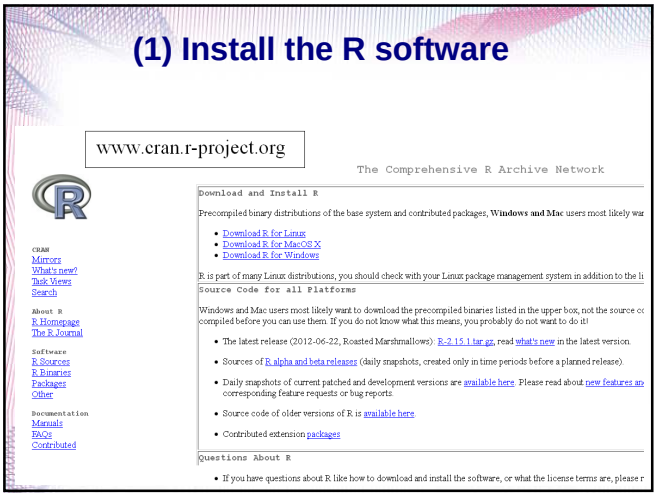
Introduction to the LRmix program of the Forensim R package

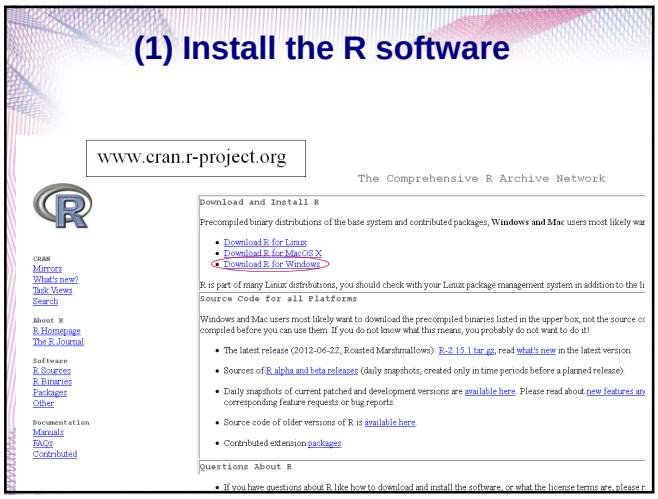
Hinda Haned
Peter Gill

For news updates subscribe to:
forensimnews@gmail.com

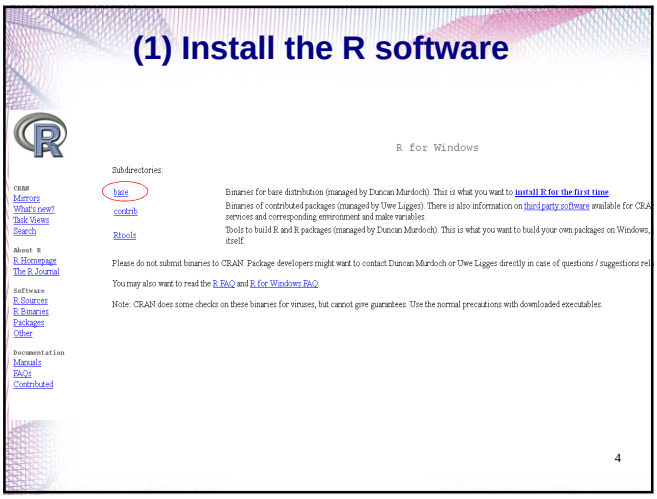
(1) Install the R software



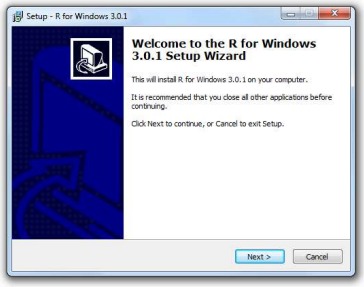
(1) Install the R software



(1) Install the R software



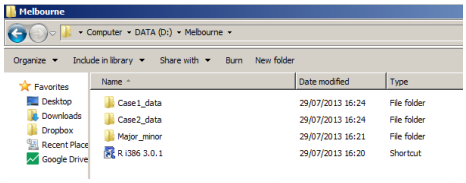
- An executable file will be downloaded automatically.
- R.3.0.1.exe
- Simply click and follow the instructions!



Press 'next' until...

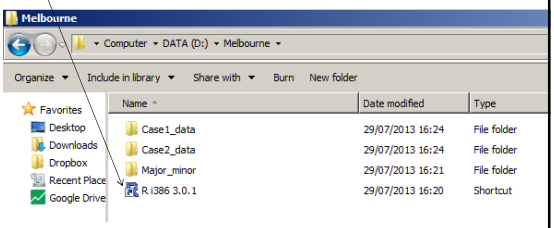
Prepare your working folder first (make sure this is set up before the ISFG workshop)

- You have been sent some data-sets in folders – place these into a working folder on your computer
- And place a short cut to R in the same folder (you can drag the R icon from your desktop)

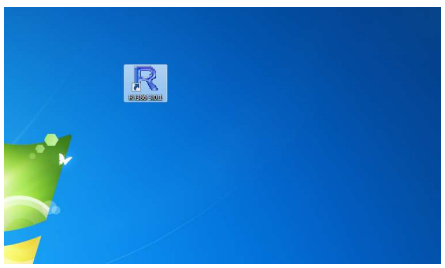


You are ready to launch R

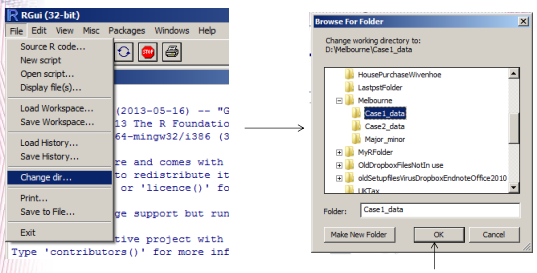
Double-Click blue icon.



Simply click on blue icon to launch R



Set directory to your folder

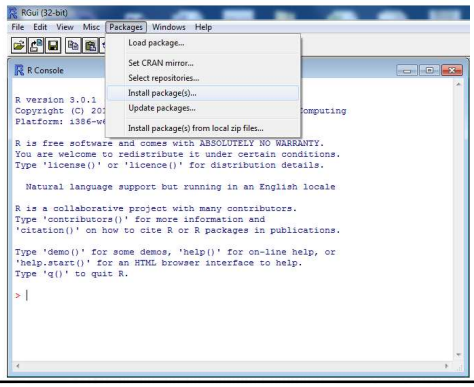


Press OK to set directory

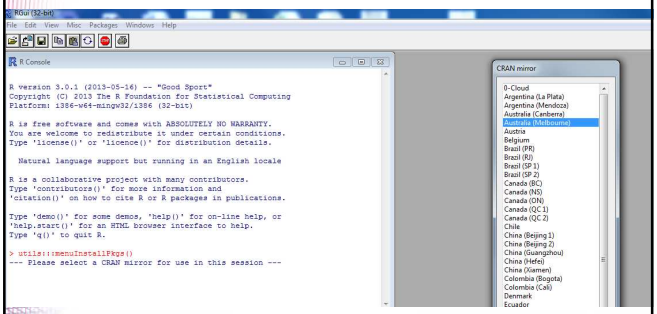
(2) Install the Forensim package

- Option 1: install the package directly from the R environment (Internet connection) - please follow this option now.
  - Also download LRmix tutorial from: http://forensim.r-forge.r-project.org/misc/LRmix.pdf
- Option 2: Install the package manually (no Internet connexion)
  - Refer to LRmix tutorial online:

