The use of computer simulation to develop new expert systems to analyse low level DNA profiles

Peter Gill, James Curran, Amanda Kirkham

Definition of low level DNA (formerly LCN)

- Low level DNA is defined by 2 conditions
  - When allele dropout may occur
  - When low level (single allele) contamination may be observed –c. 1% in our hands
- This occurs with 28 cycles too!!
- Low level DNA is not just 34 cycles (which is why we prefer to abandon the LCN term).

A typical low level result has drop-out occurring

The Forensic Science Service®

- Ideally need to take account of the probability of dropout and the probability of contamination.

	D3	VWA	D16	D2	Amel	D8	D21	D18	D19	TH	FGA
suspect	15 18	19 19	12 13	17 19	x y	15 15	30 30.2	14 17	13 15	8 9	20 26
female victim	16 16	16 18	10 12	17 17	x x	15 16	30.2 31.2	14 18	14 14	8 8	18 20
Crime stain	15 16 18	16 18	10 12 13	17	x y	15 16	31.2	18	13 14 15	8 9	18 20 26
dropout alleles		*19*		*19*			*30*	*14 17*			

- Note: If there are alleles that have dropped out of the crime stain DNA profile – this is not neutral evidence, it is evidence that favours the defence i.e. the probability of the evidence if the suspect is a contributor is less than 1
- This is why it is preferable to use likelihood ratios for low level DNA analysis

Before I go on.....

The Forensic Science Service®

- Lets try to understand the biochemistry first before we try to understand what to do about calculations
- In particular.....

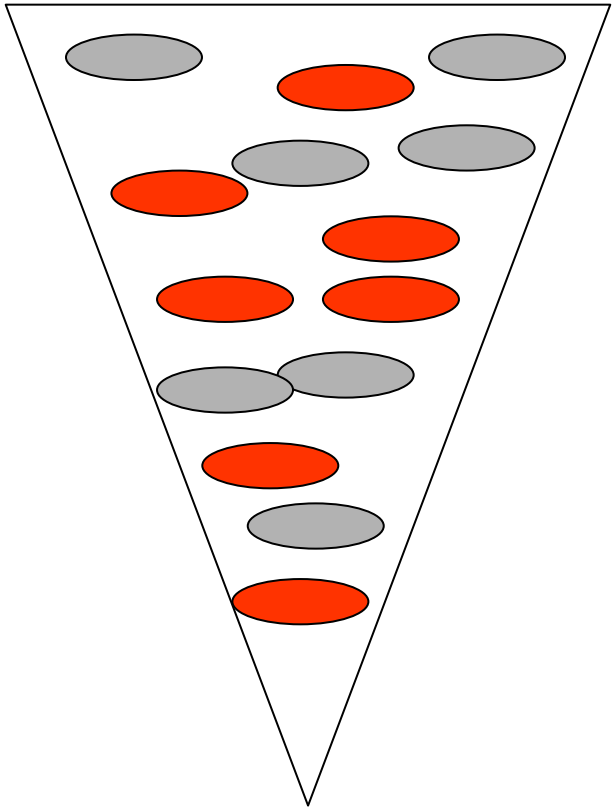
What is low level DNA exactly? – why is it different to conventional DNA??

The Forensic Science Service®

- To answer this we need:
  - To define causes of heterozygous imbalance
  - To define causes of allele drop-out
  - By comparing the results of a computer simulation model with real data
- And to show how a good understanding of the complex biochemical processes that underpin PCR can assist towards developing a robust way to interpret low level DNA profiles

Consider a tube containing 50ul with 7 cells that have been extracted. There are 7 A alleles and 7 B alleles at a particular locus. Take half of the sample in an aliquot of 25ul – how many alleles are extracted?

