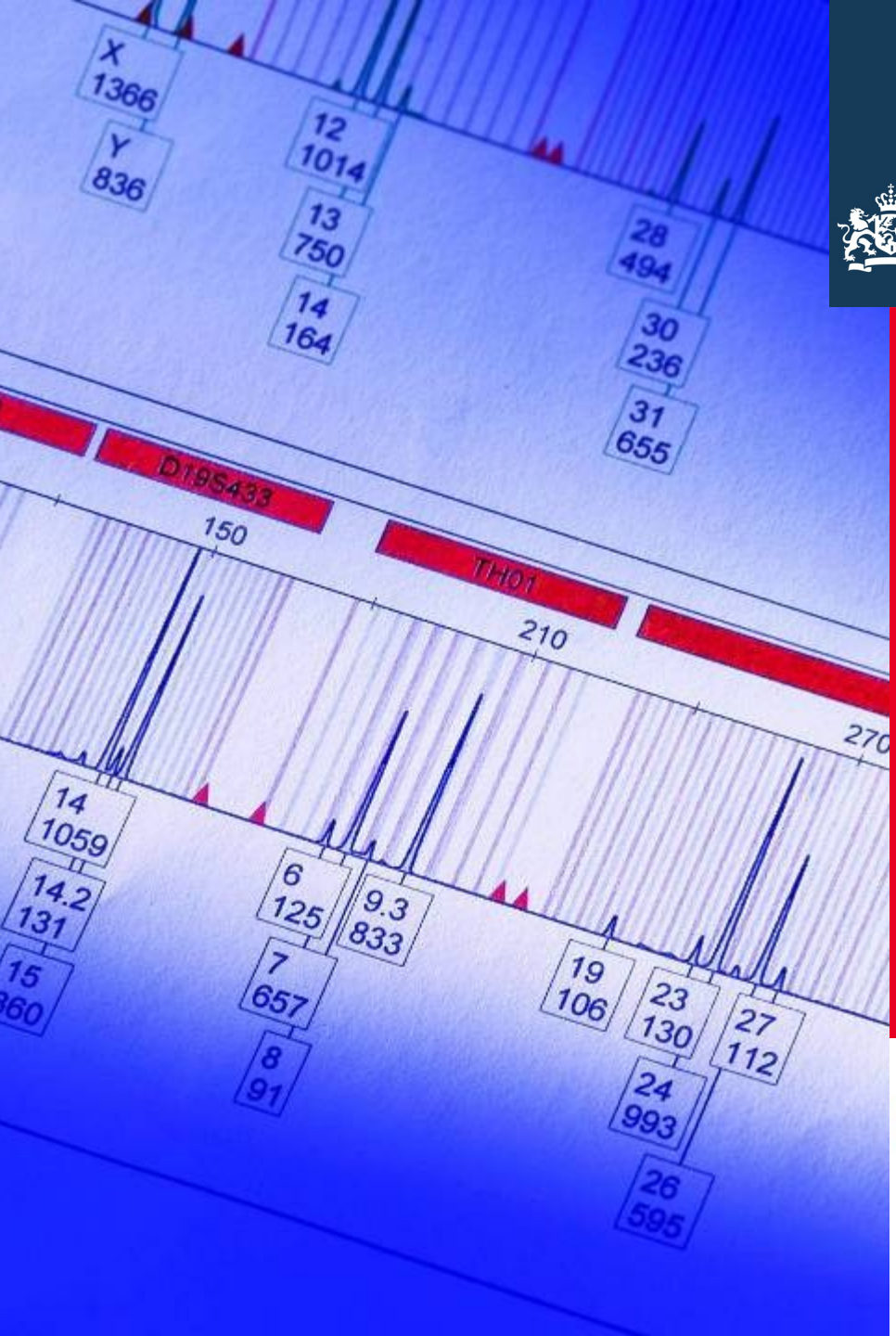


Netherlands Forensic Institute
Ministry of Security and Justice

Evaluating drop-out probabilities

Hinda Haned

h.haned@nfi.minvenj.nl

Forensic Genomics
Consortium Netherlands

Estimating the probability of dropout

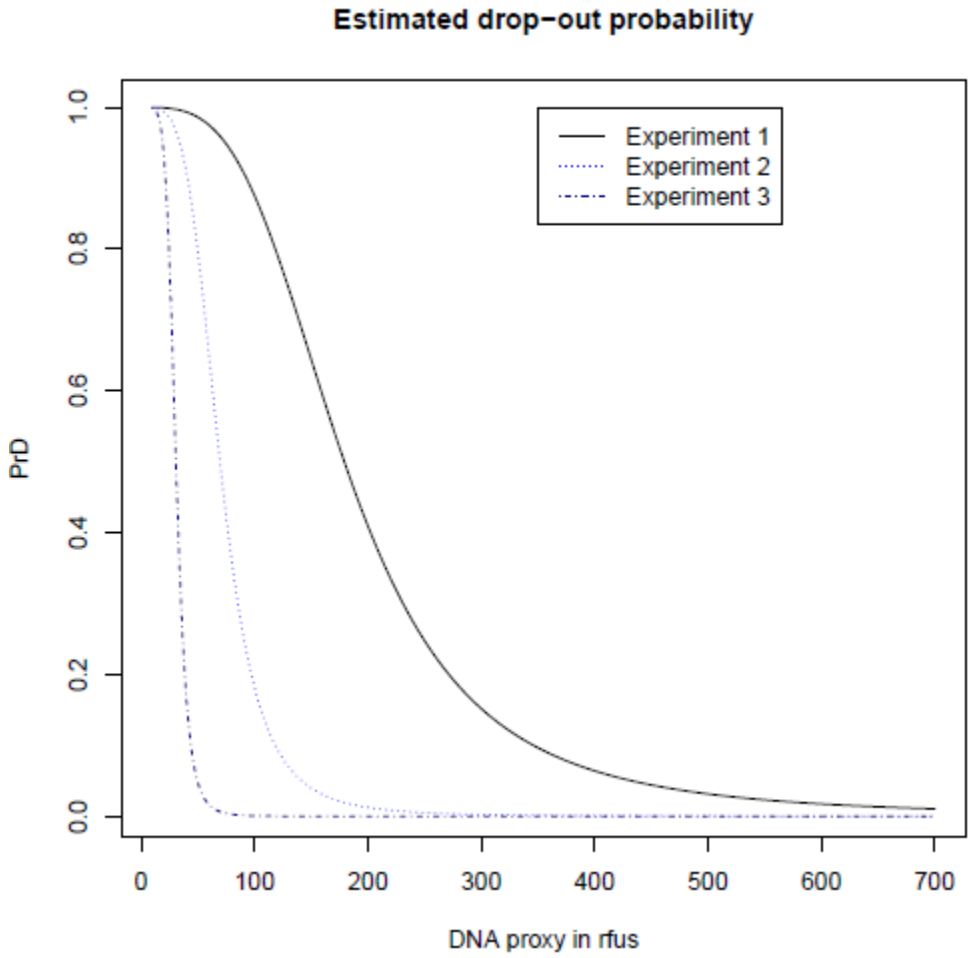
Given the observed profiles, what is the probability that an allele at a given locus has dropped-out?