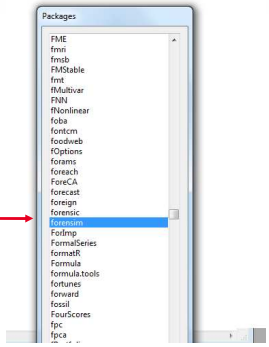
(2) Install the Forensim package



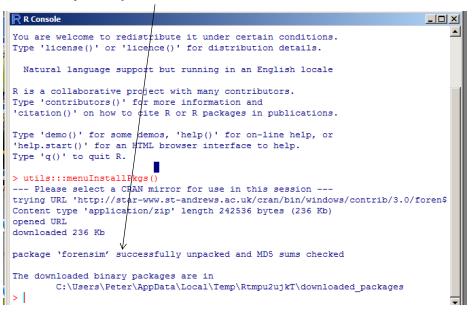
Choose mirror in Melbourne (or wherever you happen To be at the time)



Choose package forensim



Your screen should look something like this
Make sure you have a message:" 'forensim' successfully unpacked"

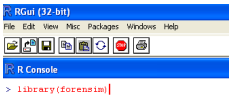


- Please try to get this far, and make sure you bring a laptop with the R program and files preloaded as described in the previous slides.
- This will save us a lot of time if you can do this.
- if you have a problem up to here, please contact me for advice: peterd.gill@gmail.com
- For those who are interested, you may wish to attempt to start an analysis of the first case
- Continue to the next slide to do this

(3) Load the Forensim library

Type the following code in the R console:

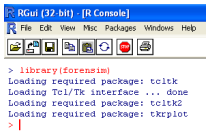
library(forensim)



(3) Load the Forensim library

Type the following code in the R console:

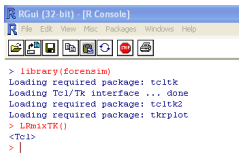
library(forensim)



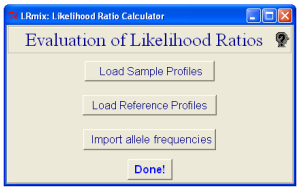
(4) Start LRmix

Type the following code in the R console:

library(forensim)
LRmixTK()



Main LRmix interface



Input files in LRmix

n.b. The data files are already in your folder

Type 1: CSV files, they are comma separated files (‘,’), and the decimal separator is the dot (‘.’)

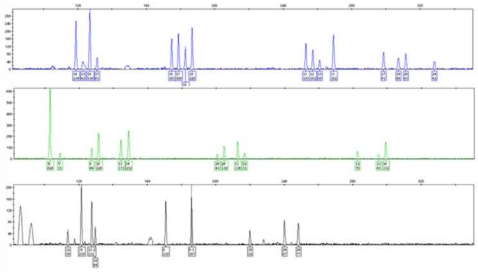
Type 2: tab separated files, they are tab separated (‘\t’, e.g. Excel), and the dot(‘.’) is the decimal separator

Never use spaces in your column-names, or in the sample-names (epg, or references)

A case example

- The crime-stain is from an epithelial swab taken from the female victim
- There are two suspects accused of sexual assault, S_1 and S_2 respectively; both deny the offence.
- This epg is classified as a low template of three or more individuals since there are multiple alleles per locus that fall within the criterion of the low template zone (between the LDT and the stochastic threshold (T))– we expect dropout may occur, but the profiles appear to be well represented.

Epg



Step 2: List the alleles with informative formatting

	Crime-stain alleles								
Marker	Allele1	Allele2	Allele3	Allele4	S1	S1	S2	S2	Unique alleles
AMEL	X	Y			X	Y	X	Y	2
D3S1358	14	16	17	(15)	16	17	15	17	4
VWA	16	17	18	19	16	18	18	19	4
D16S539	11	12	13	15	12	13	12	12	4
D2S1338	17	19	20	(24)	19	20	17	18	4
D8S1179	9	10	13	14	9	13	13	13	4
D21S11	29	31	32		28	32	30	30	5
D18S51	12	16	(15)		12	15	12	20	4
D19S433	12	14	15.2	16	12	16	12	15	5
TH01	6	9.3			6	9.3	6	9.3	2
FGA	19	24	26		19	21	20	21	5

Key:
Alleles that are shared between victim and S_1 or S_2 (green background).
Alleles that are found in the crime stain and not observed in any known individual (blue background, not applicable in this case).
Alleles that are below the detection threshold but appear to be distinct (bracketed).
Alleles that are found in the crime stain that match a known individual under *Hd* (victim) (red typeface).

Step 3: Establish the minimum number of contributors for the 'preliminary' propositions

a) The swab is from a victim (*V*). There are two suspects (S_1,S_2) under *Hp*,

b) In this example, some loci have 5 unique alleles across sets hence there is a minimum of three individuals present under *Hp*.

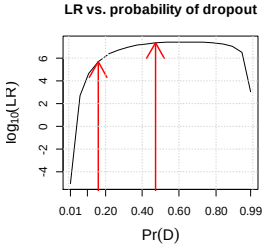
c) A similar calculation can be made under *Hd* where the sets of genotypes formed by S_1,S_2 are not used, but in our rationale, it is convenient to anchor the minimum number of contributors on *Hp* and to assume equivalence (this is revisited later in the procedure).

d) Consequently, the preliminary propositions are formulated as *Hp*=*V*,S_1,S_2 and *Hd*=*V*,*U*,*U*

Step 4: LRmix analysis

- *Hp*=*V*,S_1,S_2 and *Hd*=*V*,*U*,*U*
- The $\log_{10}(LR_{min})$= 5.66 is derived for a drop-out probability *Pr*(*D*)=0.16.
- Pr(D) value is in fact the 5 percentile calculated from an empirical distribution of the drop-out probability conditioned on the expected number of alleles observed relative to the genotype of the hypothesised contributors, the procedure is described by Haned et al (FISG 2012)

Sensitivity plot



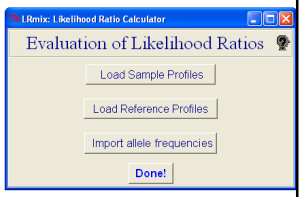
Main LRmix interface

Now we show how to:

(1) Load the crime-sample profile

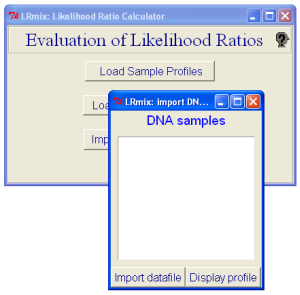
(2) Load the references (suspect/victim)

