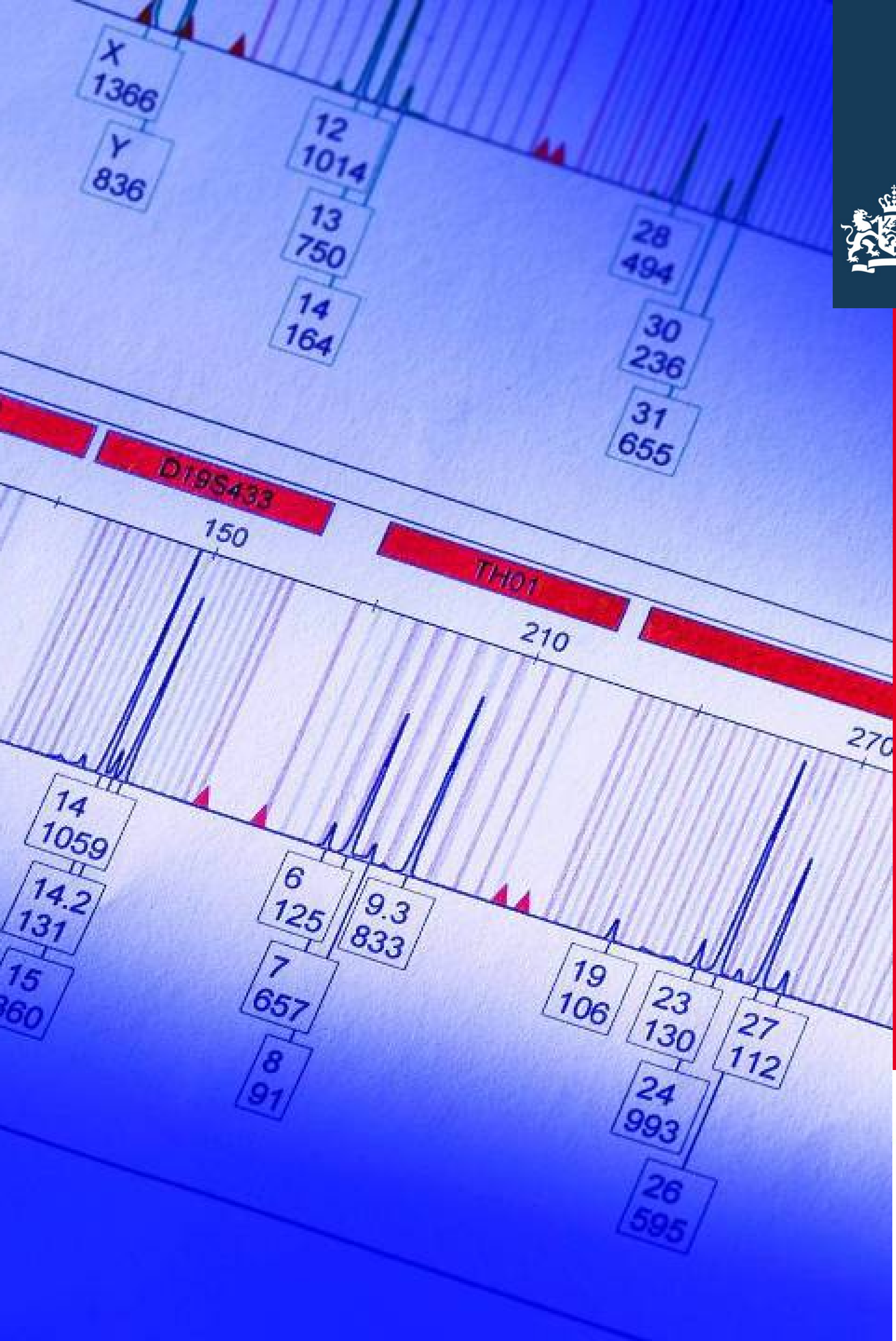


Netherlands Forensic Institute
Ministry of Security and Justice

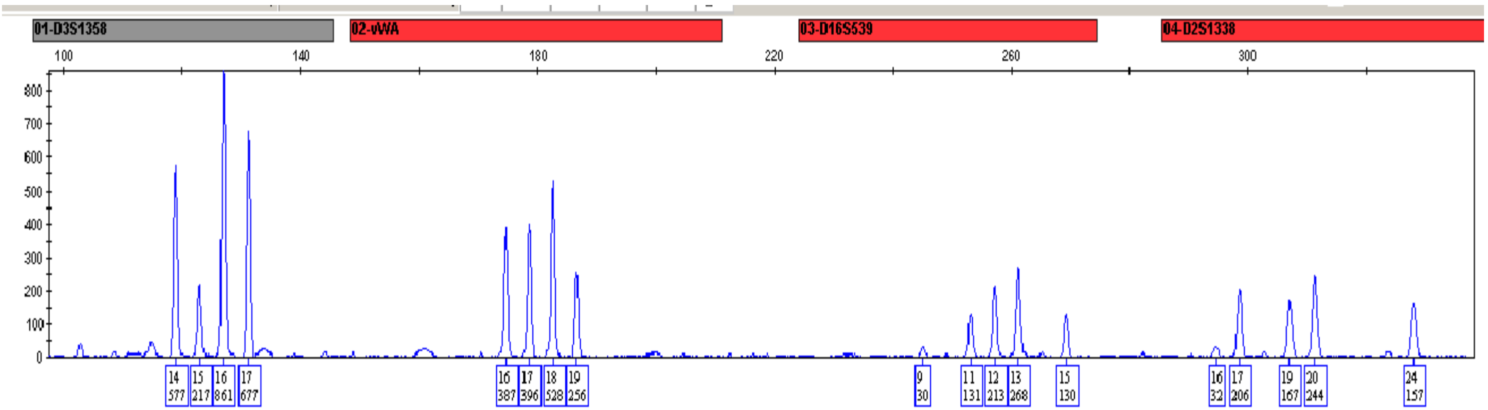
Evaluating Drop-out Probabilities

Hinda Haned

Forensic Genomics
Consortium Netherlands

A three-person mixture

