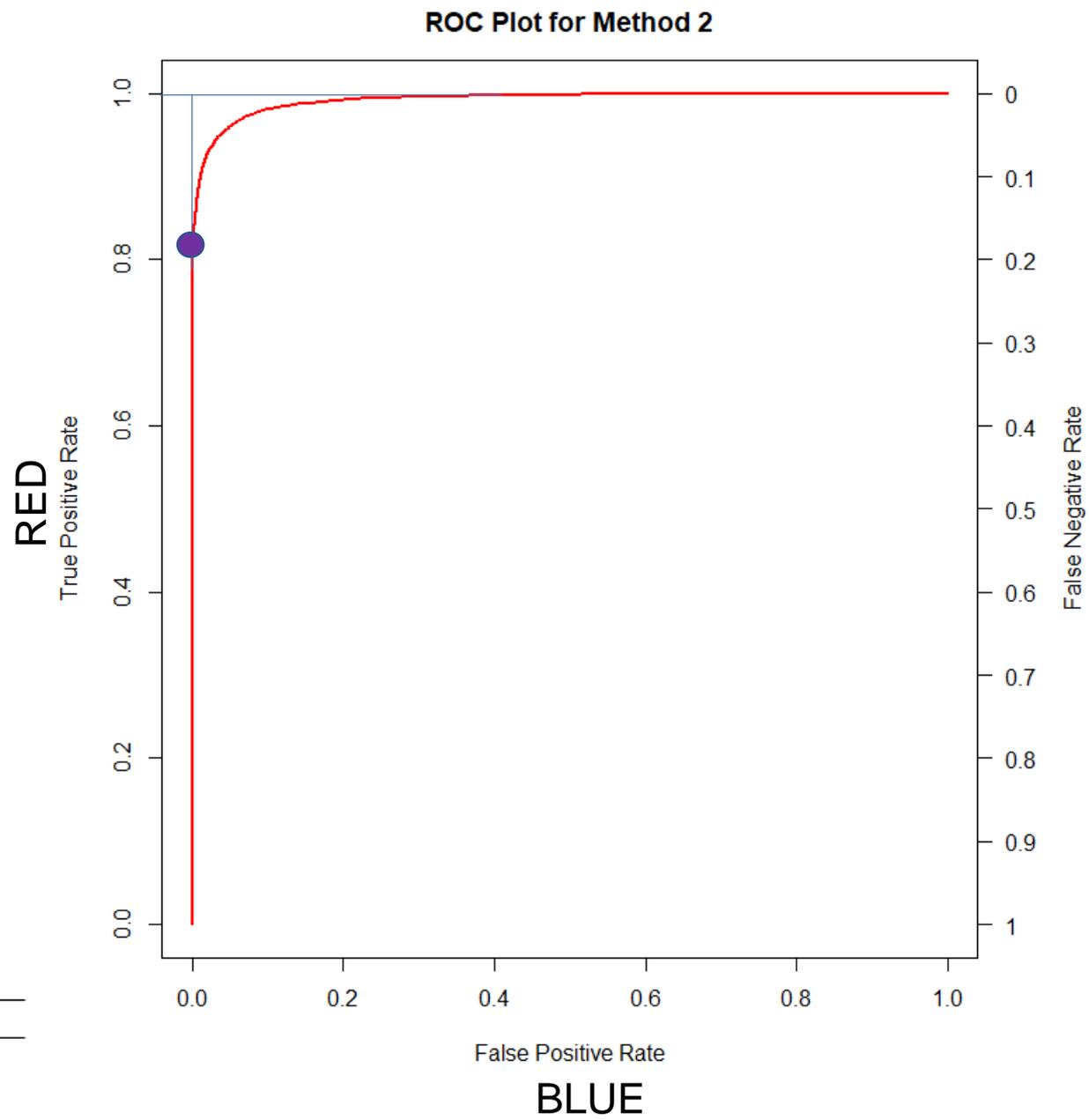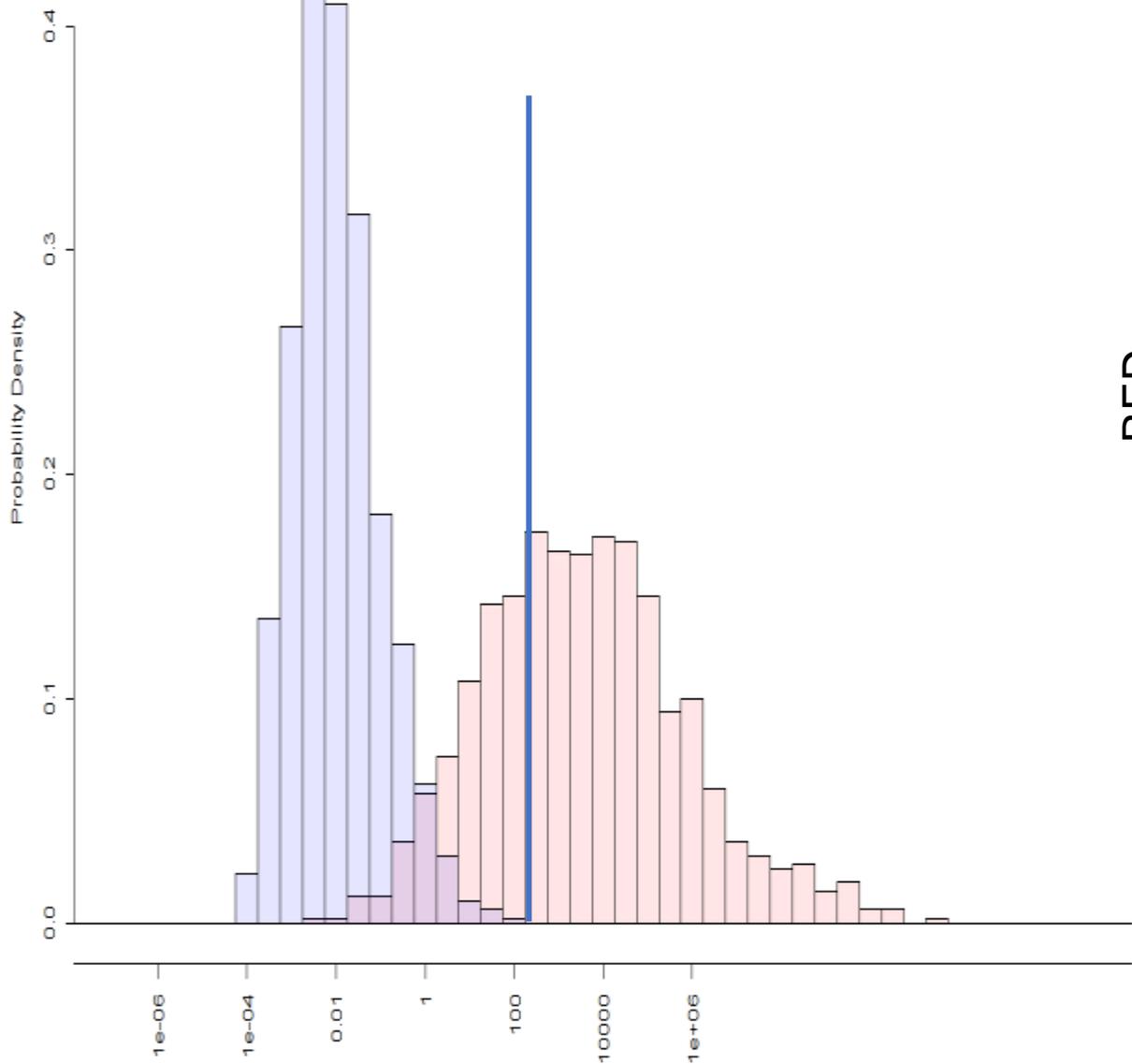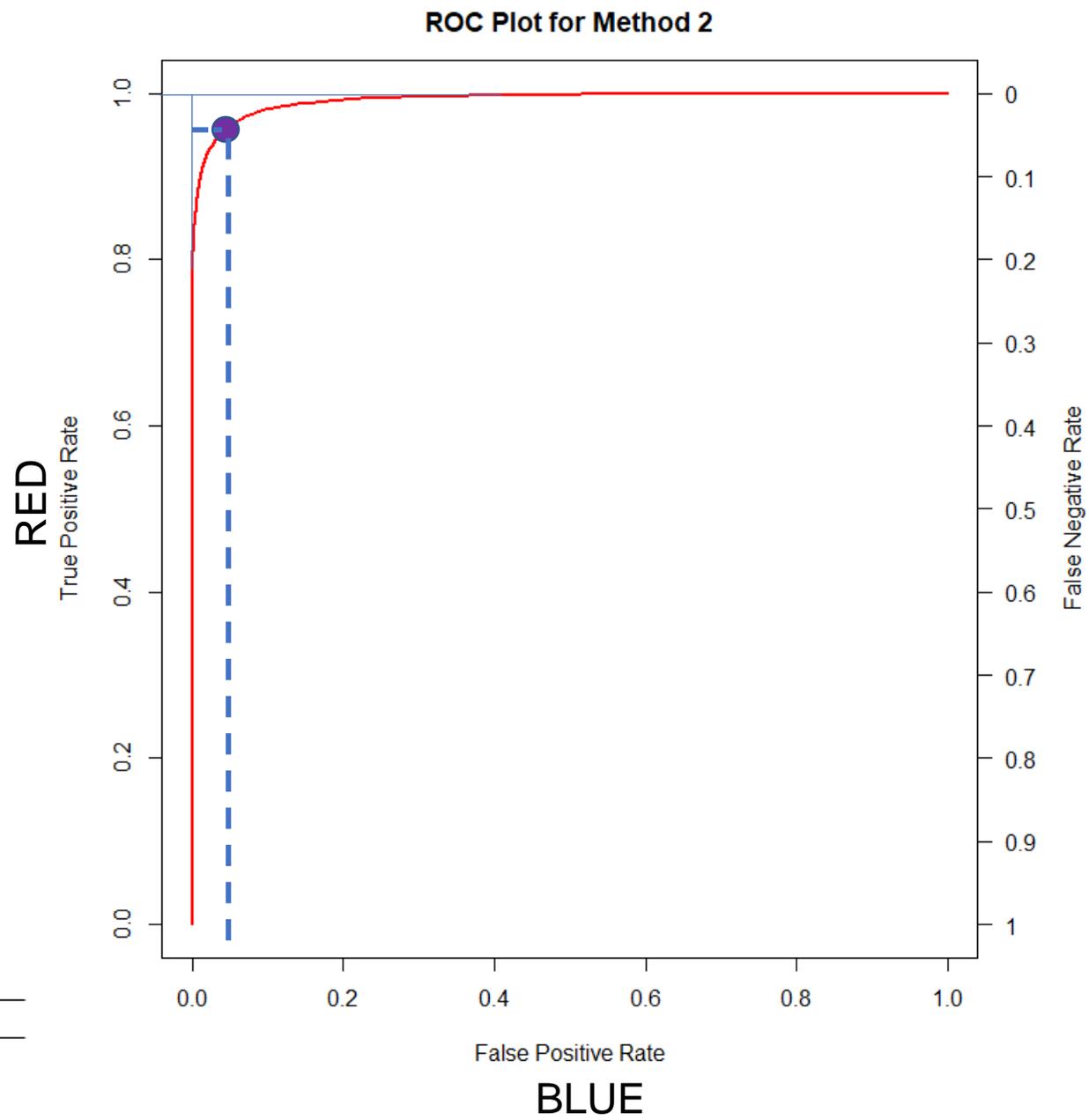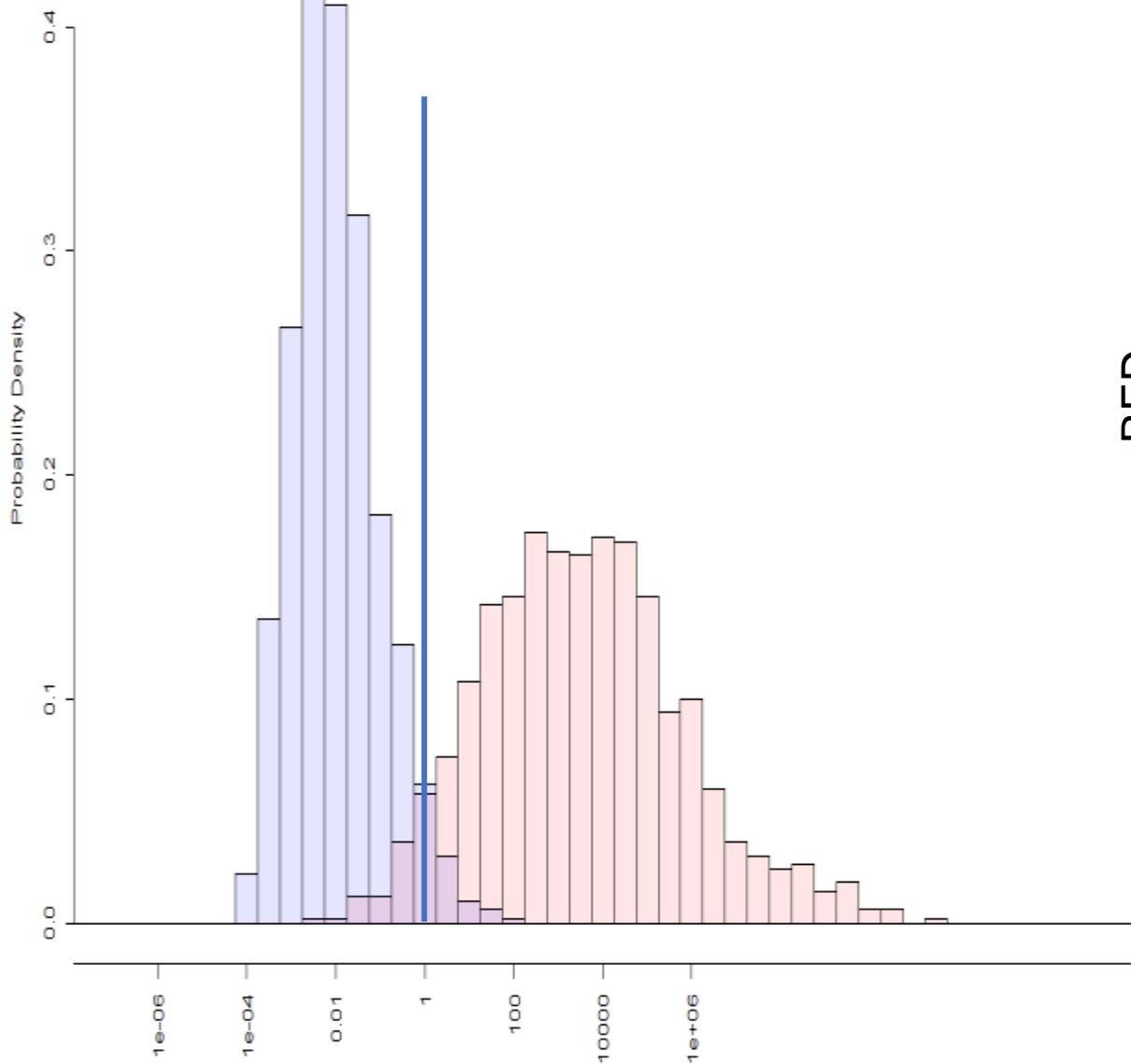