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

AAFS 2019 Workshop #10 (February 18, 2019; Baltimore, MD)

Measurement & Interpretation of DNA Mixtures: Performance Assessment Methods

Hari Iyer

Statistical Engineering Division

Information Technology Laboratory

National Institute of Standards and Technology

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

DNA: MEASUREMENT & INTERPRETATION

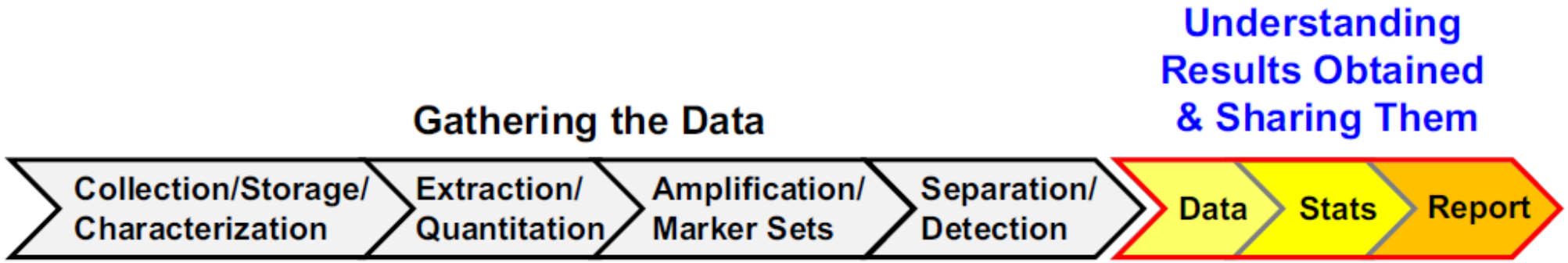


FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

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

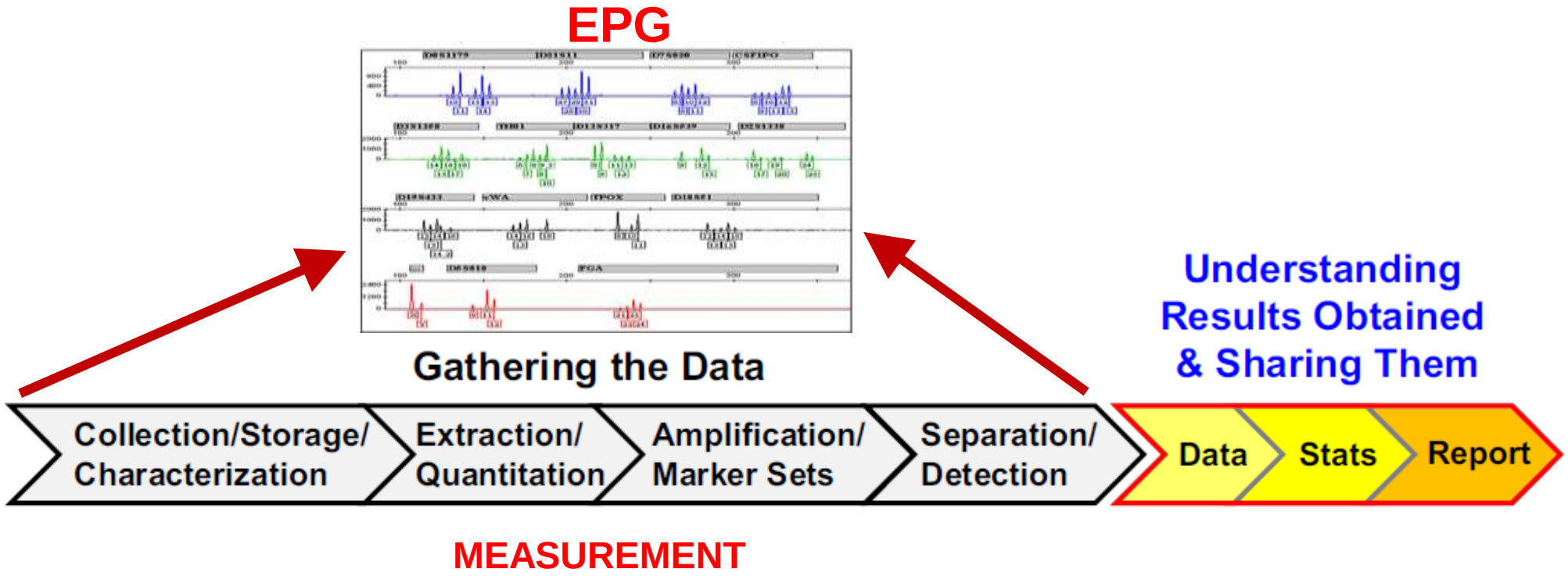
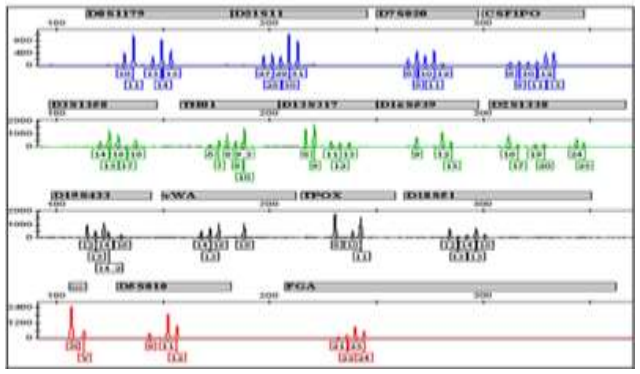


FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

JOHN M. BUTLER
National Institute of Standards and Technology
Gaithersburg, Maryland, USA

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

DNA: MEASUREMENT & INTERPRETATION



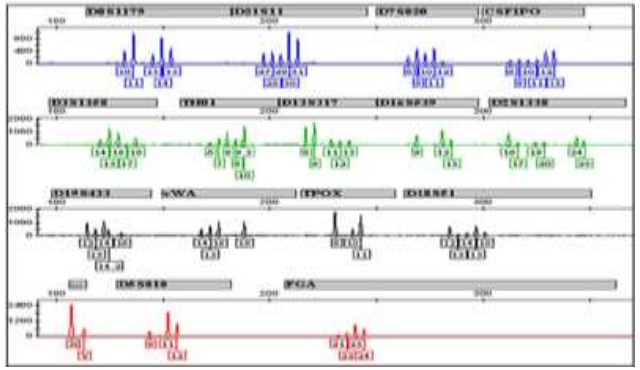
Gathering the Data

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



Gathering the Data

LR

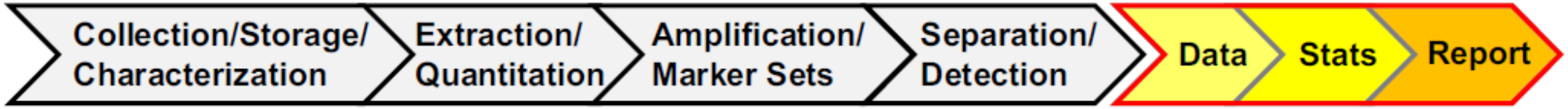


FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

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

LIKELIHOOD RATIO

H_p: DNA from POI is in the sample

H_d: DNA from POI is not in the sample

E = Expert Findings

I = Background Information prior to examining crime sample

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

DNA: MEASUREMENT & INTERPRETATION

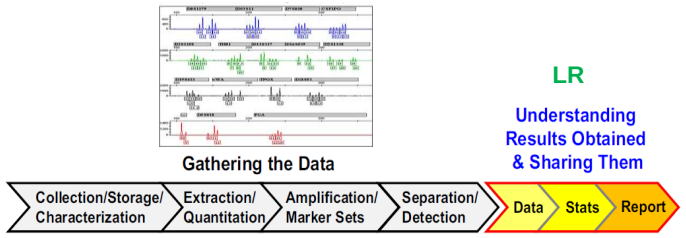
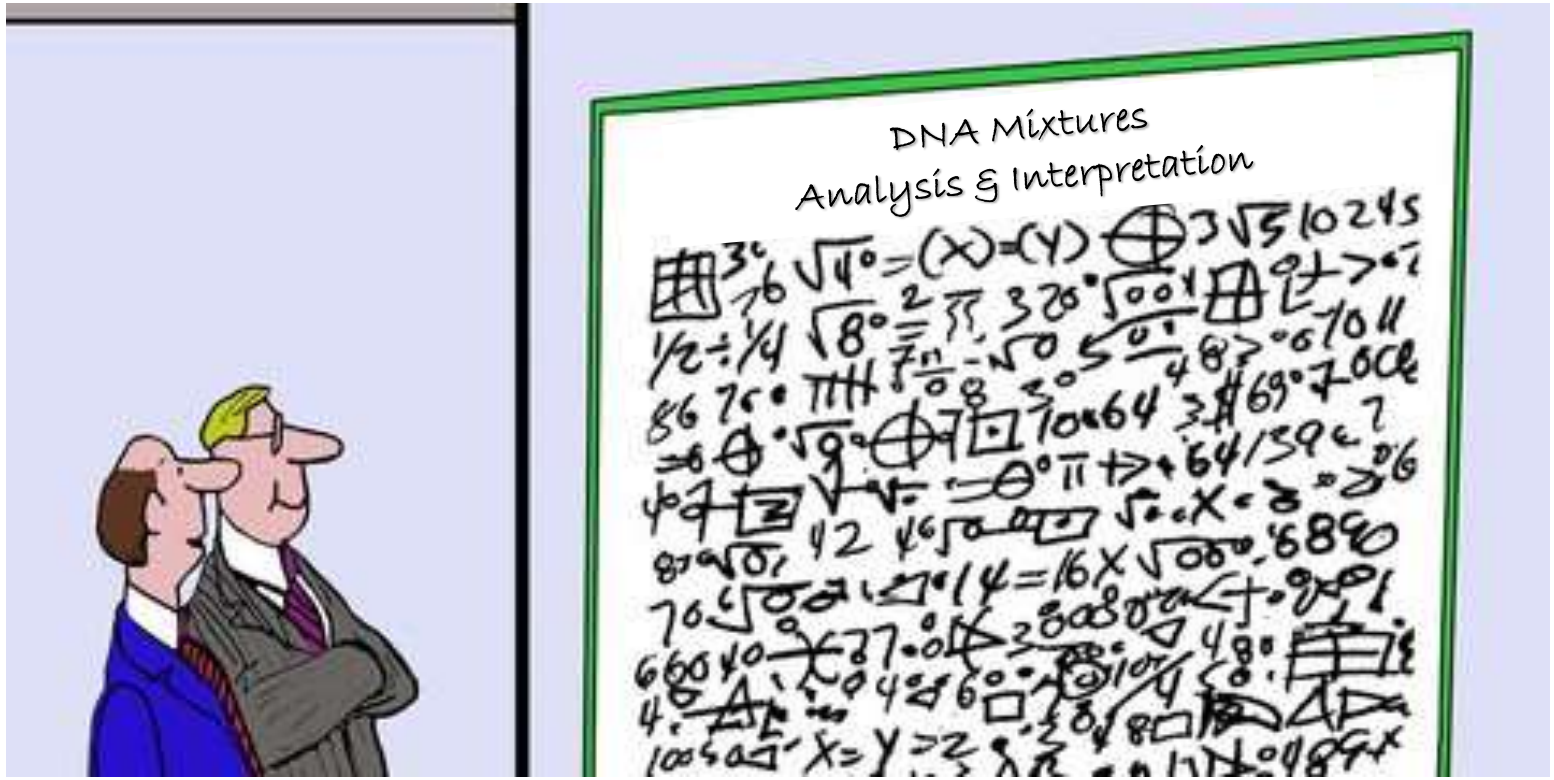


FIGURE 1.1 Steps involved in the overall process of forensic DNA typing. This book focuses on understanding the data through data interpretation and statistical interpretation.

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



RELIABILITY

HOW RELIABLE ARE THE RESULTS?

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.

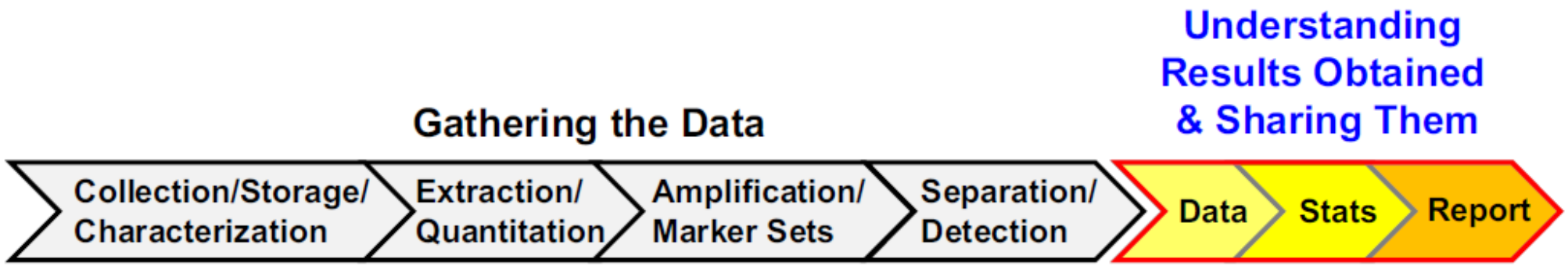
*Public confidence in **the reliability of** a company's financial statements depends on investors perceiving the company's auditors as being independent from the company.*

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

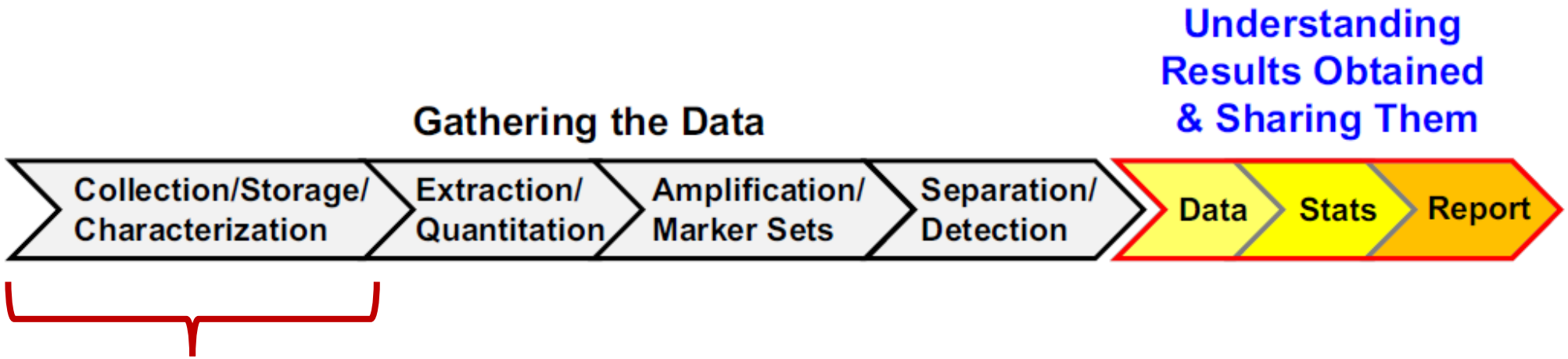
Reliability

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

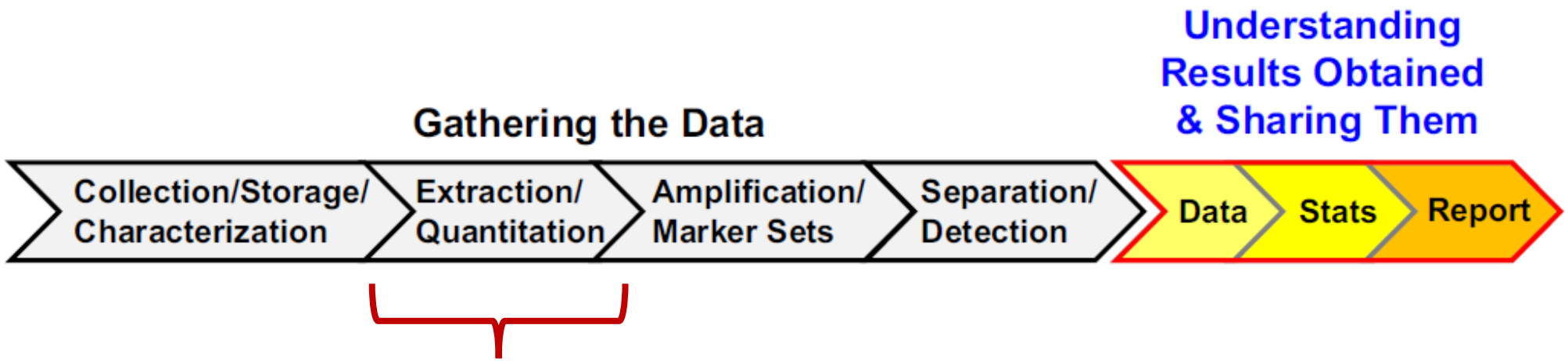
System Reliability vs Component Reliability



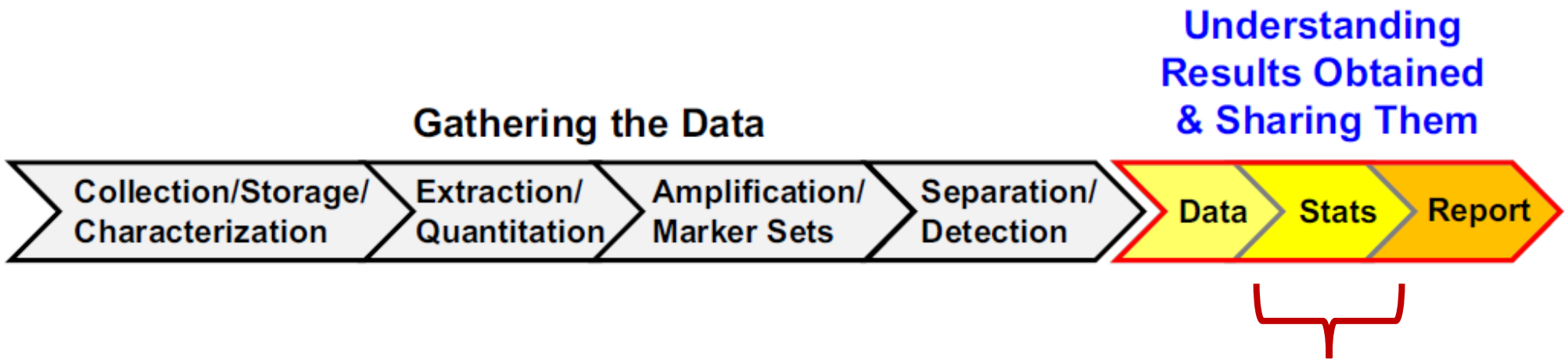
System Reliability vs Component Reliability



System Reliability vs Component Reliability



System Reliability vs Component Reliability



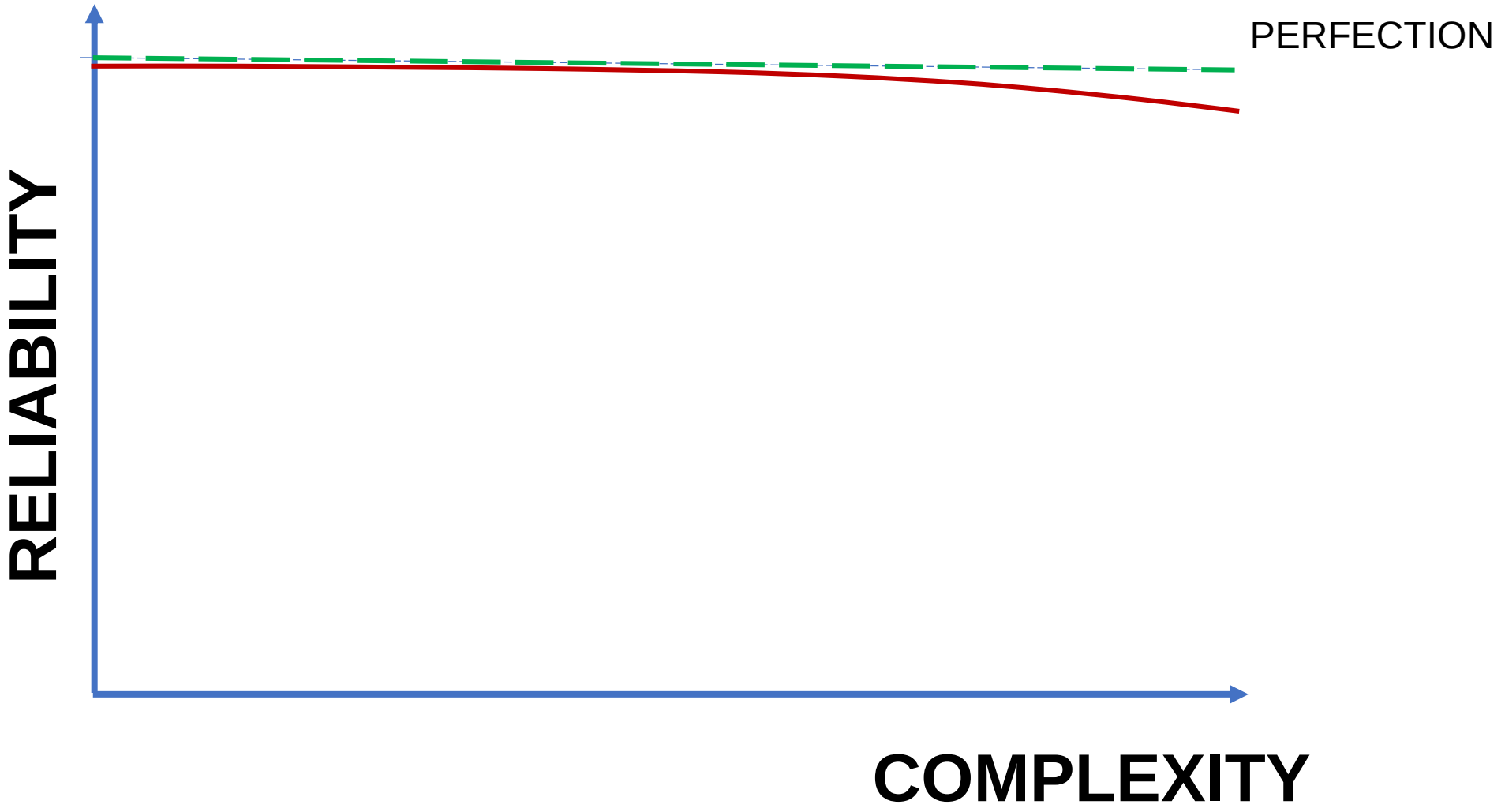
Some Factors Affecting Reliability of an LR System

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 laboratory specific parameters for use in the PG model
4. Software Implementing the PG Model
   a) Choice of numerical methods for computing LR (MCMC, Numerical Integration)
   b) Choice of burn-in cycles and number of ‘accepts’
      OR numerical integration parameters (such as grid size)