Recent contributions based on the logistic regression suggest predicting the probability of dropout from a measure of the quality of the DNA profiles, based on peak heights:

- DNA proxy model: based on an estimation of the amount of DNA contributed to the stain Tvedebrink et al. FSIG, 2009
- Surviving peak heights model: based on heterozygote dropout Gill et al, FSIG, 2009

Logistic regression

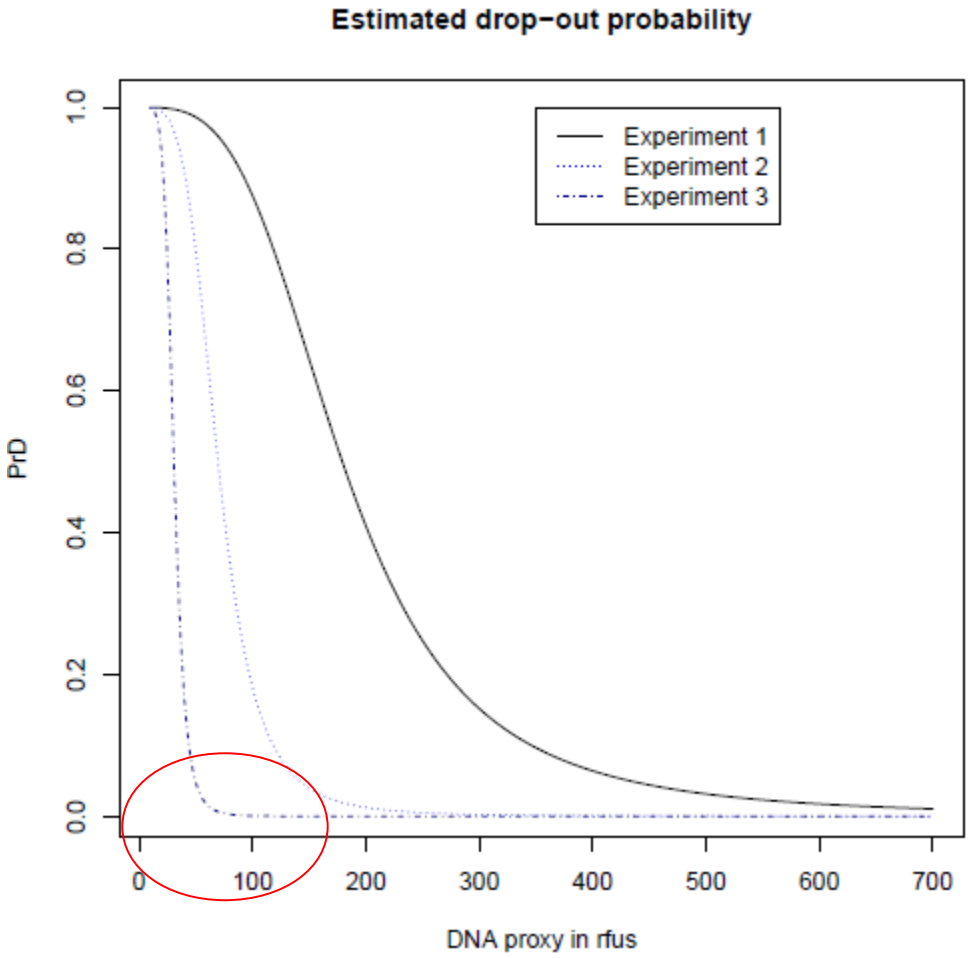
- Repeated experiments yield slightly different logistic curves describing the decay of dropout