(3) Load your allele frequencies

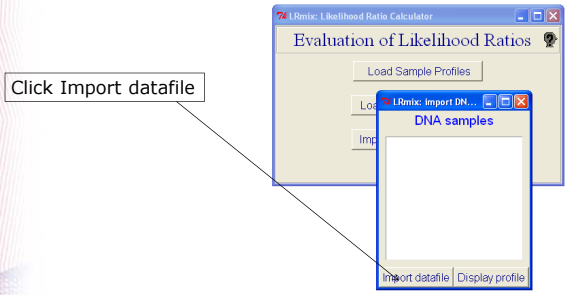


(1) Load the crime-sample profiles

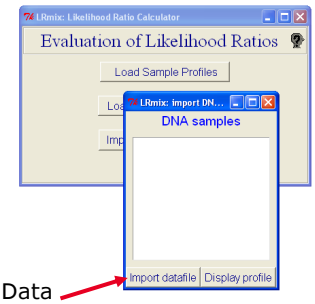
click: “Load Sample Profiles”



(1) Load the crime-sample profiles

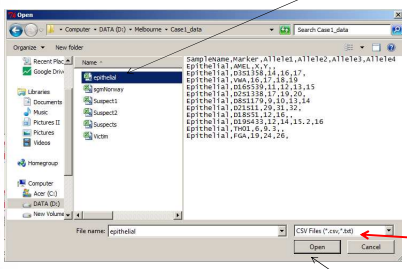


(1) Load the crime-sample profiles



Import Data

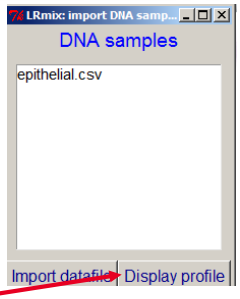
GOTO your Melbourne Case1_data folder: choose epithelial



Make sure this is set to CSV Files

Then click ‘Open’

Display the crime-sample profile

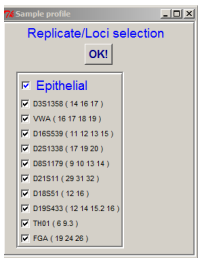


Now click ‘Display profile’,

To make sure the data are OK

If everything looks good, press OK!

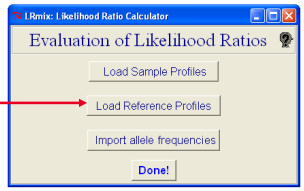
- You can select loci if you want
- But leave intact for this exercise

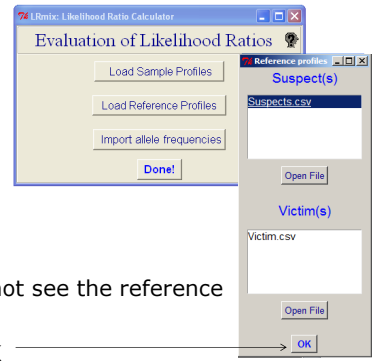


(2) Load reference profiles from your folder

suspects

victim





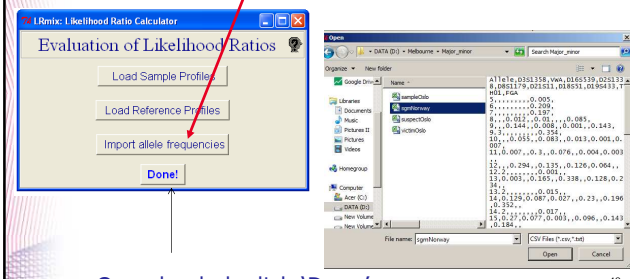
You cannot see the reference profiles

Press OK

- You cannot see the reference profiles
- The program will automatically select the loci you chose in step(1)
- If there are loci in the epg that are not given in the reference profile, the program will give an error message

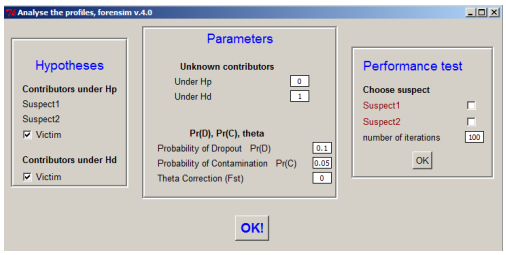
(3) Import the allele frequencies

sgmNorway



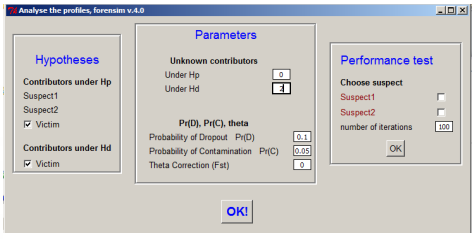
Once loaded, click 'Done'

Now you should see this



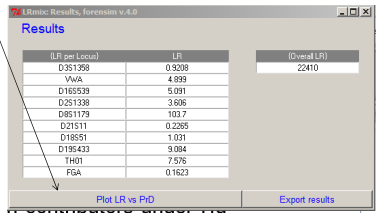
Alter the parameters

- 2 unknown contributors under Hd
- Click OK

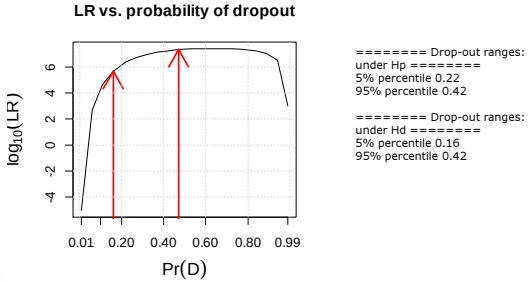


Results Table

- Now Carry out sensitivity analysis – click on button



Result of sensitivity analysis



The red arrows delineate the reasonable range for Pr(D). The LR≈10^6 .

Case evaluation

- So far we have only done a partial evaluation
- Think about how you would further evaluate this case?
- Are the propositions reasonable?
- Would you like to evaluate any other propositions?
- What would a final statement look like?

Recap (with further explanation)

Why exploratory?

- The purpose is not to give a 'black-box' answer because there is no definitive answer
- All of the answers are conditional hence the function of the 'expert' is to explore the various possibilities, on behalf of the prosecution and defence.
- Some generalisations are possible
- The 'process' used to interpret complex DNA profiles is provided in this talk
- Consider a minor/minor(s) contributors in the following epg. We could regard this as a typical LTDNA profile

Step 1: examine the epg

- And Consider the case circumstances
- Is it a mixture?