The Forensic Science Service®



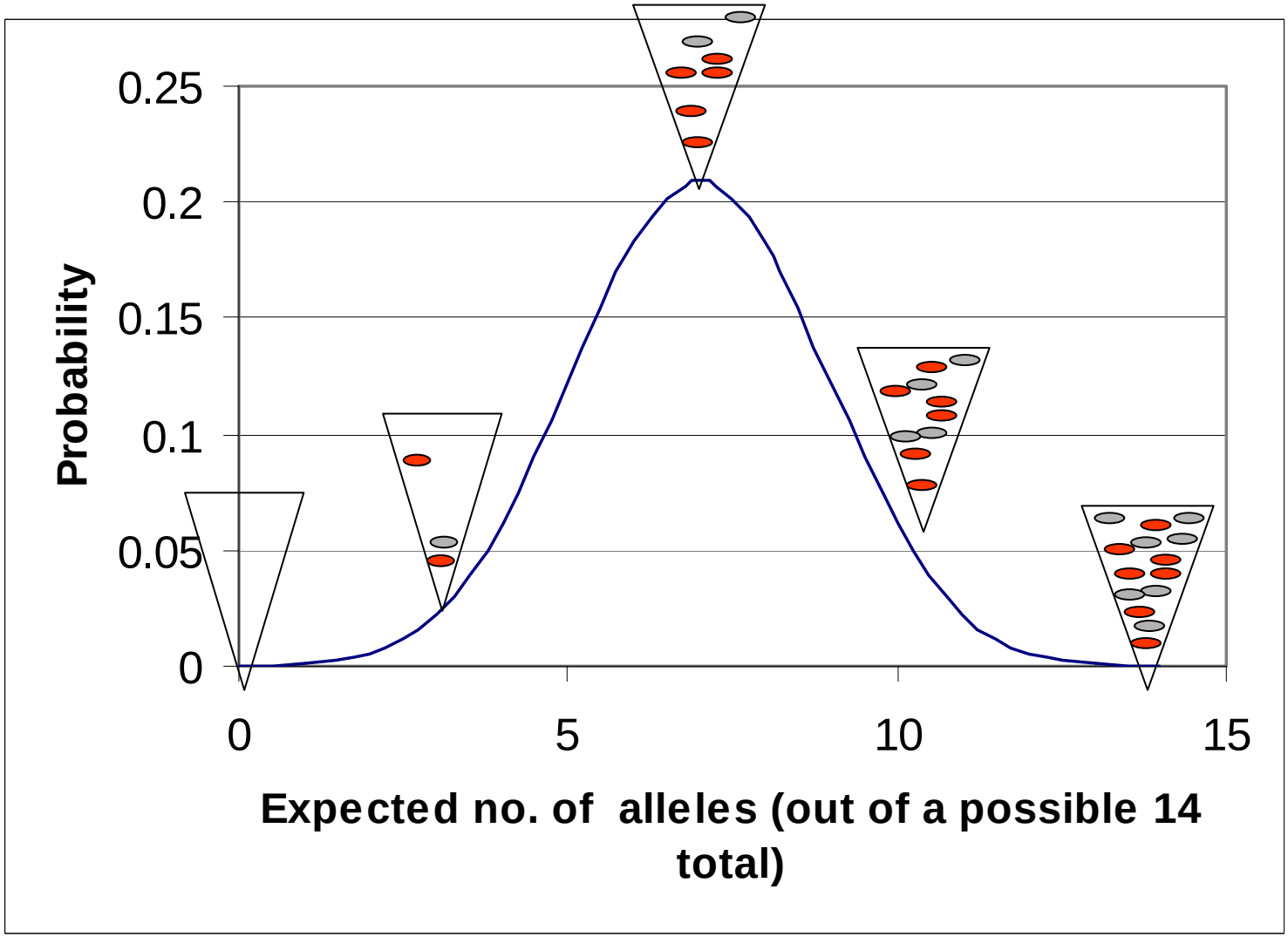
How many molecules are recovered in an aliquot if continuous samples are taken?