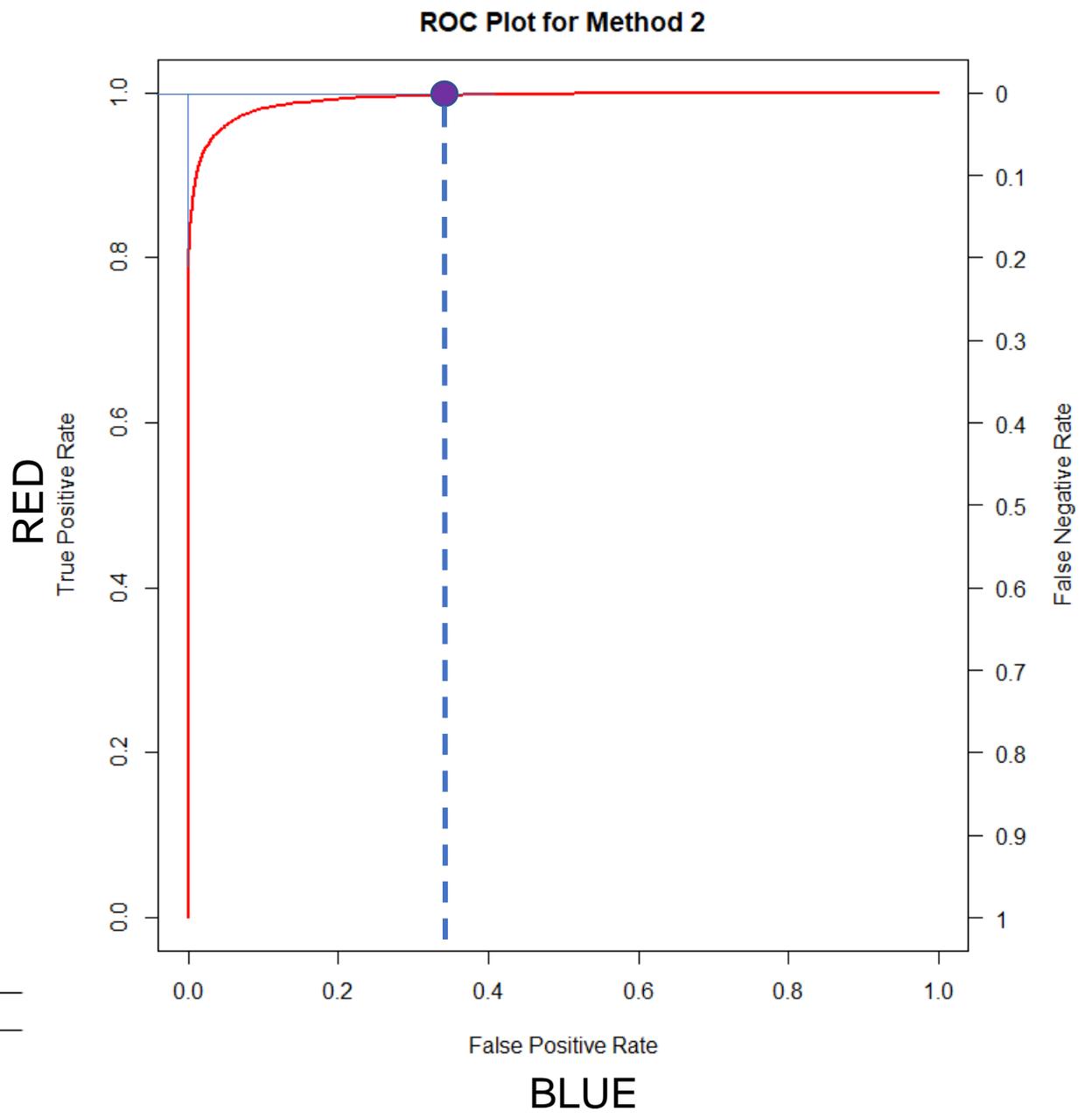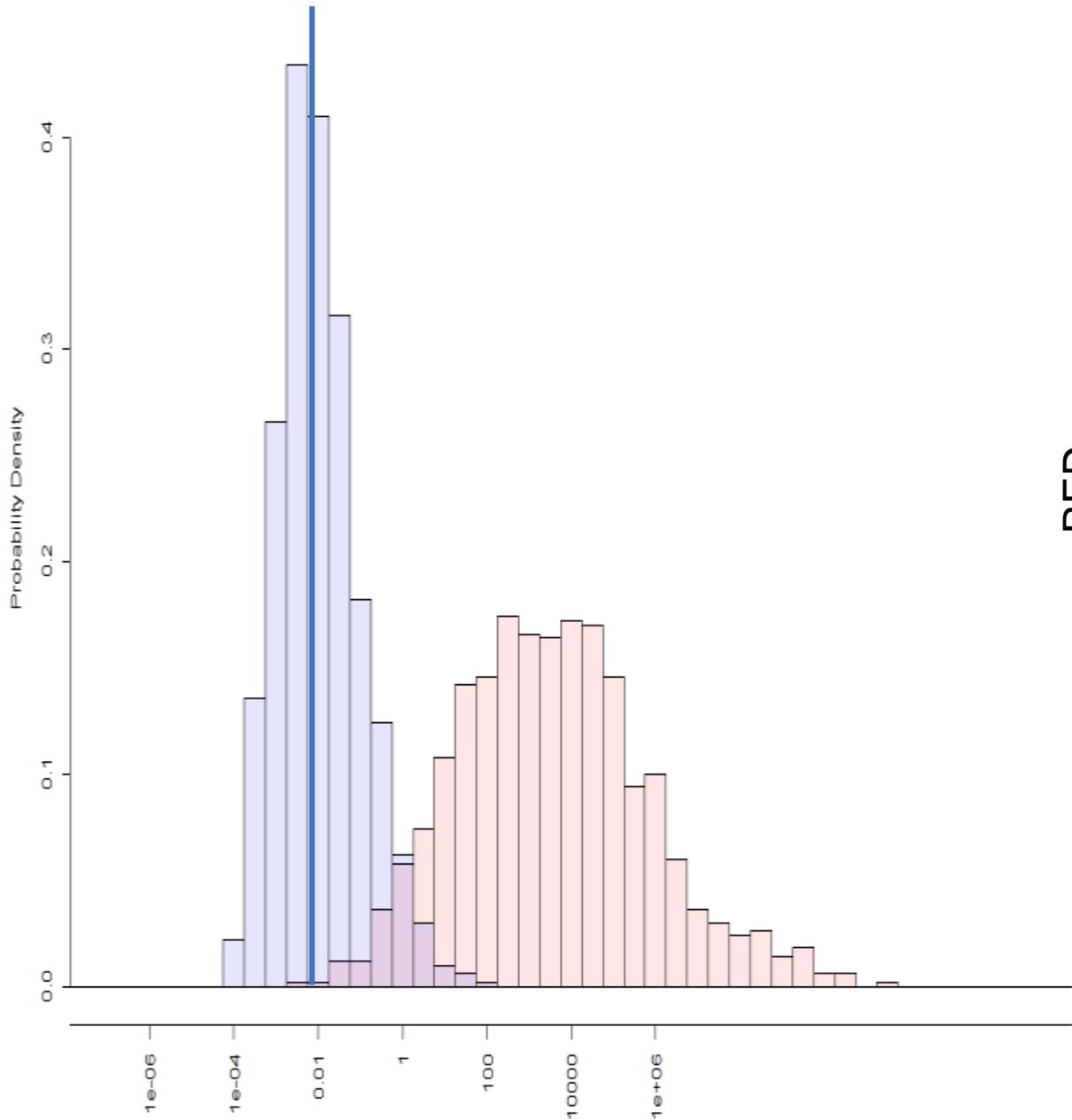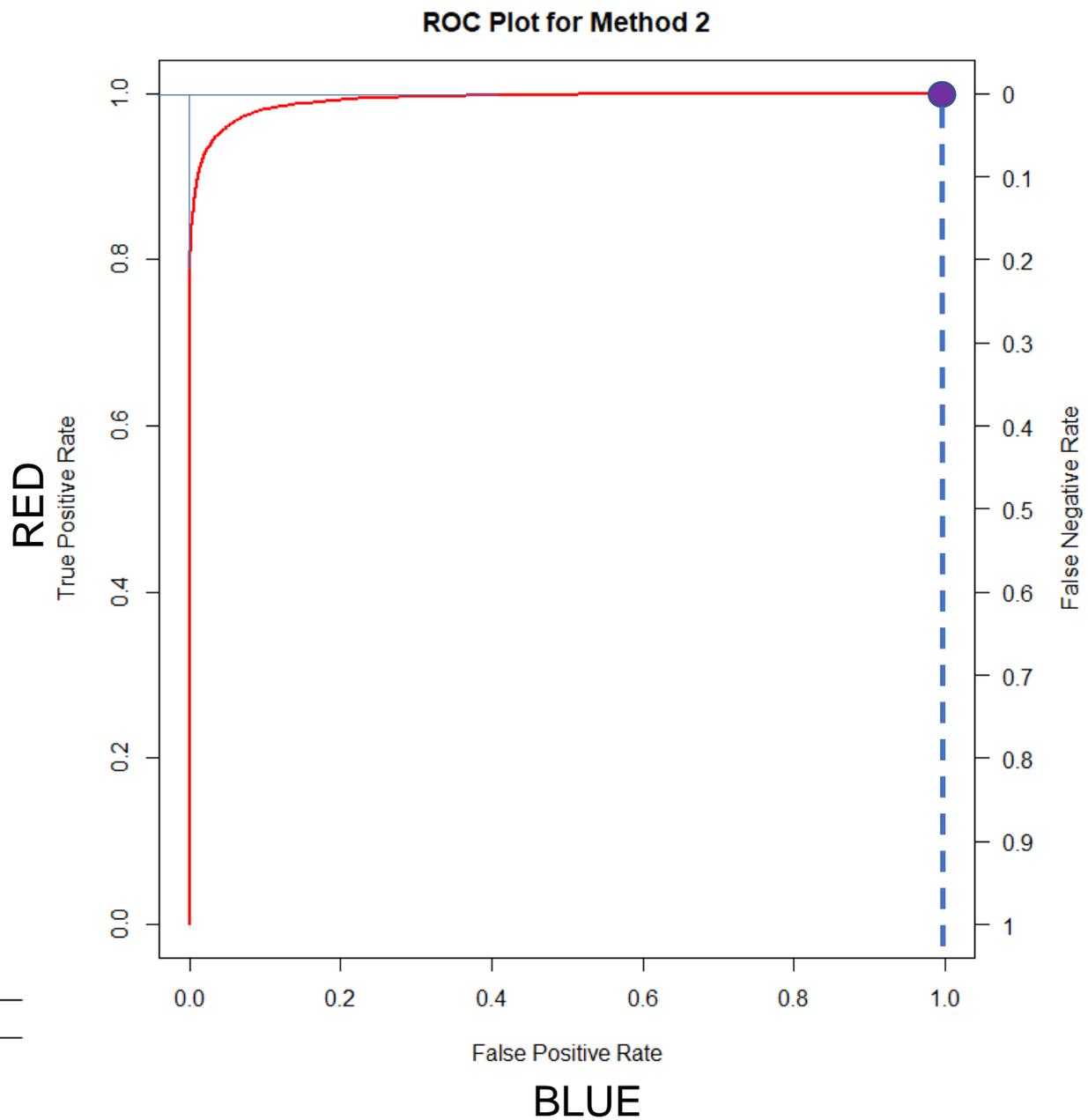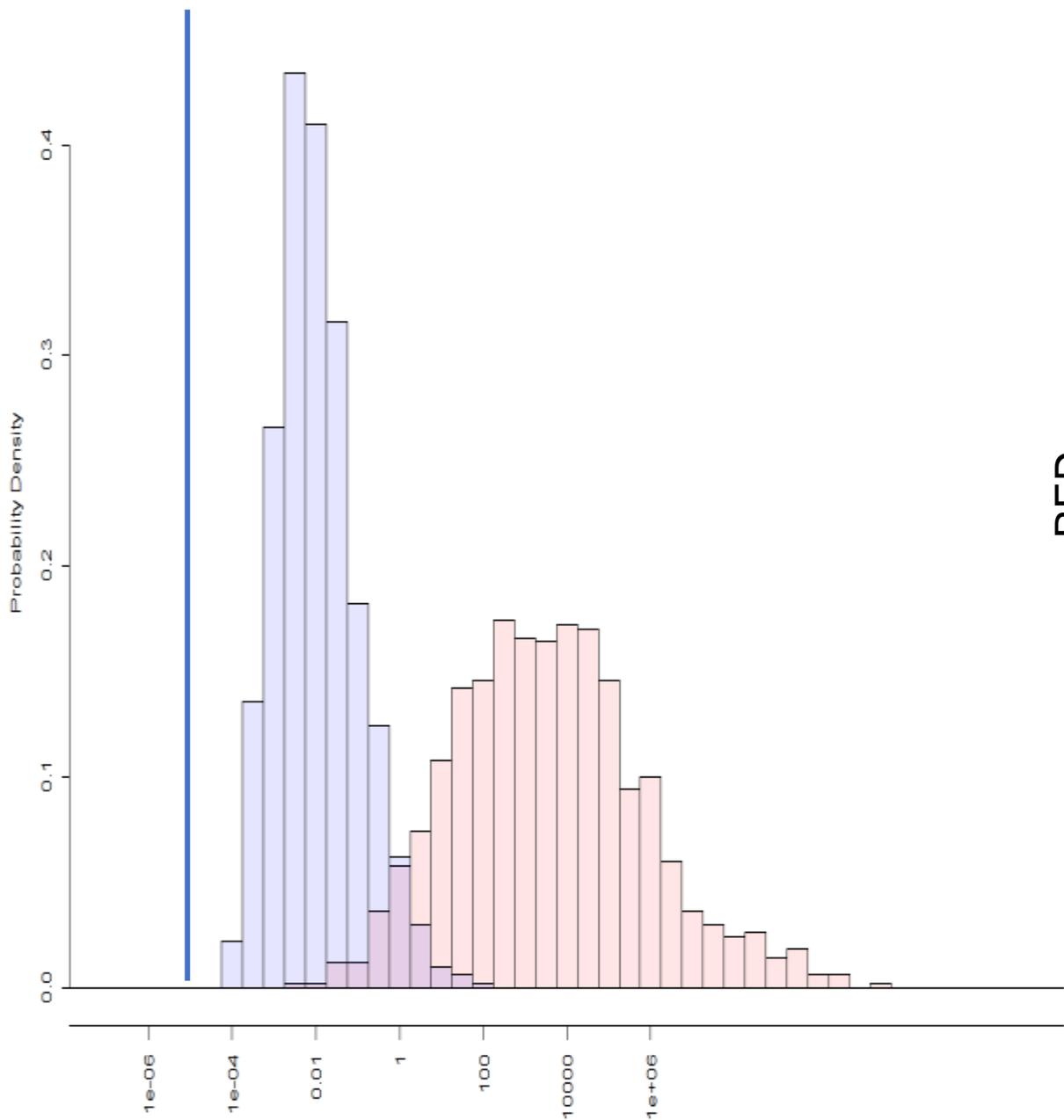
ROC Plot for Method 1