Logistic regression

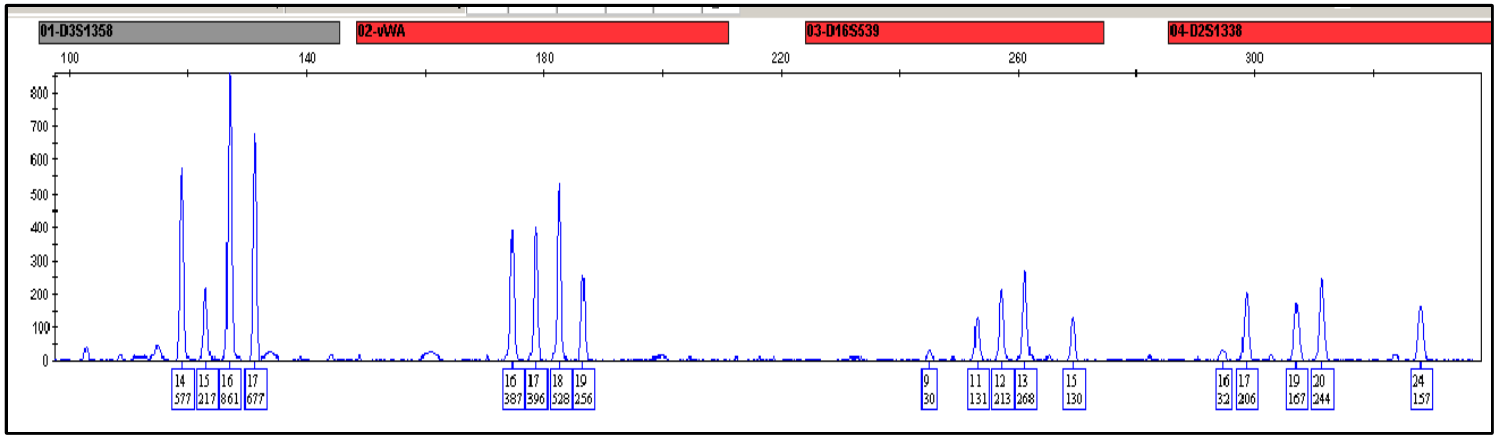
- Repeated experiments yield slightly different logistic curves describing the decay of dropout

- Critical range in LTDNA samples: 50-200 rfus (NGM system)



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

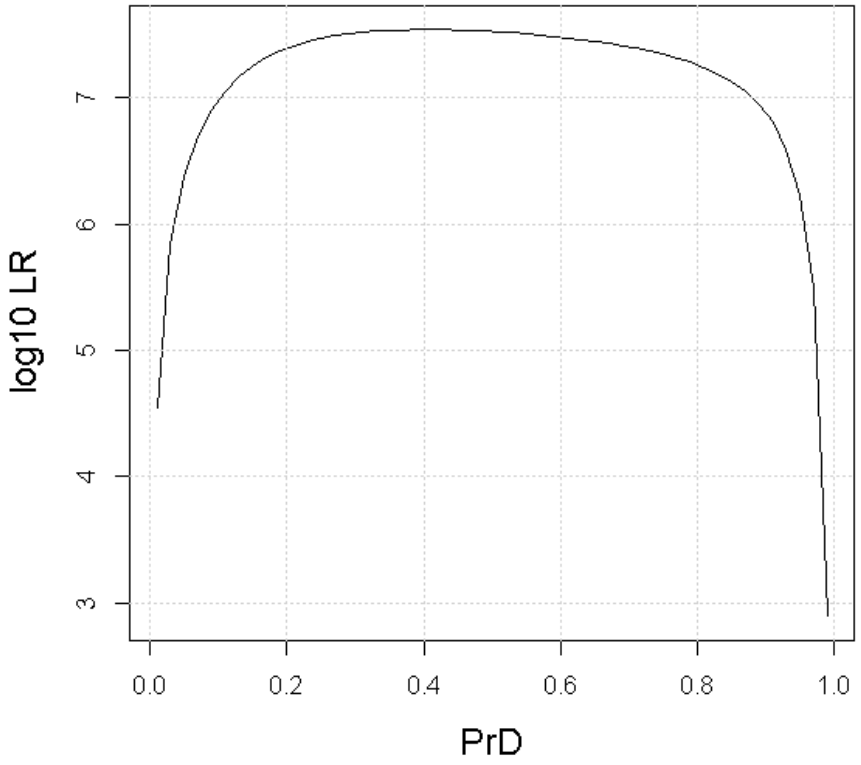
- Hp: the victim, suspect 1 and one unknown contributed to the sample
- Hd: the victim and two unknowns contributed to the sample