EPG

Case circumstances:

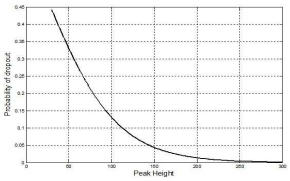
- Epithelial swab from female victim (V)
- Sexual assault with two suspects under Hp (S1, S2)

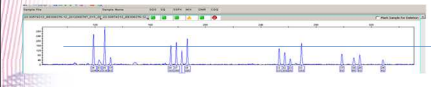
Step 1

- What kind of mixture is it?
- Choose from following:
  - Major/minor?
  - Even?
- Do we expect drop-out?
  - (compare with logistic regression)

A typical low template profile showing PrD range relative to thresholds

Check the peak heights against logistic regression to work out if drop-out is expected





Stochastic T: PrD≈0

LOD: PrD≈0.35

Change in philosophy

- With the old methods we had to 'filter' alleles and there were many restrictions about the kind of analysis that could be undertaken
- The new method can evaluate profiles without filtering alleles and are not restricted by numbers of contributors etc.
- Consequently, we are able to devise simple rules that can be followed to produce an LR.
- The questions shift towards "what are the propositions that should be considered"
- The role of the RO now becomes a facilitator of the court going discussion by following a logical process

Step 2: Make a table of alleles in the case-stain and the known contributors

- A format is suggested in the next slide
- Note that the procedure here differs from the Clayton guidelines since we must condition the hypotheses using all the evidence under Hp – so this means that the reference samples are evaluated concurrently with the crime-stain
- However, all alleles are included so long as they are above LOD

Step 2: List the alleles with informative formatting

	Crime-stain alleles								
Marker	Allele1	Allele2	Allele3	Allele4	S1	S1	S2	S2	Unique alleles
AMEL	X	Y			X	Y	X	Y	2
D3S1358	14	16	17	(15)	16	17	15	17	4
VWA	16	17	18	19	16	18	18	19	4
D16S539	11	12	13	15	12	13	12	12	4
D2S1338	17	19	20	(24)	19	20	17	18	4
D8S1179	9	10	13	14	9	13	13	13	4
D21S11	29	31	32		28	32	30	30	5
D18S51	12	16	(15)		12	15	12	20	4
D19S433	12	14	15.2	16	12	16	12	15	5
TH01	6	9.3			6	9.3	6	9.3	2
FGA	19	24	26		19	21	20	21	5

Key:
Alleles that are shared between victim and S_1 or S_2 (green background).
Alleles that are found in the crime stain and not observed in any known individual (blue background, not applicable in this case).
Alleles that are below the detection threshold but appear to be distinct (bracketed).
Alleles that are found in the crime stain that match a known individual under *Hd* (victim) (red typeface).

Count the number of *unique alleles* in the 'set' in order to decide the number of contributors

Step 3: Establish the minimum number of contributors for the 'preliminary' propositions

a) The swab is from a victim (*V*). There are two suspects (S_1,S_2) under *Hp*,

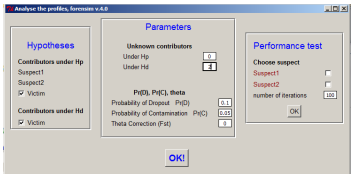
b) In this example, some loci have 5 unique alleles across sets hence there is a minimum of three individuals present under *Hp*.

c) A similar calculation can be made under *Hd* where the sets of genotypes formed by S_1,S_2 are not used, but in our rationale, it is convenient to anchor the minimum number of contributors on *Hp* and to assume equivalence (this is revisited later in the procedure).

d) Consequently, the preliminary propositions are formulated as *Hp*=*V*,S_1,S_2 and *Hd*=*V*,*U*,*U*

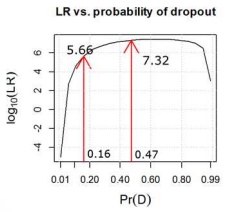
Step 4: Evaluate the first scenario

- The proposition under Hp is S1,S2,V
- The proposition under Hd is U1,U2,V
  - *Note we could also use U1,V under Hd – no need for Hd to agree on the same number of contributors*
  - (swab from female victim so this appears in Hp, Hd)



Sensitivity plot evaluation

- Plot the LR relative to all values of PrD
- Calculate lower and upper bounds in order to decide a reasonable range
- Report the lowest value (to be conservative)



We have got this far with our analysis

- Next we need to ask questions about whether the results themselves are robust?
- What sort of questions should you being asking?

Step 5: Case re-evaluation and simplification of the propositions

Although a probative LR favouring *Hp* has resulted from the preliminary analysis, this has incorporated both suspects S_1 and S_2 under *Hp*.

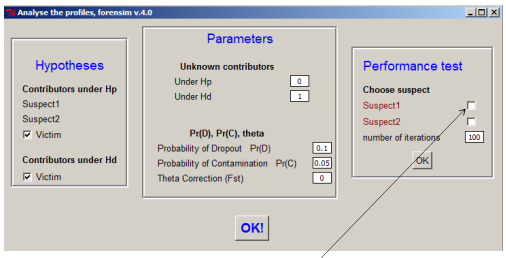
However, the likelihood ratio itself does not provide any indication about the relative *weighting* of the two contributions provided by S_1, S_2 to the actual LR result.

Consequently, the next step in the analysis is to *dissect* the propositions into their constituents in order to establish the weighting and to establish the consequent probative value of the evidence per contributor under *Hp*.

Step 5: Non-contributor test

- Why are we doing this?
- The process is *exploratory*
- So what will happen if we replace a suspect with a random man?
- We would expect the LR to be very low (an exclusion!!)
- Therefore, the non-contributor test is a measure of *robustness* and we consider this to be an important part of model *validation*

Run Test

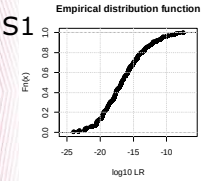


Click here – and click OK to start simulation

Comparison of non-contributor plots

There are two suspects – so we do two non-contributor plots – a) replace S1 with r.m. (x1000) and b) replace S2 with r.m. (x1000)

S1

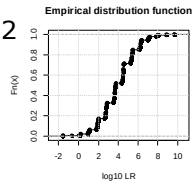


quantile" "value"
"min" "-24.0269"
"0.01" "-23.2479"
"0.05" "-21.4325"
"0.5" "-16.7792"
"0.95" "-10.5699"
"0.99" "-8.4826"
"max" "-7.4584"

b

Original LR=5.66

S2



"quantile" "value"
"min" "-1.591"
"0.01" "0.126"
"0.05" "1.0629"
"0.5" "3.7167"
"0.95" "7.0392"
"0.99" "7.9833"
"max" "9.6998"

6/2/12

Step 5:Summarise the results

- The calculated LR(log10)= 5.6
- The non-contributor plot for S1 can be summarised using the one percentile, the median and the 99 percentile (-23,-16,-8)
- The non-contributor plot for S2 can be summarised in the same way: (+0.1,+3.7,+7.9)
- This means that the model is insensitive to S2 because the same result can be achieved with random man!!

6/2/12

What does this mean?