ROC Plot for Method 1

ROC Plot for Method 2

ROC Plot for Method 2

ROC Plot for Method 2

ROC Plot for Method 2

Error rates

Fig. 2. Receiver operating characteristic (ROC) plot where the rate of false positives (FP) (along horizontal axis) and true positives (TP) (along vertical axis) are plotted as a function of LR thresholds. The plot shows the results for the maximum likelihood estimation method (MLE) and the conservative method (CONS) for both LRmix and EuroForMix. The points on the curves show the FP and TP rates for different LR thresholds.
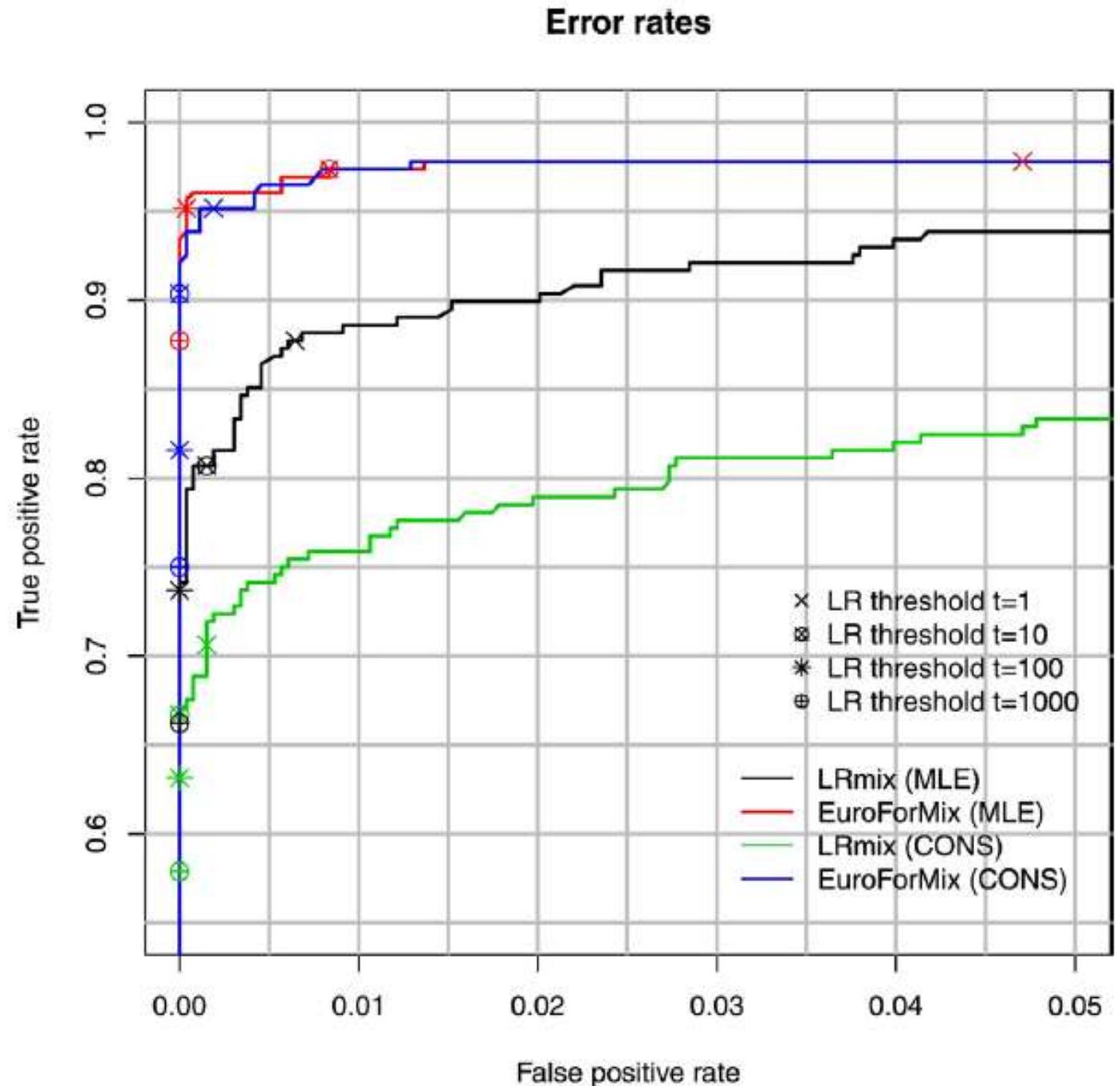
Some Take-aways

1. Primary focus should be on LR <u>system reliability</u>
    *   Improvement in System Reliability can be accomplished by improving component reliabilities
    *   Optimal strategies for this may focus on those components that affect system reliability the most. (Sensitivity analysis)

Component Reliability ≠ System Reliability

2.   Even if component reliabilities are deemed satisfactory, system reliability must be checked. The proof of the pudding is in the eating of it.

3. When examining published results from reliability studies, determine whether the experiment focused on selected parts of the system or on the entire system. If the focus was only on a subset of the components rather than the entire system claims of reliability cannot be properly interpreted.

4. There is no single correct LR. Does that mean we are justified in reporting any value for any given case? No !   Some LR systems are less reliable than other LR systems. **LR systems that are well-calibrated and have high discriminating power are what we need**.

Some Take-aways

5. One LR system may appear to be as reliable as another based on aggregate measures. However, a system can perform better than another in selected scenarios and vice versa.

6. Different LR systems, even those regarded as equally reliable, will disagree in any given casework situation. The magnitude of this disagreement is crucial information for triers of fact. The disagreement, in a given case, needs to be studied and reported.

Item 6 was a key point made in the paper "Likelihood Ratio as Weight of Forensic Evidence: A Closer Look" (2017). There were at least 3 rebuttal papers or letters to the editor. We believe that they did not address any of our main concerns.

Likelihood Ratio as Weight of Forensic Evidence: A Closer Look

Steven P. Lund and Hari Iyer

Statistical Engineering Division, Information Technology Laboratory,
National Institute of Standards and Technology,
Gaithersburg, MD 20899, USA

steven.lund@nist.gov
hari@nist.gov

Commentary: Likelihood Ratio as Weight of Forensic Evidence: A Closer Look

Colin Aitken[1]*, Anders Nordgaard[2], Franco Taroni[3] and Alex Biedermann[3]

[1] School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom, [2] National Forensic Centre (Sweden), Linköping, Sweden, [3] School of Criminal Justice, Université de Lausanne, Lausanne, Switzerland

A response to "Likelihood ratio as weight of evidence: A closer look" by Lund and Iyer

Simone Gittelson [a] ✉, Charles E.H. Berger [b], Graham Jackson [c, d], Ian W. Evett [e], Christophe Champod [f], Bernard Robertson [g], James M. Curran [h], Duncan Taylor [i, j], Bruce S. Weir [k], Michael D. Coble [l, m], John S. Buckleton [k, n]

Letter to the Editor—The Roles of Participants' Differing Background Information in the Evaluation of Evidence*

Sir,

There is a continuing debate about the role of background information in the evaluation of evidence. There is concern that different participants in the legal process have different background information and that this causes difficulty in the evaluation of evidence. We argue that although forensic scientists and fact finders have different information, no difficulty is caused in the evaluation

$\{I_{a\backslash b}, I_{b\backslash a}, I_{ab}\}$ are mutually exclusive and their union is I; I may be written as $\{I_{a\backslash b}, I_{b\backslash a}, I_{ab}\}$ or $\{I_{a\backslash b} \cup I_{b\backslash a} \cup I_{ab}\}$.

Assume now that A is a fact finder whose background information is formally independent of the scientific evidence E and has no effect on the probability of E and that B is a forensic scientist whose background information is formally independent of the propositions H_p and H_d and has no effect on the probability

Key Questions to Ask When Interpreting Reliability/Validation Studies

- Is the study focusing on one specific component of the system?

- Is the study addressing end-to-end (system) reliability?

- How were the test samples (ground truth known) selected?

- How many independent subject profiles were considered?

- Were the H_d true profiles synthetically generated OR were they obtained from a random sample of subjects from a relevant population?

- Does the conclusion state that "a process is **reliable**" or is it giving **measures of reliability**?

Relevance Considerations

Sheila

DNA

- DNA is one of the greatest advances in forensic science

- Its use as a biometric marker has had many successes – identification of mass casualties, missing persons and individual identification in tragic situations. It is also powerful in enabling identification of sources of blood or semen

- DNA transfers, and therefore we have a method of establishing links with crime scenes

- In instances where the source of DNA is known i.e. blood, semen, saliva some inference can be made as to the relevance of the crime scene sample

- The same inference cannot be made with certainty when dealing with invisible stains

- When speaking of new sensitive methods DNA profile may not be directly associated with the evidential body fluid that is "apparently" analyzed." Gill, P. (2001) Application of low copy number DNA profiling. *Croatian Medical Journal* 42(3): 229-232
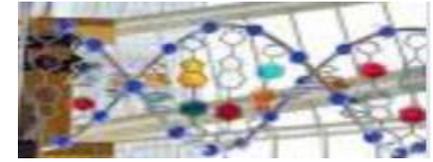
Development of DNA

Gill, P., et al. (2000). "An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA." Forensic Sci Int. **112**(1): 17-40.

Increased sensitivity – low template

Further sensitivity and increased number of loci

Increase in number of complex mixtures

Use of probabilistic genotyping
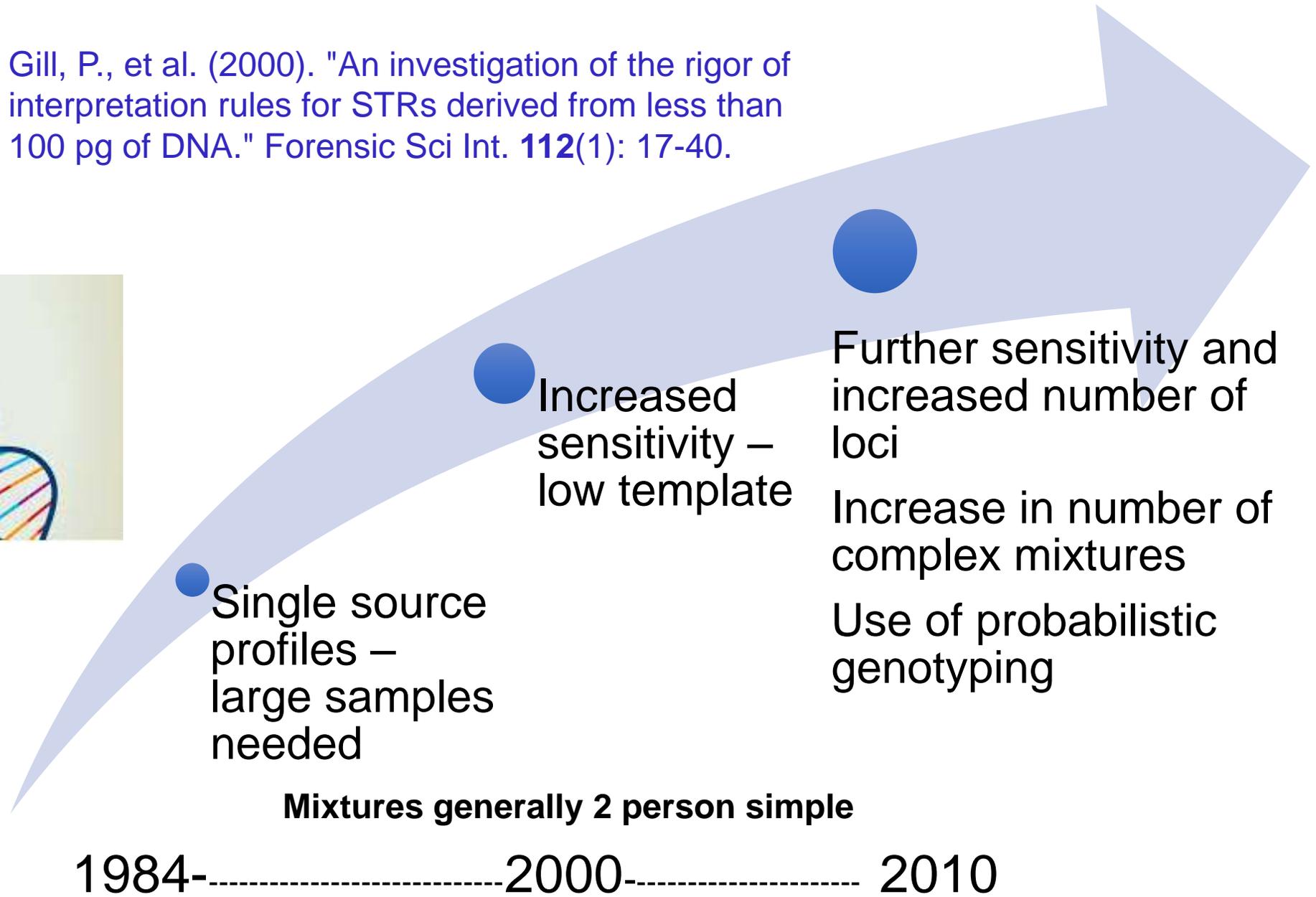
Single source profiles – large samples needed