- sample: epithelial cells recovered from the victim of an assault
- two suspects are detained by the police



A three-person mixture

Hypotheses

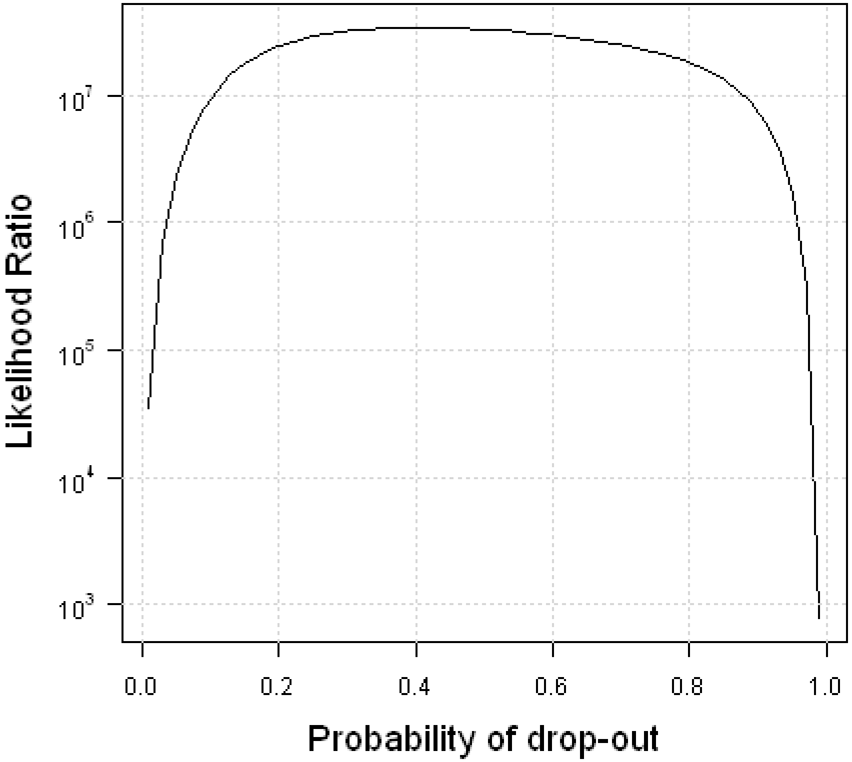
- Hp: the victim, suspect 1 and one unknown contributed to the sample
- Hd: the victim and two unknowns contributed to the sample