➔ 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 **suspect1** available

Monte Carlo method

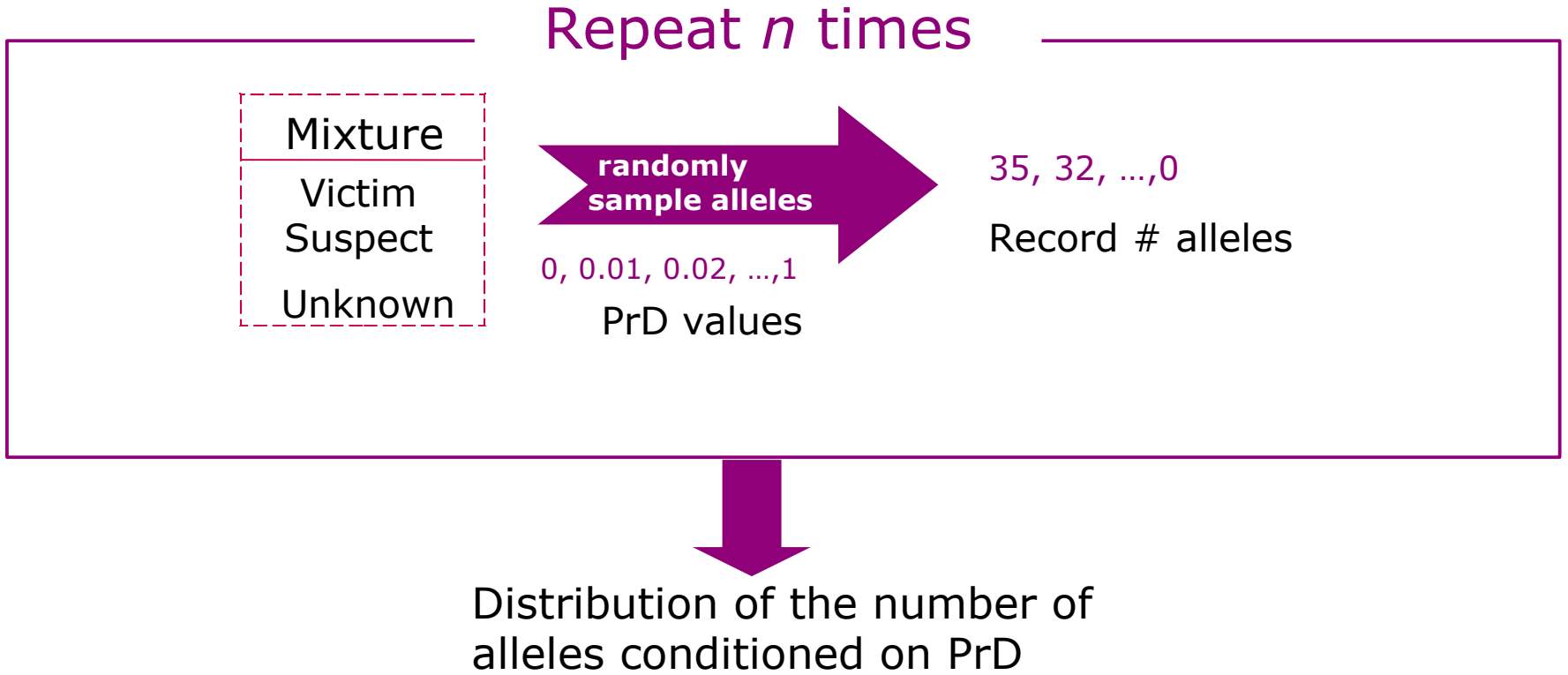
Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Questioned sample properties:

- three-person mixture: SGM+
- **33** alleles observed in the epg
- profiles of **victim** and **suspect1** available

➡ Simulate a large number of mixtures that have these properties

Monte-Carlo simulation procedure



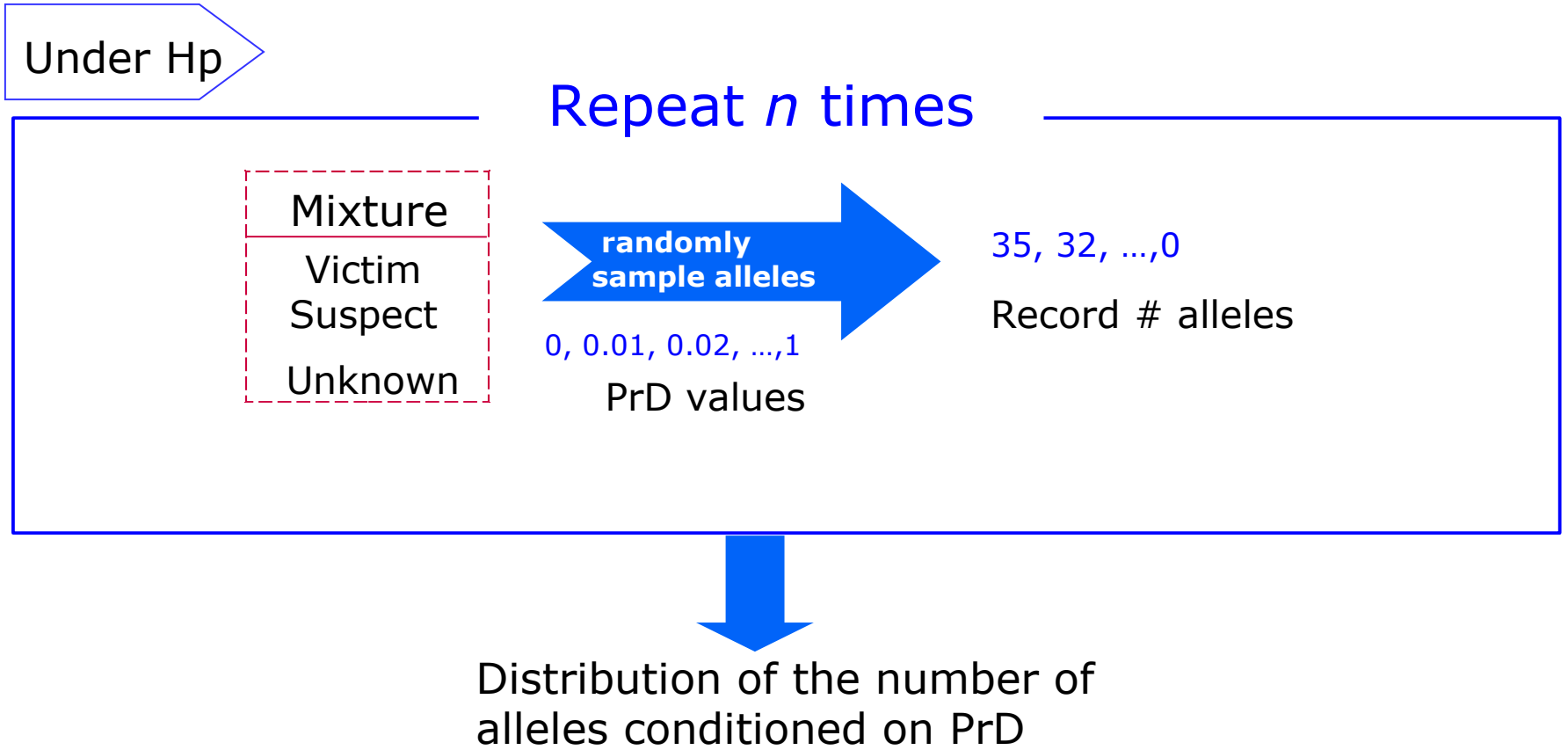
Important note!