Mixtures generally 2 person simple

1984-----------------------------------2000------------------ 2010
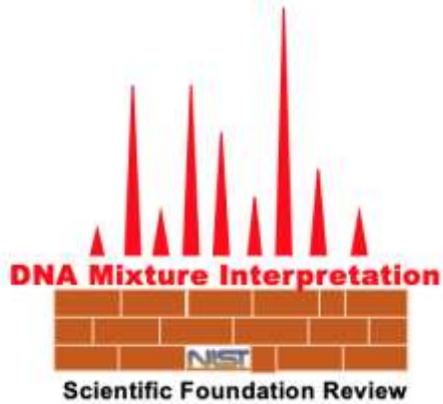
Foundation Study
on DNA Mixtures
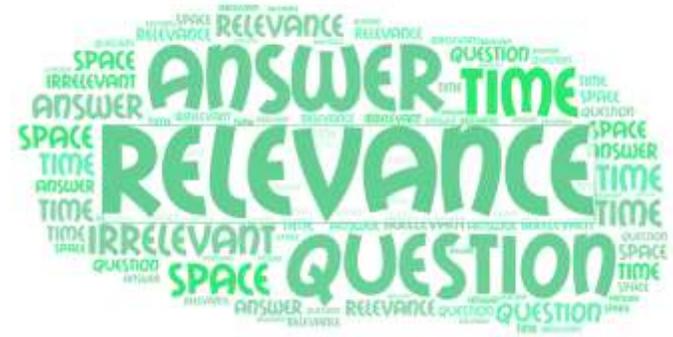
- Increased sensitivity had two immediate consequences
- Need to deconvolute mixtures which were more prevalent because more DNA was detected
- Information needed on transfer to help assess the relevance of the recovered DNA
- Most of the literature is concentrated on the first point
- The second point is particularly important for mixtures when at least some of the contributing genotypes are likely to be irrelevant

As part of the Foundation study, I studied the literature on transfer and persistence and the aim of this presentation is to give you an overview of my studies
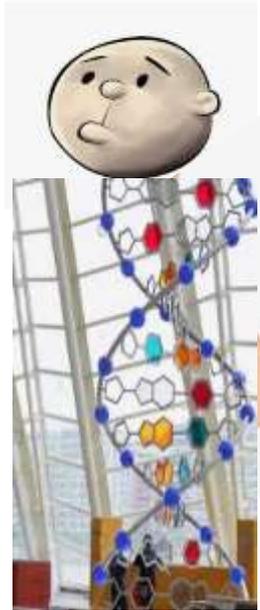
Mr. Blue

Mr. Red

- We need to ensure that we focus on the questions to be answered and not get too overwhelmed by technology
- If the most appropriate question is how did the DNA get there, we need to be sure we do not use inappropriate information to answer that question
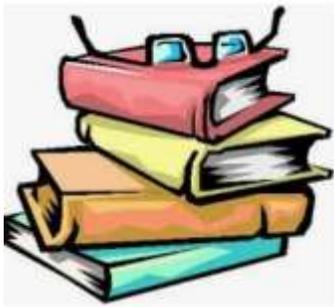
Literature 5 papers published in 2000 increased to 35 articles on the topic in 2015 Kokshoorn et al. 2018.
8 review articles since 2002

Numbers of publications
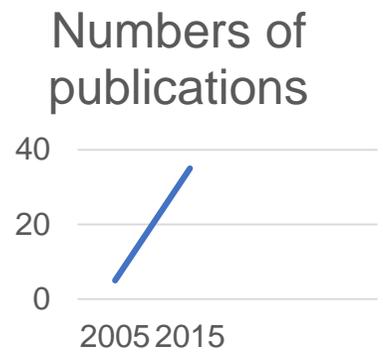
- Description of DNA as trace (2002) **Wickenheiser** R.A.. Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *J Forensic Sci,* 47**,** 442-50

- Emphasis on evaluation (2012) Biedermann, A. and Taroni, F. Bayesian networks for evaluating forensic DNA profiling evidence: a review and guide to literature. *Forensic Sci. Int. Genet.* 6(2):147-57.

- Risks identified quantity not reliable (2013) Meakin G. &Jamieson A. . DNA transfer: review and implications for casework. *Forensic Sci Int Genet,* 7**,** 434-43

- Review by one of early proponents (2015)Gill, P., Hanned, H., Bleka,O., Hannson, O., Dorum,G. & Egeland, T. Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches-Twenty years of research and development. *Forensic Sci Int Genet,* 18**,** 100-17.

- Mechanisms on how DNA is transferred from donor (2018) Burrill, J., Daniel, B., and Frascione, N. A review of trace "touch DNA" deposits: variability factors and an exploration of cellular composition. *Forensic Sci. Int. Genet.* 39: 8-18.

- Promotion of activity propositions (2018)Taylor, D., Kokshoorn, B. and Biedermann, A. Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic Sci. Int. Genet.* 36: 34-49.

- Comprehensive review of all aspects of transfer (2018) van Oorschot, R.A.H., Szkuta, B., Meakin, G.E., Kokshoorn, B. and Goray, M. (2018) DNA transfer in forensic science: a review. *Forensic Sci. Int. Genet.* 38: 140-166

- Scheme for more systematic approach to data collection (2019) Gosch, A. and Courts, C. (2019) On DNA transfer: the lack and difficulty of systematic research and how to do it better. *Forensic Sci. Int. Genet.* 40: 24-36.

There is a growing body of knowledge about how and how often DNA transfers, but significant knowledge gaps remain.

- Information about how DNA transfers described as though we have clarity on how and what transfers

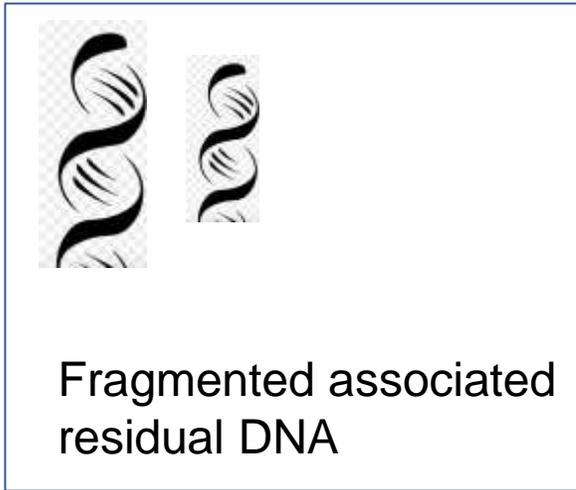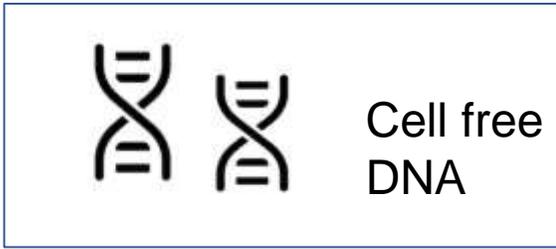Useful to consider what we know and what we do not know

- Lot of information on how variables affect transfer and persistence
- Donor qualities – shedder
- Moisture
- Activities
- Background DNA
- Substrate characteristics

Information not robust or repeatable
Developed with time
Different sensitivities – different outcomes

DNA transferred from person – not sure of source

Cell free DNA

Fragmented associated residual DNA

Transferred exogenous nucleated cells

self

Endogenous nucleated cells

Non-self

Anucleate corneocytes
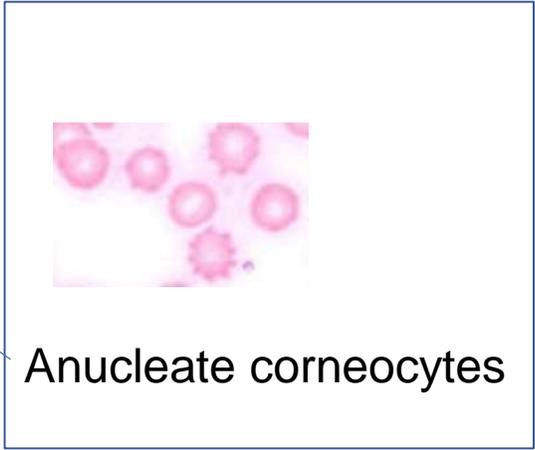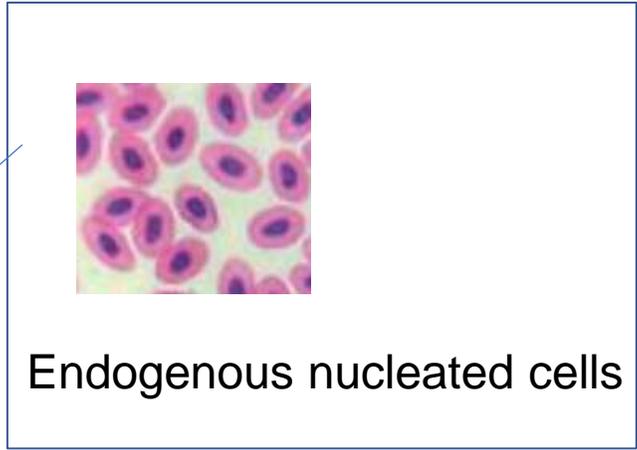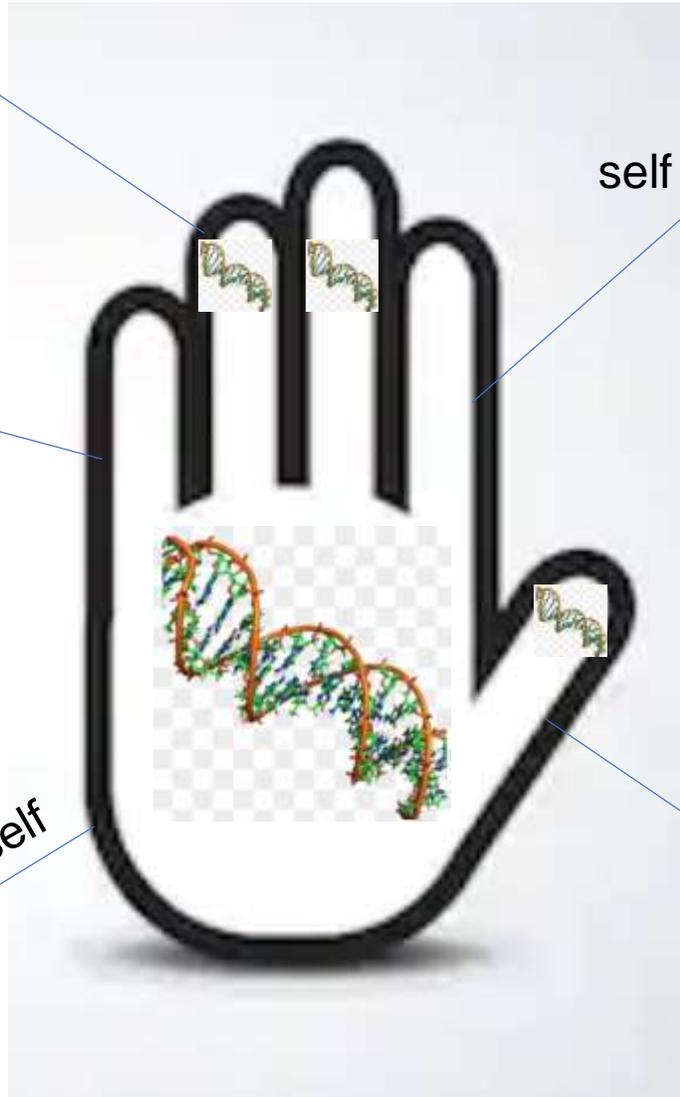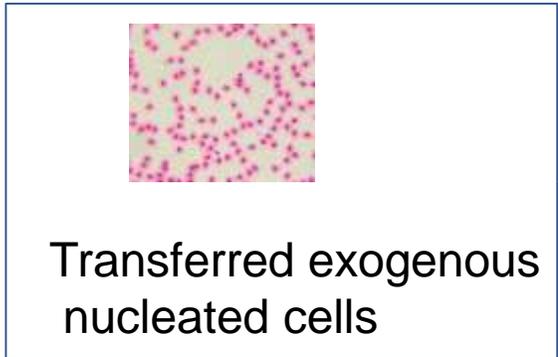
Possible sources of transferred DNA

Copied from image in Burrill et al (2018)

Burrill, J., Daniel, B. and Frascione, N. (2018) 'A review of trace "Touch DNA" deposits: Variability factors and an exploration of cellular composition', *Forensic Sci Int Genet,* 39, pp. 8-18.