- Beware complex propositions – the relative weightings of the S1,S2 'contributions' are not reflected in the likelihood ratio
- Therefore complex propositions must be simplified and qualified before they can be reported
- The non-contributor plot is a useful adjunct to verify the likelihood ratio (define limitations of the model) and also provides an additional way to think about the results (court-friendly)

6/2/12

Step 6: Simplify the propositions

- So far we don't have evidence for S2 under Hp
- So we need to think about different propositions in order to reevaluate the evidence
- There seems to be good evidence under Hp for S1

6/2/12

New table with S1

Marker	Allele1	Allele2	Allele3	Allele4	S1	S1	No of unique alleles
AMEL	X	Y			X	Y	2
D3S1358	14	16	17	(15)	16	17	3
VWA	16	17	18	19	16	18	4
D16S539	11	12	13	15	12	13	4
D2S1338	17	19	20	(24)	19	20	4
D8S1179	9	10	13	14	9	13	4
D21S11	29	31	32		28	32	4
D18S51	12	16	(15)		12	15	3
D19S433	12	14	15.2	16	12	16	4
TH01	6	9.3			6	9.3	2
FGA	19	24	26		19	21	4

Analysis

Visual examination of the evidence (table 2) revealed that S_1 has more matching alleles than S_2; furthermore the crime stain could be explained under *Hp* if it was a simple mixture of *V* and S_1 (with three *dropped-out* alleles).

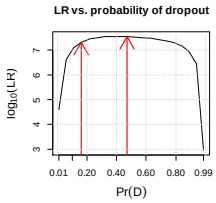
Individual S_2 is not required at all in the analysis, since there are no missing alleles observed in the crime stain (*Hp*=*V*,S_1).

Although the number of unique alleles reduces the number of contributors to two, in order to be consistent, three contributors are evaluated and the propositions are simplified to: *Hp*=S_1,*V*,*U* and *Hd*=*V*,*U*,*U*.

(*note the LR is much larger if two contributors are analysed under Hp and Hd – data not shown, hence the choice of three contributors is demonstrably conservative*).

Hp=S1,V,U and Hd=V,U,U

The new $\log_{10}(LR_{min})$= 7.32; $Pr(D_{min})$=0.16

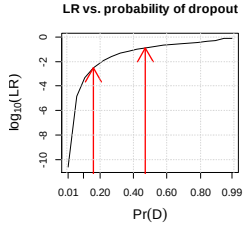


Now determine the S2 effect

Hp=S_2,*V*,*U*; *Hd*=*V*,*U*,*U*.

$Pr(D_{min})$=0.16

$\log_{10}(LR_{min})$= -2.6 which is clearly 'exclusionary'



Step 7: Non-contributor performance (Np) tests summary

Np tests can be used to support the conclusion that evidence supporting S_1 is 'inclusionary' whereas evidence supporting S_2 is 'exclusionary'

	Three person mixture			Non-contributor performance
Hp	*Hd*	Random man substituted	$\log_{10}$(LR)	percentiles
S_1,S_2,*V*	*V*,*U*,*U*	S_1	5.5	(-21,-15,-7)
S_1,S_2,*V*	*V*,*U*,*U*	S_2	5.5	(+0.17,+4.2,+8.2)
S_1,*V*,*U*	*V*,*U*,*U*	S_1	7.2	(-10,-5,+0.14)
S_2,*V*,*U*	*V*,*U*,*U*	S_2	-3	(-10,-5,+0.14)

Principles to follow when evaluating complex sets of hypotheses

Conditioning rules (a)

- Conditioning hypotheses are defined by the casework circumstances
- Remember to evaluate the hypotheses based on the number of contributors derived from the unique number of alleles in the 'set' observed in the epg: i.e. the sum of alleles of known contributors and the sum of alleles of the crime-stain(s) under Hp (to maximise)
- Do not use the *drop-in* principle to 'explain away' additional contributors

Conditioning rules (b)

- If there are two or more 'suspects' under Hp then the hypothesis should be simplified i.e. evaluate: S1,V,U in addition to S1,S2,V
- It is important to explore the likelihood ratio by use of the non-contributor plot.
- In the S1,S2,V example we show that the LR is very insensitive to S2 (random man still gives a high LR)

Summary of results

- Case circumstances
  - Both S1 and S2 are suspects of sexual assault and a sample is taken from the victim. We condition on the victim under Hd
  - No evidence for S2 in the crime stain [even though a three person evaluation with S1,S2 under Hp gives a high LR= log10(5)
  - Advice: Simplify propositions if there are two suspects - always evaluate them separately.

LRmix practical session - case 2

Peter Gill
Hinda Haned

Case details

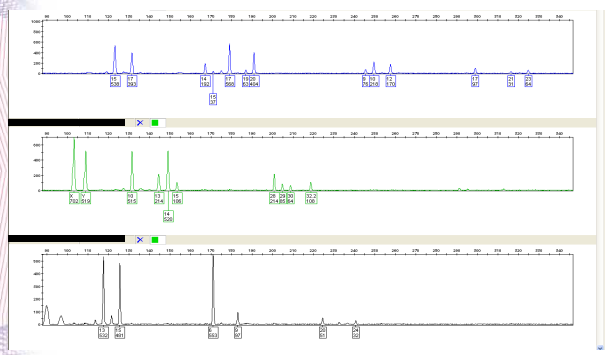
- Murder case with a male victim killed in a fight
- There are five suspects that are apprehended by police and DNA profiled
- Is there evidence of that any of the suspects' DNA is at the crime-scene?

Crime-Stain R1

Recovered from victim's ankle, analysed for (presumed) epithelial cells.

Note: there were 5 separate crime stains in this case, but for simplicity we consider just one of these

EPG (case stain R1) SGM

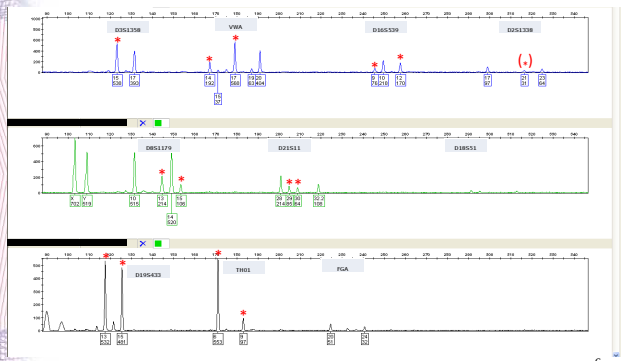


Profiles (the LRmix input)

SampleName	Marker	Allele1	Allele2	Allele3	Allele4	Allele5
R1	AMEL	X	Y			
R1	D3S1358	15	17			
R1	VWA	14	17	19	20	(15)
R1	D16S539	9	10	12		
R1	D2S1338	17	23			
R1	D8S1179	10	13	14	15	
R1	D21S11	28	29	30	32.2	(21)
R1	D18S51					
R1	D19S433	13	15			
R1	TH01	6	9			
R1	FGA	20				

Note only >50rfu alleles recorded and victim alleles highlighted in red
() alleles below 50rfus but distinct on epg
15 victim's alleles

EPG showing victim's alleles



List of suspect genotypes (note there are five suspects in this case)

Marker	S1	S1	S2	S2	S3	S3	S4	S4	S5	S5
AMEL	X	Y	X	Y	X	X	X	X	X	Y
D3S1358	17	18	15	16	16	18	17	18	15	17
VWA	16	19	16	17	15	18	14	17	17	20
D16S539	10	13	12	12	9	11	9	12	10	12
D2S1338	19	23	18	21	17	19	20	25	17	23
D8S1179	13	15	11	13	10	13	12	13	10	14
D21S11	28	30	30	32.2	28	29	31	31	28	32.2
D18S51	15	15	14	18	14	17	14	15	14	19
D19S433	14	15	14	14	14	16	14	15.2	13	15
TH01	9	9	8	9	7	9	7	7	6	6
FGA	21	21	24	24	22	24	20	20	20	24

How many contributors?