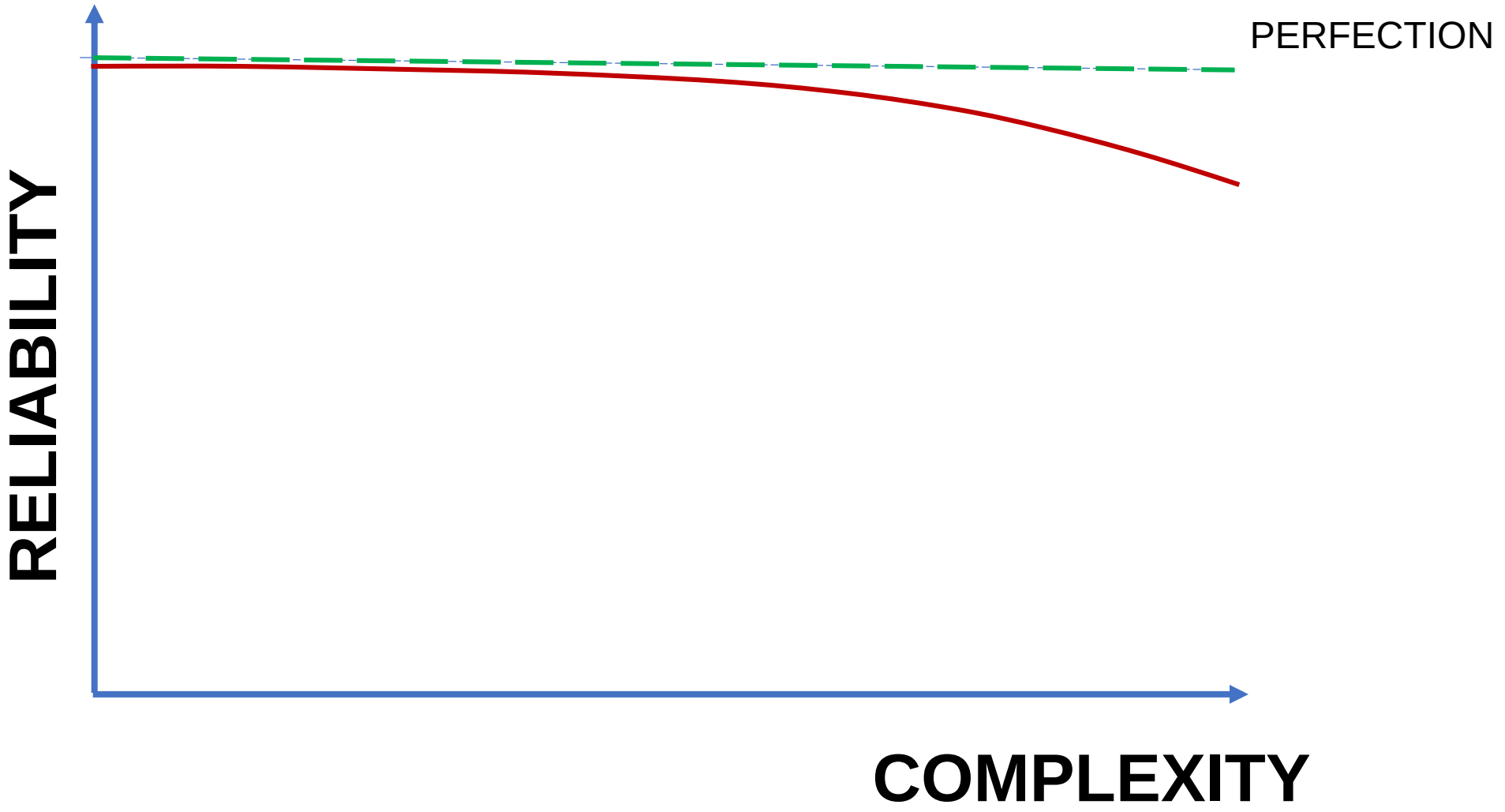
FACTOR SPACE

RELIABILITY vs COMPLEXITY

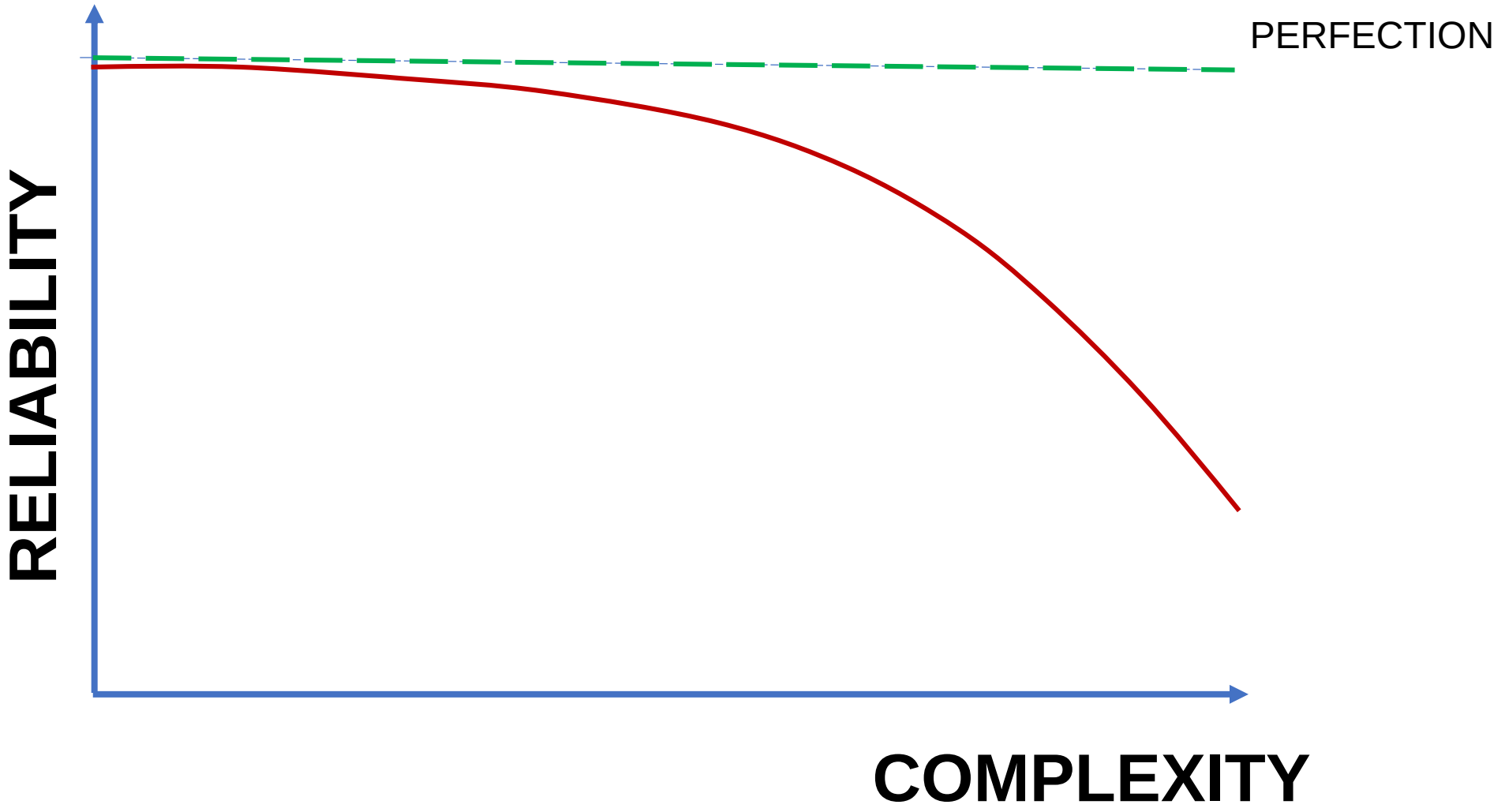
RELIABILITY vs COMPLEXITY



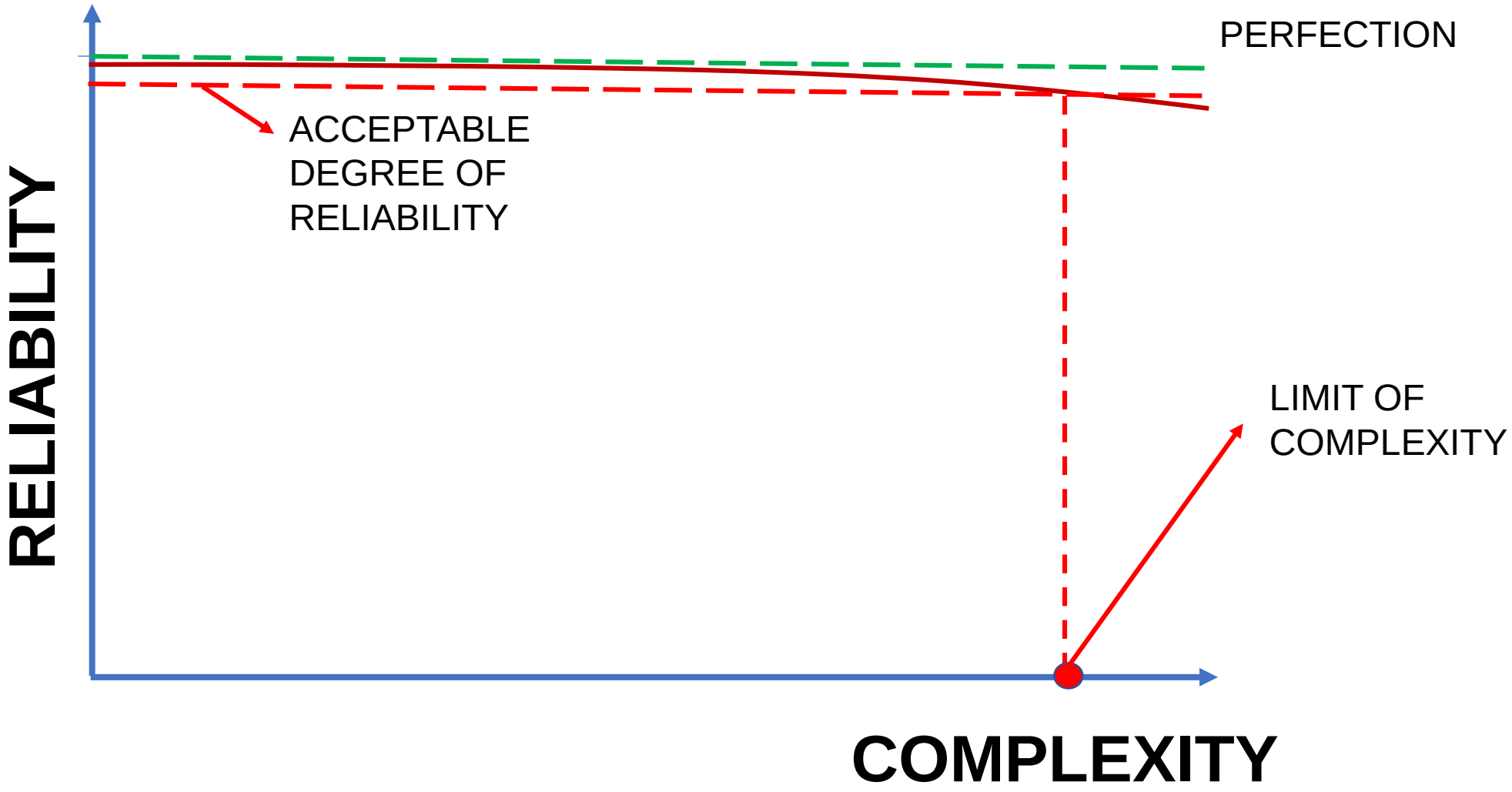
RELIABILITY vs COMPLEXITY



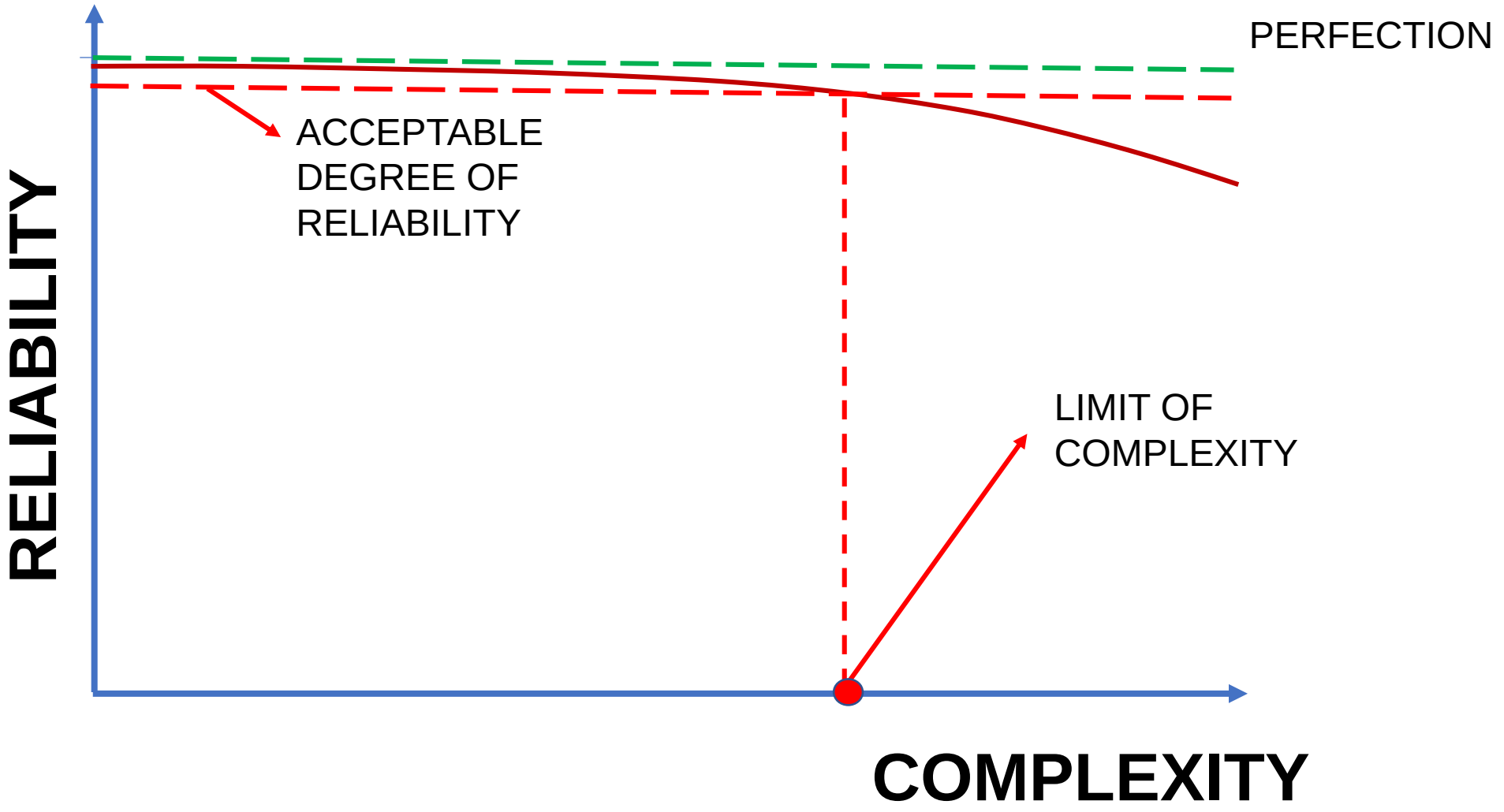
RELIABILITY vs COMPLEXITY



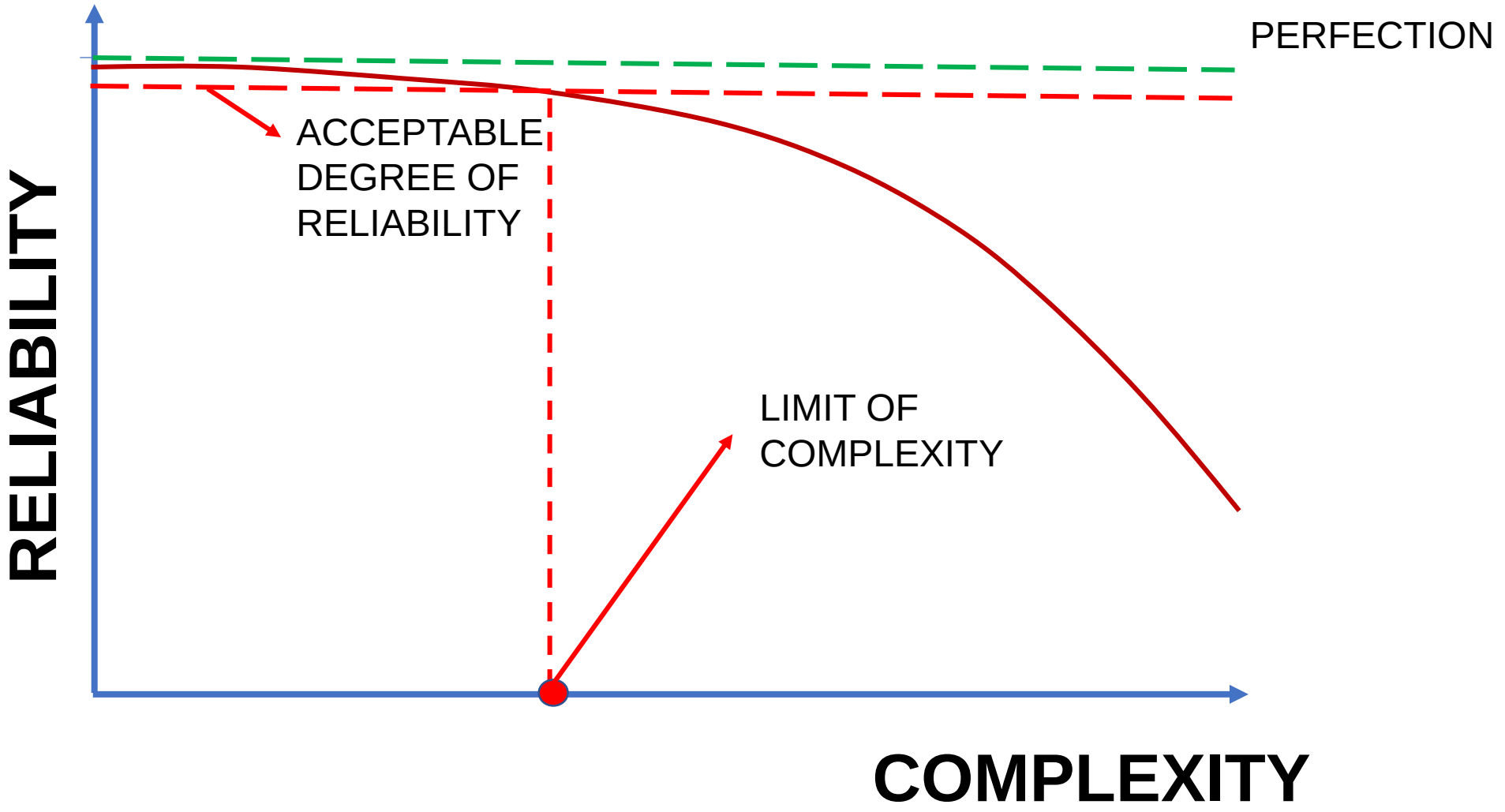
RELIABILITY vs COMPLEXITY



RELIABILITY vs COMPLEXITY



RELIABILITY vs COMPLEXITY



WHY DISCUSS RELIABILITY?

- NAS Report
- PCAST Report

WHY DISCUSS RELIABILITY?

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?

Will the lab results be close enough to the 'true value' that differences between the reported values and the true value would be inconsequential to the outcome of this case?

WHY DISCUSS RELIABILITY?

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

148 mg/dl, 175 mg/dl, 215 mg/dl,

375 mg/dl, 450 mg/dl

WHY DISCUSS RELIABILITY?

WHY DISCUSS RELIABILITY?

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

146 mg/dl, 151 mg/dl, 162 mg/dl,

155 mg/dl, 166 mg/dl

WHY DISCUSS RELIABILITY?

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

146 mg/dl, 151 mg/dl, 162 mg/dl,

155 mg/dl, 166 mg/dl

WHAT IS THE TRUE VALUE?

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 911c

Cholesterol

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).

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

WHY DISCUSS RELIABILITY?

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

146 mg/dl, 151 mg/dl, 162 mg/dl,

155 mg/dl, 166 mg/dl

REFERENCE VALUE = 100 mg/dl !!

WHY DISCUSS RELIABILITY?

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

146 mg/dl, 151 mg/dl, 162 mg/dl,

155 mg/dl, 166 mg/dl

REFERENCE VALUE = 100 mg/dl !!

DEVIATIONS: 46, 51, 62, 55, 66

WHY DISCUSS 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?

Forensic Science International: Genetics 35 (2018) 156–163

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

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

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).

			Other evaluations	
Labs	LR value	Hypothesis	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?

Forensic Science International: Genetics 37 (2018) 143–150

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

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

Page 145

DNA•VIEW®,
EuroForMix and
STRmix

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 ?

THERE IS NO TRUE LR

Then what do the LRs offered by different labs supposed to mean?

EMPIRICAL ASSESSMENT OF LR SYSTEMS

There are two aspects to judging the reliability of a 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 x is x times more likely to occur under H_p than under H_d.

- LR value of 1 occurs as often under H_p as under H_d
- LR value of 10 occurs 10 times more often under H_p as it does under H_d.
- LR value of 100 occurs 100 times more often under H_p as it does under H_d.
- LR value of 0.1 occurs 10 times more often under H_d as it does under H_p.
- And, **in general, whether a LR value of x occurs (roughly) x times more often under H_p than under H_d**.

ACCURACY/CALIBRATION

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 Hp true case or for a Hd 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.

Don’t Try This At Home !

Disclaimer:

This is only a thought experiment.

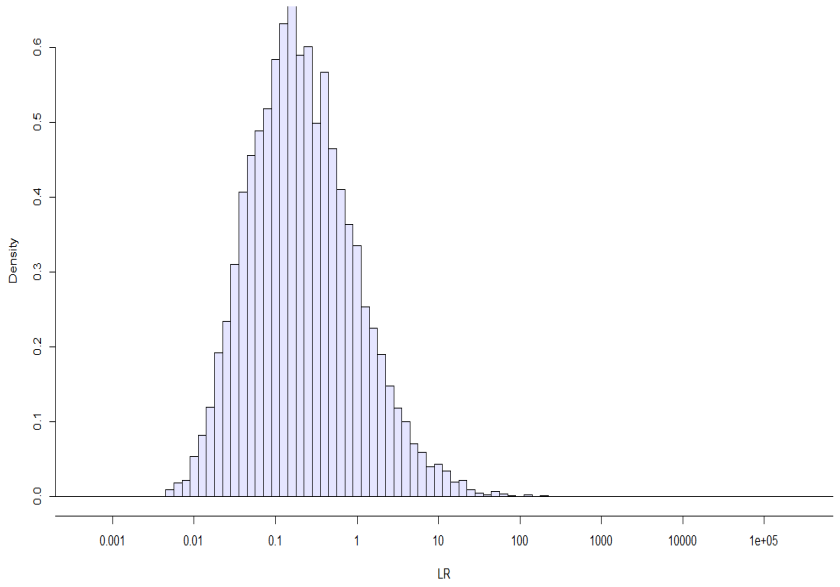
Actual assessment will require a well thought out experimental design and lots of

$$

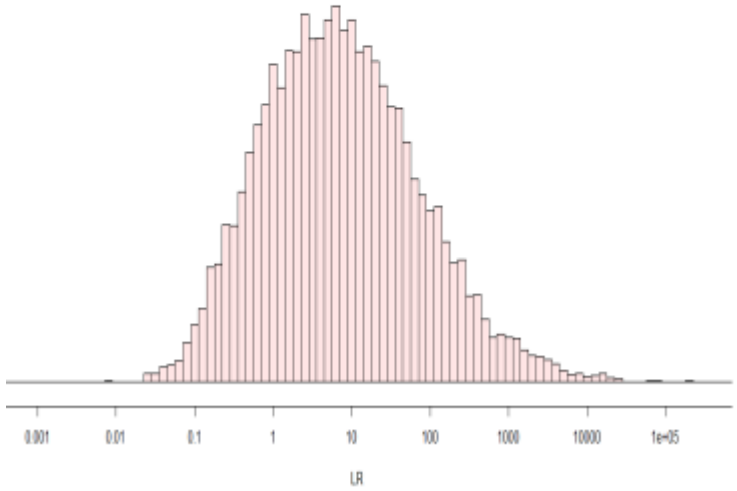
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT

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



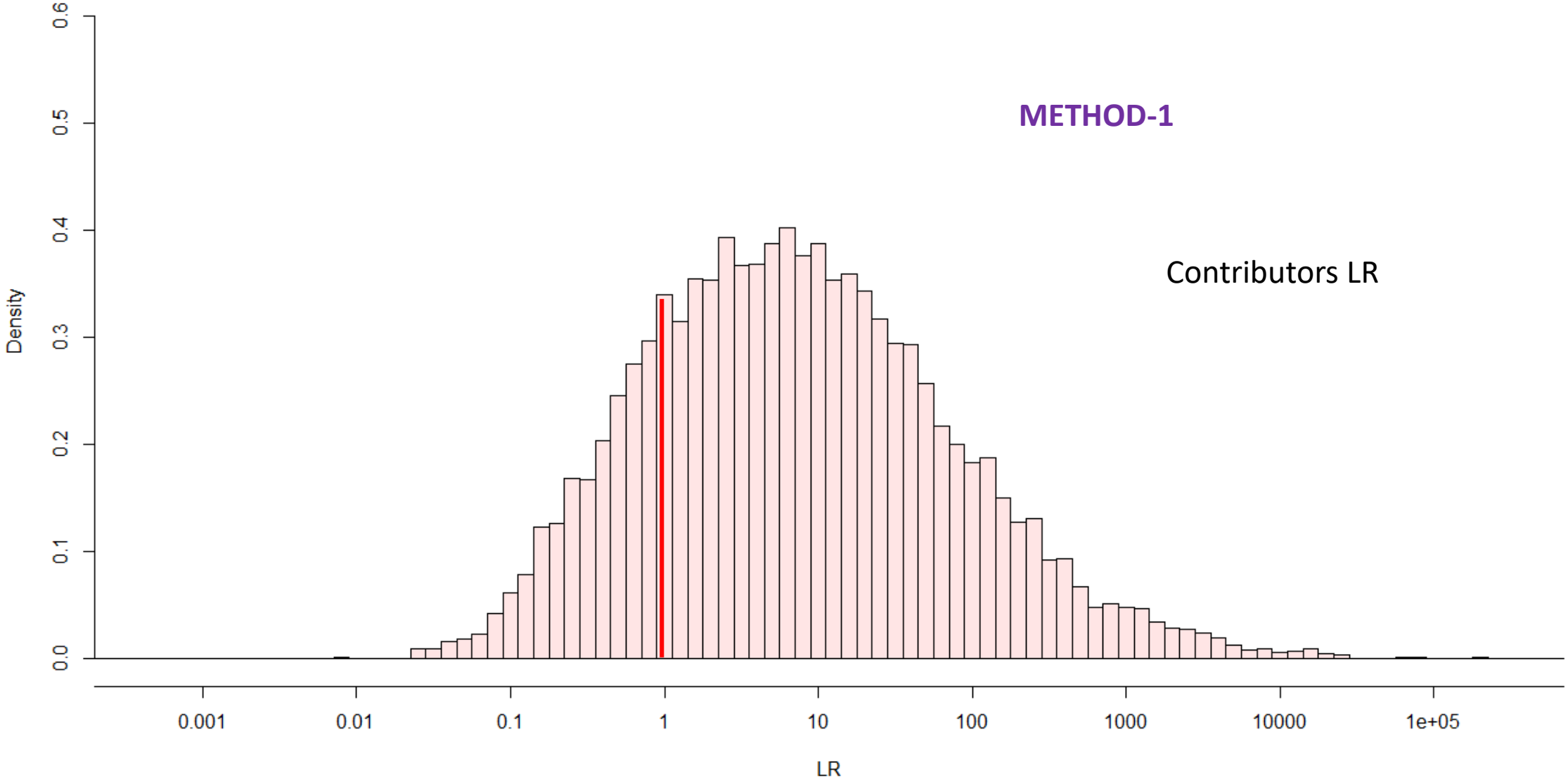
Disclaimer: This is only a thought experiment.

Actual assessment will require a well thought out experimental design.

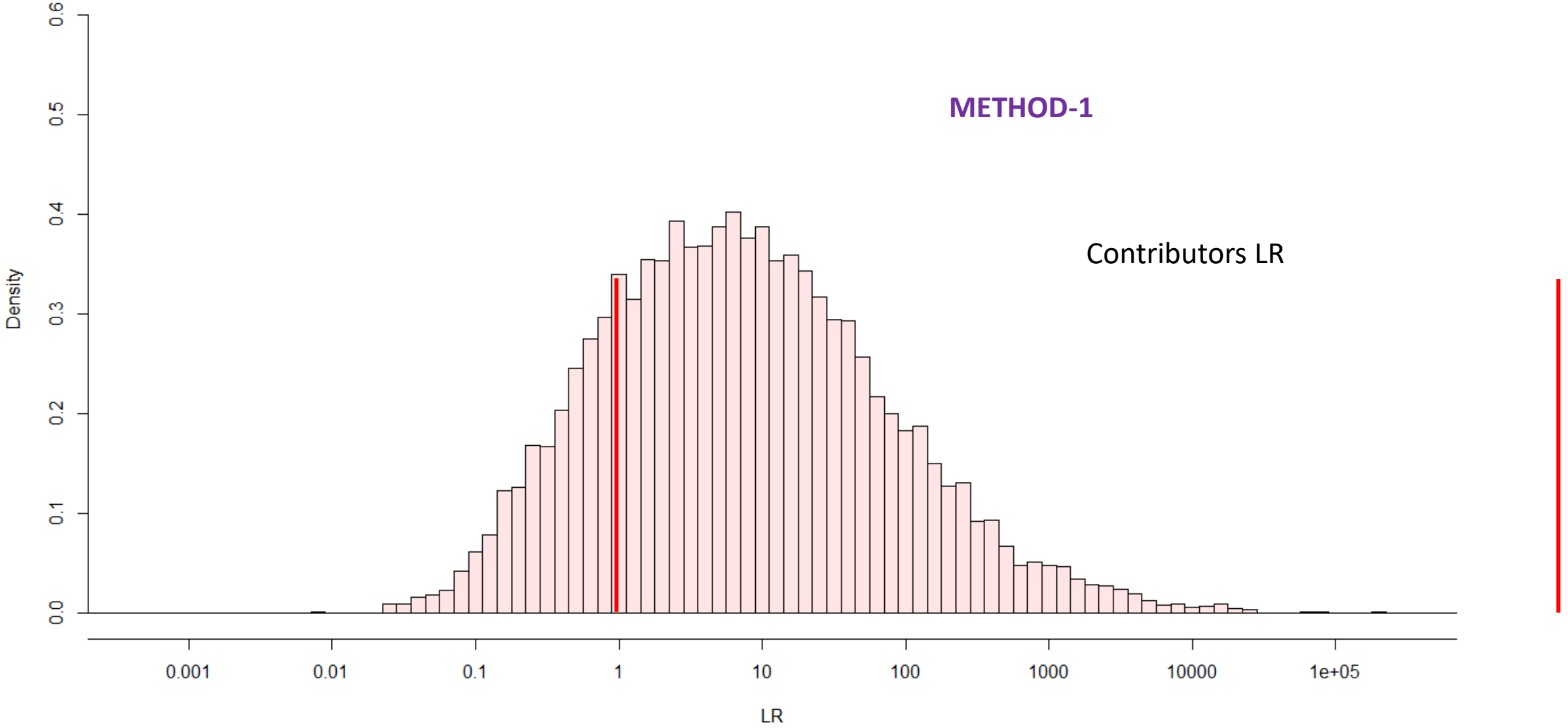
ACCURACY/CALIBRATION

- check whether a LR value of 1 occurs (roughly) equally often under H_p as well as under H_d
- check whether a LR value of 10 occurs (roughly) 10 times more often under H_p as it does under H_d.
- Check whether a LR value of 100 occurs (roughly) 100 times more often under H_p as it does under H_d.
- Check whether a LR value of 0.1 occurs (roughly) 10 times more often under H_d as it does under H_p.
- And, **in general, whether a LR value of x occurs (roughly) x times more often under H_p than under H_d**.