DNA undergoes many interactions between crime scene and production of EPG –not sure of when

Pre –
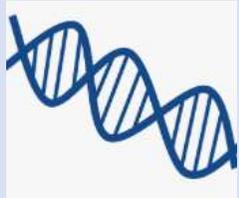
Opportunities for deposits of DNA from those with legitimate access

Background DNA

Crime
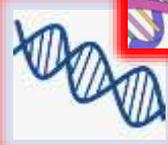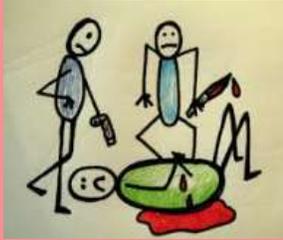
Transfers to and from scene relevant to crime

Pre-laboratory

CRIME SCENE - DO NOT CROSS

Potential for many additions and subtractions from scene until DNA recovered

Laboratory
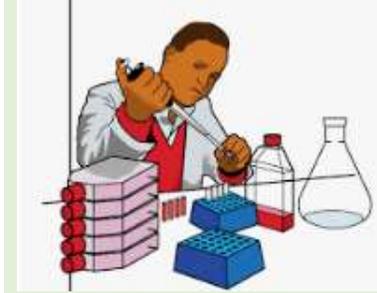
Contamination possible

Expectation that DNA represented in EPG contains DNA from crime scene

Extraction, PCR and production of EPG

Timeline

Shedder status

Studies on factors that affect variability of donor

Lot of literature on this topic

Sex, age, washing, activities, individual

Continuum with some high and low
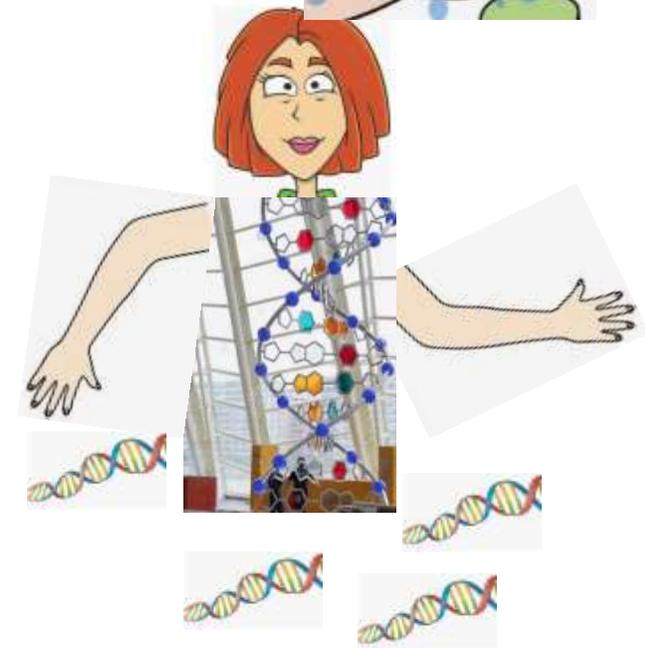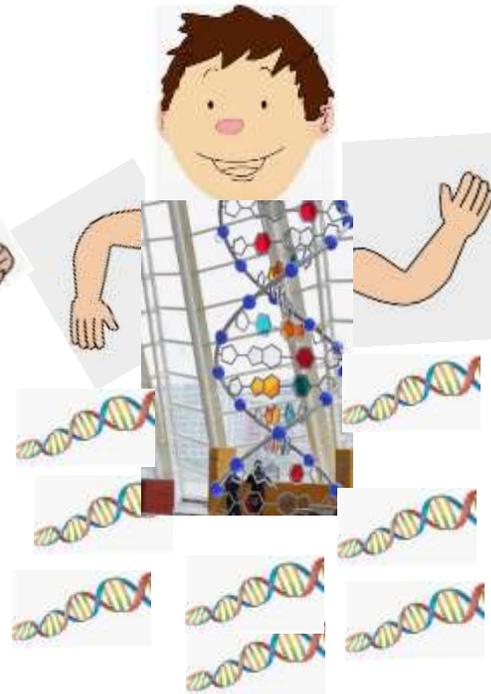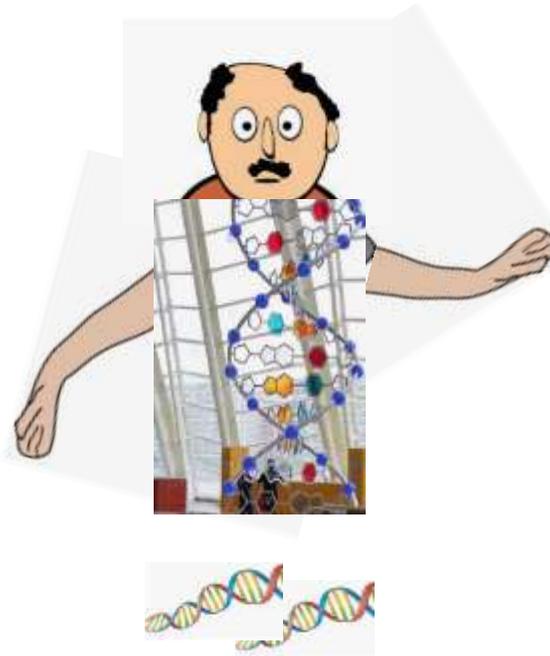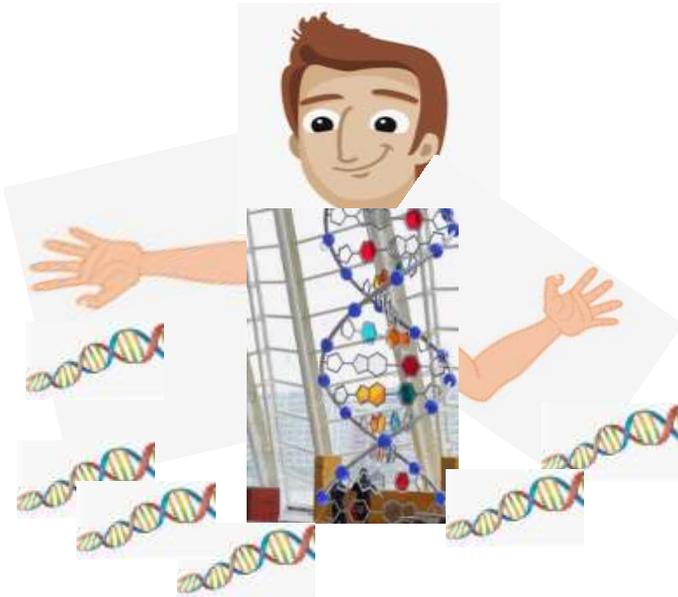
Shedder status and Investigation of Self and Non-Self

Last person to handle

- 240 handprints deposited by 10 individuals;
- Analyzed for differences in DNA quantity and type of profile at different times on different days;
- Inter-personal variation higher than intra-personal. 0.05-5 ng per deposit;
- Mainly 2 person mixtures or non interpretable; non-self in 79% Non-self usually the minor component ;
- On rare occasion when non-self was major, they were associated with poor depositor/shedder;
- 7/240 self excluded

Goray, M., Fowler, S., Szkuta, B. and van Oorschot, R. A. H. (2016) 'Shedder status-An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time', *Forensic Sci Int Genet,* 23, pp. 190-196.

Transfer studies - substrate
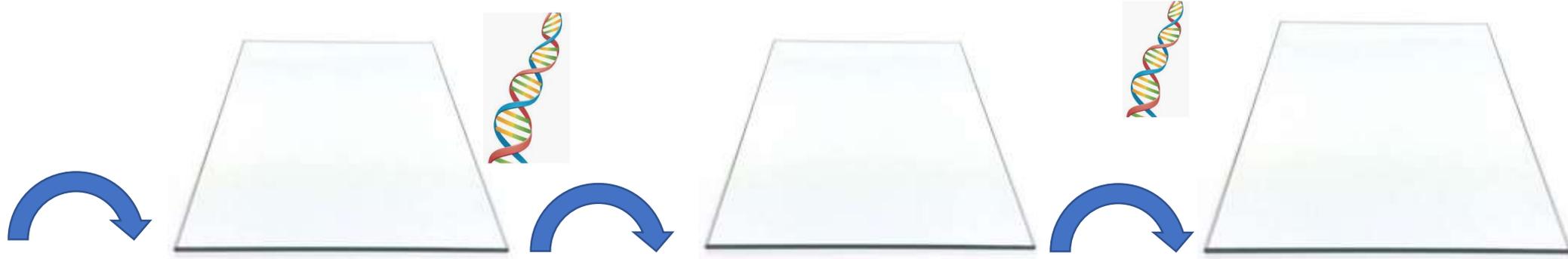
Less transferred to hard non-porous surface but lost more quickly from
Higher amount transferred to soft porous surface but less transferred
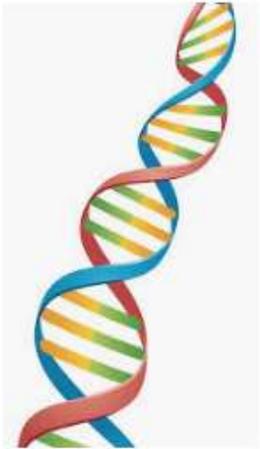
Passive , pressure and friction

Increasing amount of transfer

Moisture

Friction to transfer DNA from non-porous to porous
the most efficient chain

DNA transfers affected by
1. type of substrate
2. Moisture
3. pressure

Goray, M., Eken, E., Mitchell, R. J. and van Oorschot, R. A. (2010) 'Secondary DNA transfer of biological substances under varying test conditions', *Forensic Sci Int Genet,* 4(2), pp. 62-7.
Verdon, T. J., Mitchell, R. J. and van Oorschot, R. A. (2013) 'The influence of substrate on DNA transfer and extraction efficiency', *Forensic Sci Int Genet,* 7(1), pp. 167-75.

Fewer systematic persistence studies than transfer studies

Time since deposit

Raymond, J. J., van Oorschot, R. A. H., Walsh, S. J., Roux, C. and Gunn, P. R. (2009) 'Trace DNA and street robbery: A criminalistic approach to DNA evidence', *Forensic Science International: Genetics Supplement Series,* 2(1), pp. 544-546

Profiles obtained out of doors up to two weeks with technology available in 2009

Taylor, D., Abarno, D., Rowe, E. and Rask-Nielsen, L. (2016) 'Observations of DNA transfer within an operational Forensic Biology Laboratory', *Forensic Sci Int Genet,* 23, pp. 33-49.

Profiles built up over a period of time in laboratory setting again highlighting need for elimination databases

Helmus, J., Zorell, S., Bajanowski, T. and Poetsch, M. (2018) 'Persistence of DNA on clothes after exposure to water for different time periods-a study on bathtub, pond, and river', *Int J Legal Med,* 132(1), pp. 99-106.

Results demonstrate that DNA could still be recovered from clothes exposed to water for more than 1 week. Difference between winter and summer but worth checking even when exposed to water

Persistence of one user following another

- Original user detected vast majority of experiments
- Varied depending on
  - Duration of use by second person
  - Substrate
  - Original handler – shedder status
  - Activities /action
  - Duration of use

Fonneløp, A. E., Johannessen, H. and Gill, P. (2015) 'Persistence and secondary transfer of DNA from previous users of equipment', *Forensic Science International: Genetics Supplement Series,* 5(Supplement C), pp. e191-e192.

Study with computer and mouse

Original user detected and transferred to second user up to 8 days
Differences in ability to detect initial user reasoned to be due to shedder s
Major profile not always last user

Persistence on burglary tools

Moderate use of mock borrowed tools demonstrated a material-dependent persistence. In total, outcome depends mainly on the nature of contact, the handle material, and the user- specific characteristics.
Detected profile is DNA from the last person who handled barehanded

Pfeifer, C. M. and Wiegand, P. (2017) 'Persistence of touch DNA on burglary-related tools', *Int J Legal Med,* 131(4), pp. 941-953.

Trend in studies to mimic casework in a broad manner

Experiments with knives

Experiments with knives to check if indirect transfer from person shaking hands with handler is detectable – yes **handler main profile**; 13 /20 5/20 **secondary transfer main profile**.

Cale, C. M., Earll, M. E., Latham, K. E. and Bush, G. L. (2016) 'Could Secondary DNA Transfer Falsely Place Someone at the Scene of a Crime?', *J Forensic Sci,* 61(1), pp. 196-203.

Is DNA of nearby person detected in stabbing cases and how much is transferred;
DNA of person handling knife major or single profile 83%; Profiles too complex 5% ; **Observers profiles not detected**.

Samie, L., Hicks, T., Castella, V. and Taroni, F. (2016) 'Stabbing simulations and DNA transfer', *Forensic Sci Int Genet,* 22, pp. 73-80

Is profile of regular user detected as well as stabber and hand shaker
Regular user persisted for at least a week; Non-donor DNA co-deposited 5% -declined with time.

Meakin, G. E., Butcher, E. V., van Oorschot, R. A. H. and Morgan, R. M. (2015) 'The deposition and persistence of indirectly-transferred DNA on regularly-used knives', *Forensic Science International: Genetics Supplement Series,* 5: e498-e500.

Part of one of the studies of Secondary Transfer to Wood Glass and Metal

P1

Door handle
Good source

P2

Nitrile gloves – good vectors

Transfer via person or object

11ng

Cloth at crime scene

10 transfer chains for door handle; average of 55ng deposited on handle; 64% transferred to gloves; 32% transferred to cloth.

Fonnelop, A. E., Egeland, T. and Gill, P. (2015) 'Secondary and subsequent DNA transfer during criminal investigation', *Forensic Sci Int Genet,* 17, pp. 155-162.

Research likely influenced by observations on investigation of Meredith Kercher murder resulting in miscarriage of Justice

Gill, P. (2016) 'Analysis and implications of the miscarriages of justice of Amanda Knox and Raffaele Sollecito', *Forensic Sci Int Genet,* 23, pp. 9-18.