The hypothesized contributors change under Hp and under Hd:

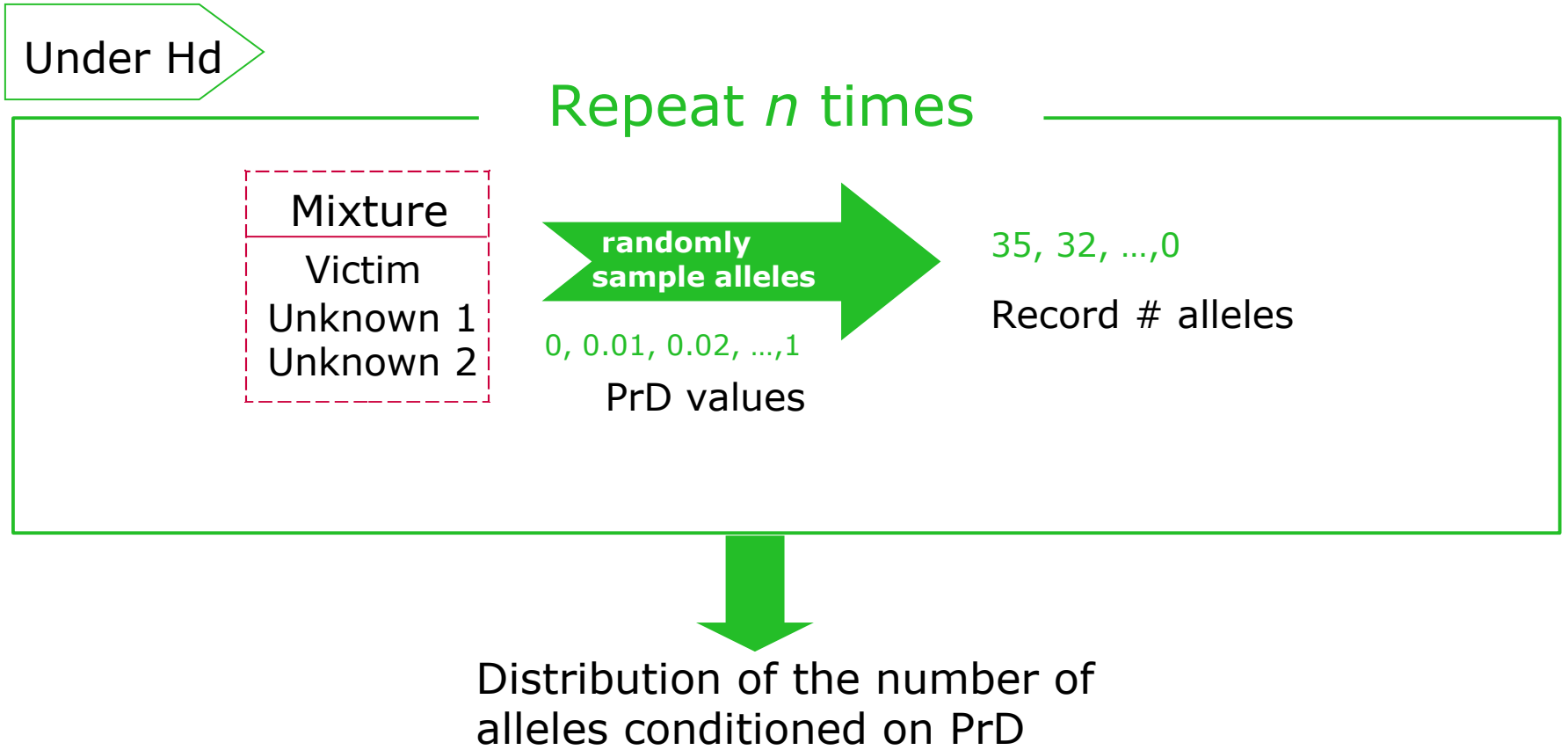
- Derive distribution of the numbers of alleles under Hp and under Hd separately