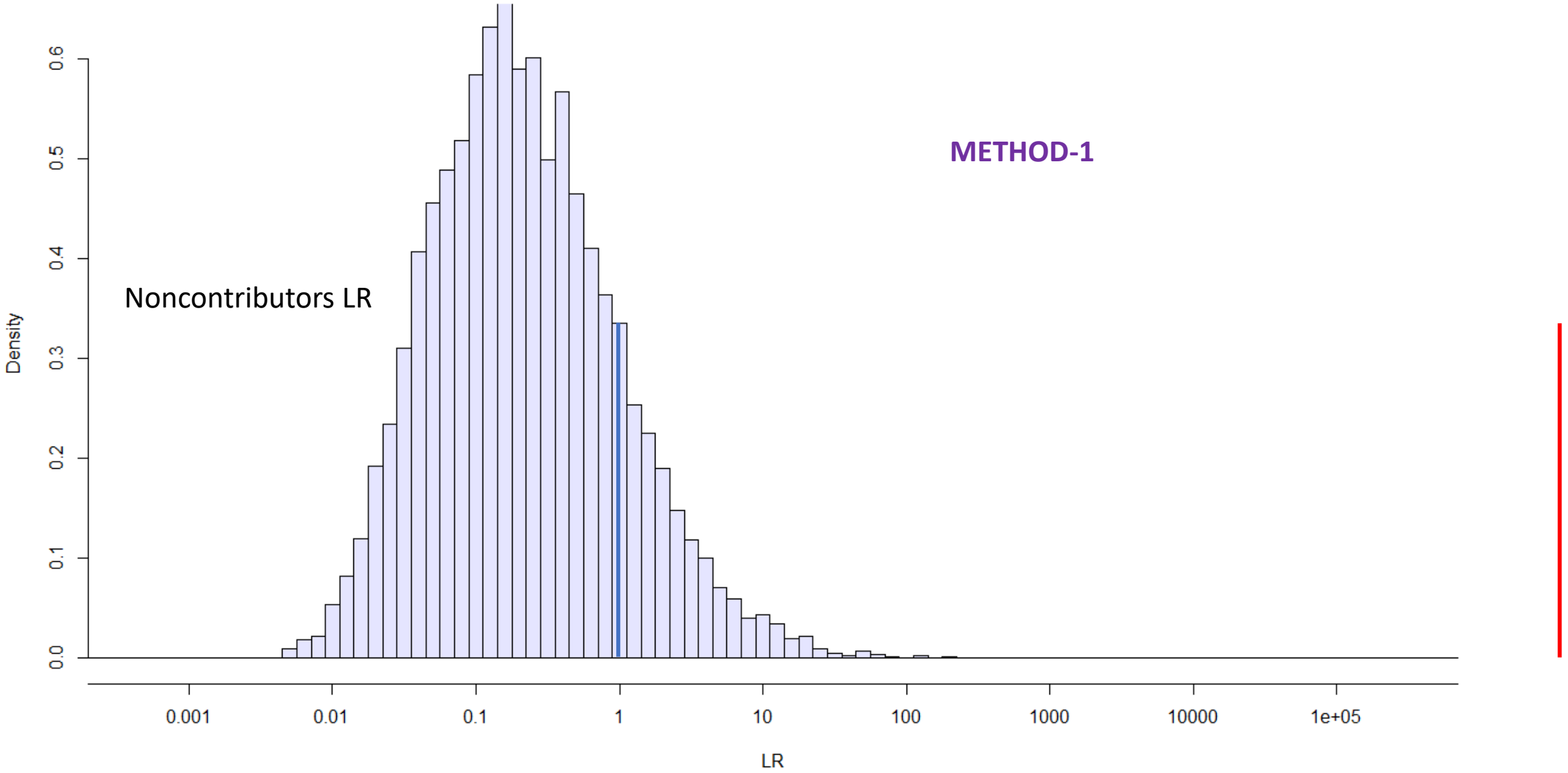
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



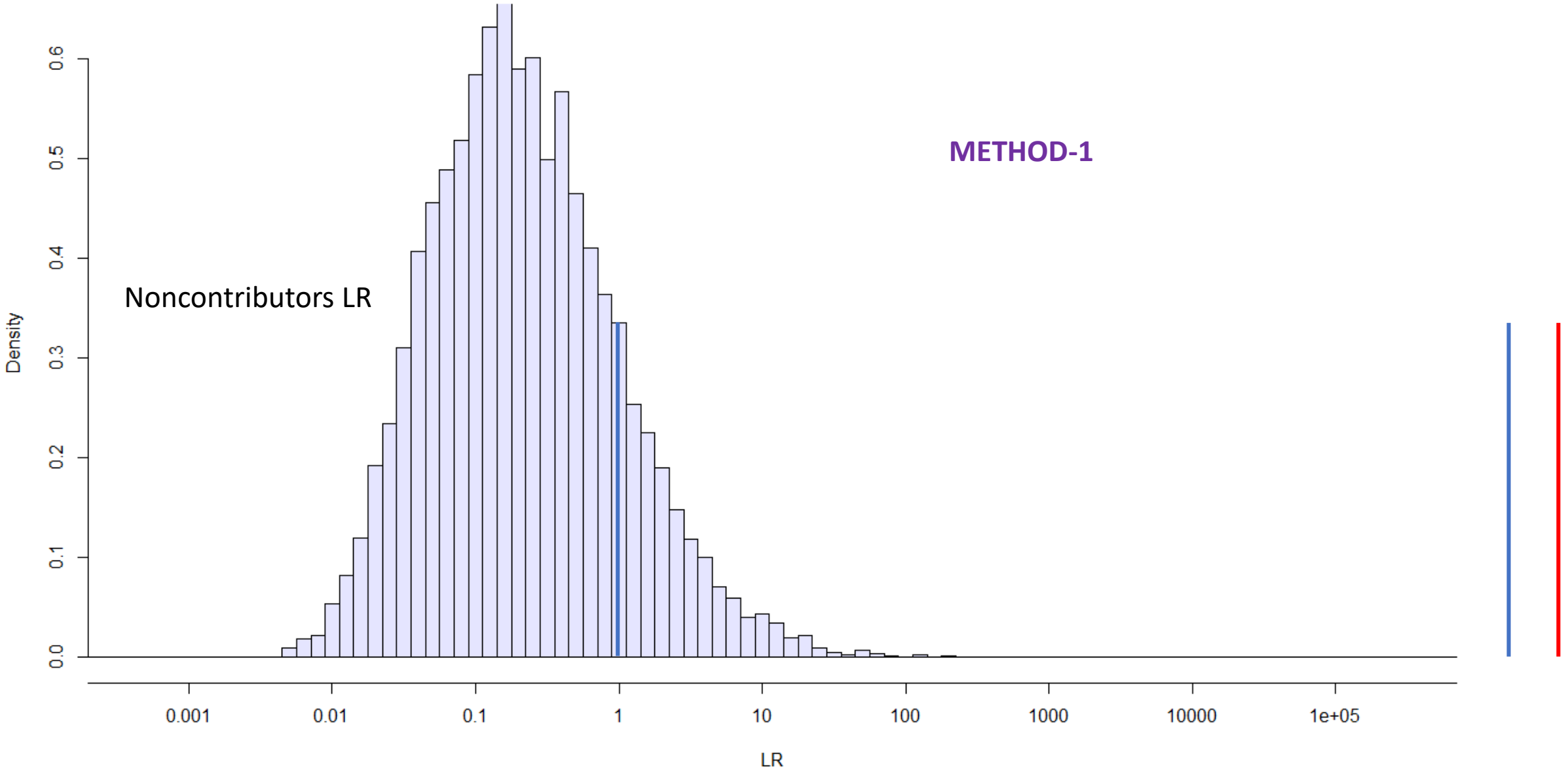
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



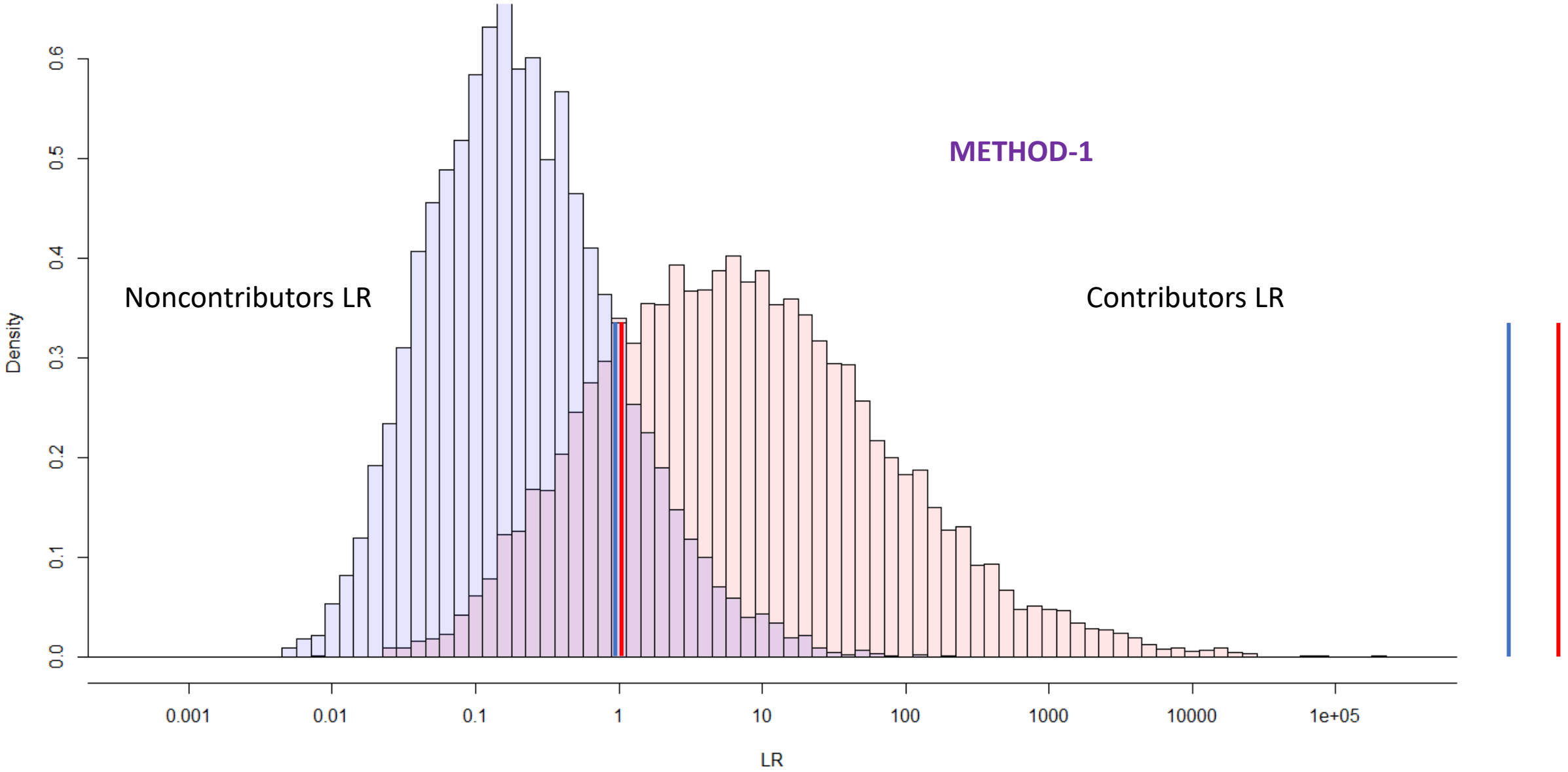
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



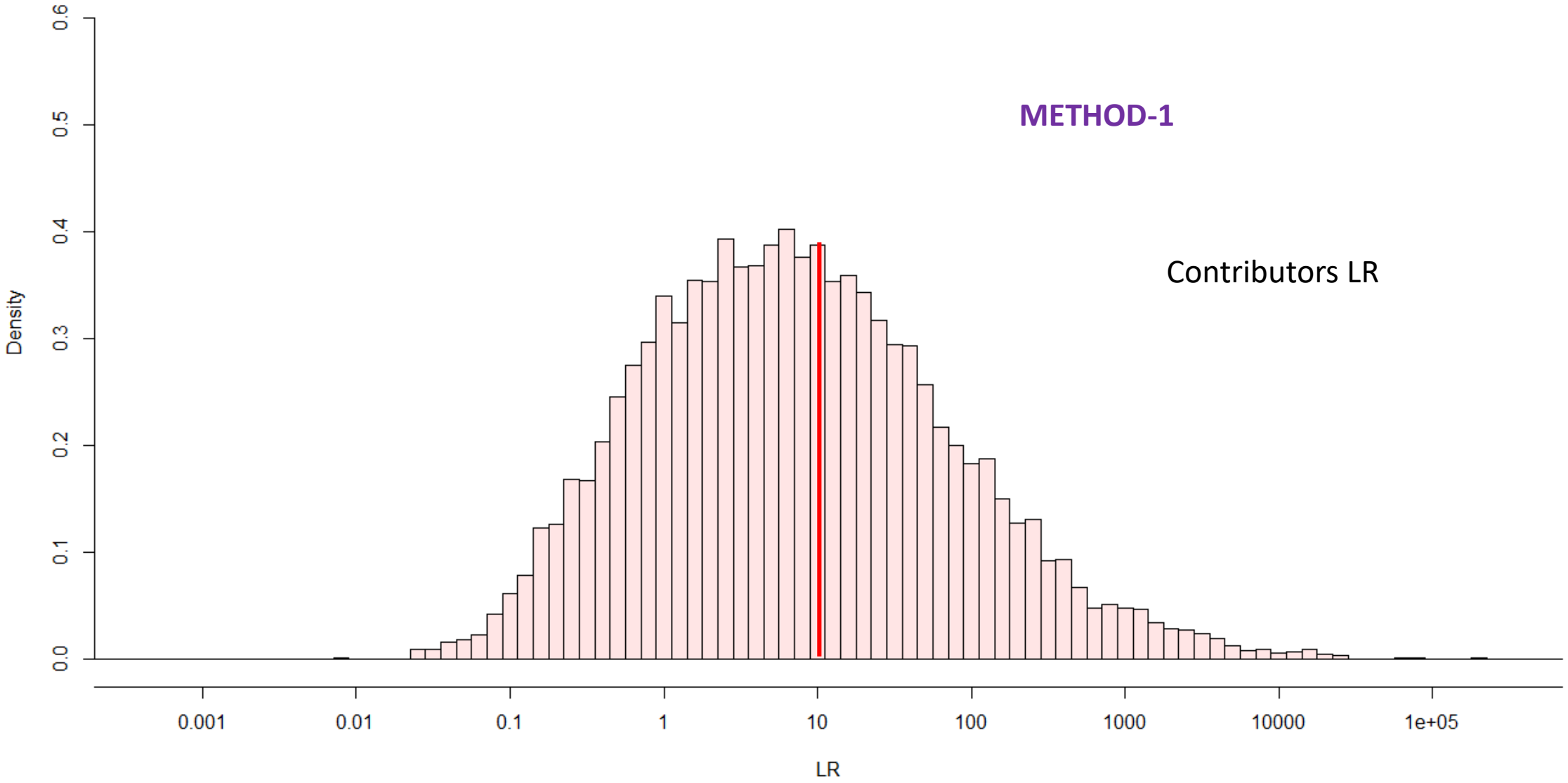
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



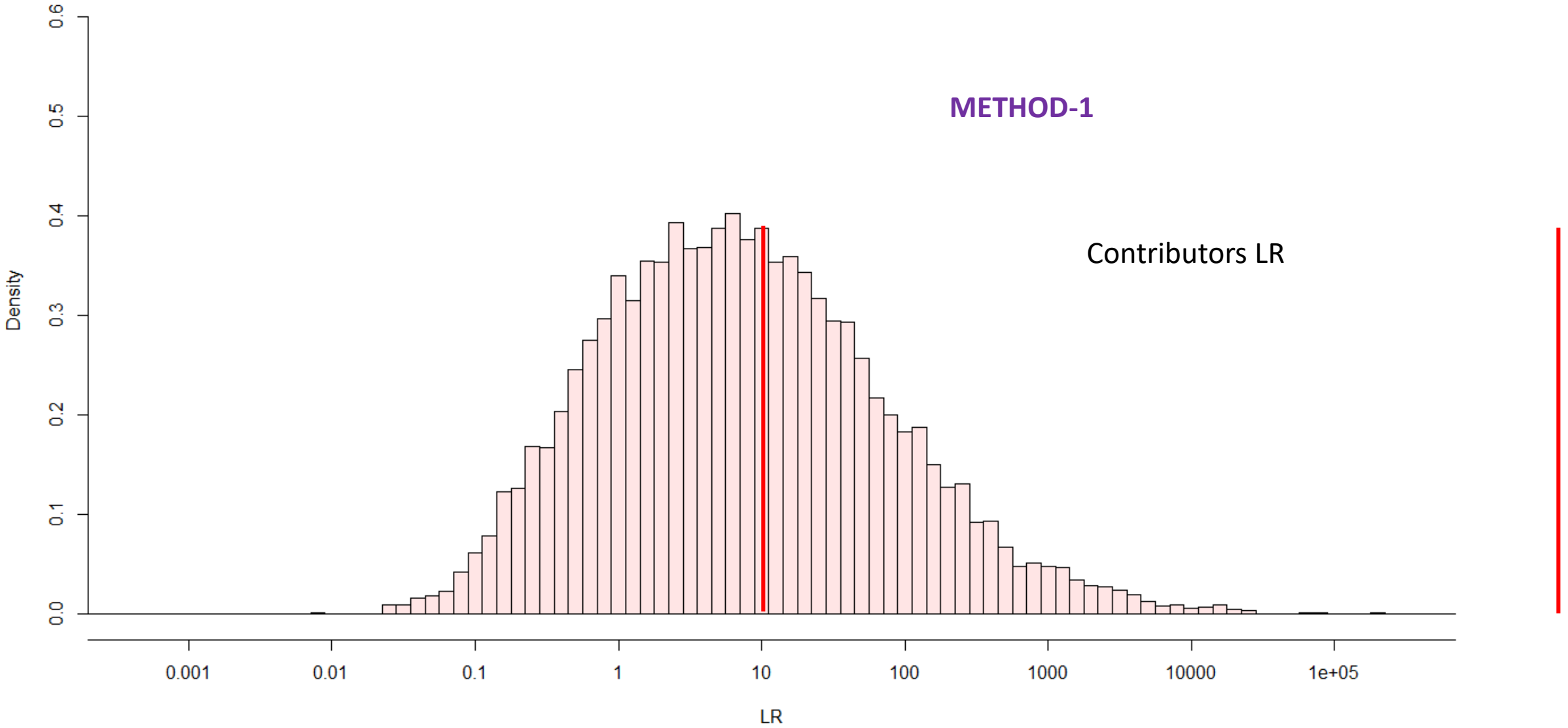
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



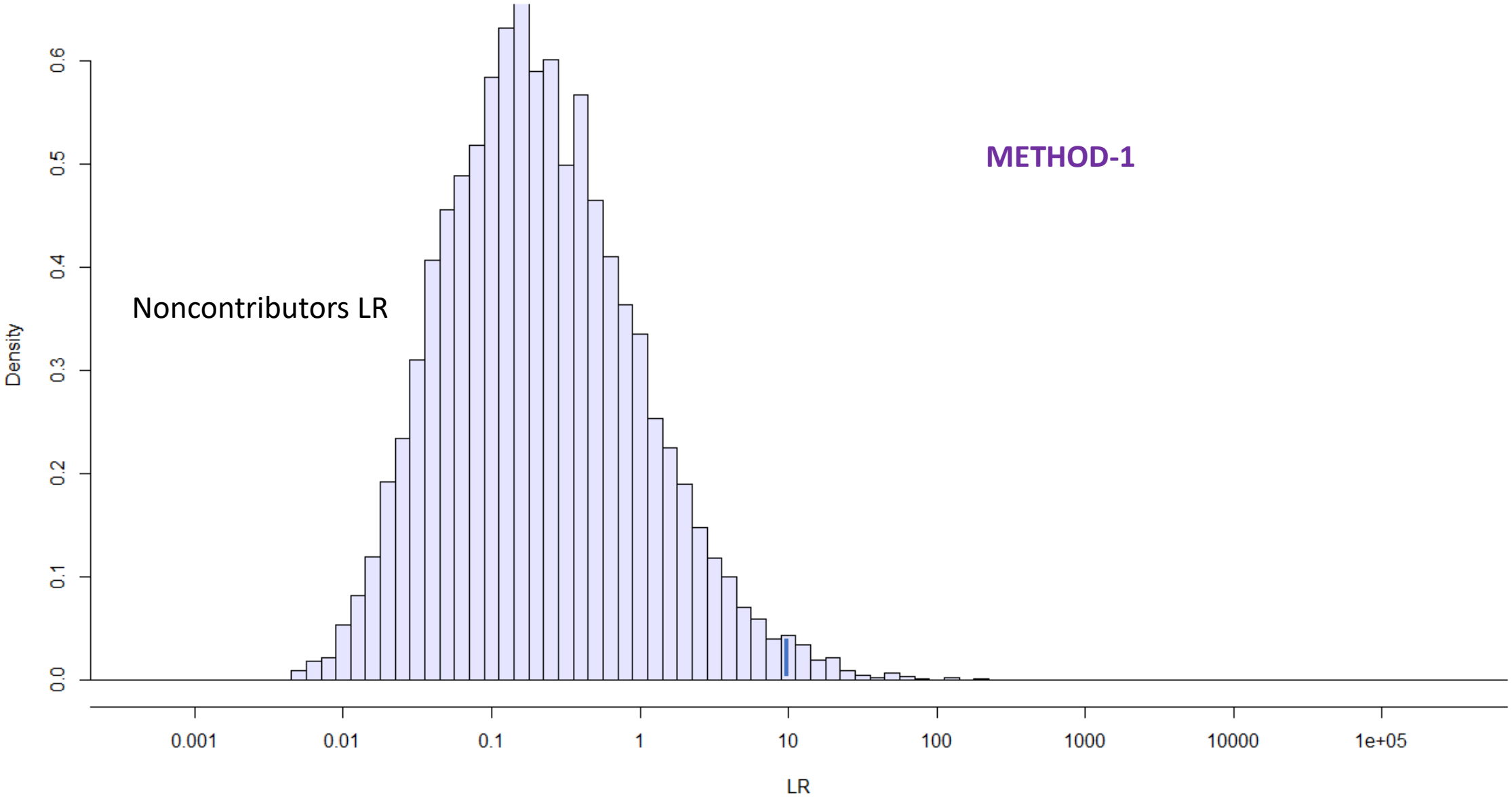
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



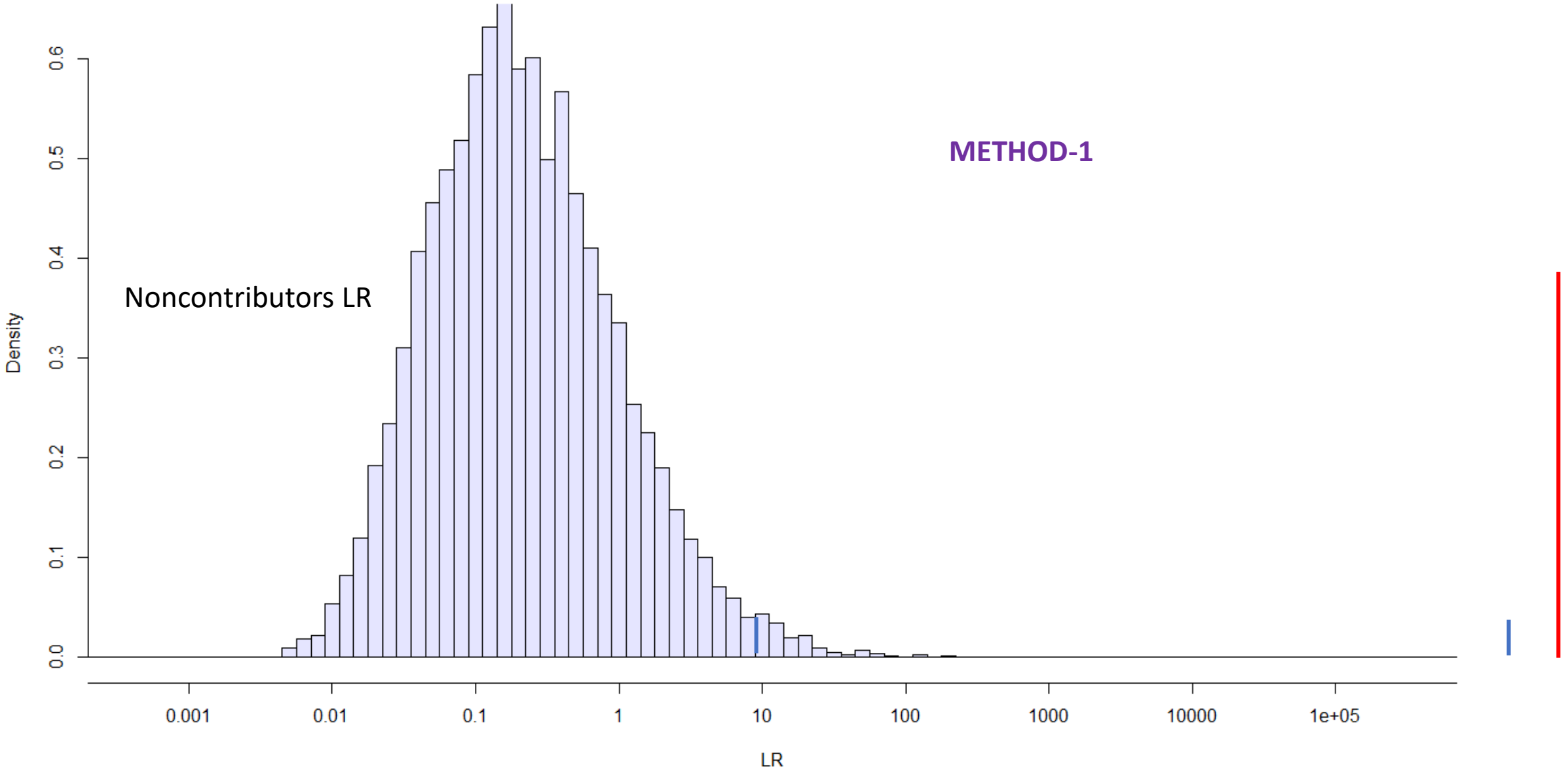
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



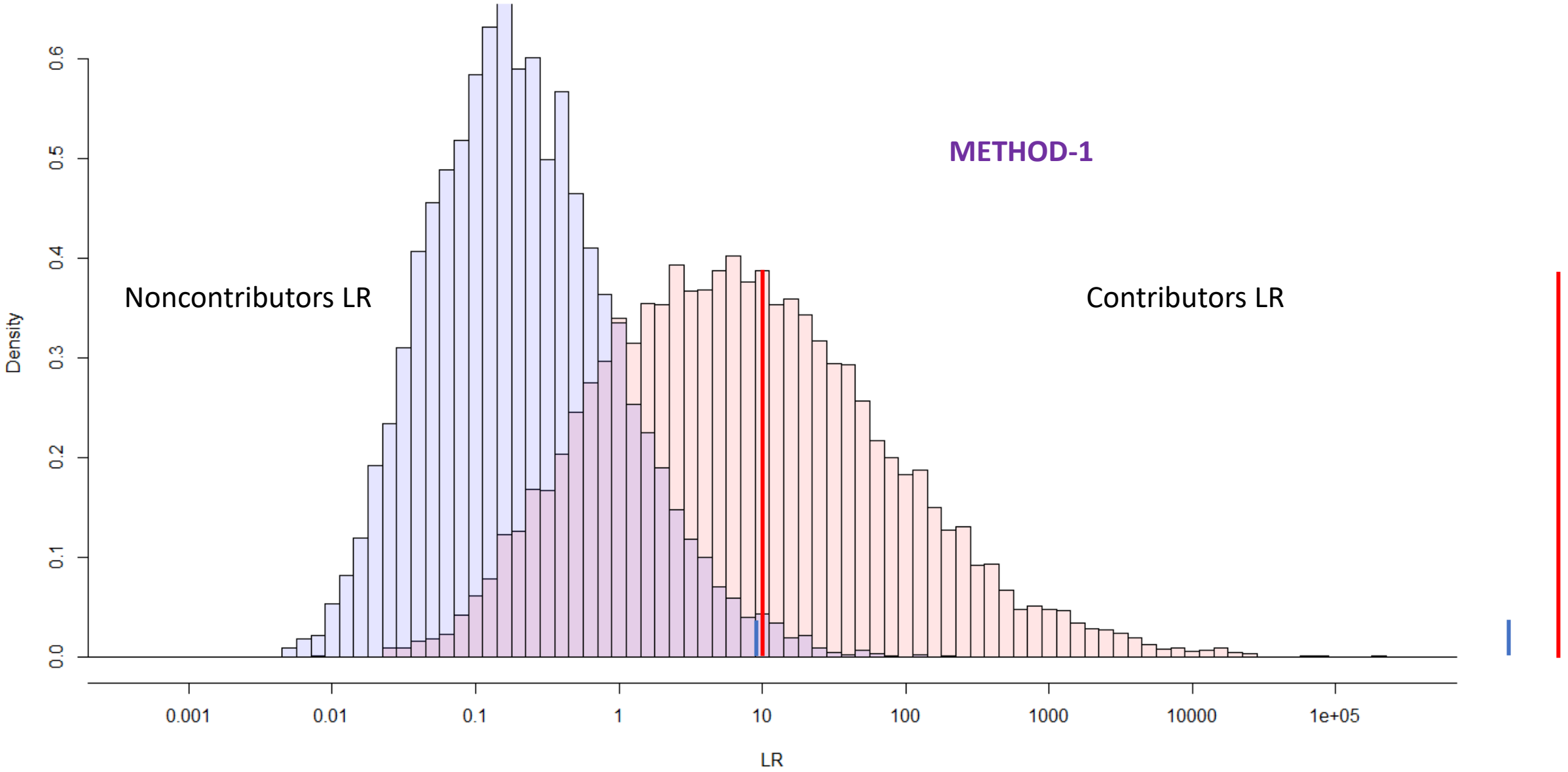
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