➔ Evaluation of the two hypotheses using likelihood ratios

Sensitivity analysis - LRmix

LRs range from 10^3 to 10^7:

- $10^3 \Rightarrow$ PrD=0.99
- $10^7 \Rightarrow$ PrD=0.41



Available methods

- Experimental mixtures (Perez et al, Coratian Med J, 2011)
  - the levels of drop-out, based on large sets of DNA mixtures obtained in different conditions
  - DNA quants
  - Number of contributors
  - Ratio of contribution

- Maximum likelihood principle LoComation software (Gill et al 2007)
  - Drop-out probabilities that maximize the probability of observing the questioned epg

Available methods

- Experimental mixtures (Perez et al, Coratian Med J, 2011)
  - the levels of drop-out, based on large sets of DNA mixtures obtained in different conditions
  - DNA quants
  - Number of contributors
  - Ratio of contribution

- Maximum likelihood principle LoComation software (Gill et al 2007)
  - Drop-out probabilities that maximize the probability of observing the questioned epg

Methods derive estimates from empirical distributions

Qualitative approach to the estimation of PrD

Relies on:

- the number of alleles observed in the sample
- the genotypes of the hypothesized contributors under H

? What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

Qualitative approach to the estimation of PrD

Relies on:

- the number of alleles observed in the sample
- the genotypes of the hypothesized contributors under H

? What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

➡ What is the distribution of the number of alleles for the questioned sample, conditioned on PrD?

Qualitative approach to the estimation of PrD

? What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

We don’t know the probabilities of drop-out, but we can evaluate the drop-out probabilities that could have led to a mixture similar to the one we are investigating

Qualitative approach to the estimation of PrD

? What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

We don’t know the probabilities of drop-out, but we can evaluate the drop-out probabilities that could have led to a mixture similar to the one we are investigating

Build the empirical distributions of the numbers of alleles, conditioned on the probabilities of dropout ranging in [0,1] using Monte-Carlo simulations

Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Questioned sample properties:

- three-person mixture: SGM+
- **33** alleles observed in the epg
- profiles of **victim** and **suspect** available

Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Questioned sample properties:

- three-person mixture: SGM+
- **33** alleles observed in the epg
- profiles of **victim** and **suspect** available

➡ Simulate a large number of mixtures that have these properties

Important note!

The hypothesized contributors change under Hp and under Hd:

- Derive distribution of the numbers of alleles under Hp and under Hd separately

- Yields two distributions, one under Hp and one under Hd

Monte-Carlo simulation procedure

Hp

Mixture 33 alleles

Victim
Suspect
Unknown

Hd

Mixture 33 alleles

Victim
Unknown
Unknwon

Step 1: simulate 1000 mixtures

Victim
Suspect
fixed → 36 alleles

Unknown
By randomly sampling from the Dutch DNA frequencies → 17 alleles

Mixture #1
51 distinct alleles

Mixture #2
50 distinct alleles

…

Mixture #1000
53 distinct alleles

Step 2: Apply drop-out

Mixture #1
51 alleles

Victim
Suspect
Unknown

drop-out

Pr(D)	# surviving alleles
0.01	51
0.02	50
…	…
0.50	25
…	
0.99	1

Repeat the procedure 1000 times

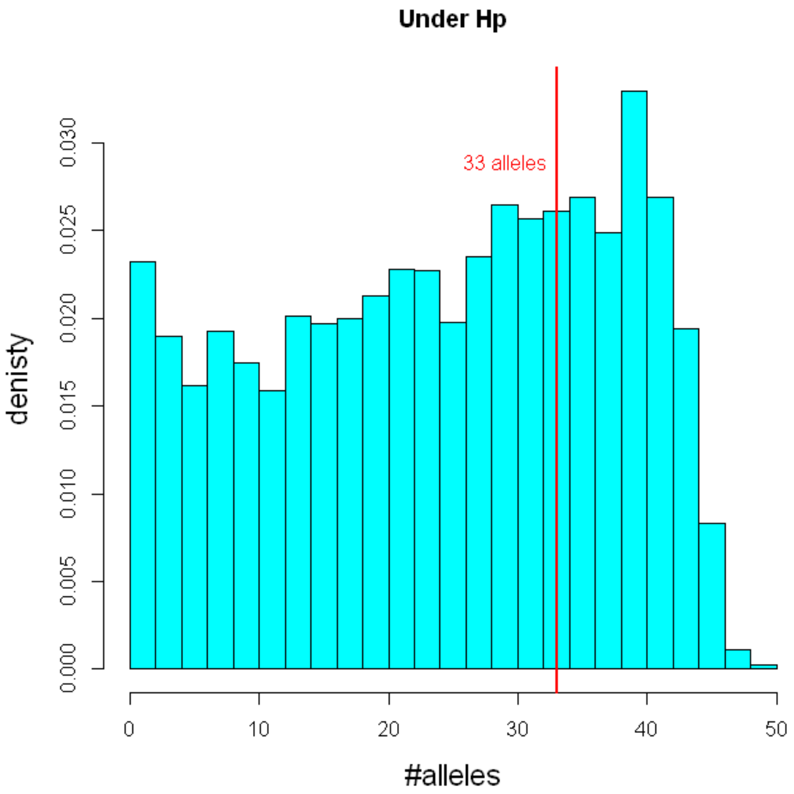
- Simulate 1000 mixtures
- For each mixture, record the number of alleles obtained after the simulated drop-out procedure

Repeat the procedure 1000 times

PrD	sim1	sim2	sim3	sim4	sim5	sim6	sim7	sim8	sim9	sim10	…
0,01	51	50	49	50	53	54	45	47	51	51	…
0,03	50	50	44	44	51	39	47	51	44	50	…
0,05	46	44	44	40	44	41	41	50	41	51	…
0,09	38	48	43	41	42	33	40	41	40	38	…
0,17	38	37	33	44	37	51	46	39	36	44	…
…	…	…	…	…	…	…	…	…	…	…	…
0,89	7	6	9	6	5	3	7	10	9	7	…
0,91	6	8	9	7	8	5	4	5	3	3	…
0,93	8	4	3	3	7	8	5	2	4	7	…
0,95	5	0	7	4	3	4	2	3	8	4	…
0,97	1	3	0	1	2	2	0	0	2	2	…
0,99	0	1	1	1	0	3	0	0	2	0	…

We look at the distributions of the numbers of alleles obtained in the simulation procedure