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

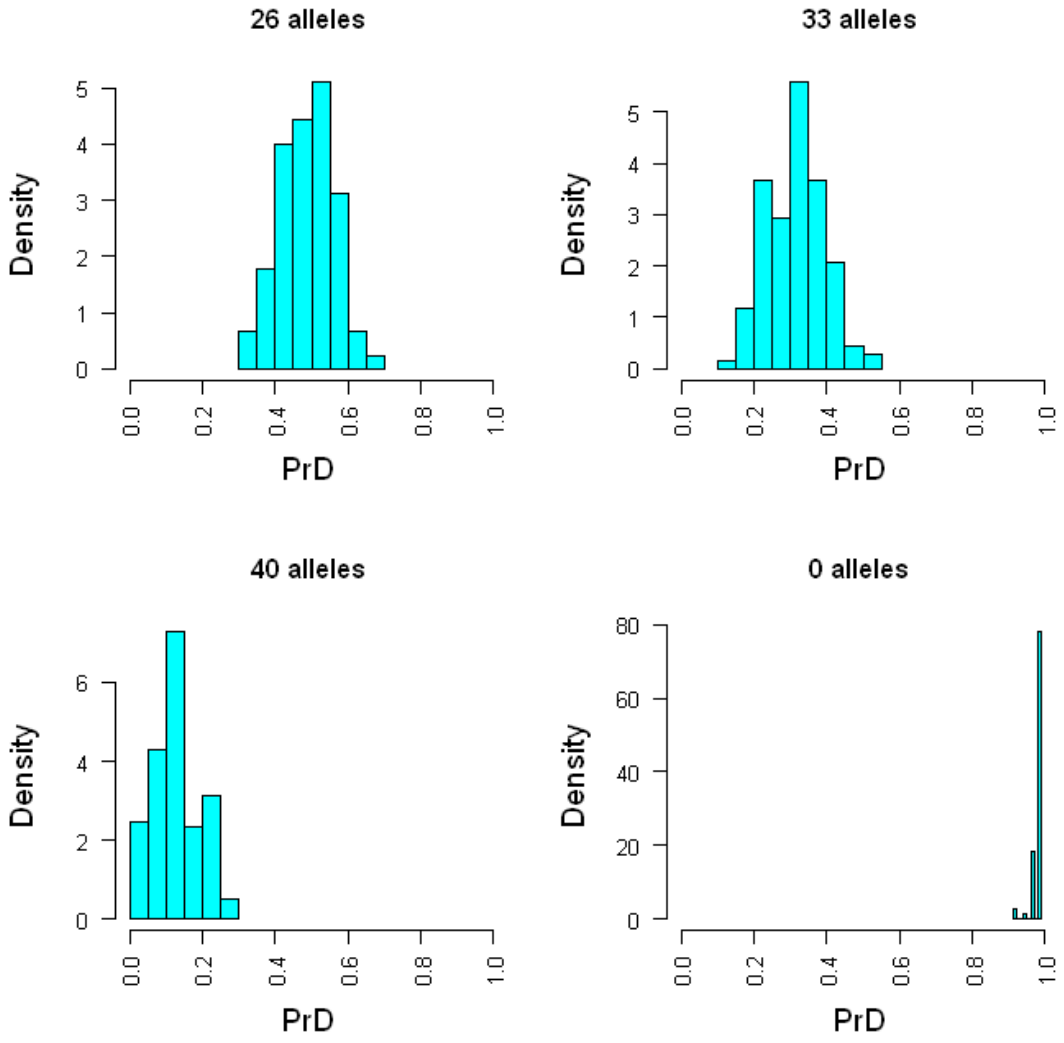
Monte-Carlo simulation procedure



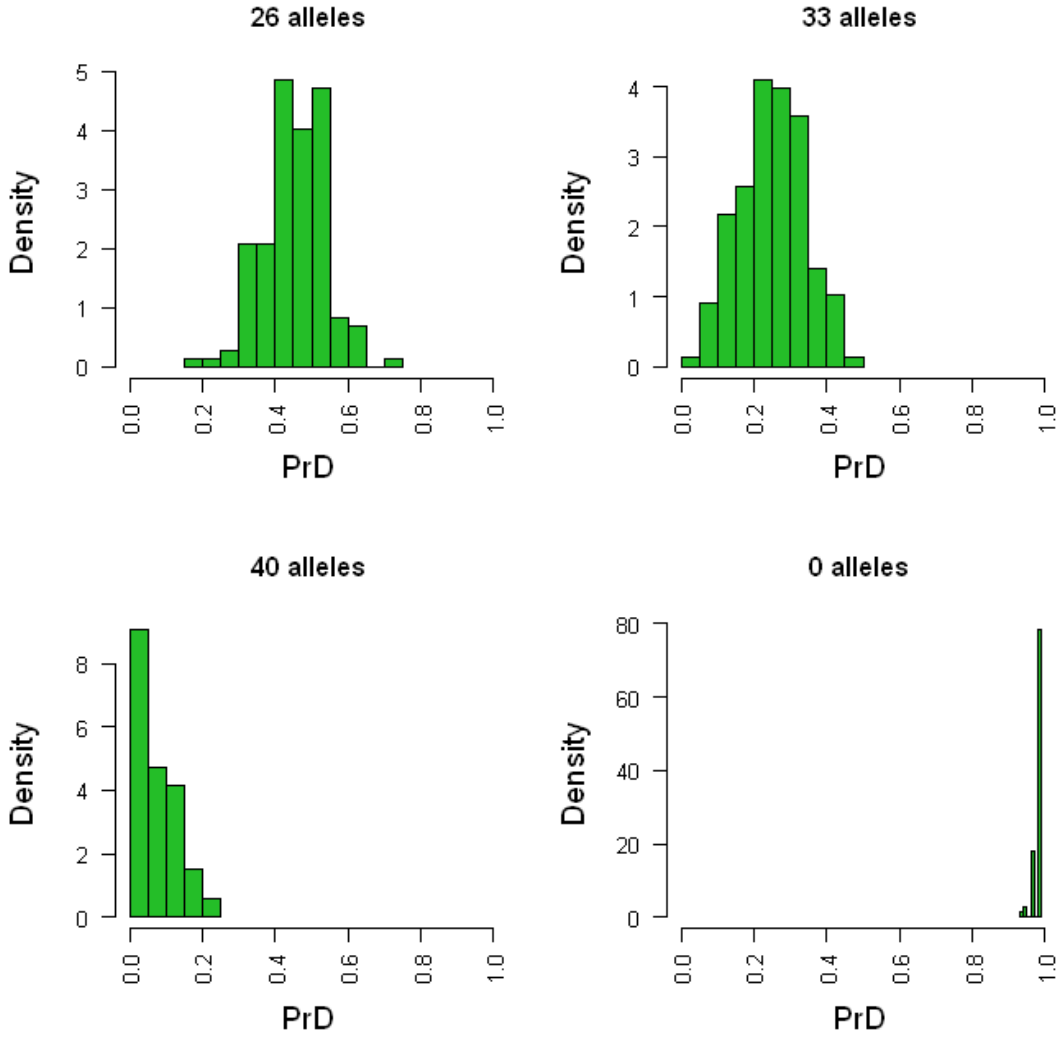
Monte-Carlo simulation procedure



Under Hp