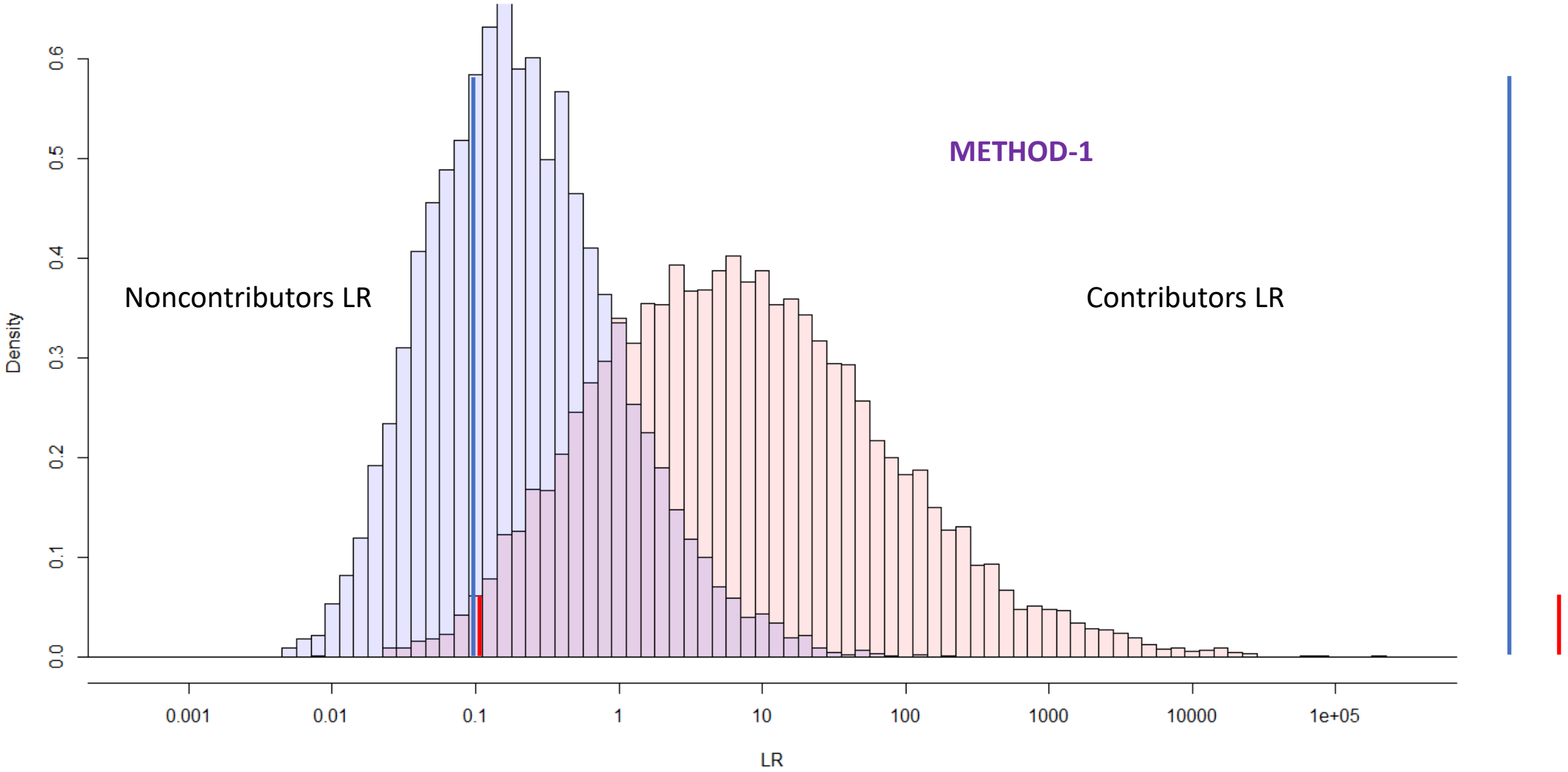
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



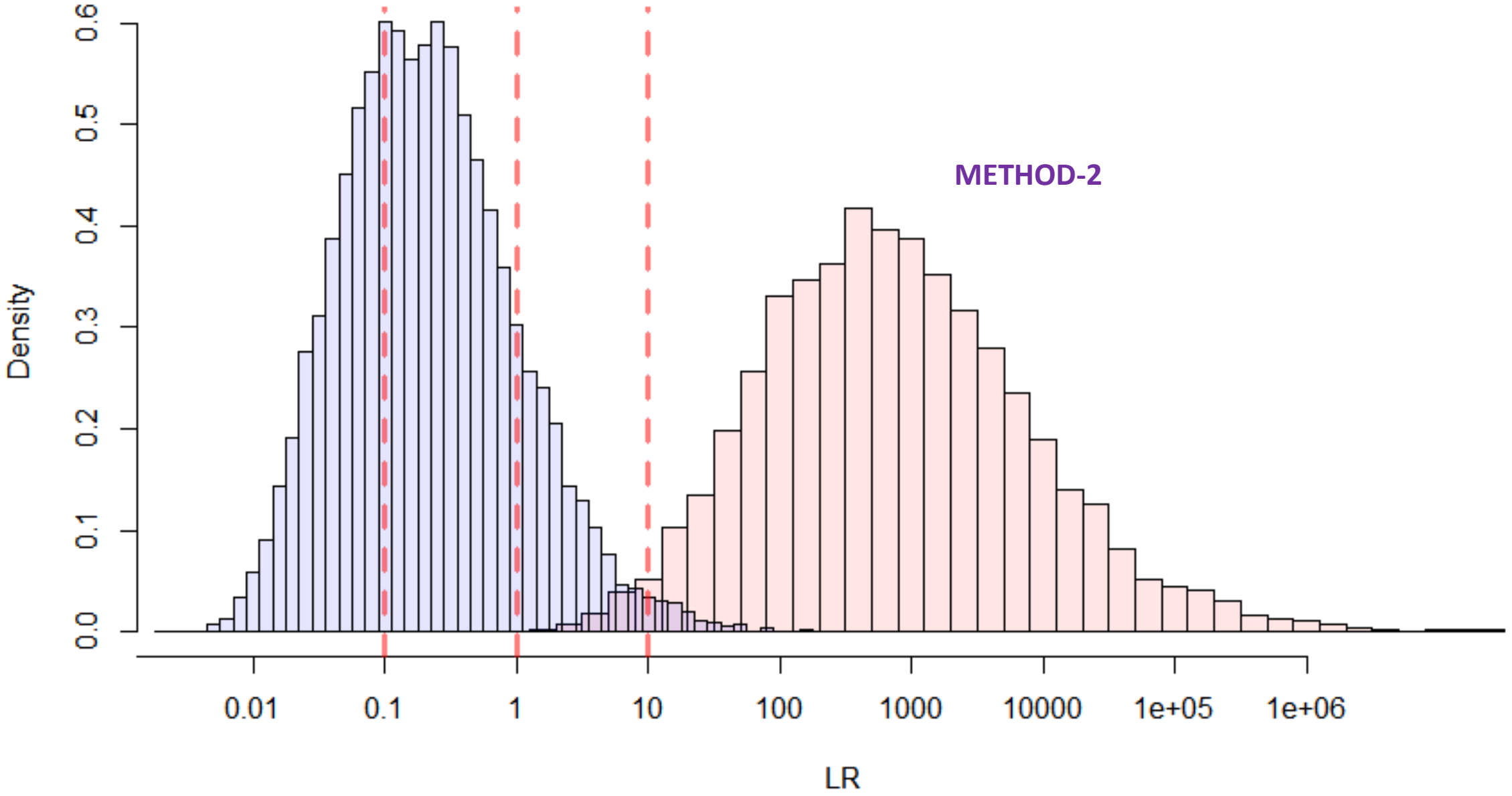
ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT



ACCURACY/CALIBRATION – EMPIRICAL ASSESSMENT

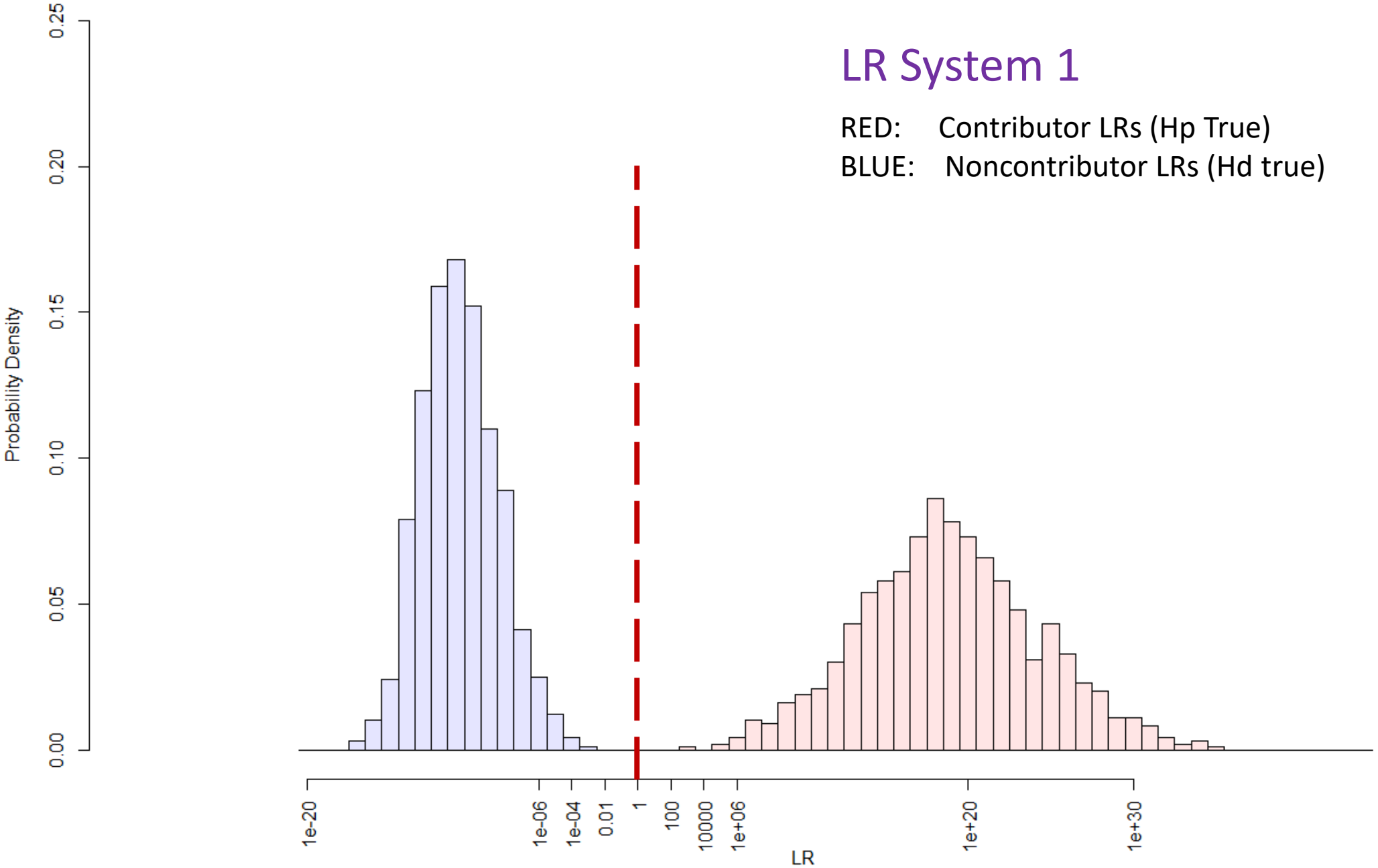


DISCRIMINATING POWER

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 and used in the interpretation?
2. Do the models used to incorporate such information represent reality adequately?

DISCRIMINATION POWER

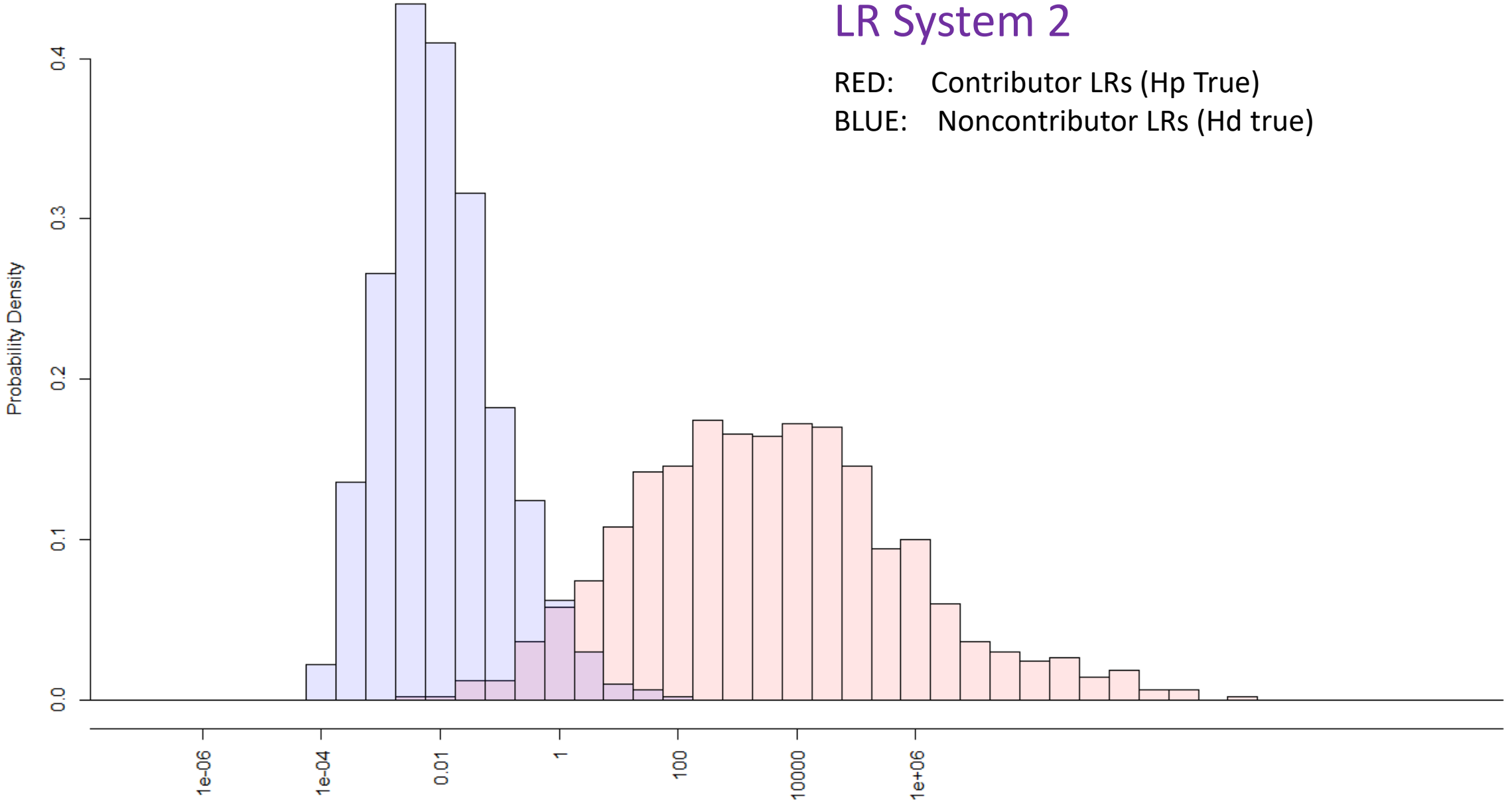


DISCRIMINATION POWER

LR System 2

RED: Contributor LRs (Hp True)

BLUE: Noncontributor LRs (Hd true)

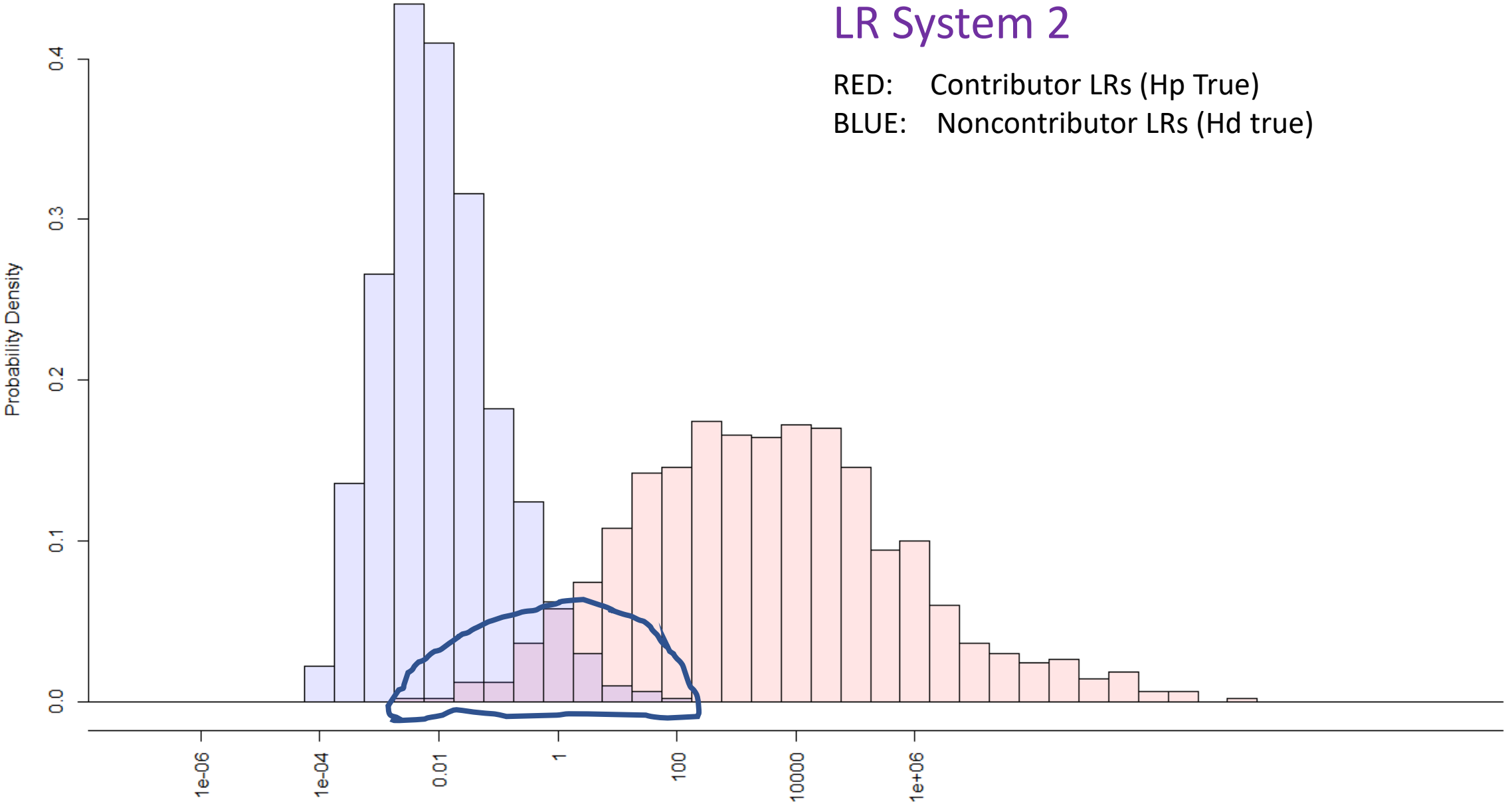


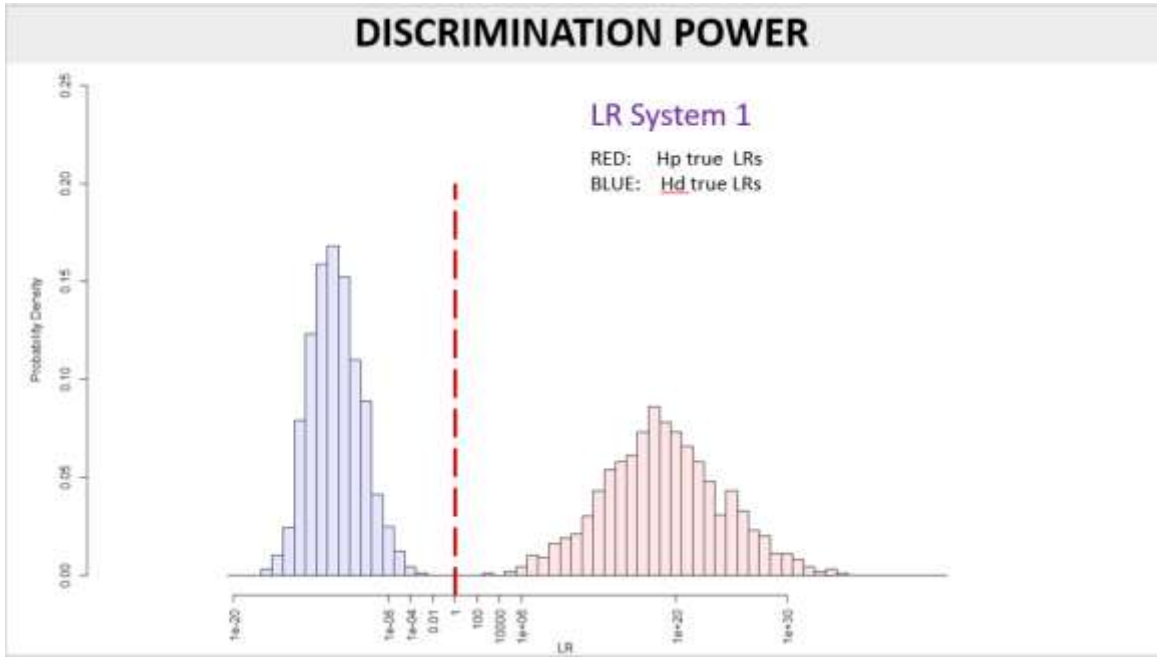
DISCRIMINATION POWER

LR System 2

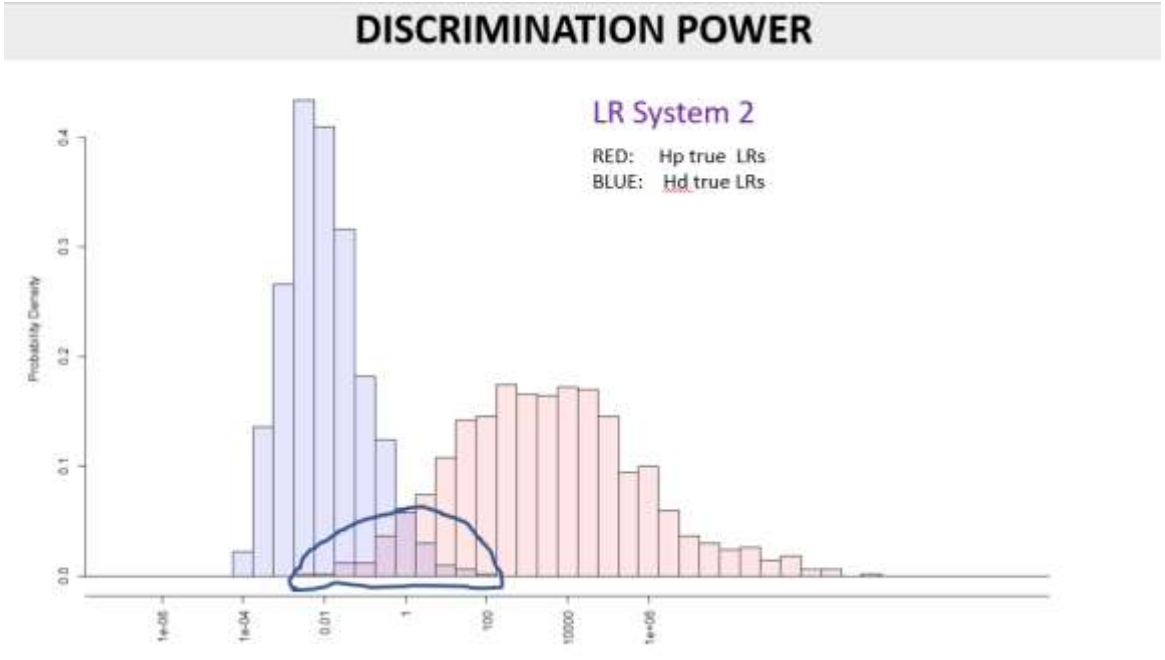
RED: Contributor LRs (Hp True)

BLUE: Noncontributor LRs (Hd true)





LR System 1



LR System 2

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

TIPPETT PLOTS

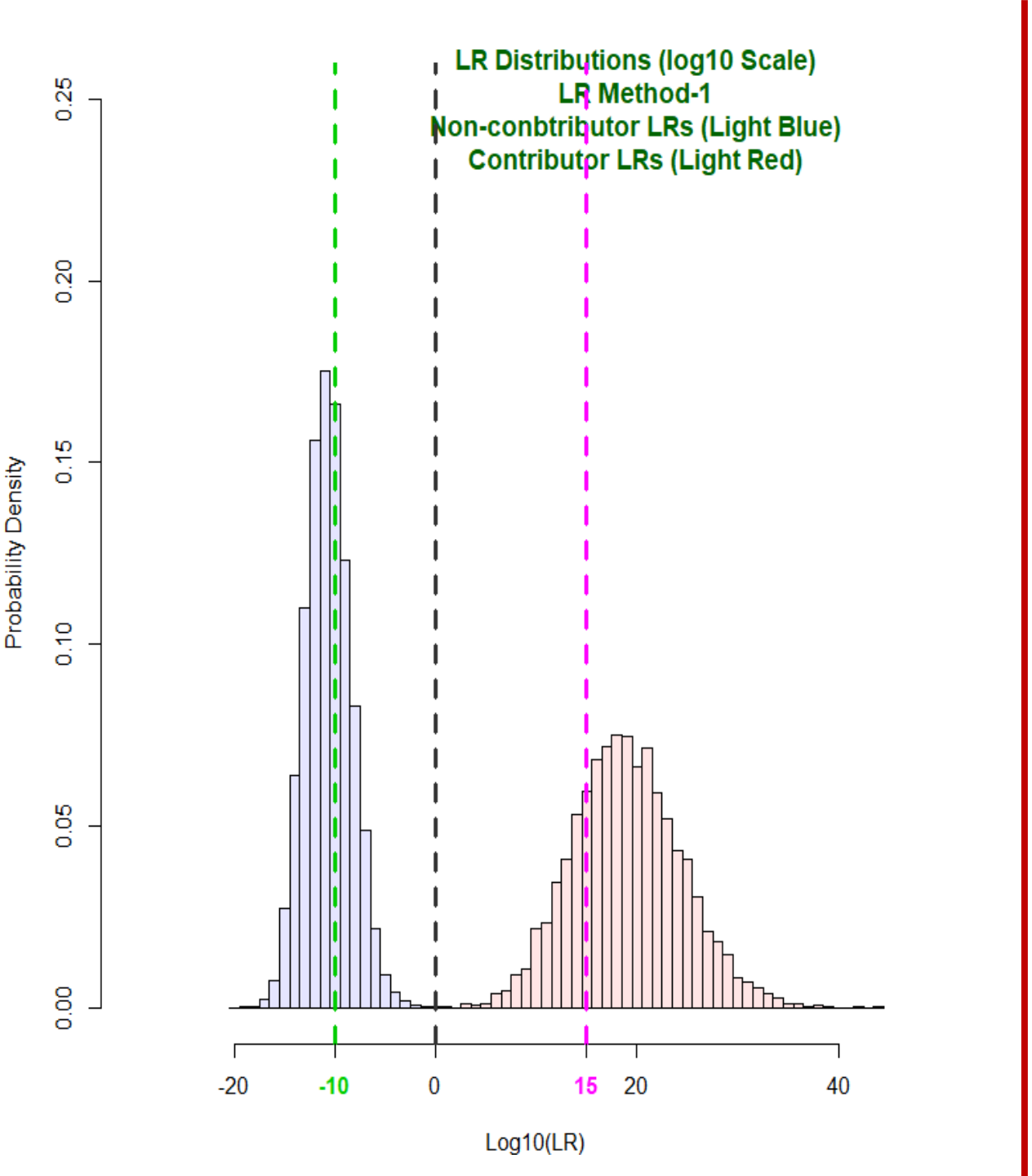
Statistical analysis of STR data

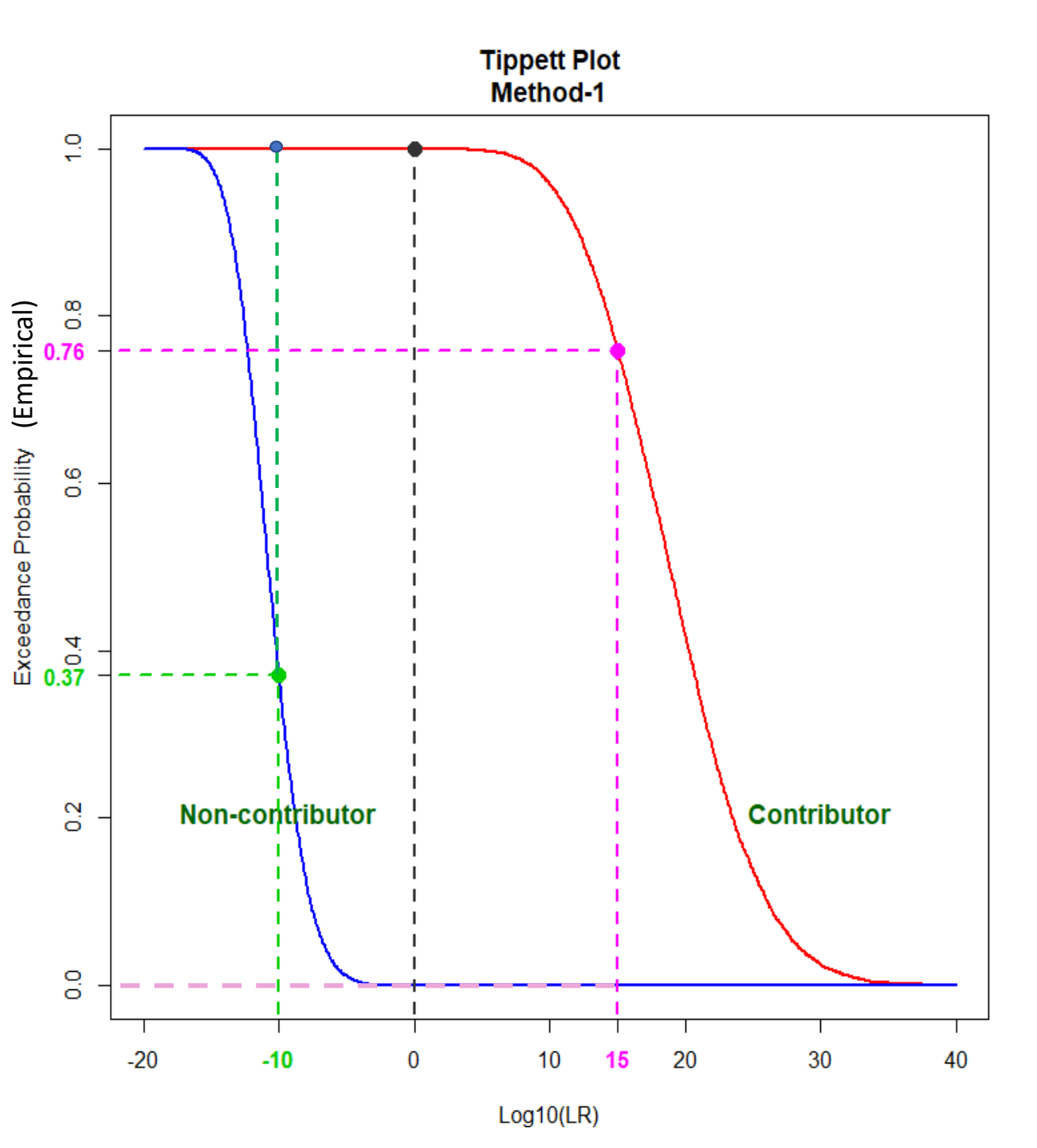
I.W.Evett[1] and J.S.Buckleton[2]