Distribution of the numbers of alleles among the 1000 mixtures

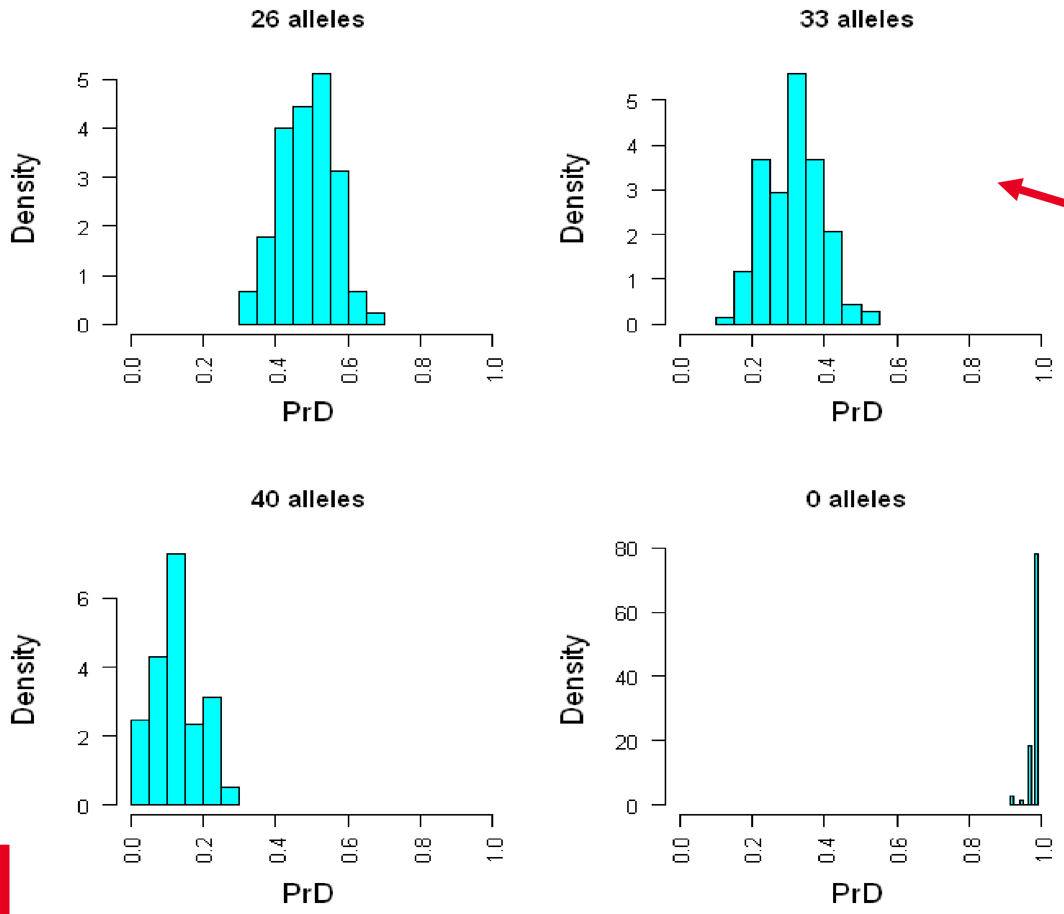


We are only interested in those where we obtained 33 alleles (the number observed in the epg).

Now we look at the distribution of the numbers of alleles and the corresponding drop-out probabilities

Distribution of the numbers of alleles among the 1000 mixtures

Hp



We are only interested in those where we obtained 33 alleles (the number observed in the epg).

The same procedure is carried out under Hd

Victim } fixed → 19 alleles

Unknown 1
Unknown 2 } By randomly sampling from the Dutch DNA frequencies → 35 alleles