26 alleles
33 alleles
40 alleles
0 alleles
Density
PrD
0.0
0.2
0.4
0.6
0.8
1.0

Under Hd

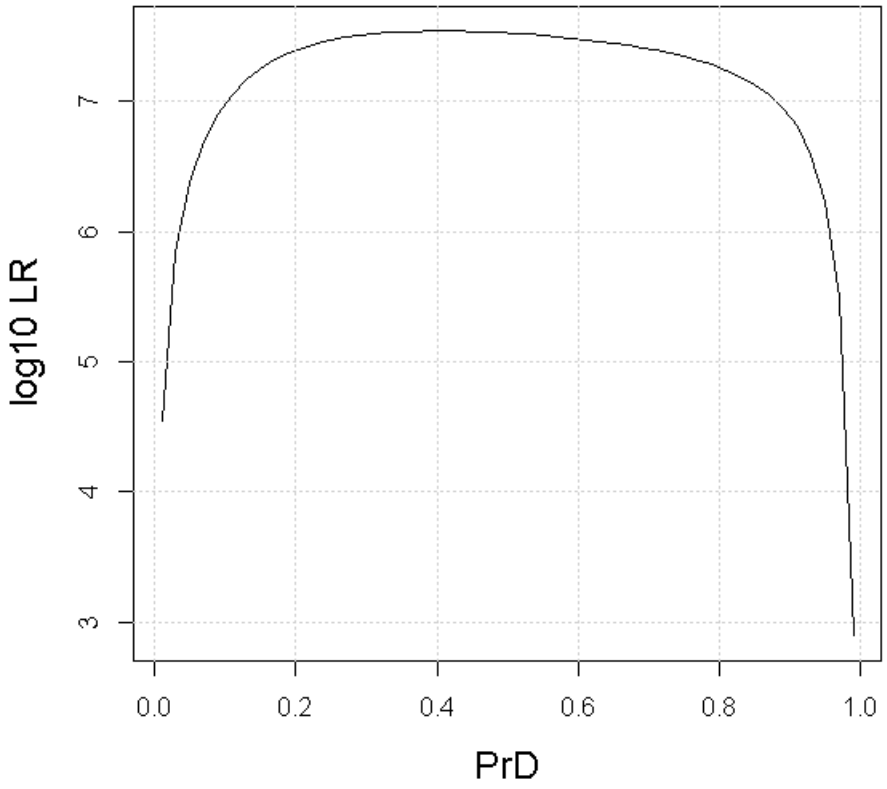

26 alleles
Density
PrD
33 alleles
Density
PrD
40 alleles
Density
PrD
0 alleles
Density
PrD