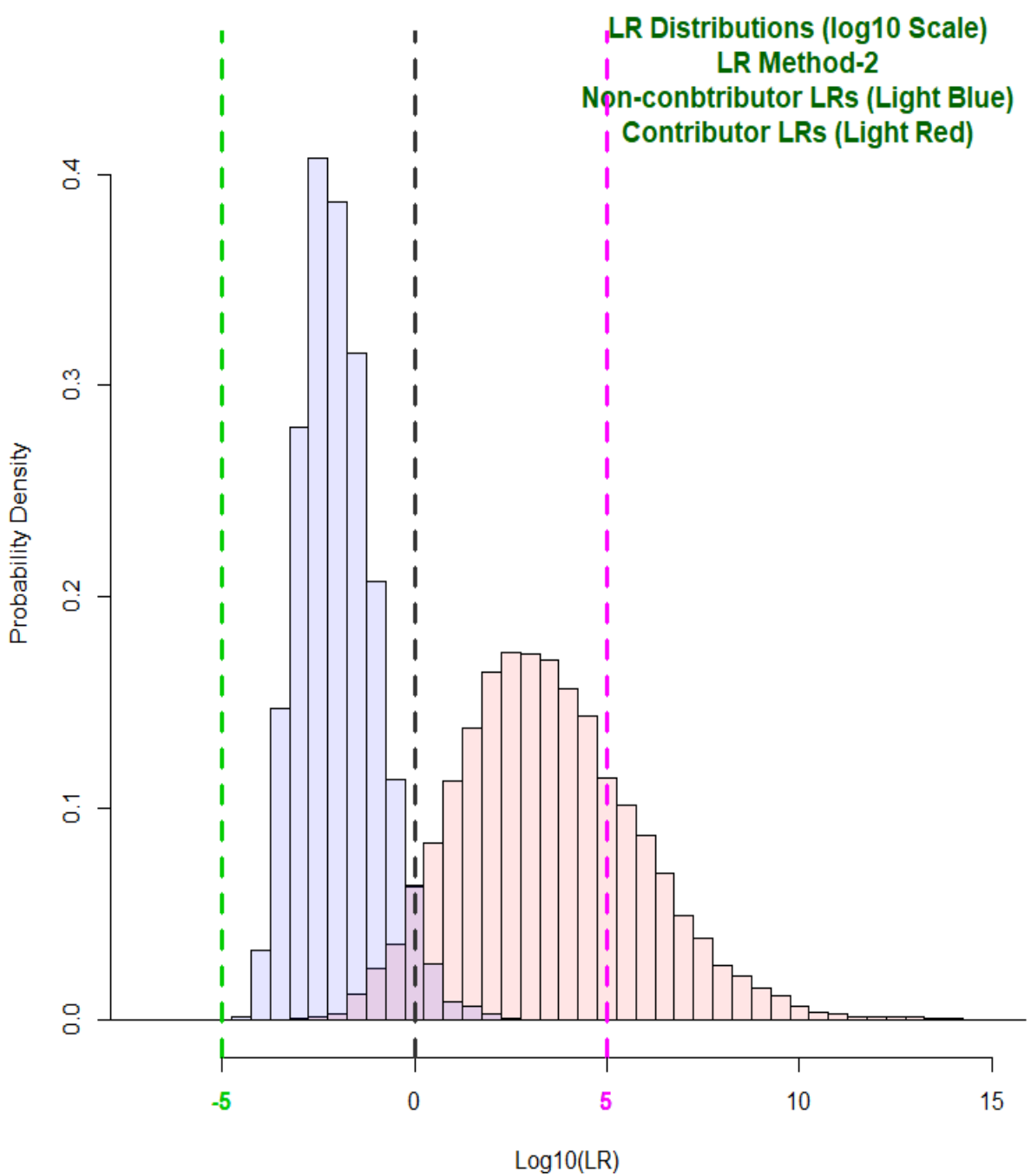
Advances in Forensic Haemogenetics 6

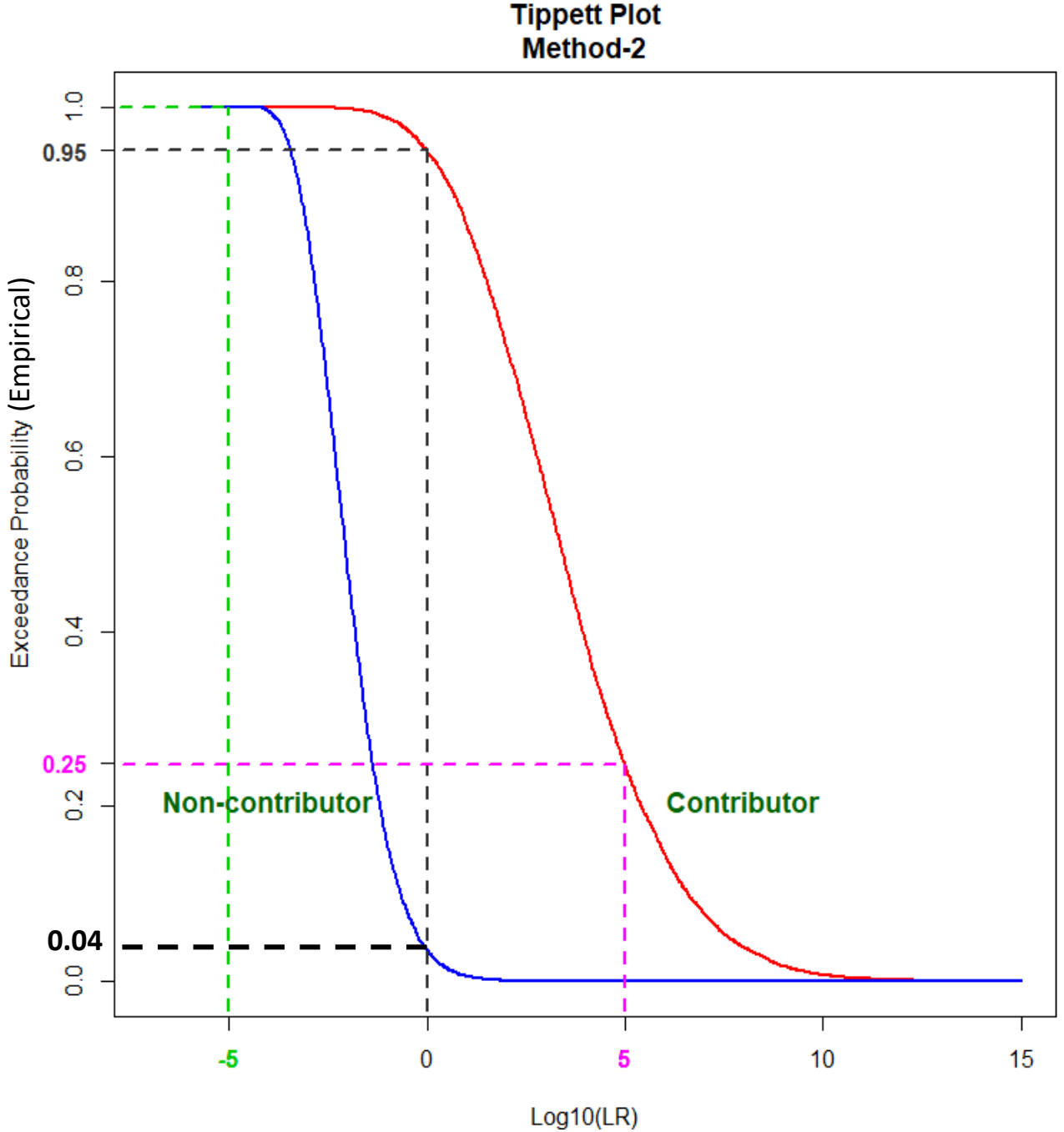
16th Congress of the
International Society for Forensic Haemogenetics
(Internationale Gesellschaft für forensische Hämogenetik e.V.)
Santiago de Compostela, 12–16 September 1995

Tippett CF, Emerson VJ, Fereday MJ, Lawton F and Lampert SM (1968) The evidential value of the comparison of paint flakes from sources other than vehicles. J For Sci Soc 8: 61-65.

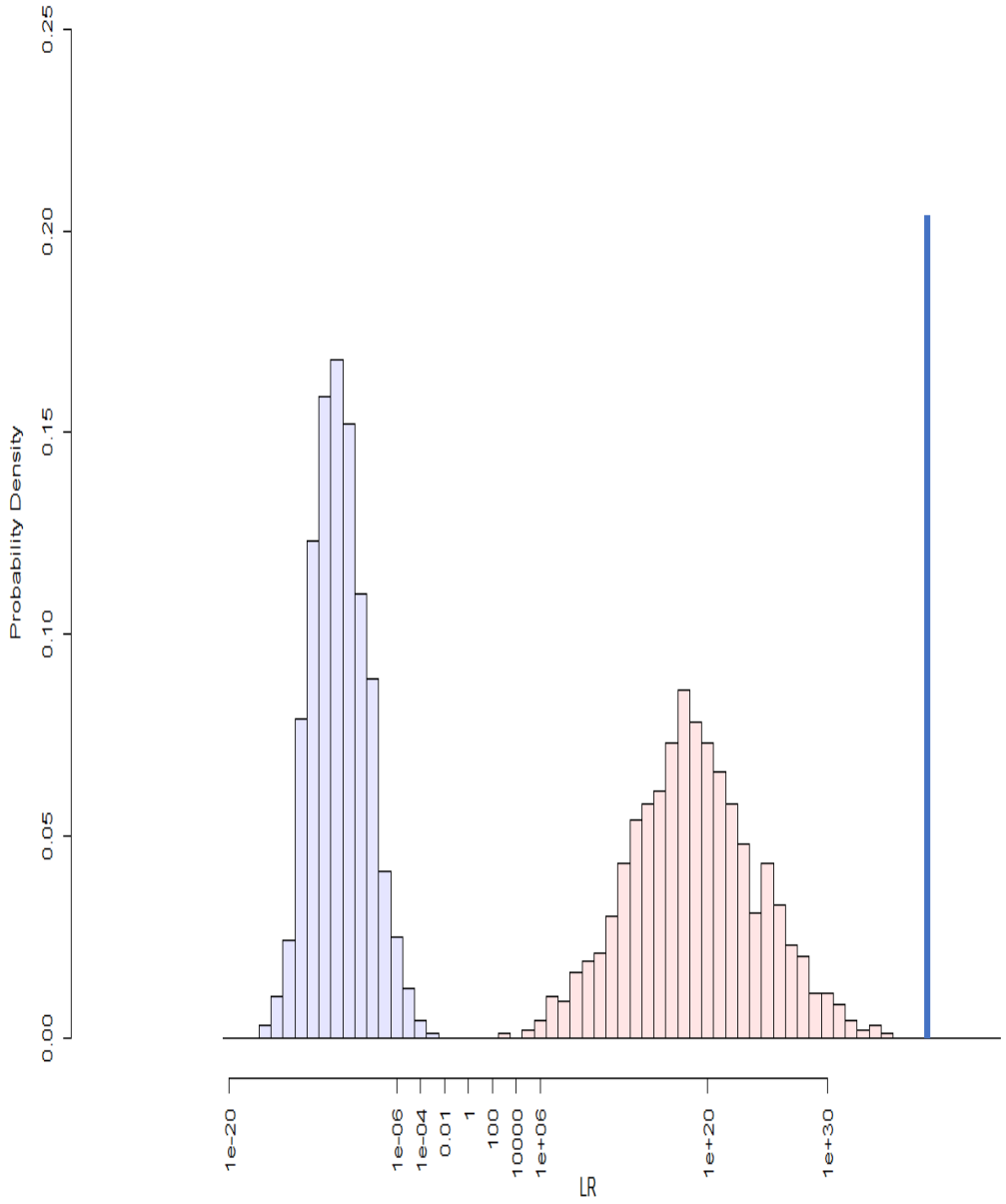

LR Distributions (log10 Scale)
LR Method-1
Non-conbtributor LRs (Light Blue)
Contributor LRs (Light Red)
Probability Density
0.00
0.05
0.10
0.15
0.20
0.25
-20
-10
0
15
20
40
Log10(LR)

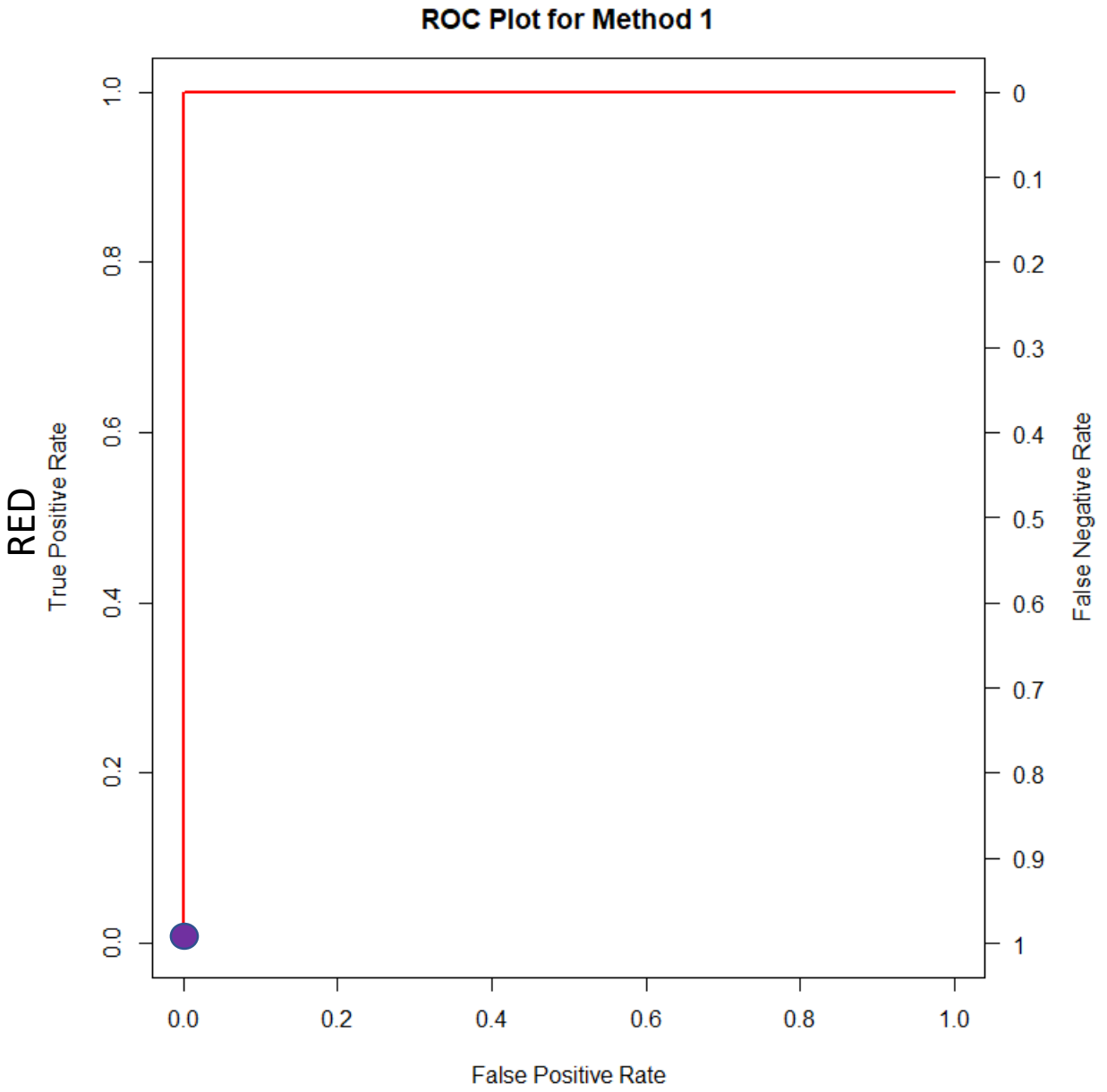

Tippett Plot
Method-1
Exceedance Probability (Empirical)
0.0
0.2
0.4
0.6
0.8
1.0
0.76
0.37
Non-contributor
Contributor
-20
-10
0
10
15
20
30
40
Log10(LR)


LR Distributions (log10 Scale)
LR Method-2
Non-conbtributor LRs (Light Blue)
Contributor LRs (Light Red)
Probability Density
0.0
0.1
0.2
0.3
0.4
-5
0
5
10
15
Log10(LR)


Tippett Plot
Method-2
Exceedance Probability (Empirical)
0.0
0.2
0.4
0.6
0.8
1.0
0.95
0.25
0.04
Non-contributor
Contributor
-5
0
5
10
15
Log10(LR)