- Examination of the epg suggests two contributors as best option
- But bear in mind that alleles are missing, and there could be an additional contributor to consider

Care needed to incorporate the conditioning profile into the estimate of the number of contributors

The epg may suggest two contributors, but we must take into account the 'conditioning' profile(s) in order to determine the number of contributors if Hp is true.

So in our 'first round' assessment we use:

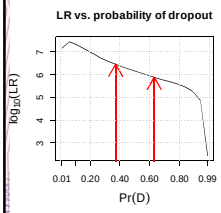
Hp: S_n + V + U

Hd: V + U +U

Hypotheses (1): three-person mixture

Hp: S_5 + V + U

Hd: V + U +U



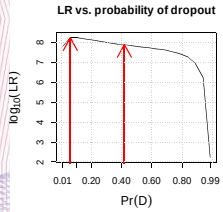
========== User parameters
Drop-in value: 0.05
Theta value: 0
======== Drop-out ranges: under Hp ========
5% percentile 0.37
95% percentile 0.63

======== Drop-out ranges: under Hd ========
5% percentile 0.37
95% percentile 0.63

==== Likelihoods & likelihood ratios =====
Pr(D) log10(LR)
0.37 6.45
0.63 5.88

Hypotheses (2): two-person mixture

- If S5 is a contributor under Hp then we can re-assess under the assumption of two-persons
- A single drop-in event is encountered in locus VWA (allele 19)
- Hp=S5,V / Hd=S5,U



========== User parameters
Drop-in value: 0.05

======== Drop-out ranges: under Hp 5% percentile 0.16
95% percentile 0.42

======== Drop-out ranges: under Hd
5% percentile 0.062
95% percentile 0.42

==== Likelihoods & likelihood ratios =====
Pr(D) l og10(LR)
0.062 8.22
0.16 8.17
0.42 7.87

Discussion on models

- Deciding the precise model to use is not straightforward and often multiple models can be used
- Number of contributors is not just a matter of observing the number of alleles in the epg. But is also dependent upon the conditioned profiles which usually include suspect and victim under Hp
- Do not use the drop-in principle as a convenience instead of invoking an additional contributor
- This is not what the parameter is designed for

Two or three contributors?

- Its best to think of the method we demonstrate as an imperfect model that generates a 'number' and we hope that this number is 'meaningful'.
- With LtDNA, stochastic effects increases the uncertainty of PrD
- We don't know (we will never know) which model is the best, all models are approximations.
- We do know that different models give different answers – so how can we deal with this issue?

Comparison of models

Recall that the current example gives:

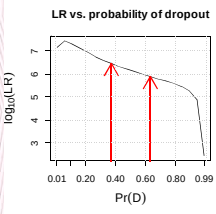
- LR= 10^5 (769,600, PrD=0.63, three contributors)
- LR= 10^7 (75240000 ,Pr(D)=0.42, two contributors)

⇨We don't follow principle that biggest number is the best as there would be a prosecution bias with this conclusion

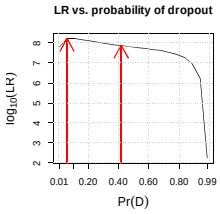
⇨Rather we ask – which model(s) is reasonable, given the case information

Sensitivity plots: Both models are reasonable so long as the PrD<0.9

Three person



Two person

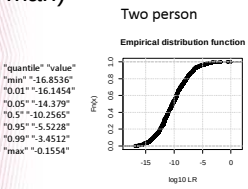


Note: same data but more contributors must reduce the Pr(D)

Testing the model

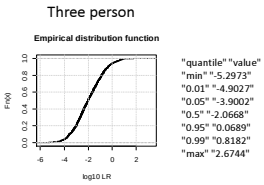
We now evaluate both models using non-contributor tests (replacing suspect with random man)

Two person



(-16,-10,-3) (see next slide for explanation)

Three person



(-5,-2,+1)

What does this mean?

The non-contributor plot can be conveniently summarized by three figures (*a*,*b*,*c*)

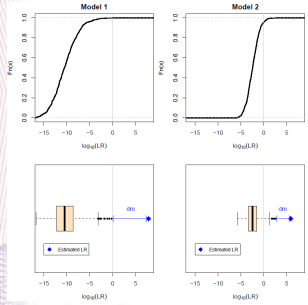
- *a*= $\log_{10}$(lower one percentile)
- *b*=$\log_{10}$ (median)
- *c*=$\log_{10}$ (upper one percentile)

So the two alternative models can be summarized as follows:

- Three persons $\log_{10}(LR)= 10^5$ (-5,-2,+1)
- Two persons $\log_{10}(LR)= 10^7$ (-16,-14,-3)

Performance of models

Defined by the discriminatory metric, distinguishing between Random man model and the estimated LR



We are interested to confirm that random man gives an answer that is much less than the observed likelihood ratio (the 'distance' is given by the discriminatory metric – but this is not used to define the 'best model').

Court reporting

- For complex models there is no right or wrong answer
- There is more than one choice.
- Also different models (e.g. TrueAllele) will give different answers, given the same conditioning and this is because the modeling assumptions are different.

Court reporting

- How sure can we be that the LR provided is meaningful?
- Random man simulation provides the necessary assurance
- Court report would follow: (next slide)

Statement

I have evaluated the proposition that Mr X is a contributor to the crime stain Y compared to the alternative proposition that Mr X is not a contributor to crime stain Y using the conditions defined in the LRmix model. These conditions are as follows:

a) Mr X and the victim are both contributors to the sample
b) An unknown person and the victim are both contributors to the sample

The evidence is 75million times more likely if the first proposition (a) is true, compared to the alternative described by (b).

Optional:
[This figure can be qualified with a test of robustness. To do this we replace Mr X with a random unrelated individual and we repeat the measurement of the likelihood ratio. We do this a total of 1000 times, with a different random individual each time.
When this was carried out the greatest likelihood ratio observed was of the order of 0. 001.

Summary of the principles

Case pre-assessment:

- Make a table of alleles
- Count the number of unique alleles to decide the minimum number of contributors across the set
- Formulate a set of propositions

Evaluation of the strength of the evidence

- Evaluate the propositions
  - Determine the LR
  - Carry out Performance test to determine the robustness of the answer
- Re-evaluate the case and the propositions if necessary
- Report the case using suggested template

Analysis of a complex case using Exploratory Data Analysis (EDA), Part 2

Peter Gill and Hinda Haned

Why exploratory?