Mixture #1
49 distinct alleles

Mixture #2
50 distinct alleles

…

Mixture #1000
48 distinct alleles

Apply drop-out

Mixture #1
49 alleles

Victim
Suspect
Unknown

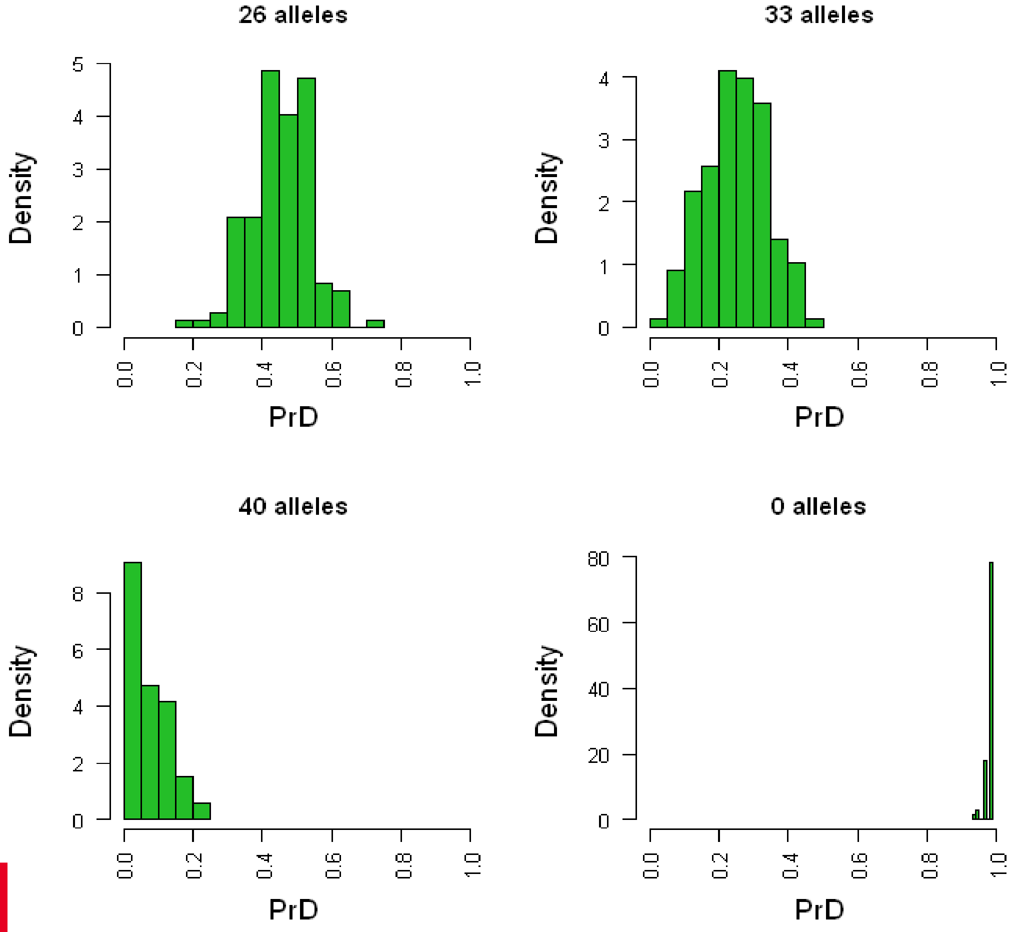
drop-out

Pr(D)	# surviving alleles
0.01	48
0.02	47
…	…
0.50	25
…	
0.99	1

The simulation procedure is repeated on a 1000 mixtures

Distribution of the numbers of alleles among the 1000 mixtures

Hd



We are only interested in those where we obtained 33 alleles (the number observed in the epg).