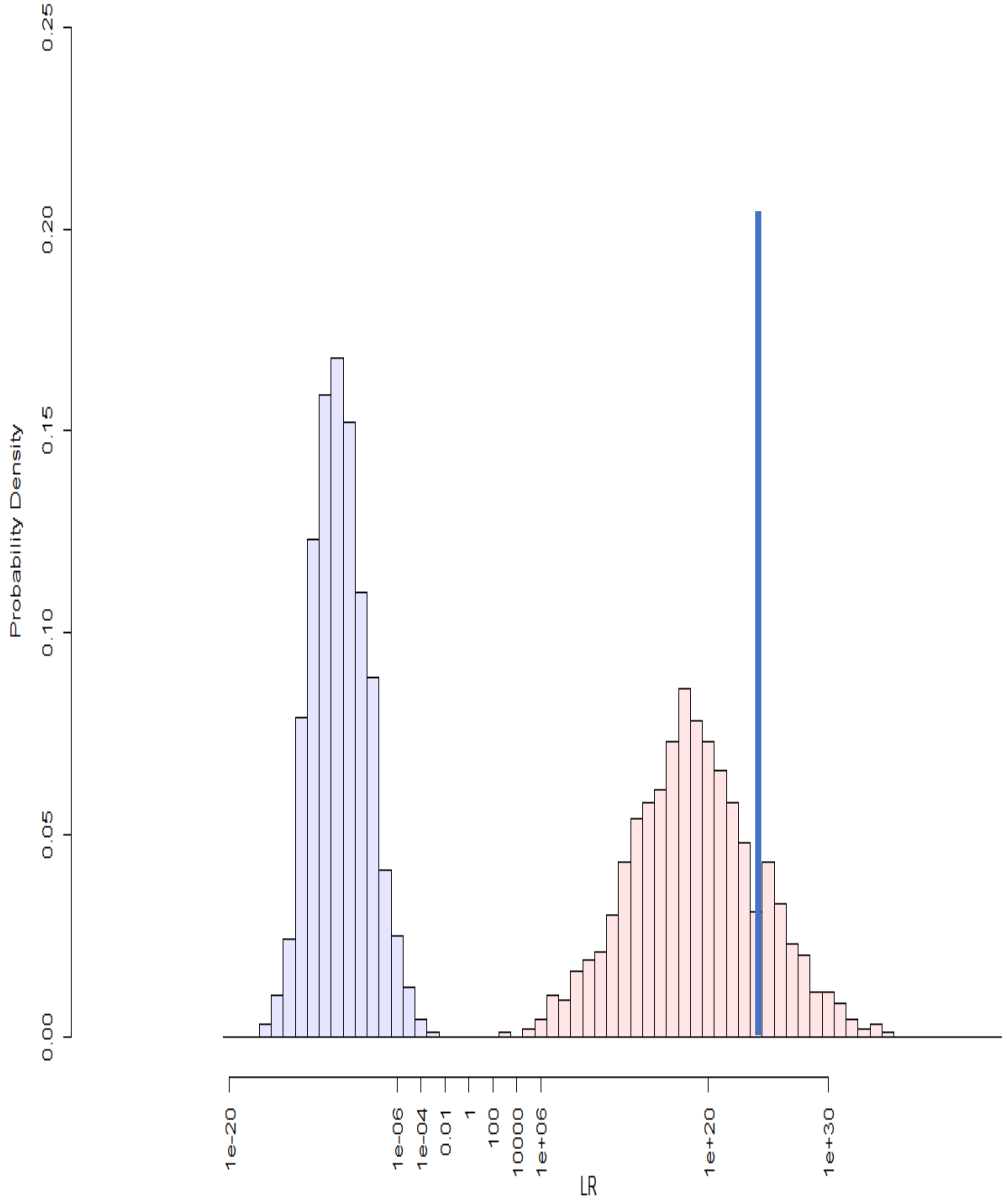
ROC = Receiver Operating Characteristic

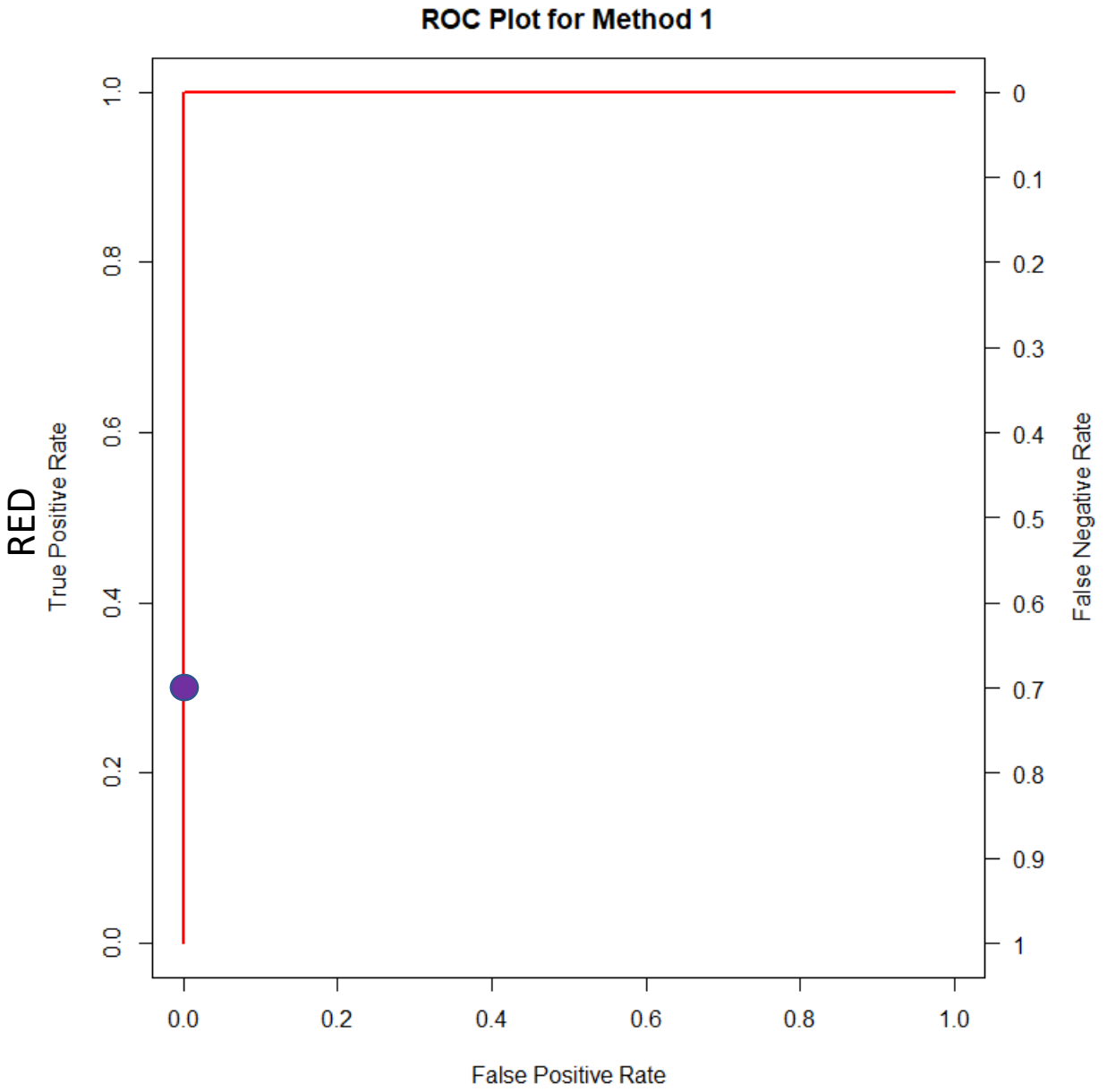




RED

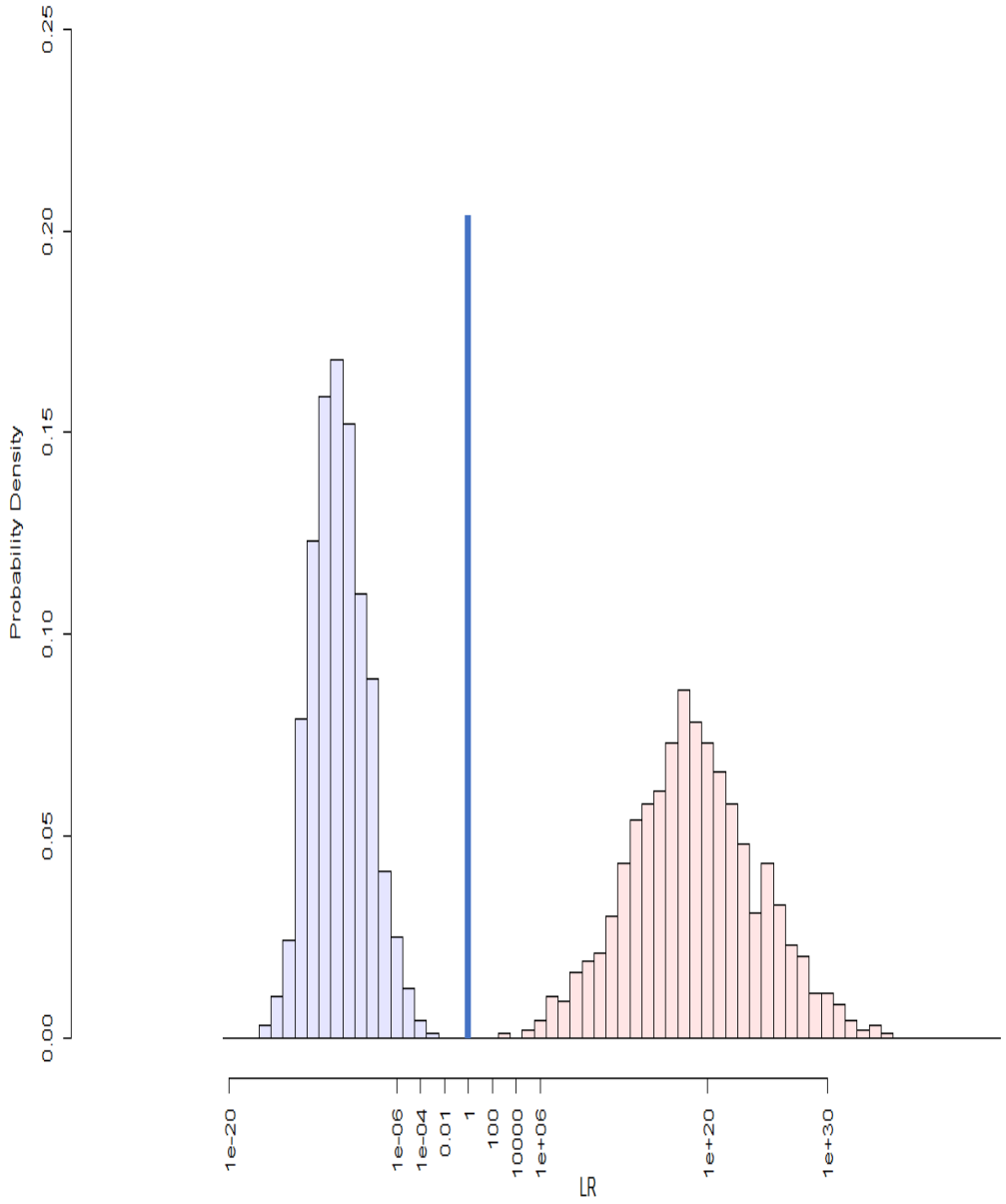
BLUE

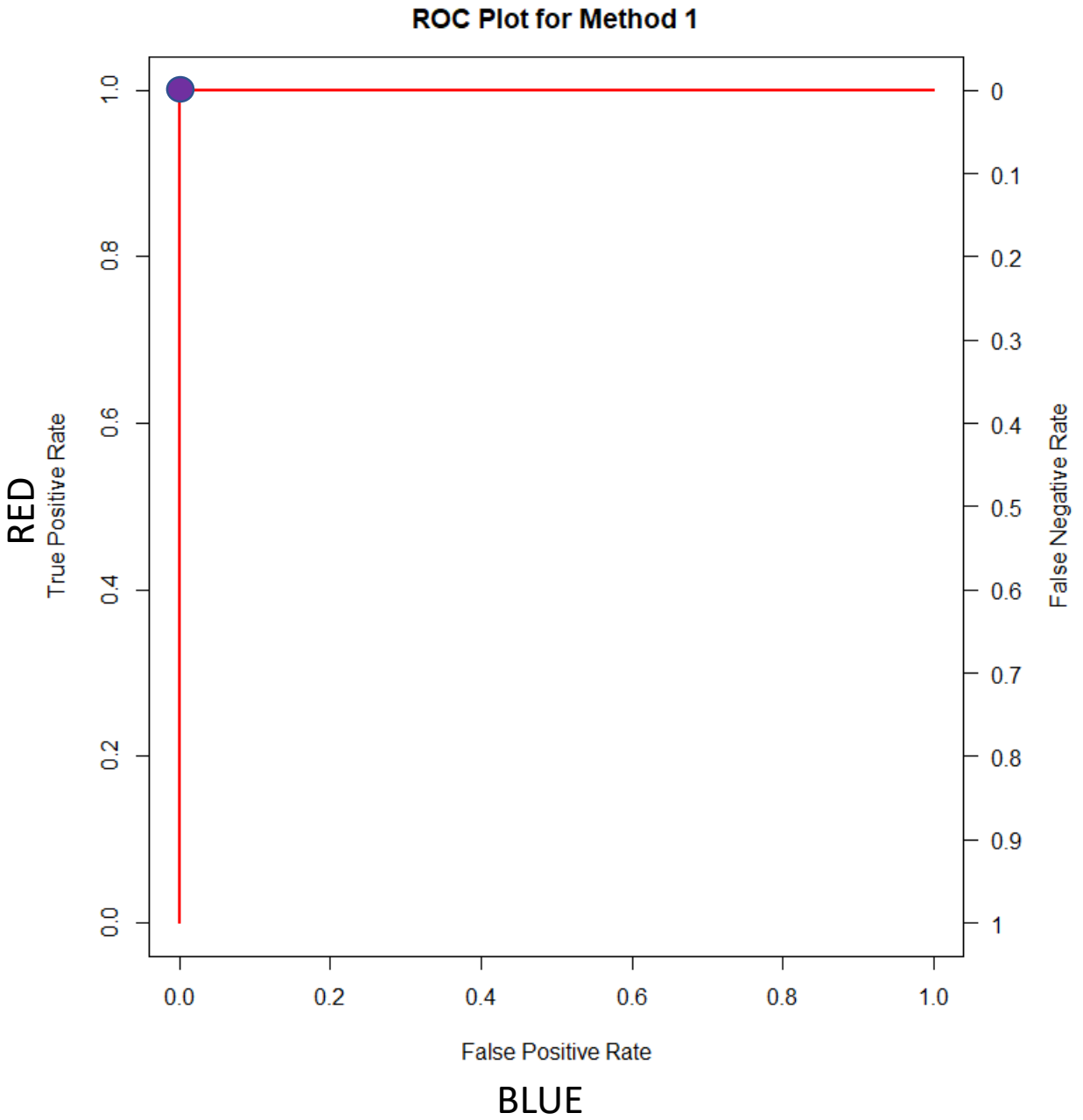

Probability Density
0.00
0.05
0.10
0.15
0.20
0.25
1e-20
1e-06
1e-04
0.01
1
100
10000
1e+06
1e+20
1e+30
LR

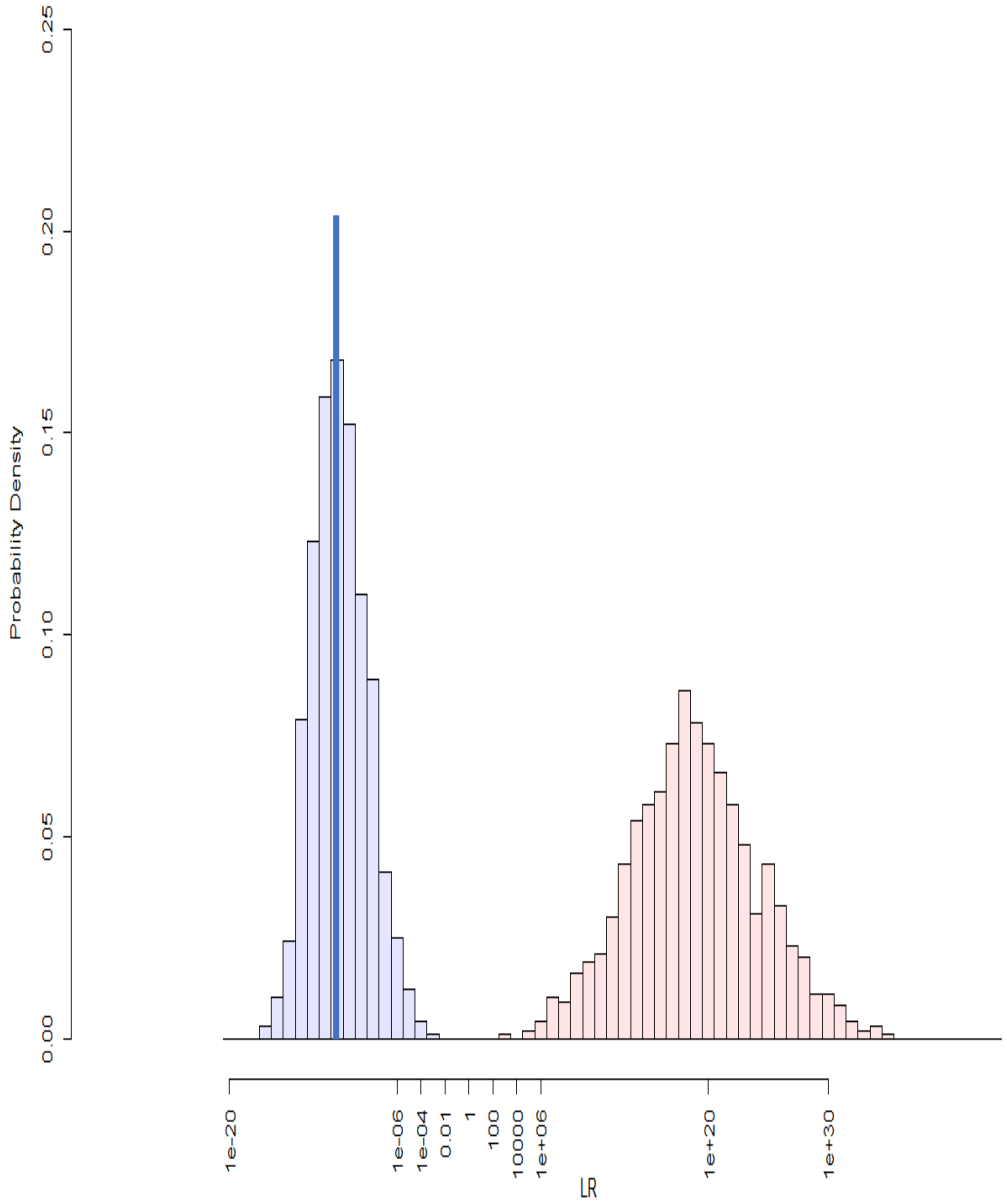

ROC Plot for Method 1
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

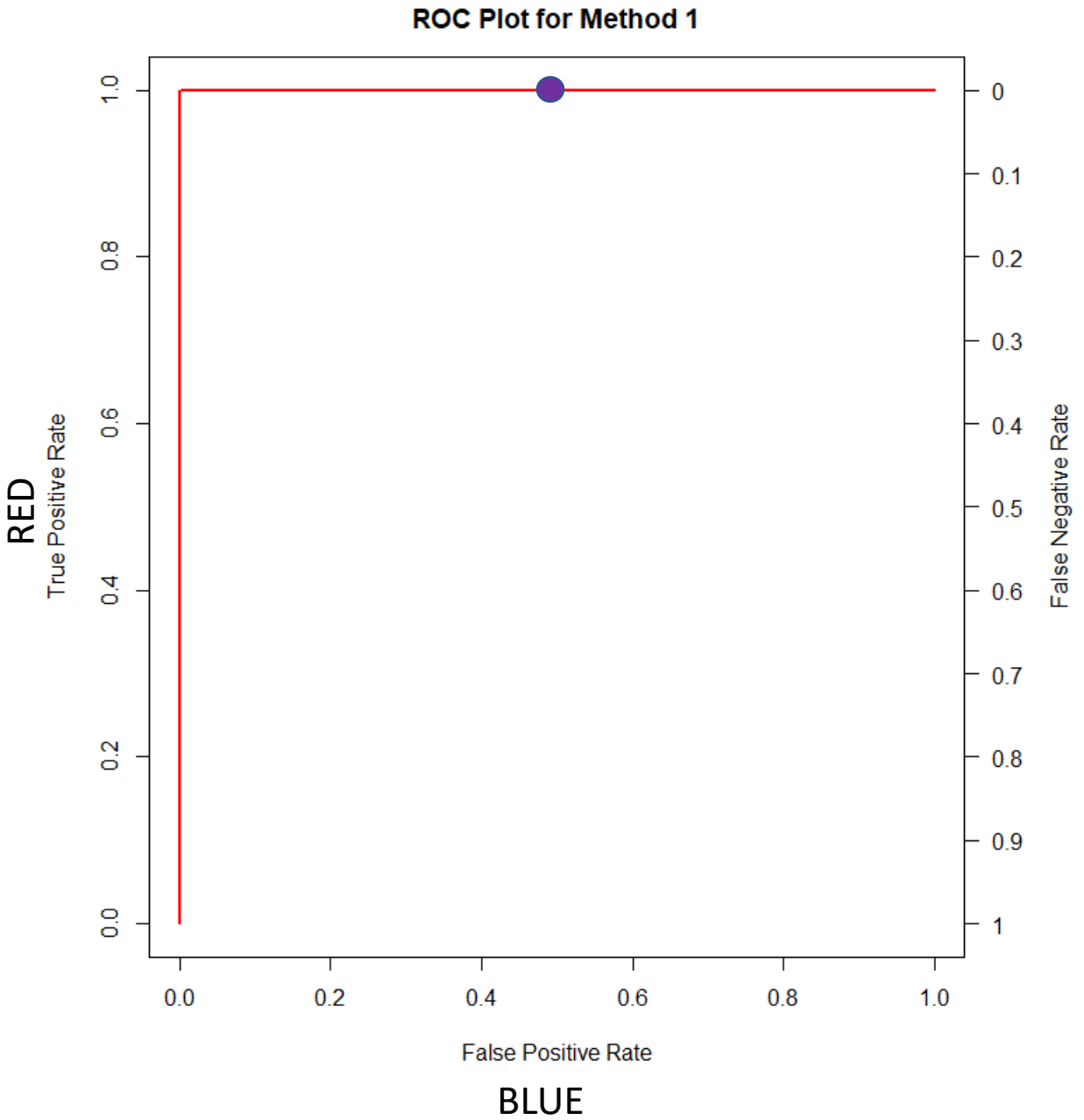
RED

BLUE


Probability Density
0.00
0.05
0.10
0.15
0.20
0.25
1e-20
1e-06
1e-04
0.01
1
100
10000
1e+06
1e+20
1e+30
LR

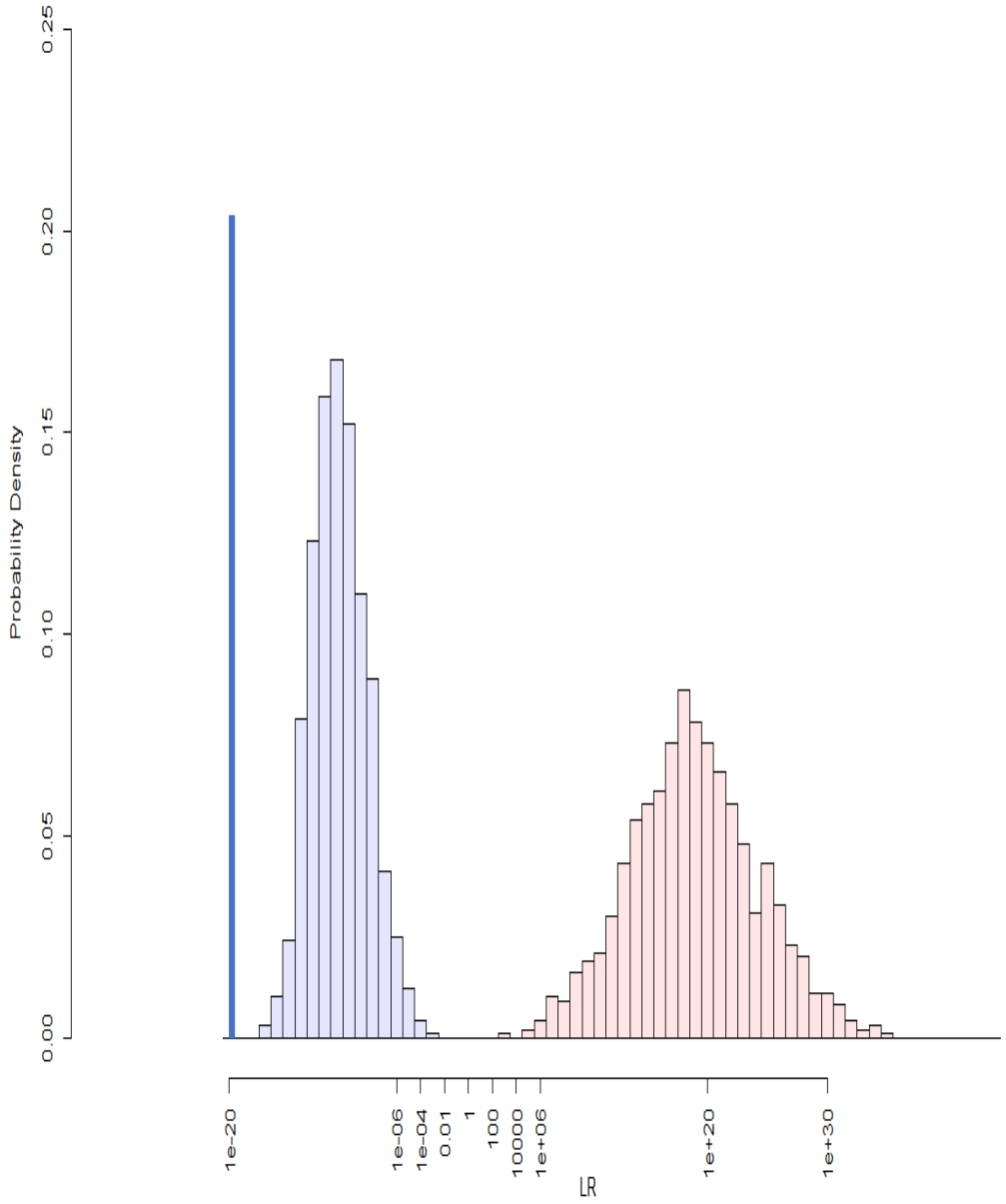

ROC Plot for Method 1
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
RED
BLUE

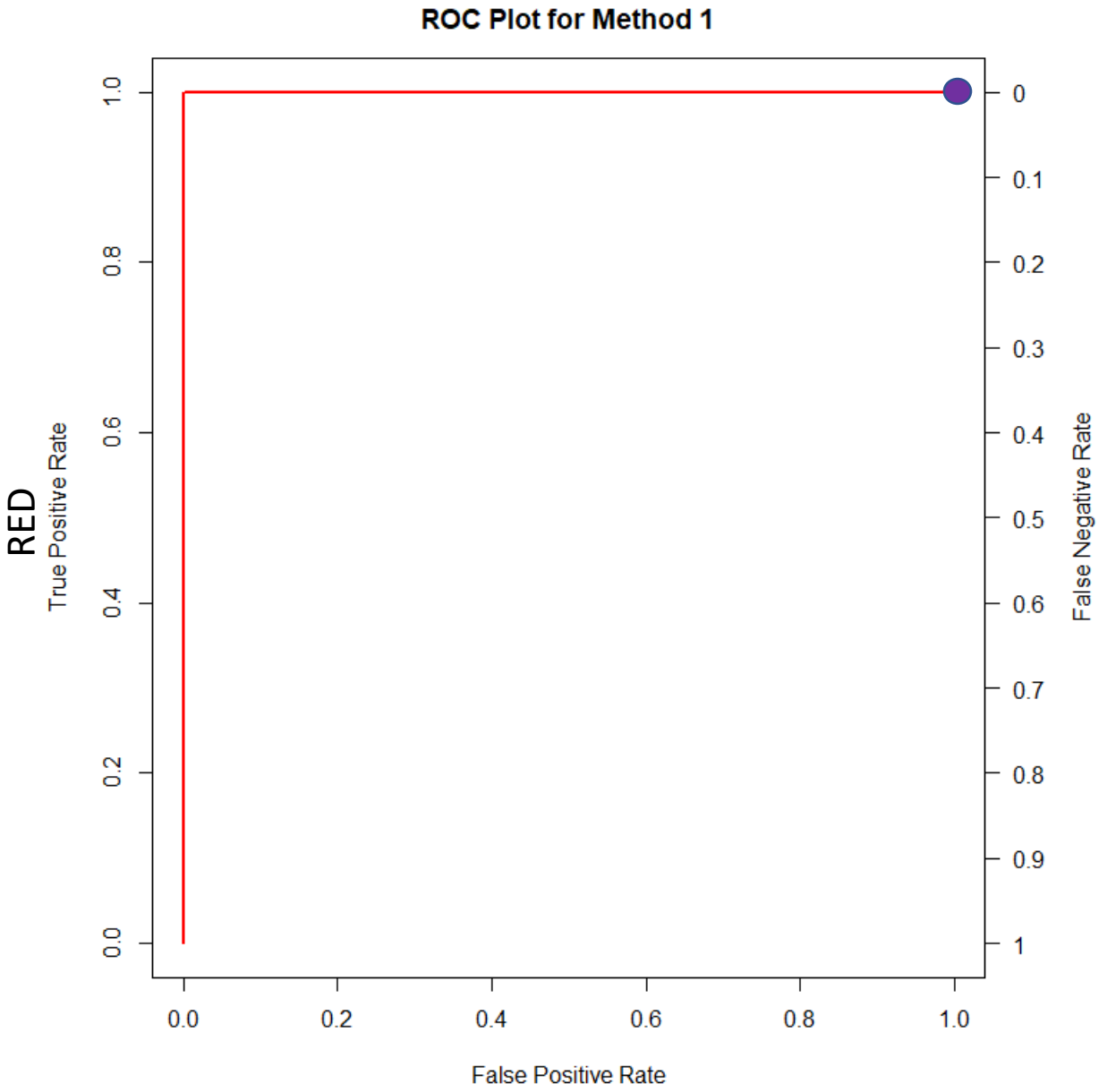

Probability Density
0.00
0.05
0.10
0.15
0.20
0.25
1e-20
1e-06
1e-04
0.01
1
100
10000
1e+06
1e+20
1e+30
LR


ROC Plot for Method 1
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

RED

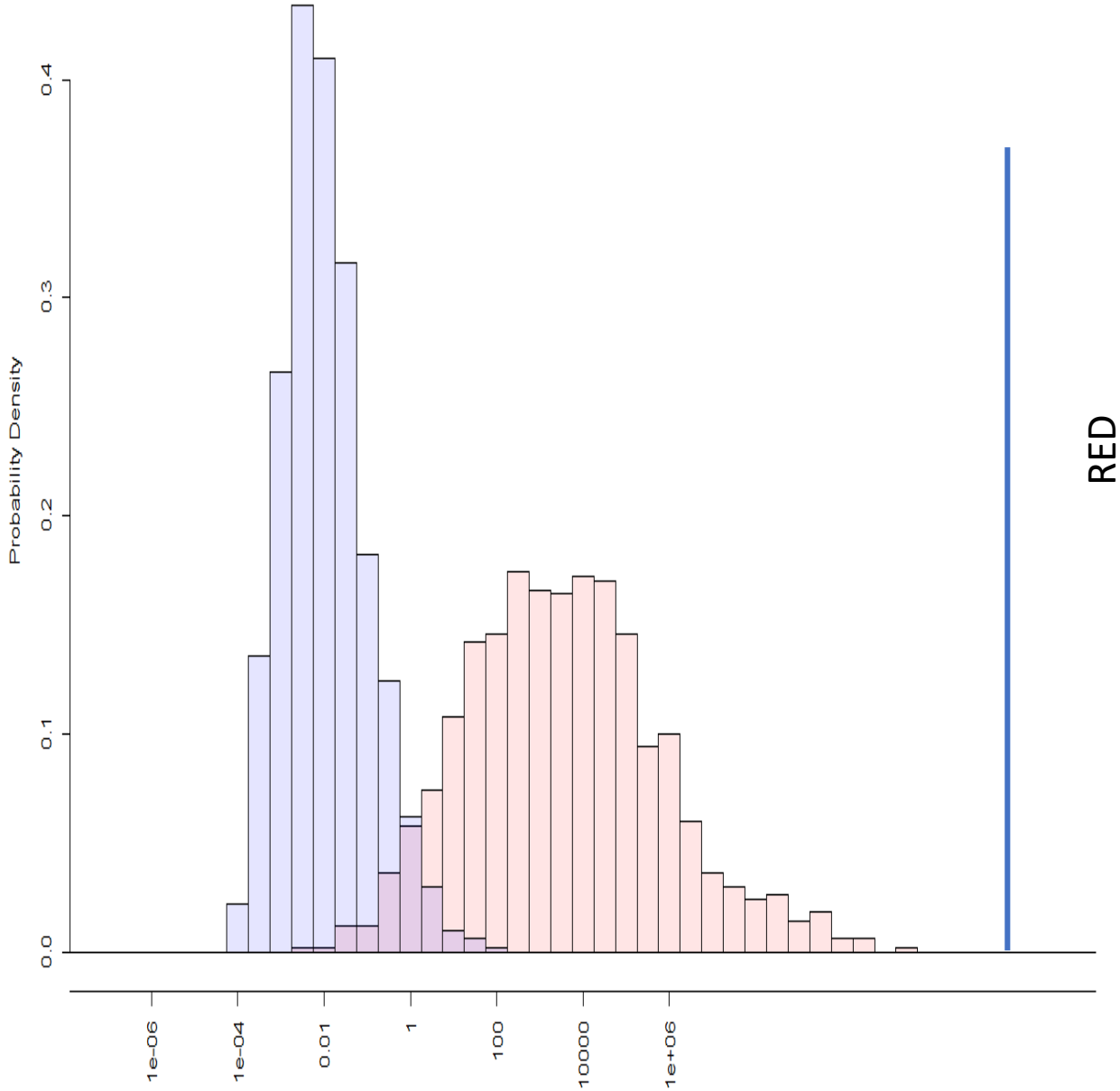
BLUE


Probability Density
0.00
0.05
0.10
0.15
0.20
0.25
1e-20
1e-06
1e-04
0.01
1
100
10000
1e+06
1e+20
1e+30
LR