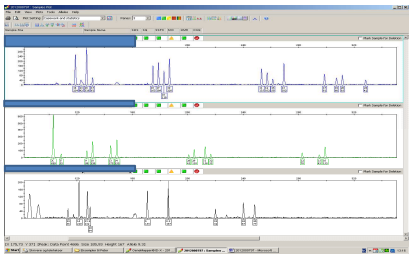
- The purpose is not to give a 'black-box' answer because there is no definitive answer
- All of the answers are conditional hence the function of the 'expert' is to explore the various possibilities, on behalf of the prosecution and defence.
- Some generalisations are possible
- The 'process' used to interpret complex DNA profiles is provided in this talk
- Consider a minor/minor(s) contributors in the following epg. We could regard this as a typical LTDNA profile

Step 1: examine the epg

- And Consider the case circumstances
- Is it a mixture?

EPG

Case circumstances:
- Epithelial swab from female victim (V)
- Sexual assault with two suspects under Hp (S1, S2)

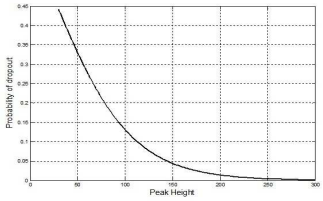


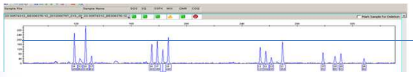
Step 2

- What kind of mixture is it?
- Choose from following:
  - Major/minor?
  - Even?
- Do we expect drop-out?
  - (compare with logistic regression)

A typical low template profile showing PrD range relative to thresholds

Check the peak heights against logistic regression to work out if drop-out is expected





Stochastic T: PrD≈0

LOD: PrD≈0.35

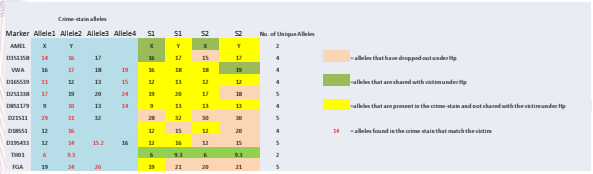
Change in philosophy

- With the old methods we had to 'filter' alleles and there were many restrictions about the kind of analysis that could be undertaken
- The new method can evaluate profiles without filtering alleles and are not restricted by numbers of contributors etc.
- Consequently, we are able to devise simple rules that can be followed to produce an LR.
- The questions shift towards "what are the propositions that should be considered"
- The role of the RO now becomes a facilitator of the court going discussion by following a logical process

Step 3: Make a table of alleles in the case-stain and the known contributors

- A format is suggested in the next slide
- Note that the procedure here differs from the Clayton guidelines since we must condition the hypotheses using all the evidence under Hp – so this means that the reference samples are evaluated concurrently with the crime-stain
- However, all alleles are included so long as they are above LOD

Case with two suspects and a victim



Crime-stain alleles

Marker	Allele1	Allele2	Allele3	Allele4	S1	S1	S2	S2	No. of Unique Alleles
AMEL	X	Y			X	Y	X	Y	2
D3S1358	14	16	17		16	17	15	17	4
VWA	16	17	18	19	16	18	18	19	4
D16S539	11	12	13	15	12	13	12	12	4
D2S1338	17	19	20	24	19	20	17	18	5
D8S1179	9	10	13	14	9	13	13	13	4
D21S11	29	31	32		28	32	30	30	5
D18S51	12	16			12	15	12	20	4
D19S433	12	14	15.2	16	12	16	12	15	5
TH01	6	9.3			6	9.3	6	9.3	2
FGA	19	24	26		19	21	20	21	5

= alleles that have dropped out under Hp

= alleles that are shared with victim under Hp

= alleles that are present in the crime-stain and not shared with the victim under Hp

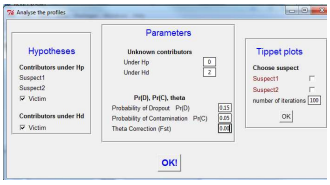
14 = alleles found in the crime stain that match the victim

Notes:

- Count the number of *unique alleles* in the 'set' in order to decide the number of contributors
- Case circumstances require consideration of S1 and S2 (three person mixture)
- The number of contributors is decided from the set of alleles (Hp)
- However, the evidence of the epg suggests two-person mixture is reasonable too (Hd).
- The mixture is low level and dropout is expected.
- But all victim alleles are observed in the mixture
- It is reasonable to condition on the victim under Hp and Hd (since this is an external swab taken from victim)
- There are no alleles in the crime-stain that are not found in S1 or S2 (i.e. no drop-in to consider under Hp)

Step 4: Evaluate the first scenario

- The proposition under Hp is S1,S2,V
- The proposition under Hd is U1,U2,V
  - *Note we could also use U1,V under Hd – no need for Hd to agree on the same number of contributors*
  - (swab from female victim so this appears in Hp, Hd)

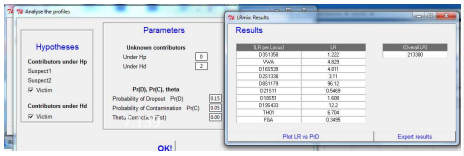


Step 5: Sensitivity plot evaluation

- Plot the LR relative to all values of PrD
- Calculate lower and upper bounds in order to decide a reasonable range
- Report the lowest value (to be conservative)

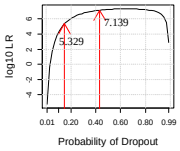
Evaluation(a) (Suspect1 and Suspect2)

- Hp=S1,S2,V
- Hd=V,U,U



LR vs. probability of dropout

"--------- drop-out ranges: under Hp------"
"5% percentile 0.21"
"95% percentile 0.43"
"--------- drop-out ranges: under Hd------"
"5% percentile 0.15"
"95% percentile 0.39"



LR=5.329

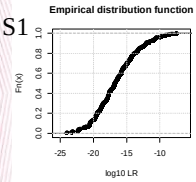
Step 6: Non-contributor plot evaluation

- Why are we doing this?
- The process is *exploratory*
- So what will happen if we replace a suspect with a random man?
- We would expect the LR to be very low (an exclusion!!)
- Therefore, the non-contributor test is a measure of *robustness* and we consider this to be an important part of model *validation*

Comparison of non-contributor plots

There are two suspects – so we do two non-contributor plots – a) replace S1 with r.m. (x1000) and b) replace S2 with r.m. (x1000)

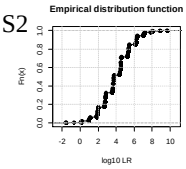
S1



"quantile" "value"
"min" "-24.0269"
"0.01" "-23.2479"
"0.05" "-21.4325"
"0.5" "-16.7792"
"0.95" "-10.5699"
"0.99" "-8.4826"
"max" "-7.4584"

b

S2



"quantile" "value"
"min" "-1.591"
"0.01" "0.126"
"0.05" "1.0629"
"0.5" "3.7167"
"0.95" "7.0392"
"0.99" "7.9833"
"max" "9.6998"

LR=5.329

Step 7:Summarise the results

- The calculated LR= 5.329
- The Tippet plot for S1 can be summarized using the one percentile, the median and the 99 percentile (-23,-16,-8)
- The Tippet plot for S2 can be summarised in the same way: (+0.1,+3.7,+7.9)
- This means that the model is insensitive to S2 because the same result can be achieved with random man

What does this mean?