5% - 95% percentiles of the distributions

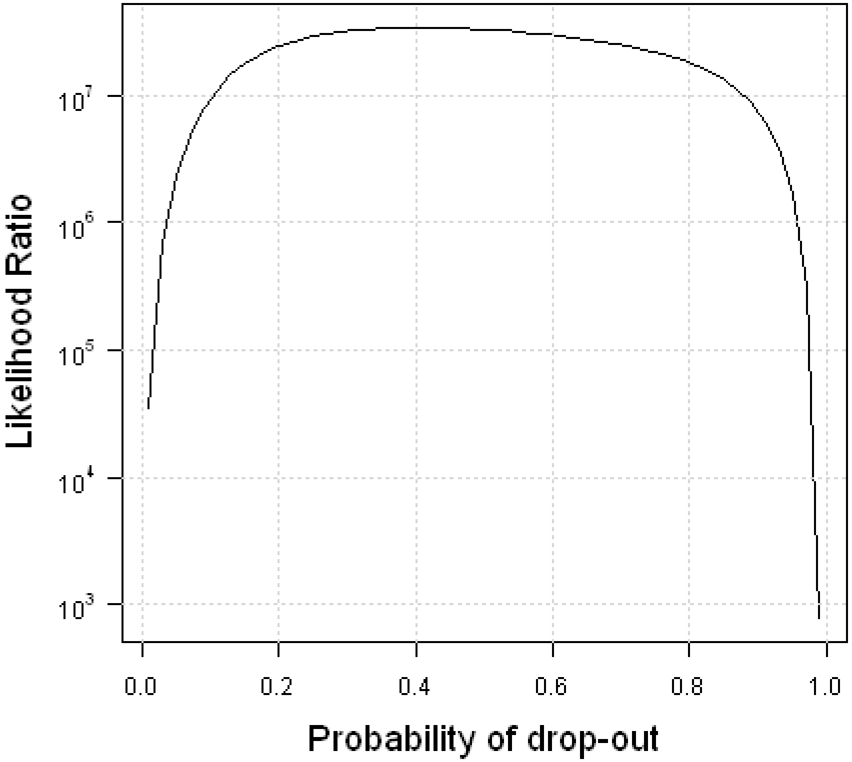
Under Hp	
5%	0.19
95%	0.45

Under Hd	
5%	0.09
95%	0.41

The drop-out estimates are given as a range: lowest-highest value

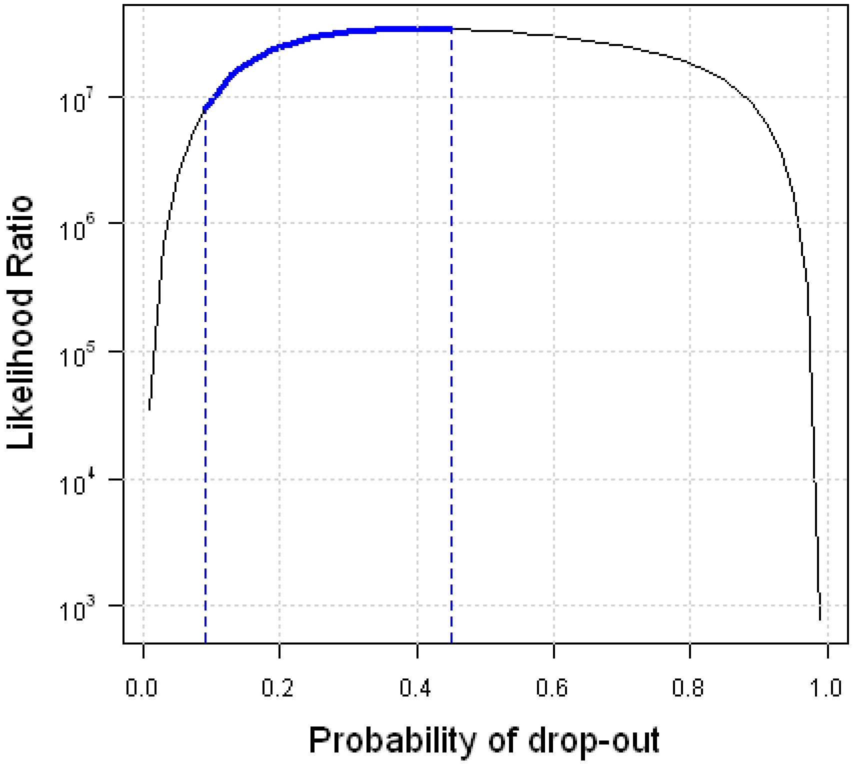
[0.09,0.45]