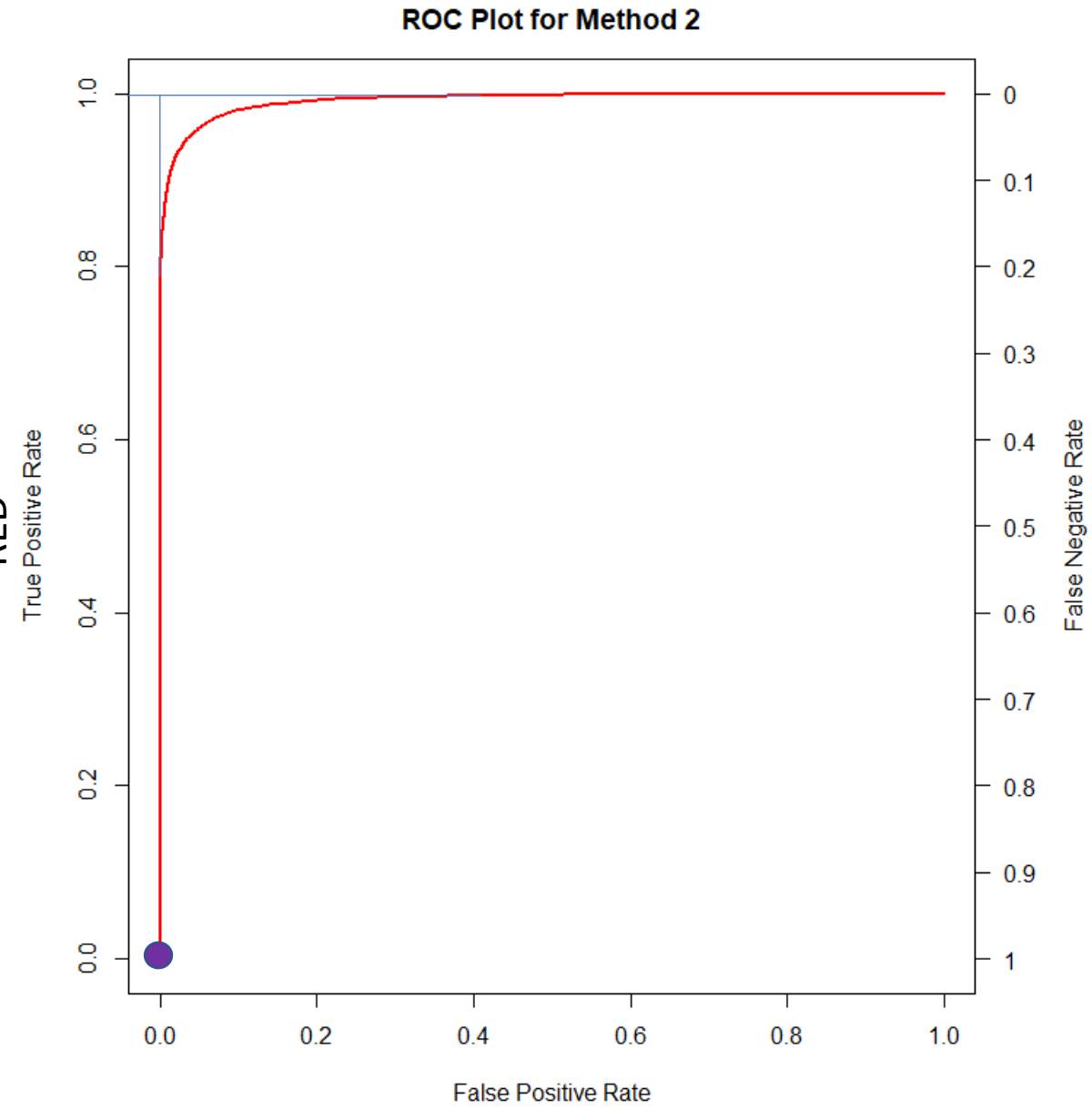

ROC Plot for Method 1
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

RED

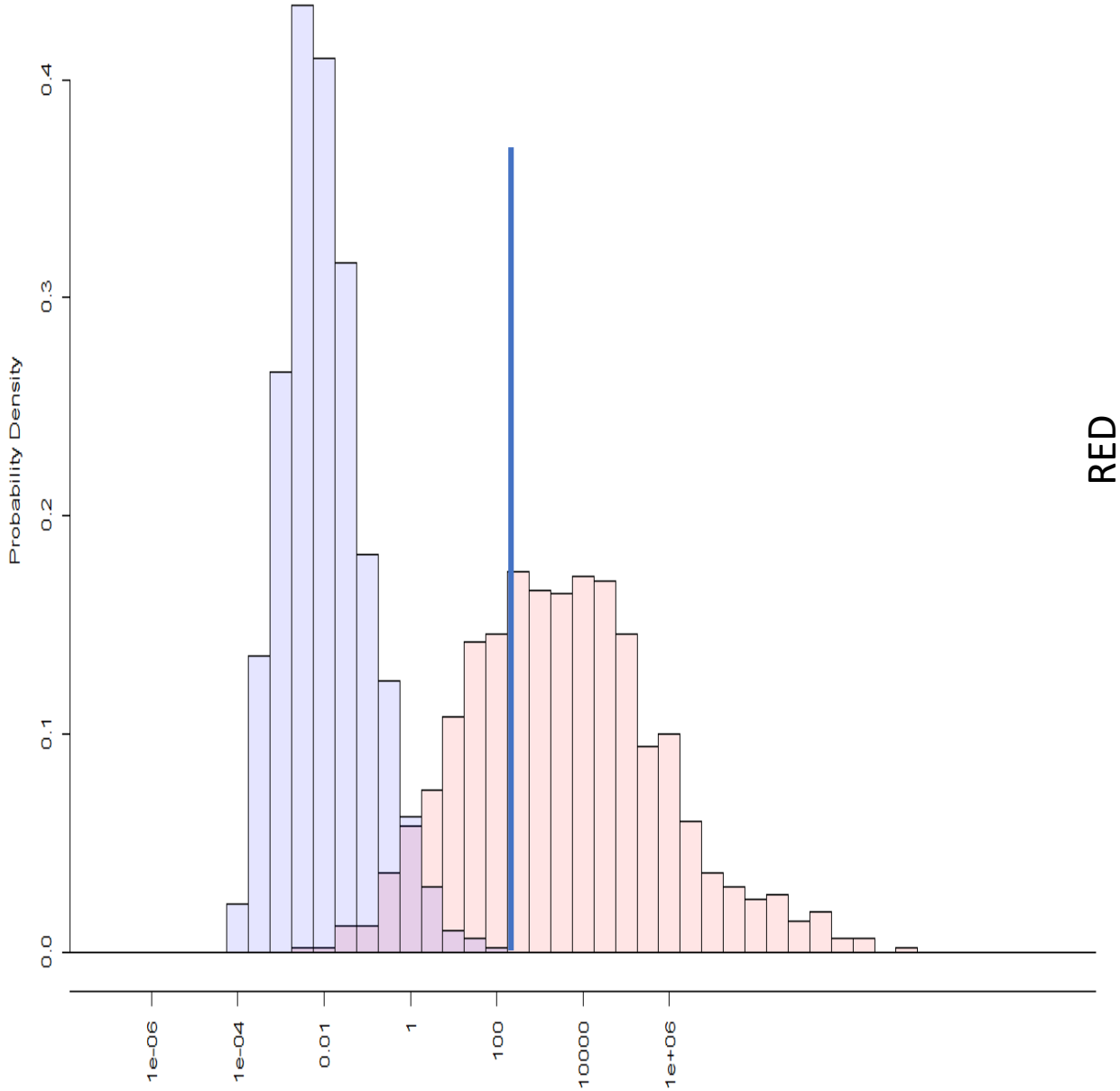
BLUE

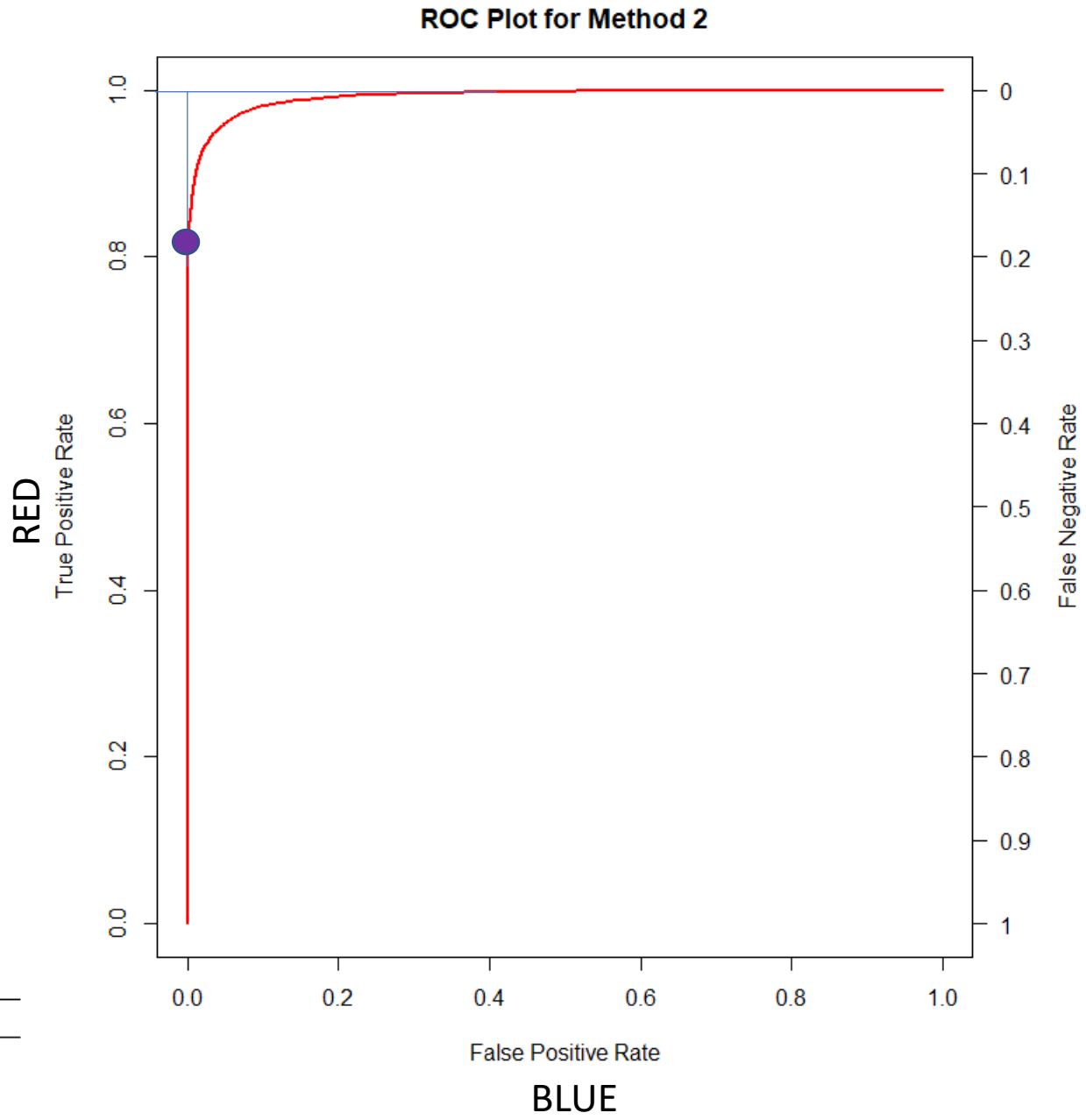

Probability Density
0.0
0.1
0.2
0.3
0.4
1e-06
1e-04
0.01
1
100
10000
1e+06

RED


ROC Plot for Method 2
True Positive Rate
False Negative Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

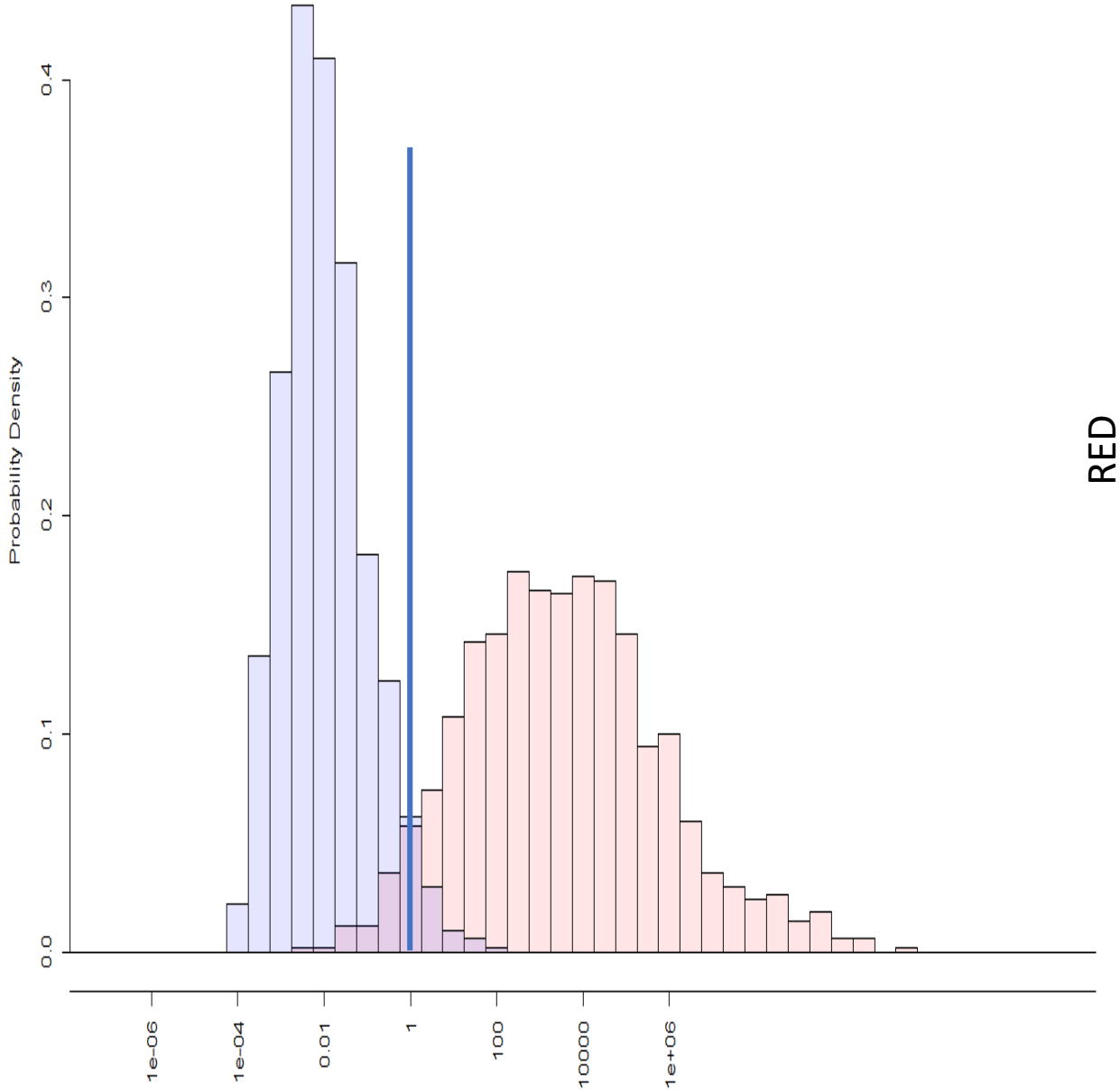
BLUE

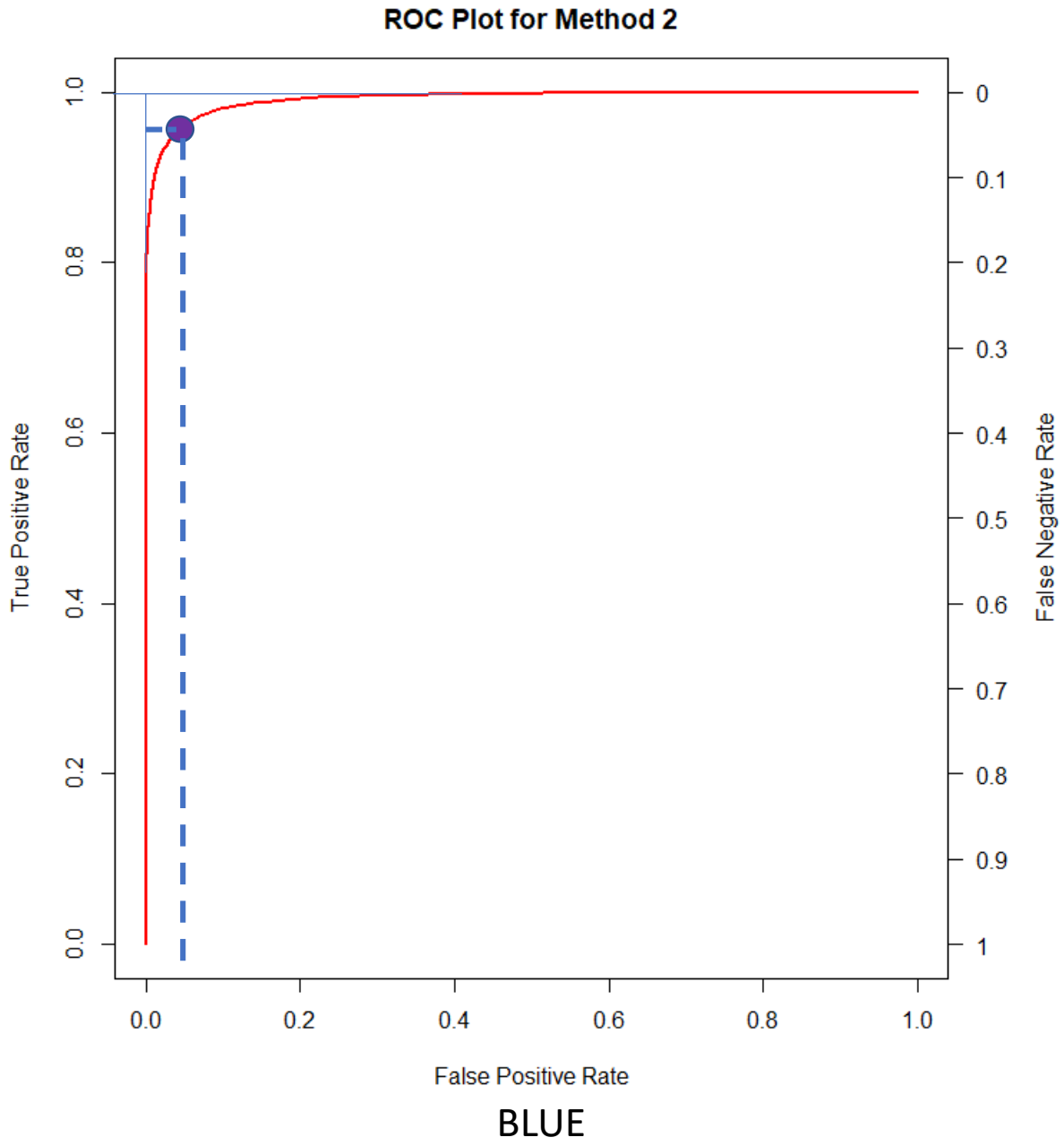

Probability Density
0.0
0.1
0.2
0.3
0.4
1e-06
1e-04
0.01
1
100
10000
1e+06

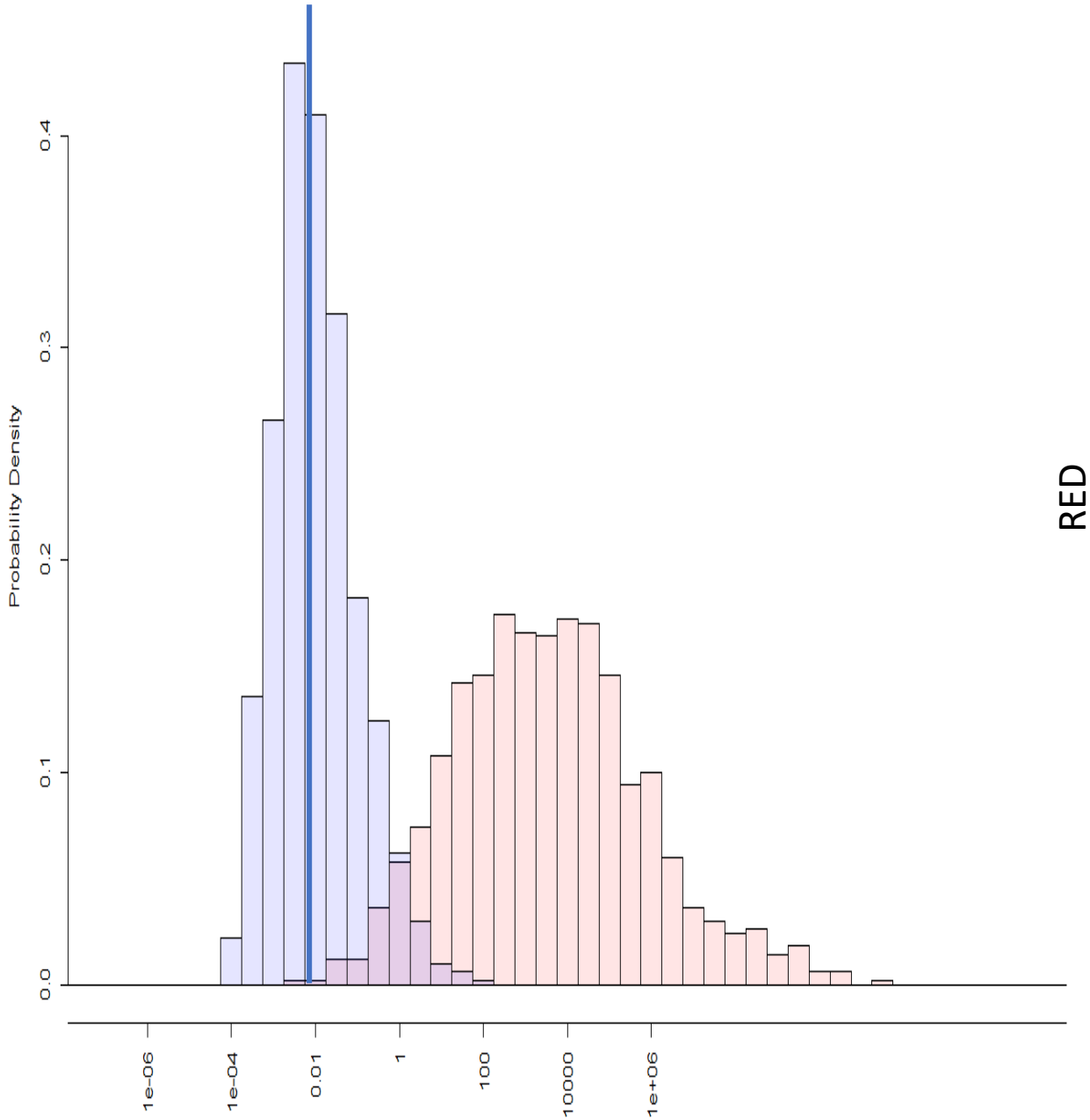

ROC Plot for Method 2
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

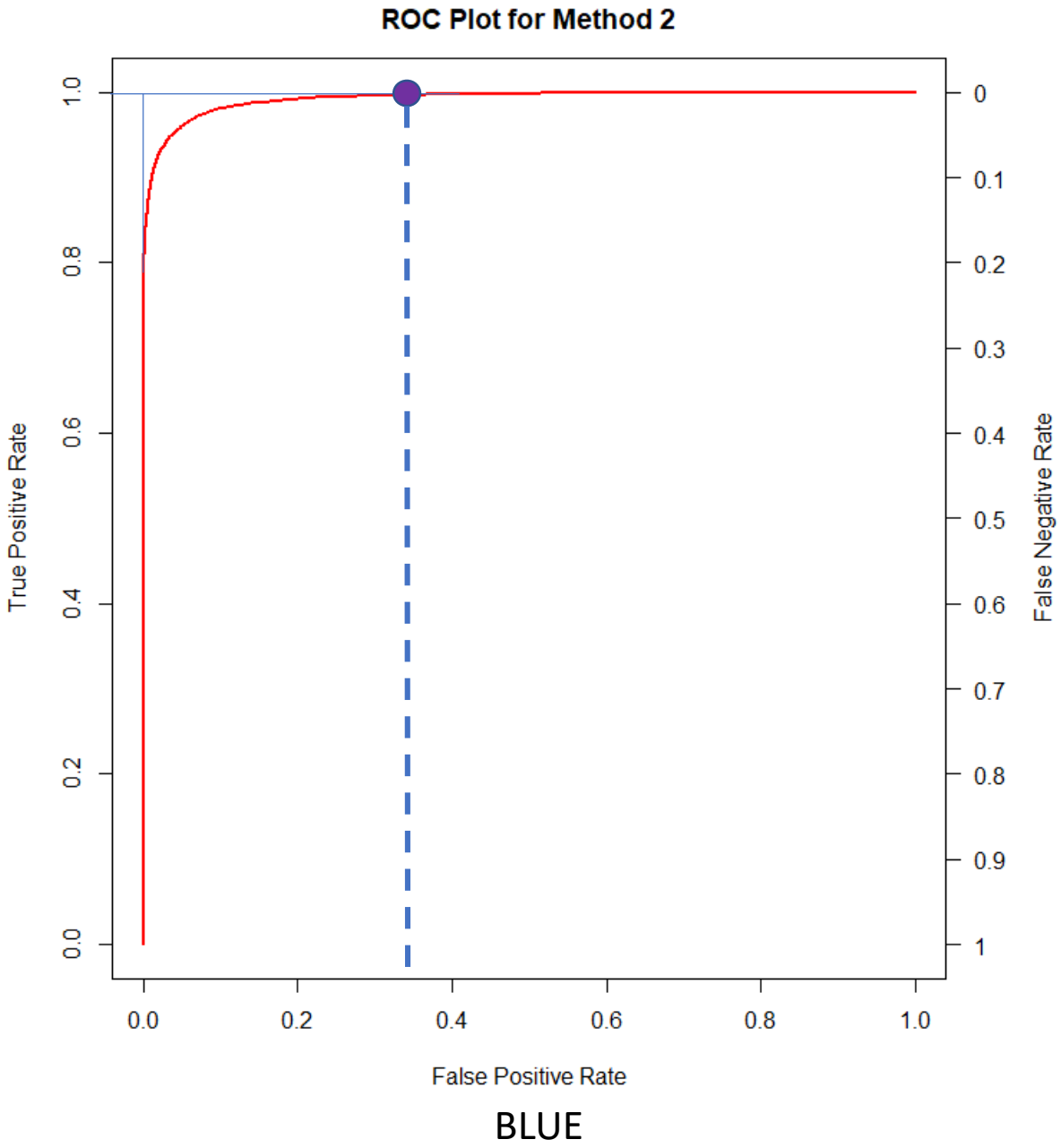
RED

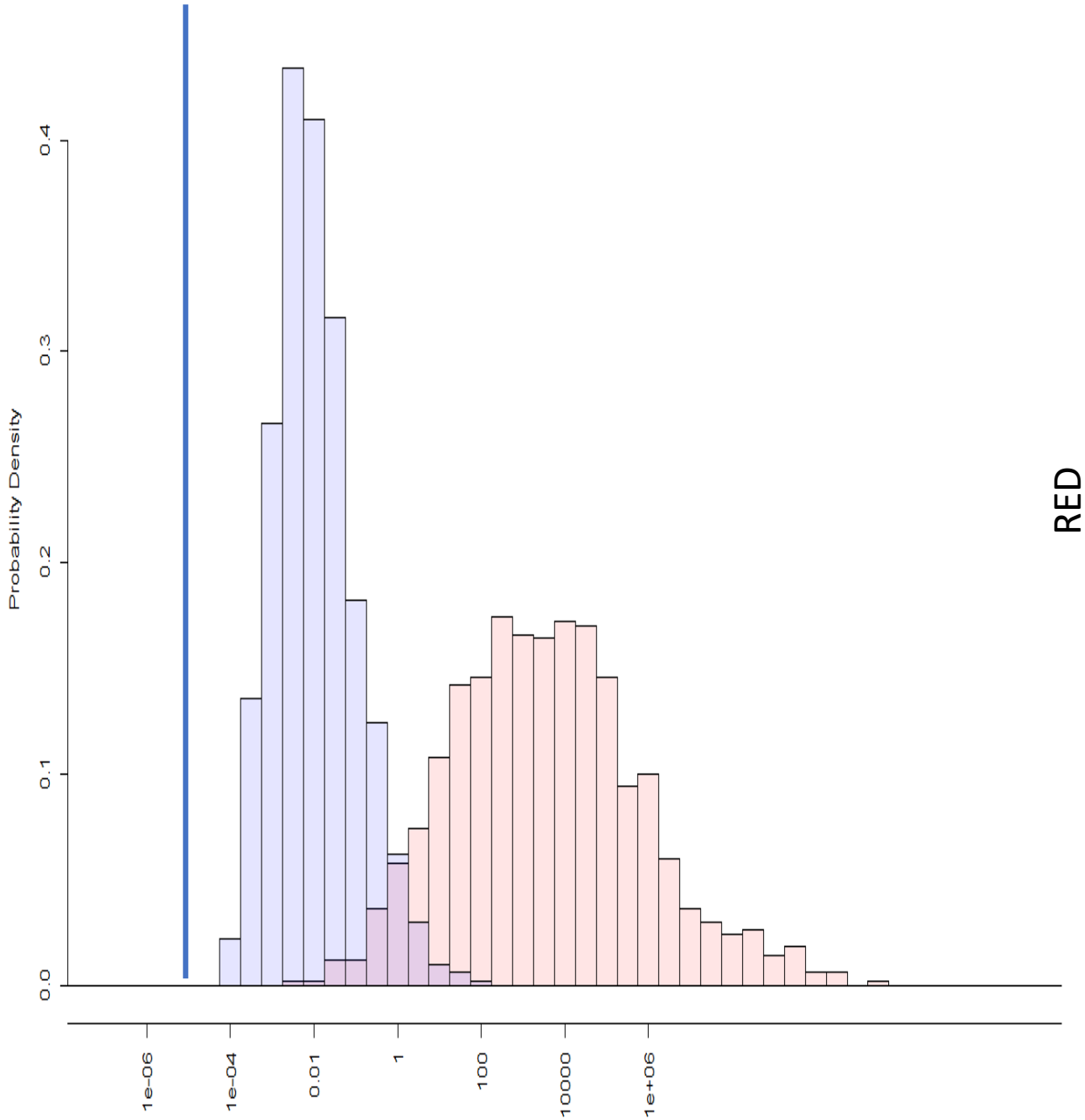
BLUE

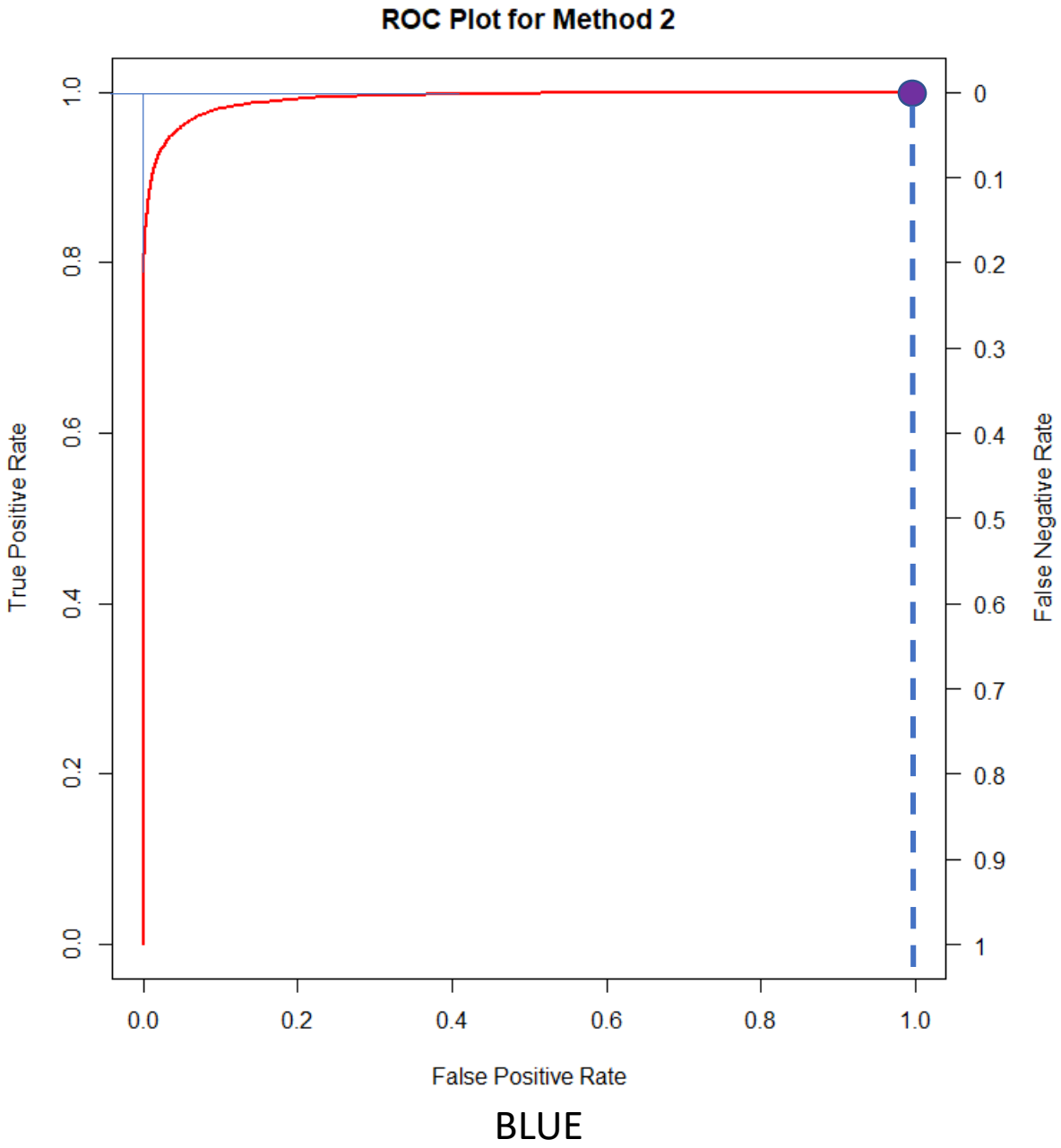

Probability Density
0.0
0.1
0.2
0.3
0.4
1e-06
1e-04
0.01
1
100
10000
1e+06


ROC Plot for Method 2
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
RED
BLUE