Washing machine experiments

DNA profiles from laundered semen stains recovered at least 8 months after deposition
•Micrograms of DNA and full DNA profiles recovered, irrespective of wash conditions
•No significant decline in DNA quantity and profile quality after multiple washes
•Both DNA sources detected on laundered T-shirt with semen stains from two donors
•Laundered semen-stained clothing should be examined in sexual assault cases

Brayley-Morris, H., Sorrell, A., Revoir, A. P., Meakin, G. E., Court, D. S. and Morgan, R. M. (2015) 'Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases', *Forensic Sci Int Genet,* 19, pp. 165-171.
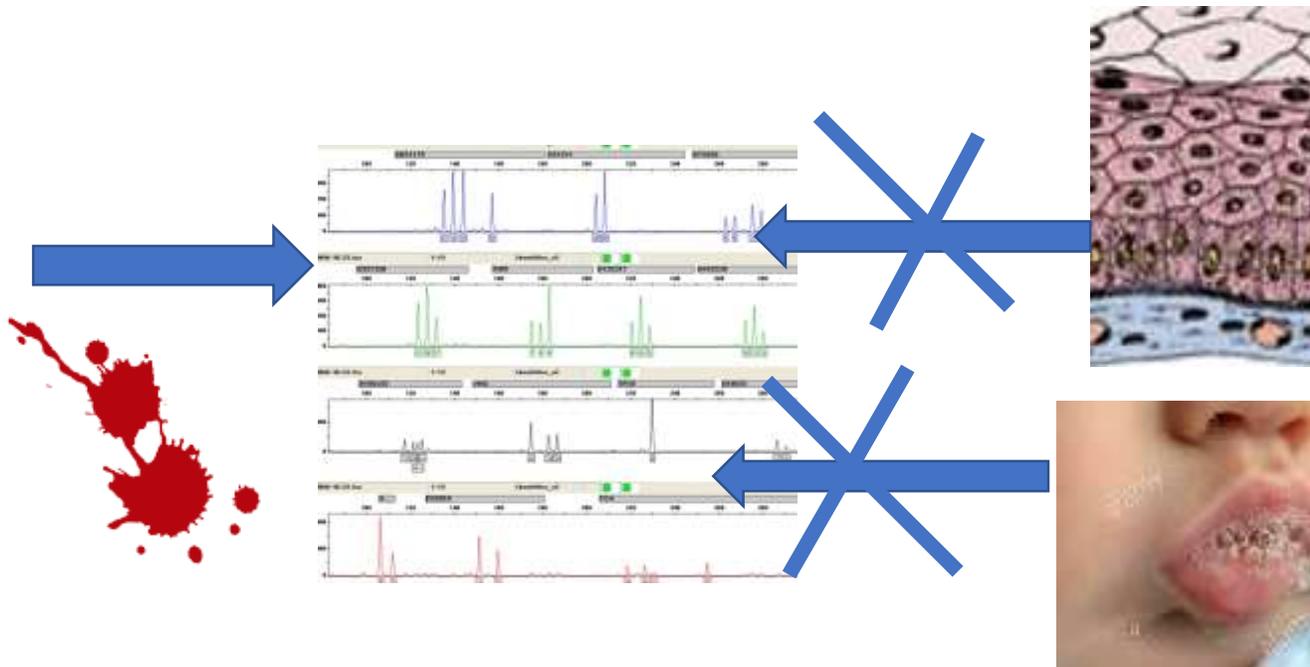
Secondary transfer of DNA from blood following washing but no usable profiles from saliva or epithelial abrasions

Kamphausen, T., Fandel, S. B., Gutmann, J. S., Bajanowski, T. and Poetsch, M. (2015) 'Everything clean? Transfer of DNA traces between textiles in the washtub', *Int J Legal Med,* 129(4), pp. 709-14.

Washing machine

	Item sampled	Microscopy	Swabs	Cuttings
Spermatozoa transfer		Positive	No DNA profiles	Some complete DNA profiles (S fraction)
Epithelial DNA transfer		n/a	No DNA profiles	Some complete DNA profiles (E fraction)
Bkground DNA on children's panties		Positive	No DNA profiles	Mixtures of family member's DNA (E fraction) / Some partial father's DNA profiles (S fraction)
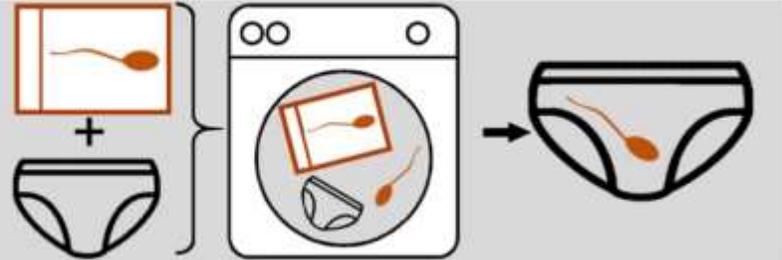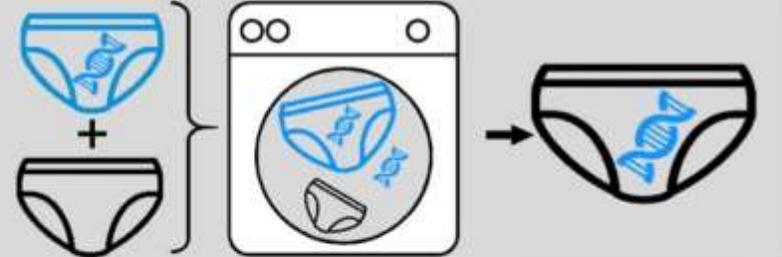
Noel, S., Lagace, K., Rogic, A., Granger, D., Bourgoin, S., Jolicoeur, C. and Seguin, D. (2016) 'DNA transfer during laundering may yield complete genetic profiles', *Forensic Sci Int Genet,* 23, pp. 240-247.

Pristine underwear washed with semen stained sheets
Background levels of DNA from family members on children's clothing

Recommended framework to ensure not relying on DNA profiles alone but use presumptive testing, sperm identification and other extrinsic factors

Study of outer clothing during regular activities

Because of the "ubiquitous nature and ease of transfer" of DNA, questions have arisen over the value of trace DNA as evidence.

- Examined outer clothing before and after washing
- DNA present from wearer, co-habiting partner and unknown
- Amount increased 8 fold after wear for a day
- More endogenous DNA on front and shoulders of shirts
- More extraneous DNA on back after day of wear
- Background DNA, not self present even in freshly washed garments
- Y profiles on female T-shirts – in one case where subject worked in a hospital, possibility of 7 males included in sample
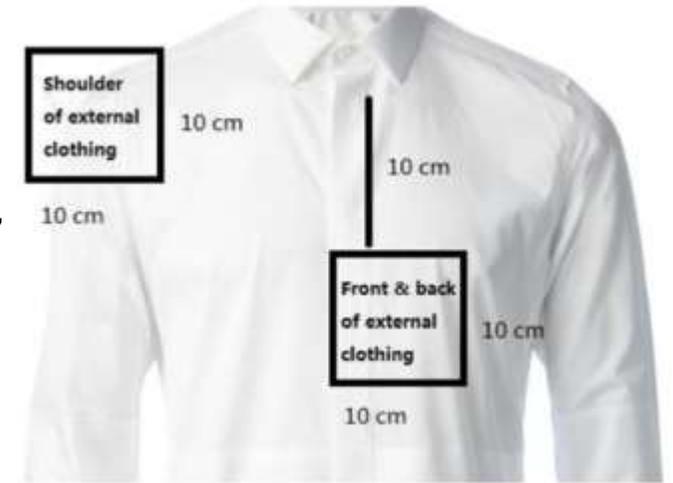
The adventitious transfer of trace DNA means that the DNA recovered in forensic casework may not always have evidentiary relevance.

RUAN, T., BARASH, M., GUNN, P. & BRUCE, D. 2018. Investigation of DNA transfer onto clothing during regular daily activities. *Int J Legal Med,* 132**,** 1035-1042

Studies to investigate factors that affect variability in casework scenarios

- Specific experiments to address various case types
  - Mixtures in sexual assault cases
  - Sexual intercourse v social contact – small number of experiments suggest you can divide between the two
  - Wearer v toucher – not as clear as might want
  - Digital penetration – number of fingernail studies
  - Particular issues in a domestic setting – ease of transfer particularly in washing machine

JONES, S., SCOTT, K., LEWIS, J., DAVIDSON, G., ALLARD, J. E., LOWRIE, C., MCBRIDE, B. M., MCKENNA, L., TEPPETT, G., ROGERS, C., CLAYSON, N. & BAIRD, A. 2016. DNA transfer through nonintimate social contact. *Sci Justice,* 56**,** 90-5.

BREATHNACH, M., WILLIAMS, L., MCKENNA, L. & MOORE, E. 2016. Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. *Forensic Sci Int Genet,* 20**,** 53-60.

MCDONALD, A., JONES, E., LEWIS, J. & O'ROURKE, P. 2015. Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. *Forensic Sci Int Genet,* 15**,** 84-9.
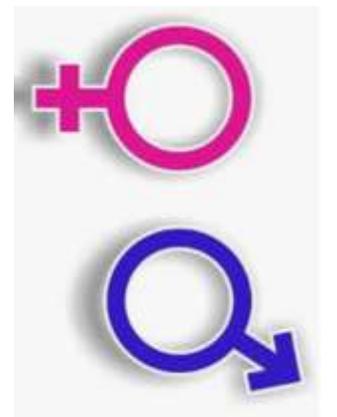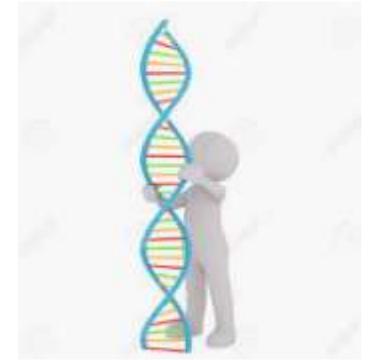
An overview from various studies

- **It is possible to handle an item without transferring any detectable DNA to it even when not wearing gloves**

  The absence of detectable DNA was noted in 11% of experiments by MANOLI,et al. 2016. and in 2.9% by GORAY et al. 2016

- **A profile may have been deposited before the crime and therefore may not be relevant to it**

  Raymond et al. 2009, Goray et al. 2015 Fonnelop et al. 2015, Oldoni et al. 2015, Oldoni et al. 2016, Meakin et al. 2015, Meakin et al. 2017, Pfeifer and Wiegand 2017

- **Detected DNA might be present due to indirect (secondary or tertiary) transfer, whether by a person or an object**

  Cale et al. 2016, Buckingham et al. 2016, Goray et al. 2016

None detected

Not my FAULT

Deposited before the crime

Secondary or higher order transfer

References from the Overview Slide
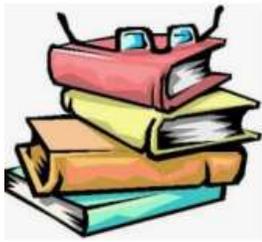
- MANOLI, P., ANTONIOU, A., BASHIARDES, E., XENOPHONTOS, S., PHOTIADES, M., STRIBLEY, V., MYLONA, M., DEMETRIOU, C. & CARIOLOU, M. A. 2016. Sex-specific age association with primary DNA transfer. *Int J Legal Med,* 130**,** 103-12.

- GORAY, M., FOWLER, S., SZKUTA, B. & VAN OORSCHOT, R. A. H. 2016. Shedder status-An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time. *Forensic Sci Int Genet,* 23**,** 190-196

- RAYMOND, J. J., VAN OORSCHOT, R. A., GUNN, P. R., WALSH, S. J. & ROUX, C. 2009. Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes. *Forensic Sci Int Genet,* 4**,** 26-33.

- GORAY, M. & VAN OORSCHOT, R. A. 2015. The complexities of DNA transfer during a social setting. *Leg Med (Tokyo),* 17**,** 82-91.

- FONNELØP, A. E., JOHANNESSEN, H. & GILL, P. 2015. Persistence and secondary transfer of DNA from previous users of equipment. *Forensic Science International: Genetics Supplement Series,* 5**,** e191-e192

- OLDONI, F., CASTELLA, V. & HALL, D. 2015. Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures. *Forensic Science International: Genetics Supplement Series,* 5**,** e300-e301

- OLDONI, F., CASTELLA, V. & HALL, D. 2016. Shedding light on the relative DNA contribution of two persons handling the same object. *Forensic Sci Int Genet,* 24**,** 148-157.

- MEAKIN, G. E., BUTCHER, E. V., VAN OORSCHOT, R. A. H. & MORGAN, R. M. 2015. The deposition and persistence of indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics Supplement Series,* 5**,** e498-e500.

- MEAKIN, G. E., BUTCHER, E. V., VAN OORSCHOT, R. A. H. & MORGAN, R. M. 2017. Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives. *Forensic Sci Int Genet,* 29**,** 38-47.

- PFEIFER, C. M. & WIEGAND, P. 2017. Persistence of touch DNA on burglary-related tools. *Int J Legal Med,* 131**,** 941-953.

- CALE, C. M., EARLL, M. E., LATHAM, K. E. & BUSH, G. L. 2016. Could Secondary DNA Transfer Falsely Place Someone at the Scene of a Crime? *J Forensic Sci,* 61**,** 196-203.

- BUCKINGHAM, A. K., HARVEY, M. L. & VAN OORSCHOT, R. A. H. 2016. The origin of unknown source DNA from touched objects. *Forensic Sci Int Genet,* 25**,** 26-33.

- GORAY, M., FOWLER, S., SZKUTA, B. & VAN OORSCHOT, R. A. H. 2016. Shedder status-An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time. *Forensic Sci Int Genet,* 23**,** 190-196.