Sensitivity analysis



Sensitivity analysis vs. plausible ranges for PrD

$LR \cong 10^7$



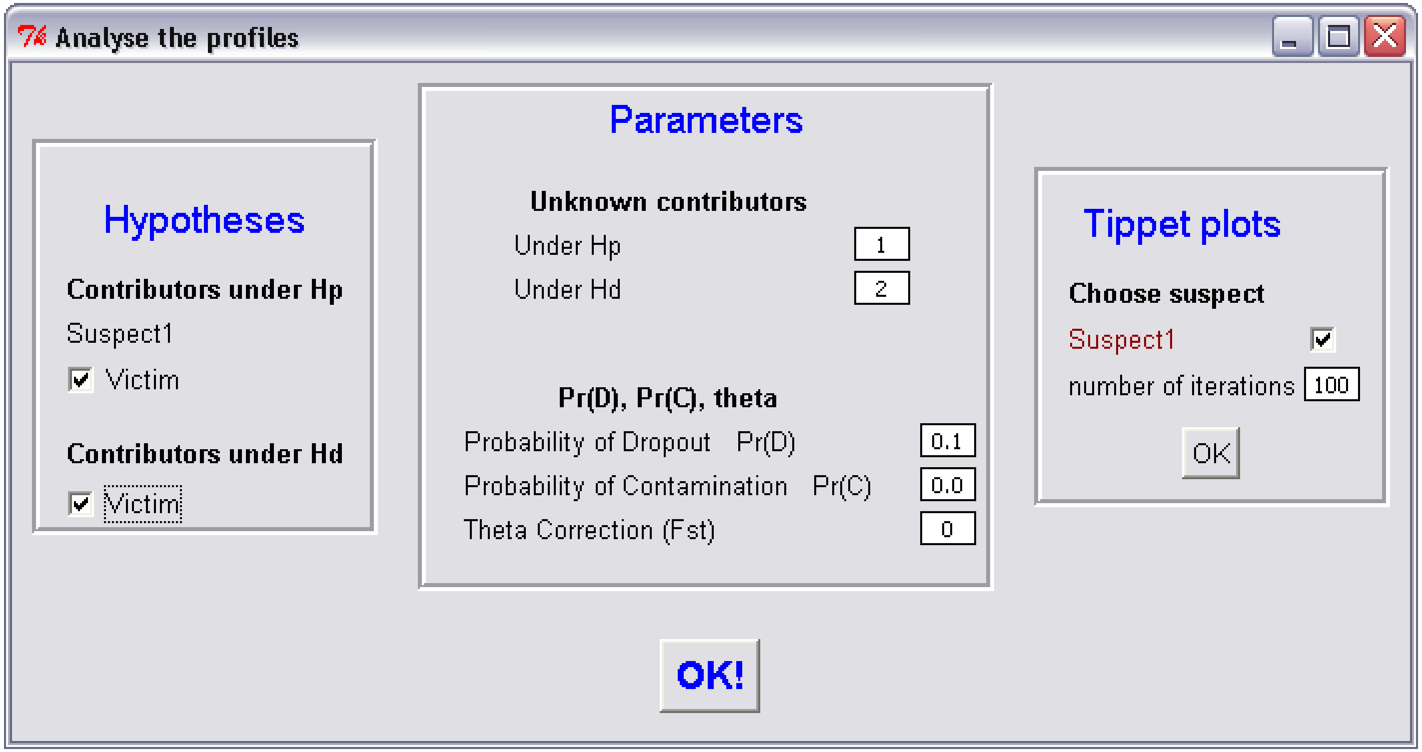
Summary

- the ranges of the drop-out probability can be evaluated separately under Hp and Hd
- avoid reporting values of drop-out that are supported by one hypothesis but not by its alternative
- qualitative data only
- peaks slightly under threshold not taken into account

Assess the uncertainty the data!

LRmix module

Available in Forensim 4.1



Hypotheses

Contributors under Hp

Suspect1

☑ Victim

Contributors under Hd

☑ Victim

Parameters

Unknown contributors

Under Hp: 1

Under Hd: 2

Pr(D), Pr(C), theta

Probability of Dropout Pr(D): 0.1

Probability of Contamination Pr(C): 0.0

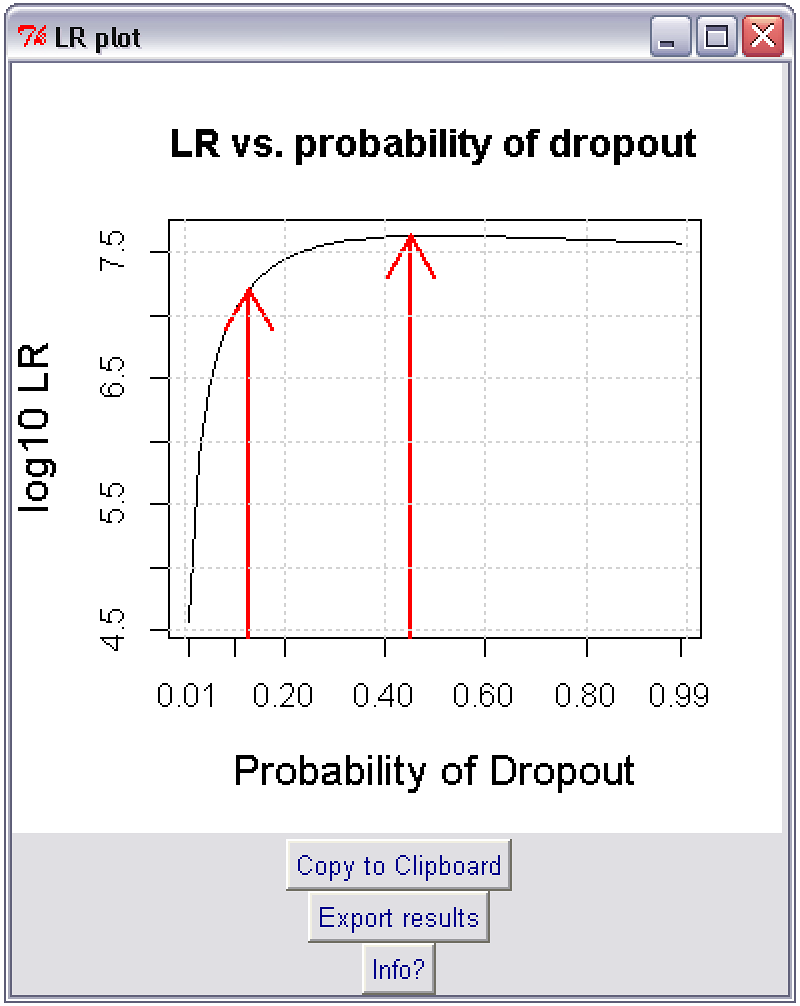
Theta Correction (Fst): 0

Tippet plots

Choose suspect

Suspect1 ☑

number of iterations: 100


LR plot
LR vs. probability of dropout
log10 LR
4.5
5.5
6.5
7.5
0.01
0.20
0.40
0.60
0.80
0.99
Probability of Dropout
Copy to Clipboard
Export results
Info?