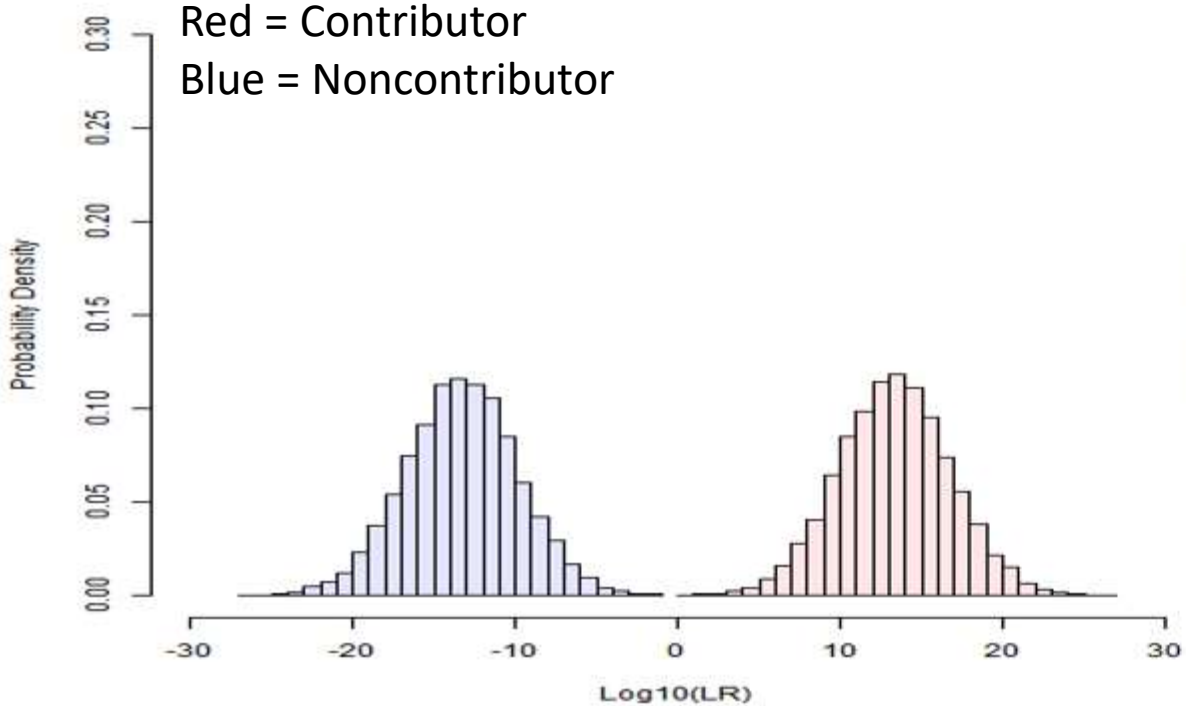

Probability Density
0.0
0.1
0.2
0.3
0.4
1e-06
1e-04
0.01
1
100
10000
1e+06


ROC Plot for Method 2
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
RED
BLUE

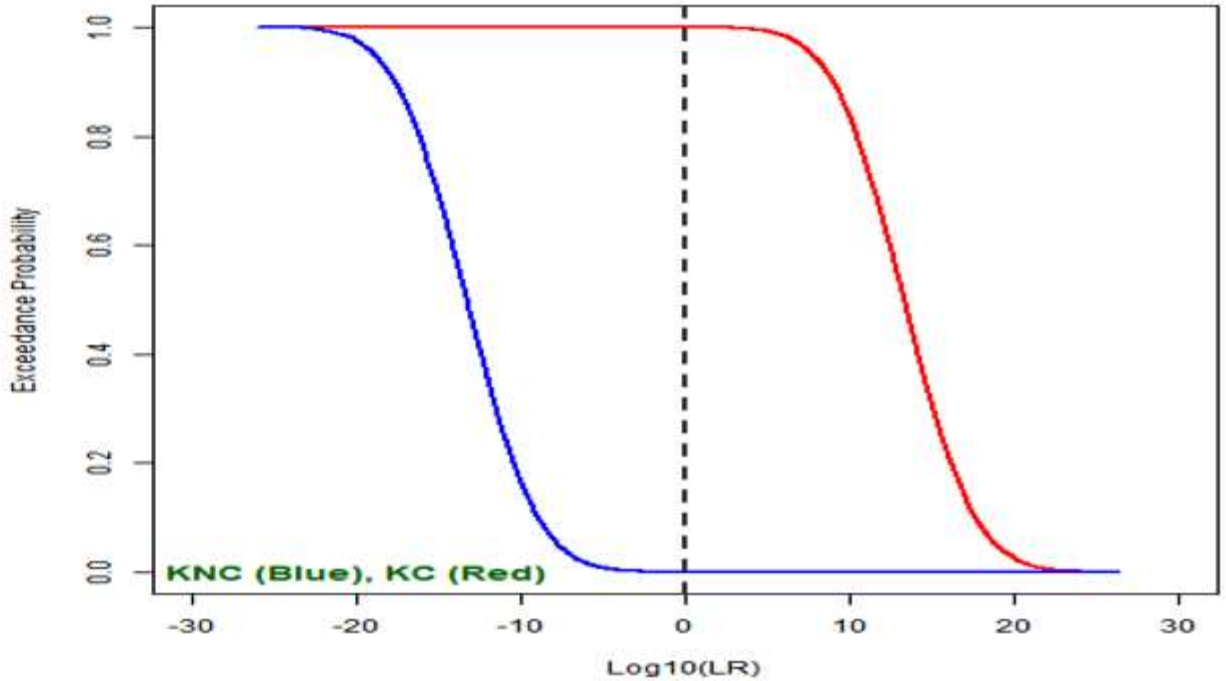

Probability Density
0.0
0.1
0.2
0.3
0.4
1e-06
1e-04
0.01
1
100
10000
1e+06

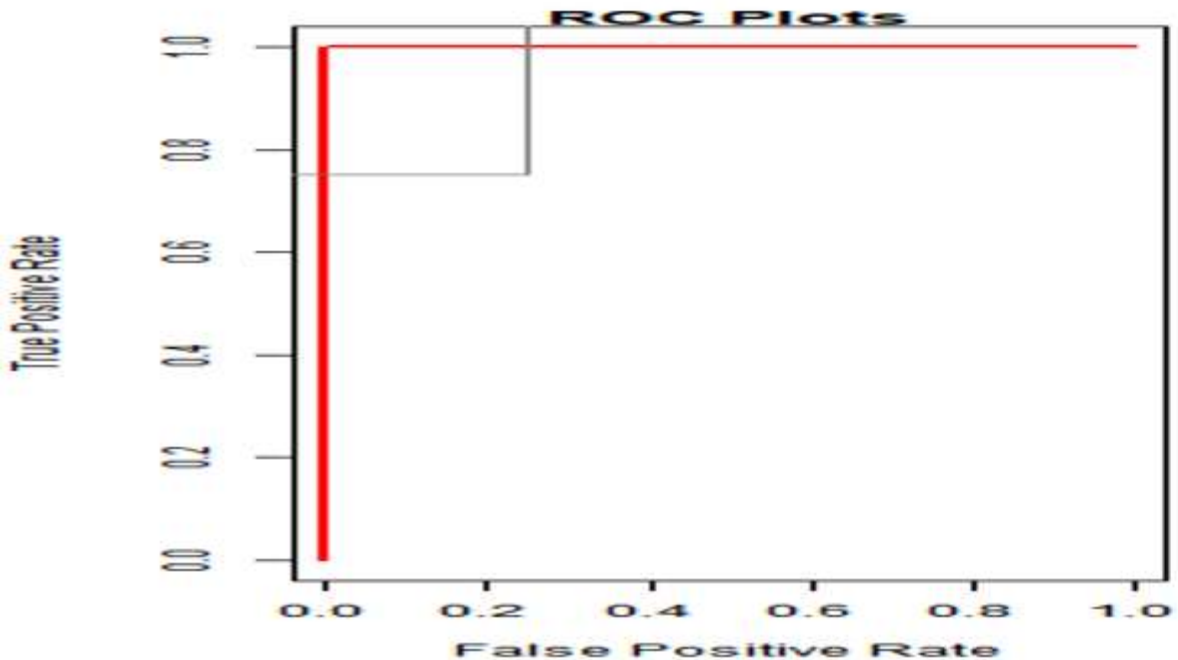

ROC Plot for Method 2
True Positive Rate
False Positive Rate
False Negative Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
RED
BLUE

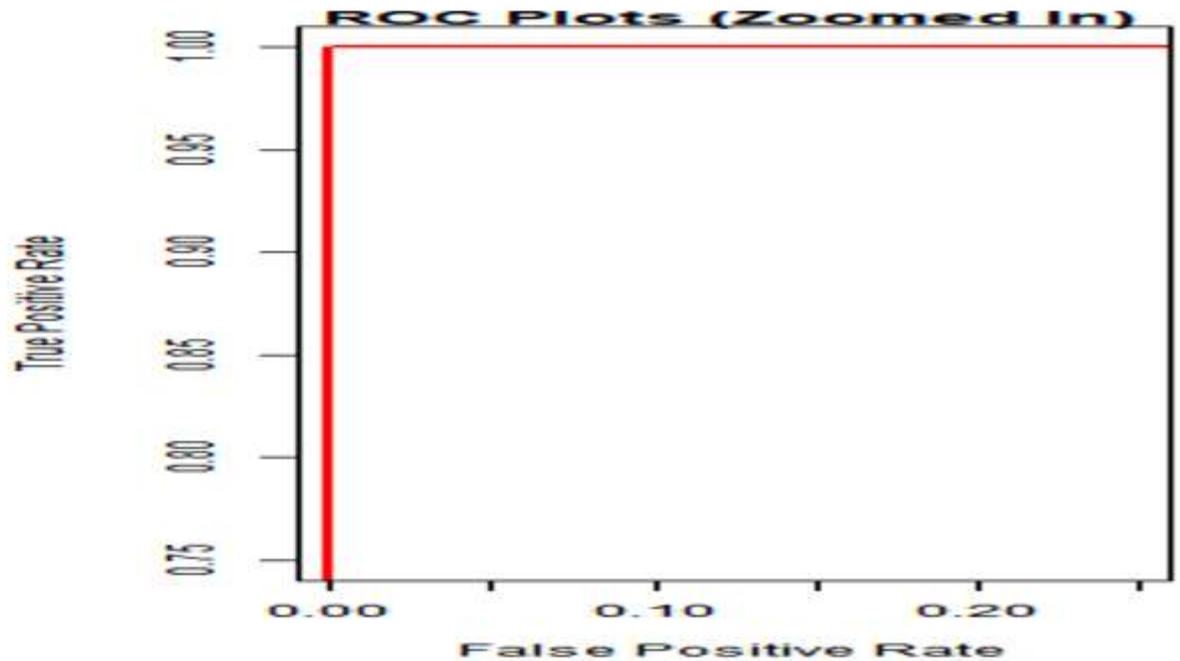
Red = Contributor
Blue = Noncontributor



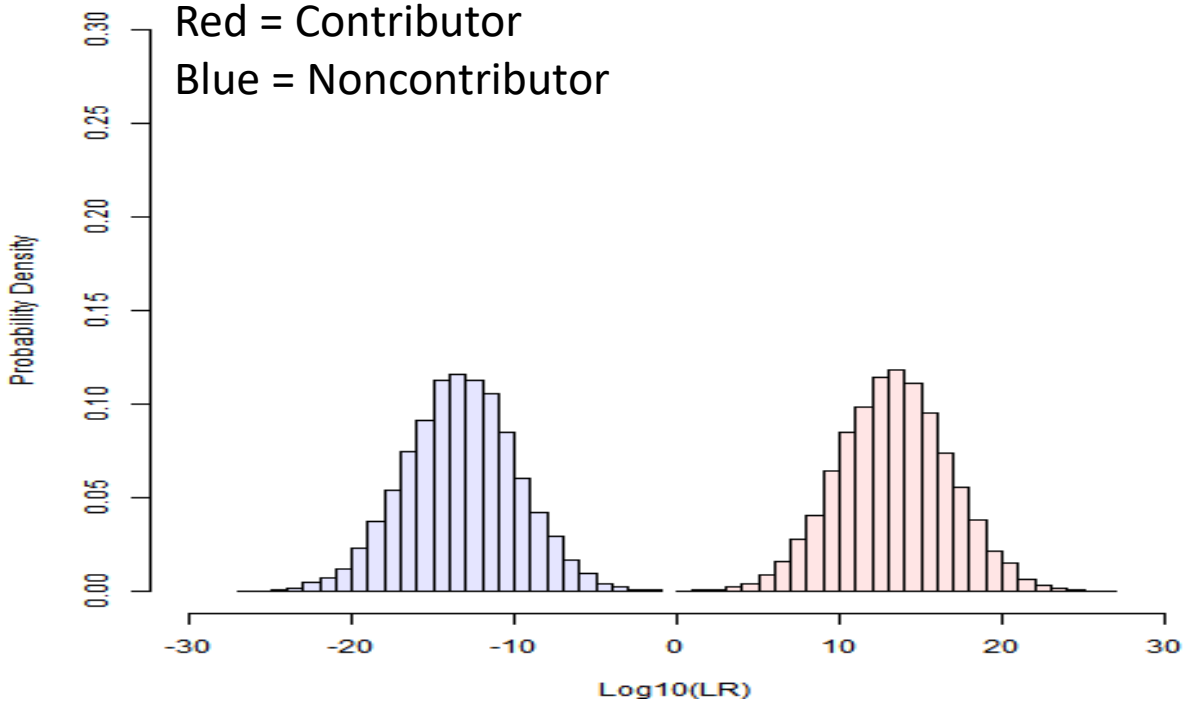
TIPPETT PLOT

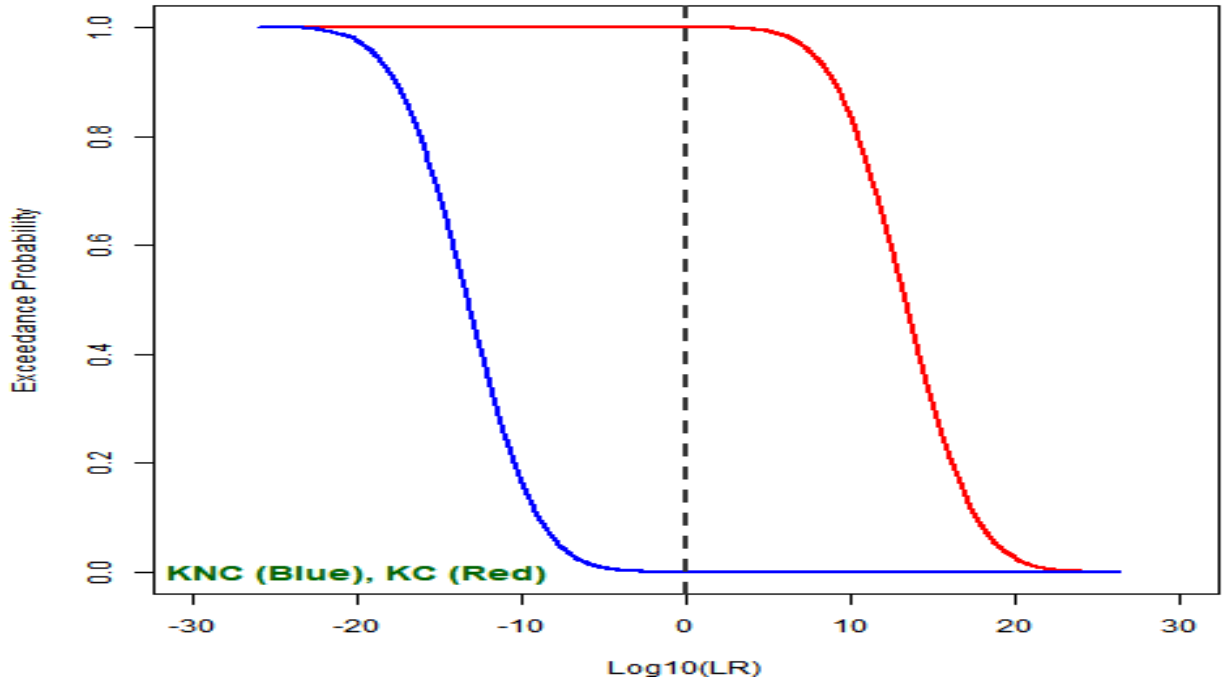


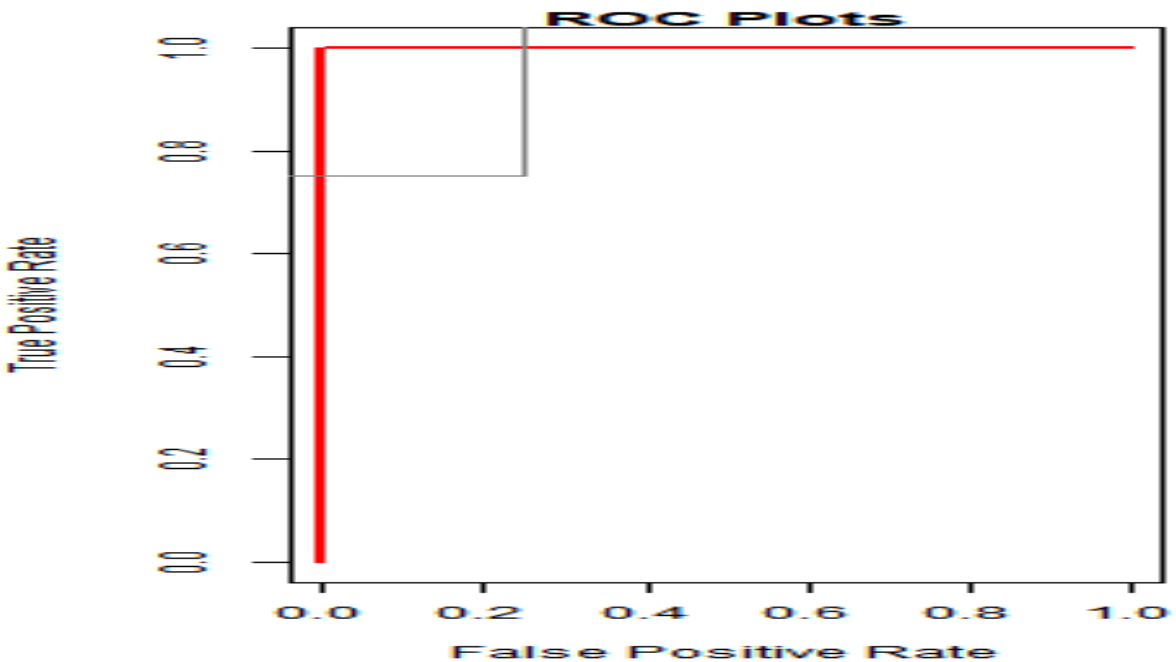


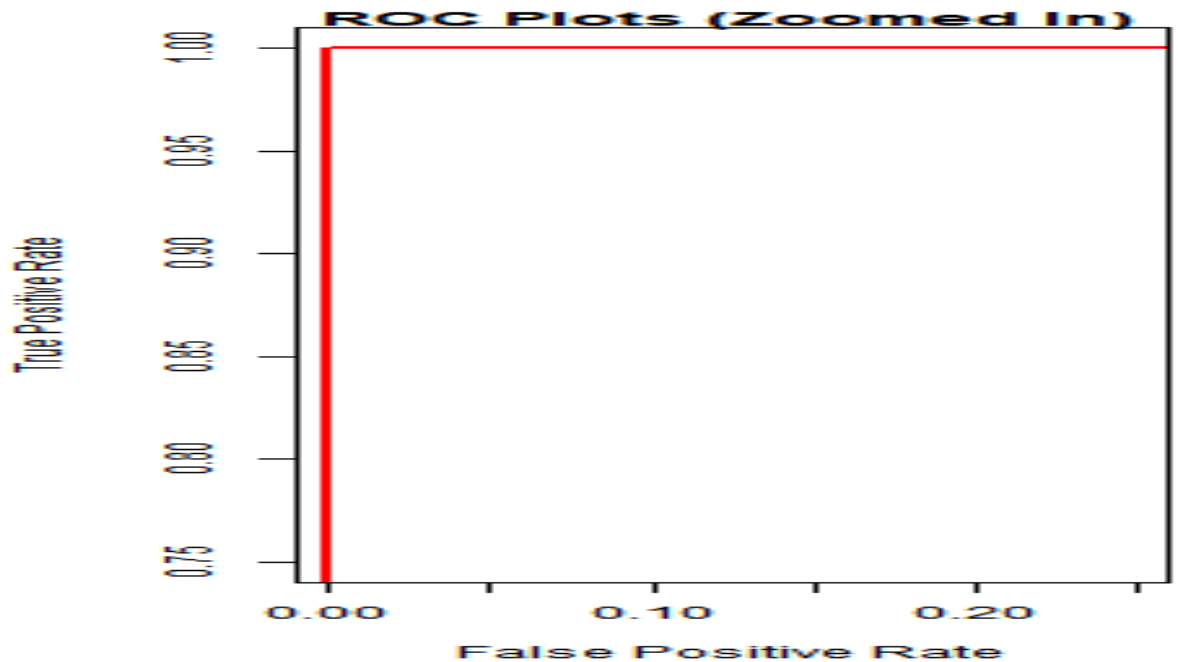


Red = Contributor
Blue = Noncontributor









Forensic Science International: Genetics 25 (2016) 85–96

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles

Øyvind Bleka [a,b,*], Corina C.G. Benschop [c], Geir Storvik [b], Peter Gill [a,d]

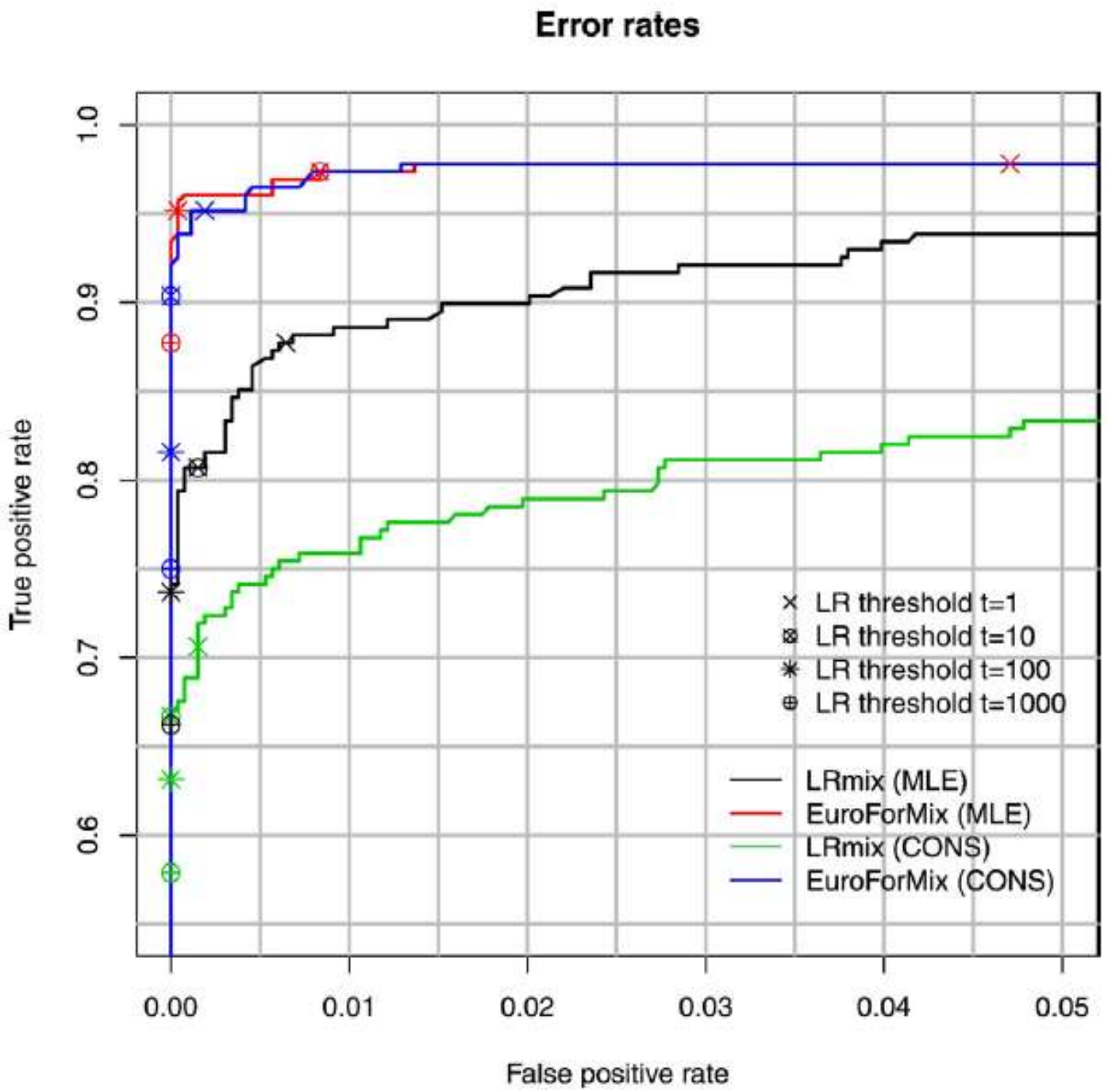


Fig. 2. Receiver operating characteristic (ROC) plot where the rate of false positives (FP) (along horizontal axis) and true positives (TP) (along vertical axis) are plotted as a function of LR thresholds. The plot shows the results for the maximum likelihood estimation method (MLE) and the conservative 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 system reliability
    - 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)

Some Take-aways

1. Primary focus should be on LR system reliability
   - 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)

2. Even if components are deemed reliable, system reliability must be checked.

The proof of the pudding is in the eating of it.

Component Reliability ≠ System Reliability

Component Reliability ≠ System Reliability

Some Take-aways

1. Primary focus should be on LR system reliability
   - 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)

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

3. When examining experimental results from reliability studies determine whether the experiment focused on selected parts of the system or on the entire system. Otherwise claims of reliability cannot be properly interpreted.

Some Take-aways

1. Primary focus should be on LR system reliability
   - 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)

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

3. When examining experimental results from reliability studies determine whether the experiment focused on selected parts of the system or on the entire system. Otherwise claims of reliability cannot be properly interpreted.

4. There is no single correct LR. **HOWEVER**, there are LR systems that perform sub-optimally relative to other LR systems.

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” (2018). There were at least 3 rebuttal papers or letters to the editor. We believe that they did not address any of our main concerns.

Volume 122, Article No. 27 (2017) https://doi.org/10.6028/jres.122.027
Journal of Research of National Institute of Standards and Technology

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

frontiers in Genetics

GENERAL COMMENTARY
published: 22 June 2018
doi: 10.3389/fgene.2018.00224

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

Keywords: likelihood ratio, value of evidence, forensic science, logarithm, forensic reporting

ELSEVIER

Forensic Science International
Volume 288, July 2018, Pages e15-e19

Rapid Communication

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]

frontiers in Genetics

GENERAL COMMENTARY
published: 22 June 2018
doi: 10.3389/fgene.2018.00224

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

Keywords: likelihood ratio, value of evidence, forensic science, logarithm, forensic reporting

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

QUESTIONS ?