5% - 95% percentiles of the distributions

Under Hp	
5%	0.19
95%	0.45

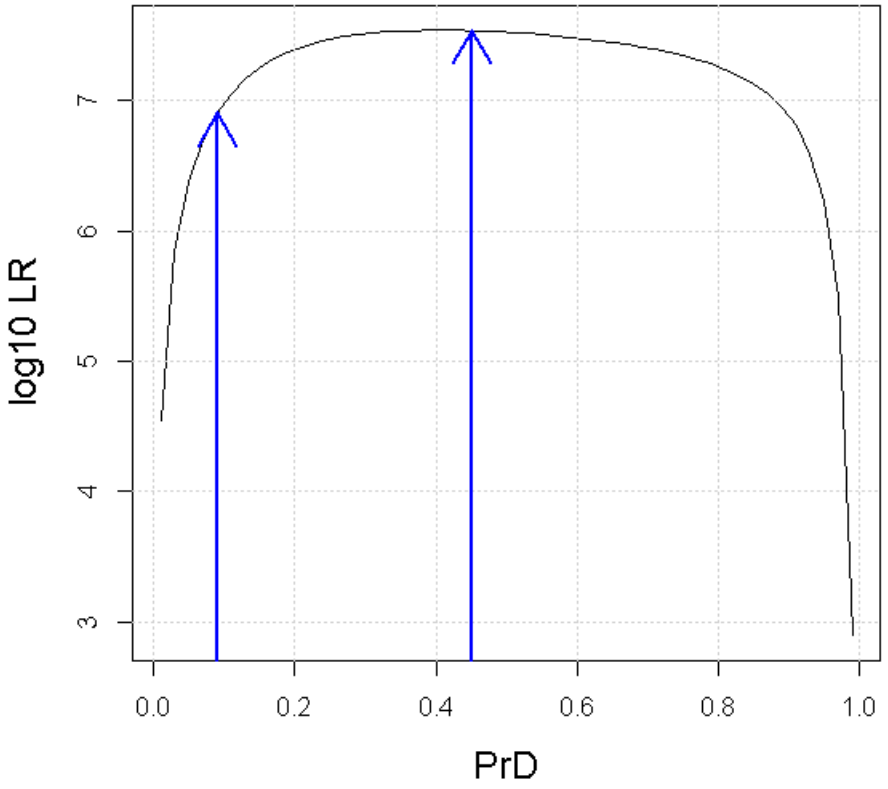
Under Hd	
5%	0.09
95%	0.41

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

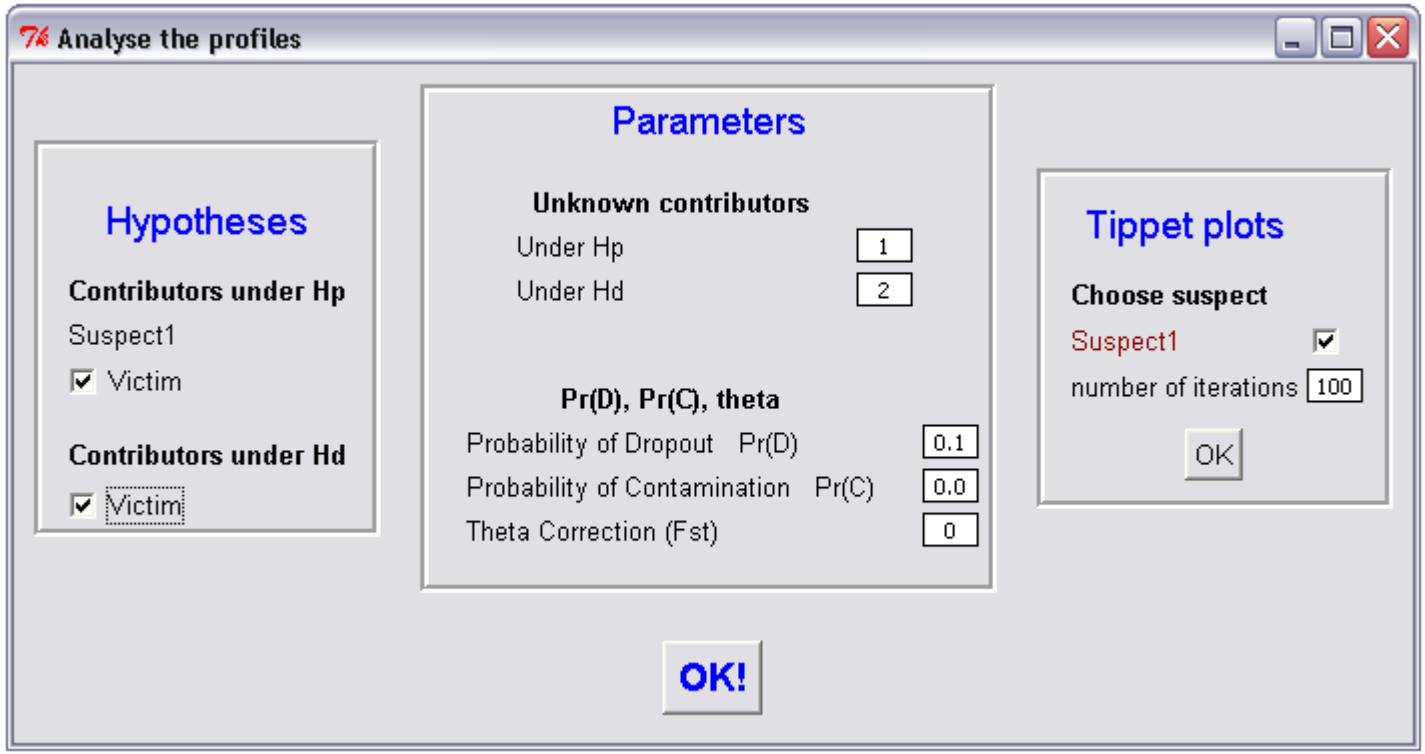
- qualitative data only

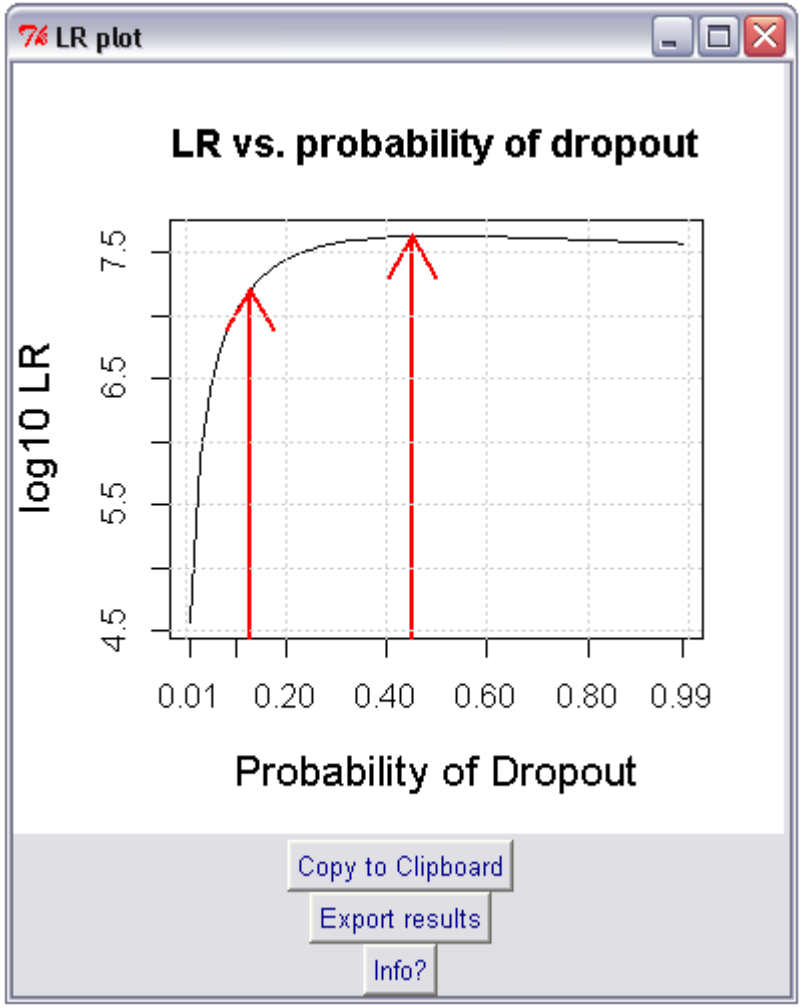
- peaks slightly under threshold not taken into account

Assess the uncertainty the data!

LRmix module

Available in Forensim 3.1




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?