- Beware complex propositions – the relative weightings of the S1,S2 'contributions' are not reflected in the likelihood ratio
- Therefore complex propositions must be simplified and qualified before they can be reported
- The non-contributor plot is a useful adjunct to verify the likelihood ratio (define limitations of the model) and also provides an additional way to think about the results (court-friendly)

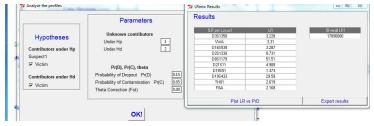
Step 8: Simplify the propositions

- So far we don't have evidence for S2 under Hp
- So we need to think about different propositions in order to reevaluate the evidence
- There seems to be good evidence under Hp for S1

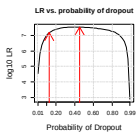
Evaluation(b) (Suspect 1)

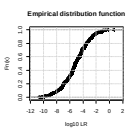
Scenario

- Hp=S1,V,U
- Hd=V,U,U



"drop-out ranges: under Hp------"
"5% percentile 0.19"
"95% percentile 0.45"
"drop-out ranges: under Hd------"
"5% percentile 0.13"
"95% percentile 0.41"



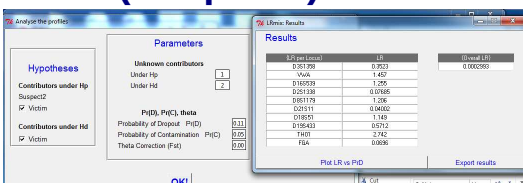


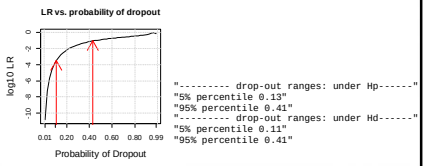
"quantile" "value"
"min" "-10.6522"
"0.01" "-10.0008"
"0.05" "-8.8262"
"0.5" "-5.199"
"0.95" "-1.8787"
"0.99" "-0.9235"
"max" "0.6591"

Reported LR=7.2(-10,-5,-0.9)

Evaluation(b) (Suspect2)

Hp=S2,V,U
Hd=V,U,U





LR= -3.5 is clearly exclusionary
Note Tippet is same as S1 previously shown

Summary (second round of analysis)

- Evaluation of S1,V,U under Hp gives Reported LR=7.2(-10,-5,-0.9)

6/2/12

Step 9: Evaluate the results and decide if new propositions are required

three person mixture			Robustness estimation	
Hp	Hd	log10(LR)	LR distribution	Random man substituted
S1,V,U	V,U,U	7.2	(-10,-5,-0.9)	S1
S2,V,U	V,U,U	-3	(-10,-5,-0.9)	S2
S1,S2,V	V,U,U	5.3	(-23,-16,-8)	S1
S1,S2,V	V,U,U	5.3	(+0.1,+3.7,+7.9)	S2

- This table summarises the Likelihood ratios
- Evidence for S2 under Hp is exclusionary
- Very strong evidence for S1 under Hp, regardless of propositions tested

- How can we evaluate these propositions further?
- If we agree under Hp that S2 is excluded, this means that the propostions can Be further simplified

- Let's return to the table of alleles in order to reassess the case

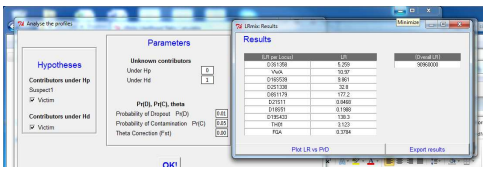
Re-evaluation (two person mixture)

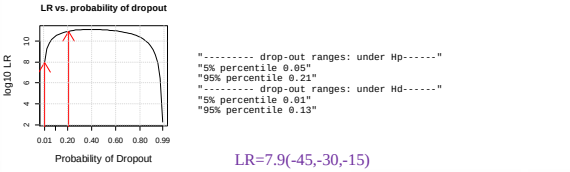
	Crime-stain alleles						
Marker	Allele1	Allele2	Allele3	Allele4	S1	S1	No. of Unique Alleles
AMEL	X	Y			X	Y	2
D3S1358	14	16	17		16	17	4
VWA	16	17	18	19	16	18	4
D16S539	11	12	13	15	12	13	4
D2S1338	17	19	20	24	19	20	4
D8S1179	9	10	13	14	9	13	4
D21S11	29	31	32		28	32	4
D18S51	12	16			12	15	4
D19S433	12	14	15.2	16	12	16	4
TH01	6	9.3			6	9.3	2
FGA	19	24	26		19	21	4

= alleles that have dropped out under Hp
=alleles that are shared with victim under Hp
=alleles that are present in the crime-stain and not shared with the victim under Hp
14 = alleles found in the crime stain that match the victim

- No more than 4 unique alleles per locus, hence 2 person mixture is reasonable
- Note that if we substitute with S2 we still have 3 loci with 5 alleles, suggesting 3-person mixture

Two person Hp=S1,V, Hd=V,U





LR=7.9(-45,-30,-15)

Comparison of results

for three different scenarios					
three person mixture			Robustness estimation		
Hp	Hd	log10(LR)	LR distribution	Random man substituted	
S1,V,U	V,U,U	7.2	(-10,-5,-0.9)	S1	
S2,V,U	V,U,U	-3	(-10,-5,-0.9)	S2	
S1,S2,V	V,U,U	5.3	(-23,-16,-8)	S1	
S1,S2,V	V,U,U	5.3	(+0.1,+3.7,+7.9)	S2	
two person mixture			Robustness estimation		
Hp	Hd	log10(LR)	LR distribution	Random man substituted	
S1,V	V,U	7.9	(-45,-30,-15)	S1	

- Several pairs of propositions were evaluated.
- Scenarios with S2 showed that the evidence was weak for this suspect
- Robustness can be measured by the 'distance' between Hp and Hd simulations
- This indicates that the most robust model is S1,V; V,U since Hp vs Hd is separated By at least 22 orders of magnitude!!
- But the reported LR=log10(7) appears to be appropriate.

Principles to follow when evaluating complex sets of hypotheses

Conditioning rules (a)

- Conditioning hypotheses are defined by the casework circumstances
- Remember to evaluate the hypotheses based on the number of contributors derived from the unique number of alleles in the 'set' observed in the epg: i.e. the sum of alleles of known contributors and the sum of alleles of the crime-stain(s) under Hp (to maximise)
- Do not use the *drop-in* principle to 'explain away' additional contributors

Conditioning rules (b)

- If there are two or more 'suspects' under Hp then the hypothesis should be simplified i.e. evaluate: S1,V,U in addition to S1,S2,V
- It is important to explore the likelihood ratio by use of the non-contributor plot.
- In the S1,S2,V example we show that the LR is very insensitive to S2 (random man still gives a high LR)

Summary of results

- Case circumstances
  - Both S1 and S2 are suspects of sexual assault and a sample is taken from the victim. We condition on the victim under Hd
  - No evidence for S2 in the crime stain [even though a three person evaluation with S1,S2 under Hp gives a high LR= log10(5)]
  - Advice: Simplify propositions if there are two suspects always evaluate them separately, replacing the other with an unknown under Hp and Hd