DNA can be transferred from one surface or person to another person or surface. The DNA present on an evidence item may be unrelated to the crime being investigated.
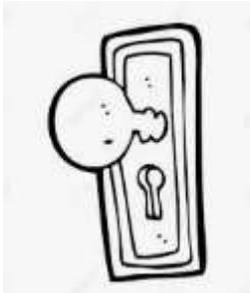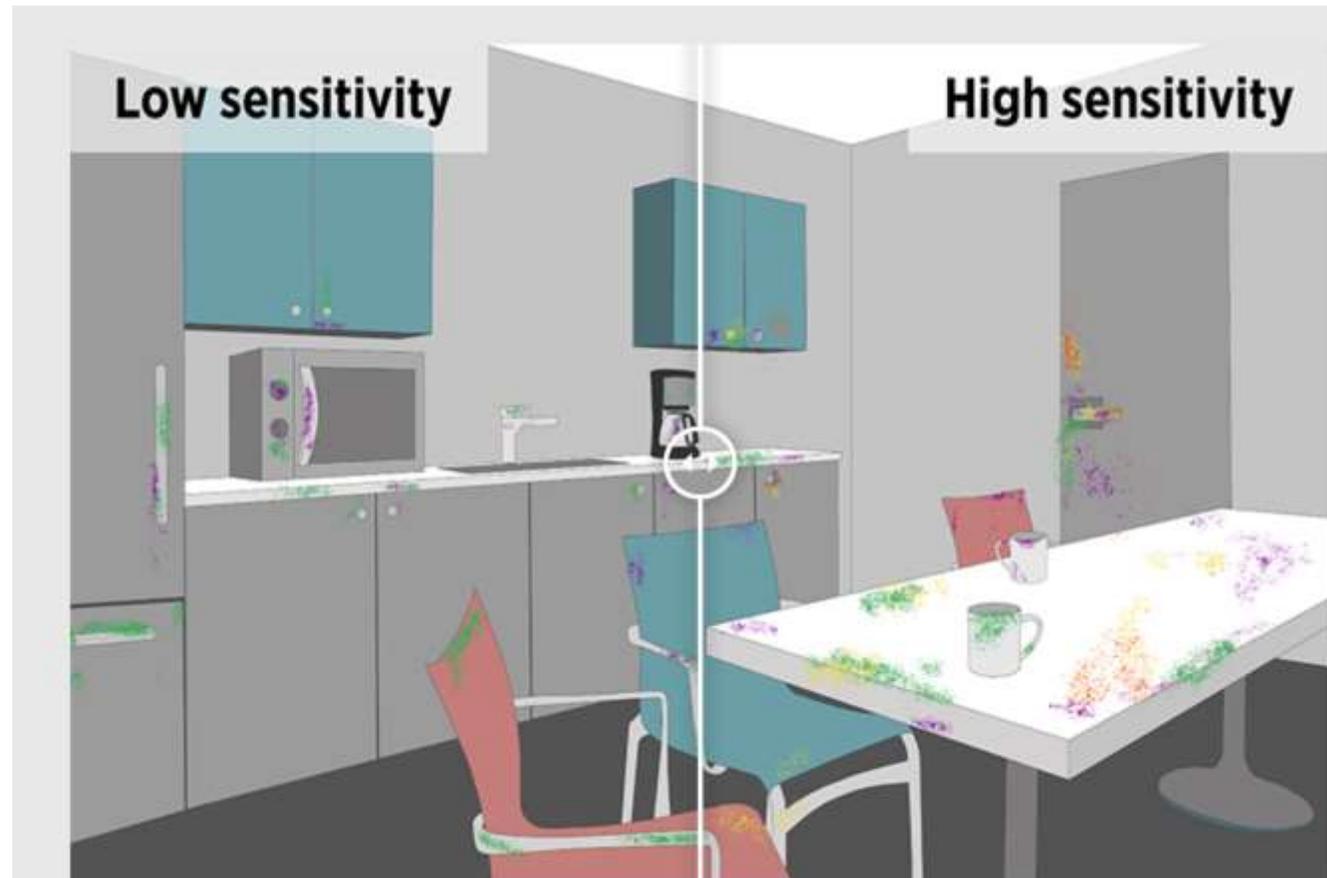

Fonnelop, A. E., Egeland, T. and Gill, P. (2015) 'Secondary and subsequent DNA transfer during criminal investigation', *Forensic Sci Int Genet,* 17, pp. 155-162.

The highly sensitive DNA methods that have become common in recent years increase the likelihood of detecting irrelevant DNA.

https://www.nist.gov/featured-stories/dna-mixtures-forensic-science-explainer

Low sensitivity   High sensitivity
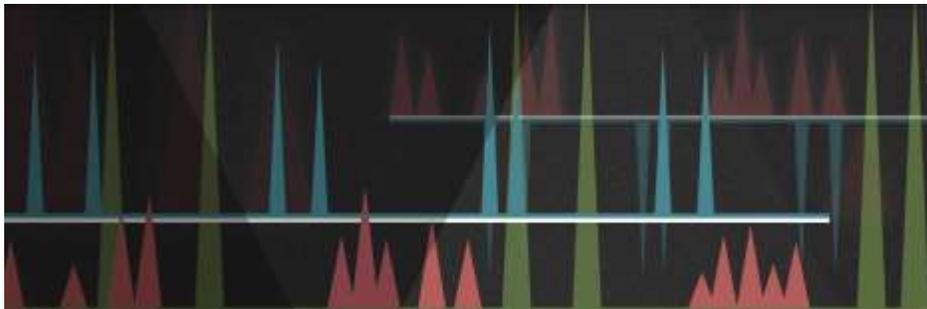
Contamination studies

Potential sources – tools and items used to safeguard evidence

Szkuta, B., Oorschot, R. and Ballantyne, K. N. (2017) 'DNA decontamination of fingerprint brushes', *Forensic Sci Int,* 277, pp. 41-50.

Szkuta, B., Harvey, M. L., Ballantyne, K. N. and van Oorschot, R. A. H. (2015) 'Residual DNA on examination tools following use', *Forensic Science International: Genetics Supplement Series,* 5(Supplement C), pp. e495-e497.

Fonnelop, A. E., Johannessen, H., Egeland, T. and Gill, P. (2016) 'Contamination during criminal investigation: Detecting police contamination and secondary DNA transfer from evidence bags', *Forensic Sci Int Genet,* 23, pp. 121-129.

GORAY, M., PIRIE, E. & VAN OORSCHOT, R. A. H. 2019. DNA transfer: DNA acquired by gloves during casework examinations. *Forensic Sci Int Genet,* 38, 167-174.

Several of the observed contacts made by the gloves were deemed high contamination risk events.

Highly sensitive methods also increase the risks of contamination. Contamination avoidance procedures should be used both at the crime scene and the laboratory. Elimination databases can help address issues of contamination.

To assess the relevance of DNA evidence, one should consider how the DNA fits into the full context of the case, including other pieces of evidence.

Tend to hear of cases where there are issues – need to use them rather than dismiss to ensure mistakes not repeated;
Jama case in Australia DNA the only evidence in a rape charge

"*In the present case, the obviously unreserved acceptance of the reliability of the DNA evidence appears to have so confined thought that it enabled all involved to leap over a veritable mountain of improbabilities and unexplained aspects that, objectively considered, could be seen to block the path to conviction*" (Vincent 2010).

INQUIRY INTO THE CIRCUMSTANCES THAT LED TO THE CONVICTION OF MR FARAH ABDULKADIR JAMA THE HONOURABLE F HR VINCENT. AO QC (2010)

Unusual case in Detroit where CODIS hit caused man to be charged in a cold case, only to find later that he had consenual sex - sample was a mixture – questions re source of other genotype?
https://www.washingtonpost.com/nation/2019/07/24/james-chad-lewis-clay-dna-detroit-rape-kit-testing-wrongful-conviction

In my opinion, there should be a red flag when DNA is the only evidence.

Not new
"It is emphasized that the relevance of the DNA evidence in a case can only be assessed by a concurrent consideration of all the non-DNA evidence." Gill 2001
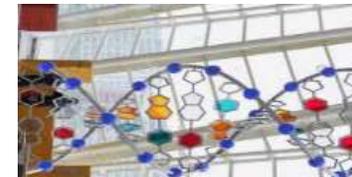
An investigative approach called Case Assessment and Interpretation, also called Evaluative Reporting, provides a framework for assessing the relevance of DNA

Inputs
Items from scene and suspects
Information
Outline of what questions to be answered

Identify what work to be attempted

Set propositions

Document expected outcomes as pre-assessment

Carry out the agreed work

Check if they need to be reassessed in light of results – unusual known materials for example

Report as sub-source or activity propositions

Scientist
Knowledge on how materials expected to behave in the given circumstances

decide what domain knowledge and context information is necessary

Cook, R., Evett, I. W., Jackson, G., Jones, P. J. and Lambert, J. A. (1998) 'A model for case assessment and interpretation', *Science and Justice,* 38(3), pp. 151-156.

$$LR = \frac{Pr\,(E/H_P\,,\,I)}{Pr\,(E/H_d\,,\,I)}$$

Hierarchy of Propositions

Proposition level	Increasing importance of context / Closer to the questions relevant to the court / Requires increasing information beyond the profile	Questions being addressed	Data needed
Offence proposition		Guilt or innocence?	Technical findings, motive, opportunity, witnesses, etc.
Activity proposition		What activity caused the DNA to be transferred?	Information about transfer and persistence
Source proposition		Can the POI be associated with a body fluid or cell type – blood, semen, saliva or epithelial?	Genotype as well as extrinsic properties, e.g. size and type of stain
Sub-source proposition		Can the POI be associated with genotype in a mixture with no information about cell type?	Only genotype considered - relevance not probed
Sub-sub-source proposition		Can the POI be associated with a part of a mixture without reference to all alleles?	Selected alleles present in a profile considered

Increased test sensitivity

All the published work based on the Case Assessment and Interpretation
Model produced by Forensic Science Service

- Results are relevant in context of the case
- Consider an alternative proposition
- Comment on the findings, ideally at activity level , not the proposition

Cook, R., et al. "A model for case assessment and interpretation." Science and Justice (1998)**38**(3): 151-156.

ENFSI GUIDELINE FOR EVALUATIVE REPORTING IN FORENSIC SCIENCE

Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)

European Network of Forensic Science Institutes

With the financial support of the Prevention of and Fight against Crime Programme of the European Union European Commission - Directorate -General Justice, Freedom and Security

A project funded by the EU ISEC 2010
Agreement Number: HOME/2010/ISEC/MO/4000001759

Science and Justice

Standards for the formulation of evaluative forensic science expert opinion

ENFSI GUIDELINE FOR EVALUATIVE REPORTING IN FORENSIC SCIENCE

A Primer for Legal Practitioners

European Network of Forensic Science Institutes

With the financial support of the Prevention of and Fight against Crime Programme of the European Union European Commission - Directorate - General Justice, Freedom and Security

A project funded by the EU ISEC 2010
Agreement Number: HOME/2010/ISEC/MO/4000001759

An introductory guide to Evaluative Reporting

Feedback that guideline was too technical

enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

wp.unil.ch/forensicdecision/files/2016/02/Champod_etal_Primer_2016.pdf

NIFS (2017) 'An Introductory Guide to Evaluative Reporting.', *secretariat.nifs @anzpaa.org.au*.

Coursera course available on-line entitled "Challenging Forensic Science; How Science Should Speak to Court" https://www.coursera.org/learn/challenging-forensic-science

DNA match statistics only provide information about the source of a DNA sample (who the DNA might have come from). They do not indicate how or when the DNA was transferred, or whether it is related to the crime.

This tendency to answer the easier question has been termed attribute substitution, "if someone doesn't know the answer to a difficult question, they will substitute and easier question (even if subconsciously) and answer that instead" (Eldridge 2019)

ELDRIDGE, H. 2019. Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*, 1**,** 24-34.

A
T

C
G
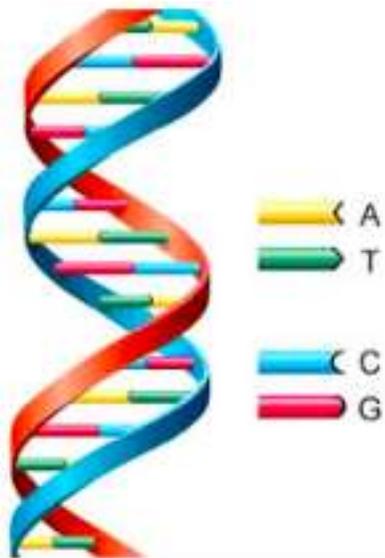
Who? ≠ Who did it?

Investigation v Evaluation

Sub-source LR

Not sufficient as stand alone

Need additional considerations

GILL et al., 2018. DNA commission ISFG-. *Forensic Sci Int Genet,* 36**,** 189-202.

Profile will be the same whether comes directly from Mr. X or via handshake or surface
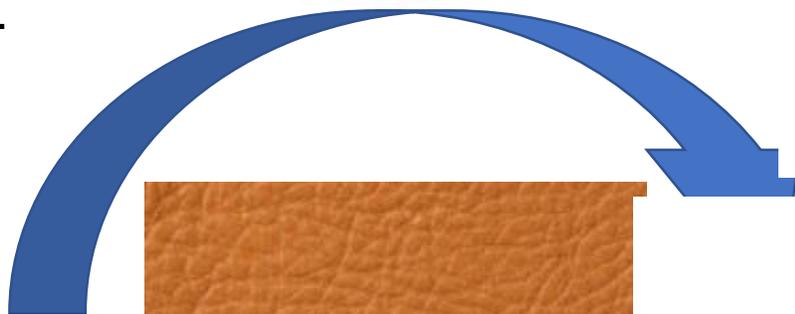
Source of DNA
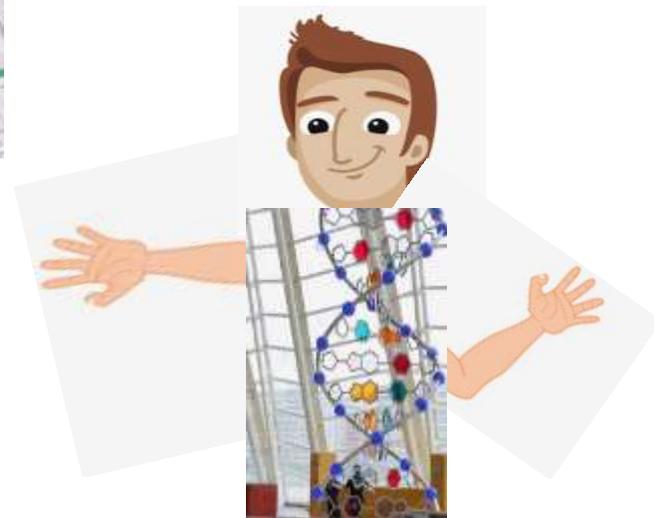
Possible mechanisms of transfer

Values for activity LRs 12-40 Peter Gill RTI webinar 1st May 2019

An early paper illustrated this using case studies showing an activity level LR of the order of 1,000 while there is an infinite LR in favor of sub-level proposition.

EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. & CHAMPOD, C. 2002. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *J Forensic Sci,* 47**,** 520-30.

Summary

- The fact that DNA transfers makes in invaluable in investigating crime
- We have little fundamental knowledge about how it transfers from a person
- Data on variables that affect transfer and persistence highlight that relevance to the crime cannot be taken for granted
- Given this knowledge, there is a risk in using the sub-source LR, or any statistic dealing with rarity of the profile, in isolation
- Mitigating strategies for reducing this risk include
  - Considering the context
  - Be particularly conscious of risk if DNA is the only evidence
  - Reduce contamination at all stages from the scene until the profile is produced
  - Use Case Assessment and Interpretation
  - Consider activity propositions as a way to address the appropriate questions
  - Separate investigation from evaluation

Thank you

Happy to take questions or comments now or later

sheilawillis6@gmail.com

Some Key Takeaways and Best Practices

John

LR Values are Relative, Subjective, and Contextual

- Decisions need to be made on the evidence (**E**) to be used in the likelihood ratio and with the contextual information (**I**) available

$$LR = \frac{\Pr(\boldsymbol{E}|H_1, \boldsymbol{I})}{\Pr(\boldsymbol{E}|H_2, \boldsymbol{I})}$$

"E" can change **based on the analytical threshold** used as well as decisions about artifacts, such as stutter products

The propositions (H_1 and H_2) can change **depending on case context** and assumptions made, such as the number of contributors

"**There are no true likelihood ratios, just like there are no true models**. Depending on our assumptions, our knowledge and the results we want to assess, different models will be adopted, hence different values for the LR will be obtained. **It is therefore important to outline in our statements what factors impact evaluation (propositions, information, assumptions, data, and choice of model)**." (Gill *et al.* 2018, *FSI Genetics* 36:189-202)

Desired Performance with a Mixture Interpretation Method
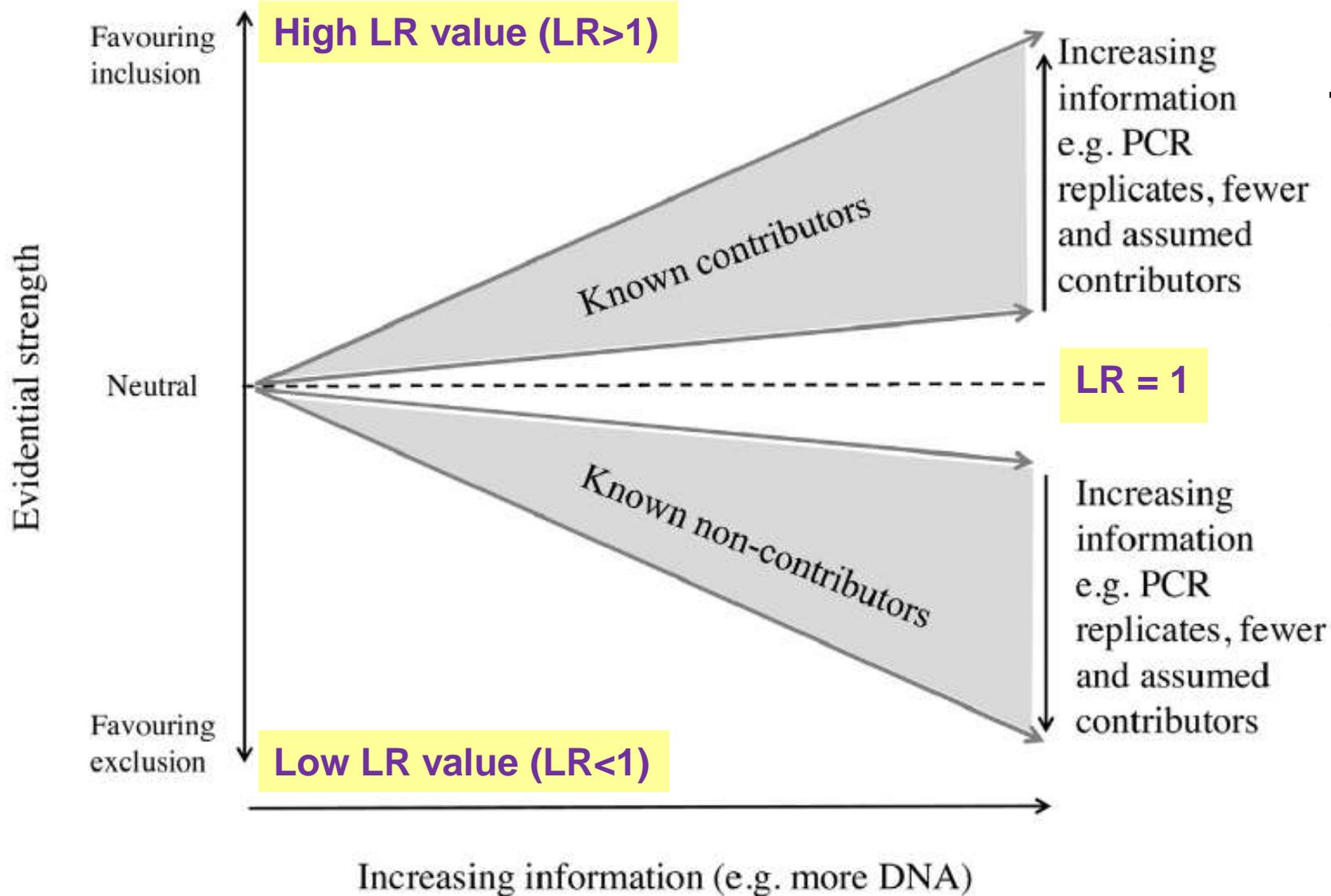
Desirable Features

1. Discrimination capacity
   (separation of known contributors from known non-contributors)

2. Calibration
   (accuracy of a specific LR value)

LR values vary based on amount of information available – **with less information, a lower LR value is obtained with a well-calibrated system**

Fig. 1 from Bright et al. (2016) Developmental validation of STRmix… *FSI Genetics* 23: 226-239

What Question Are We Answering?

- Depends on the propositions (hypotheses made)
  - Need to consider **the hierarchy of propositions** (Cook et al. 1998)

- The likelihood ratio (LR) is **the probability of getting the evidence** *if* the defendant is a contributor compared to the probability of getting the evidence *if* that defendant is not a contributor.
  - Involves assumptions as to the number of contributors in the mixture
  - LR is influenced by a number of inputs

Consider Appropriate Levels of Propositions in the Hierarchy of Propositions

- It is vital that results from one level of proposition are not used to move to another level without necessary information

- Matching DNA is not proof of guilt

- Matching DNA is not automatically relevant to the crime

Sub-source LR

Activity LR

Who? ≠ **How did it get there?**

Offense LR

Who? ≠ **Who did it?**

SWGDAM Documents

Validation Guidelines

December 2016

Interpretation Guidelines

January 2017

LR Verbal Equivalents

July 2018

UK Forensic Science Regulator

| Codes of Practice and Conduct	PGS Software Validation	DNA Mixture Interpretation
Forensic Science Regulator — *Overseeing Quality* **Codes of Practice and Conduct** for forensic science providers and practitioners in the Criminal Justice System Issue 4 October 2017 © Crown Copyright 2017 The text in this document (excluding the Forensic Science Regulator's logo) may be reproduced in any format or medium providing it is reproduced accurately, is not otherwise attributed, is not used in a misleading context and is acknowledged as Crown copyright.	**Forensic Science Regulator** — *Overseeing Quality* **Guidance** Software Validation for DNA Mixture Interpretation FSR-G-223 ISSUE 1 © Crown copyright 2018 The text in this document (excluding the Forensic Science Regulator's logo and material quoted from other sources) may be reproduced free of charge in any format or medium, providing it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and its title specified.	**Forensic Science Regulator** — *Overseeing Quality* **Guidance** DNA Mixture Interpretation FSR-G-222 ISSUE 2 © Crown copyright 2018 The text in this document (excluding the Forensic Science Regulator's logo and material quoted from other sources) may be reproduced free of charge in any format or medium, providing it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and its title specified.
October 2017	**September 2018**	**October 2018**

Q & A

John, Hari, & Sheila
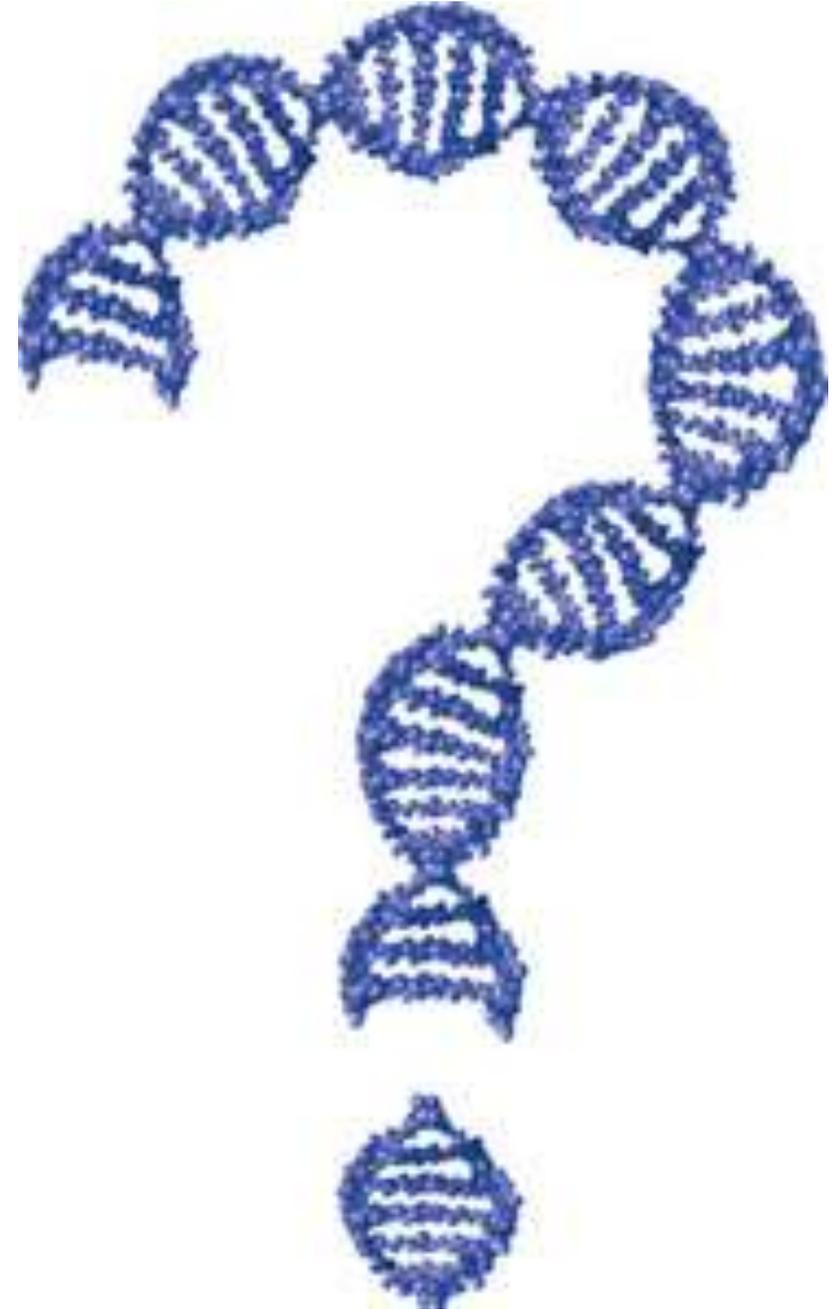
Some Common Challenges
with DNA Mixture Interpretation

- Differentiating stutter products from true alleles of another contributor

- Determining the number of possible contributors in a complex mixture

- Presenting multiple scenarios and accompanying LR values in reports and testimony

- Placing limits on the degree of complexity to examine or a lower limit on the amount of DNA to attempt amplifying and interpreting

- Obtaining sufficient training to understand probabilistic genotyping systems

- Understanding the possibilities of DNA transfer and their potential impact on a case

- Conducting validation experiments to cover the range of samples seen in casework

- Conveying to report users the meaning of results with low LR values

- Obtaining consistent results across analysts and laboratories

Thank you for your attention!

DNA Mixture Interpretation

Scientific Foundation Review

NIST FORENSIC SCIENCES
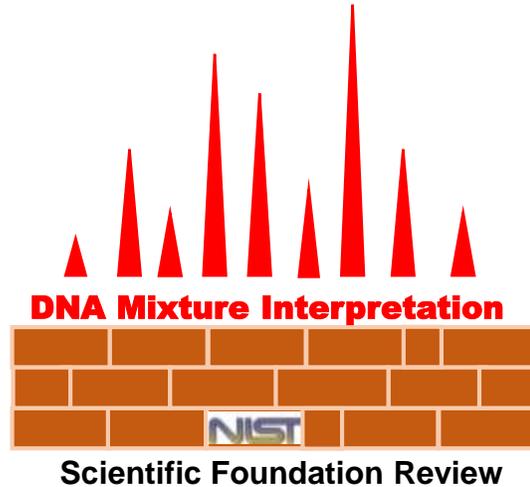
www.nist.gov/forensics

John M. Butler

301-975-4049

john.butler@nist.gov

Hari K. Iyer

301-975-2851

hariharan.iyer@nist.gov

Sheila Willis

sheilawillis6@gmail.com