The Forensic Science Service®

Pr=Bin(14,0.5)

14 trials

Pr=0.5 chance of success per trial



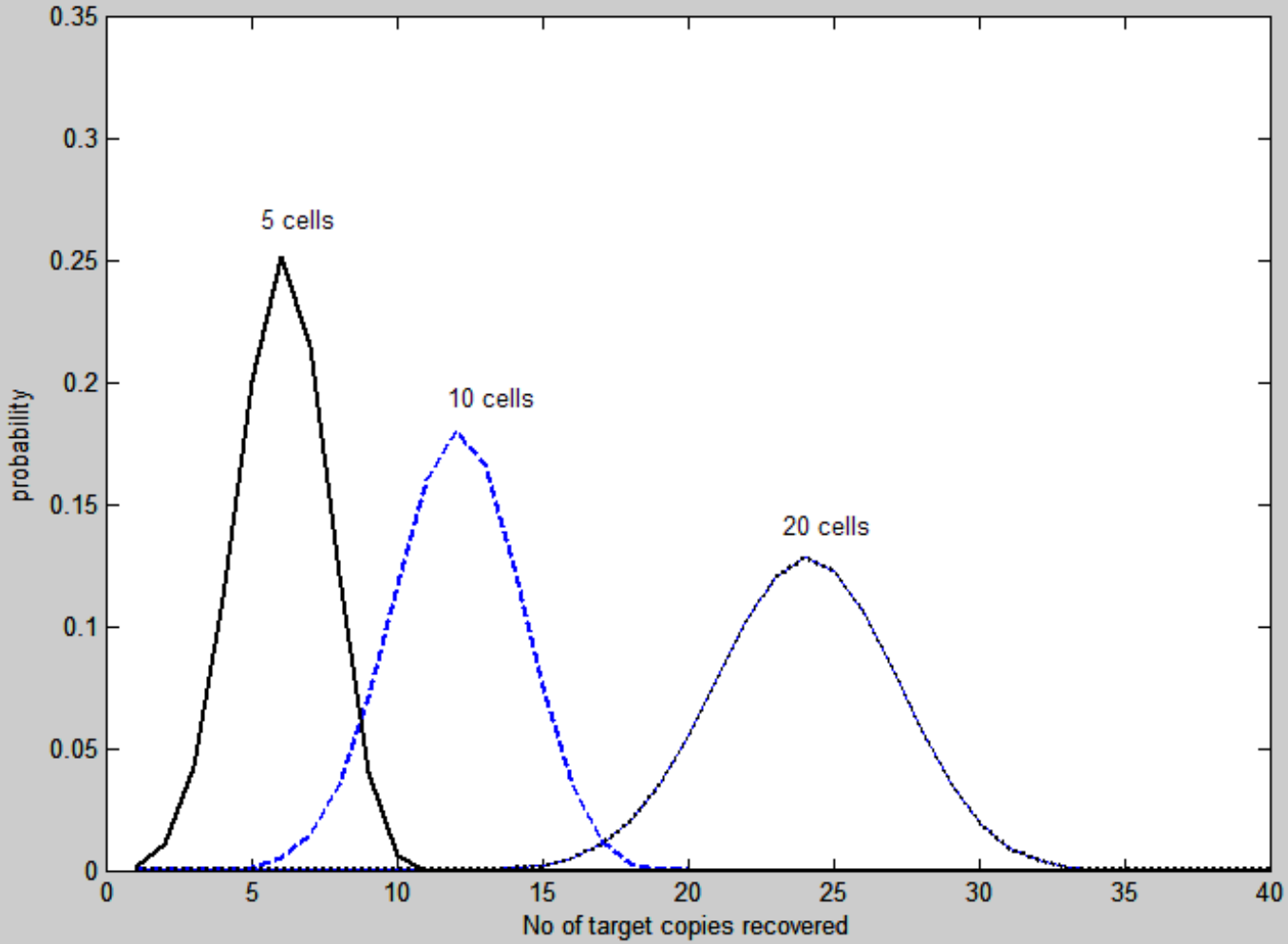
How can we simulate DNA analysis?

The Forensic Science Service®

- We start with a stain that has *n* cells and *2n* copies of DNA
- If we extract *n* cells (with any method), we can never recover *2n* copies of DNA because the efficiency of extraction is never 100%.
- We can simulate the process with a binomial random number generator.
- For a diploid heterozygote (*AB*) *nA* does not equal *nB* after extraction.

Simulation of recovery N= *5,10,20 cells respectively where extraction efficiency= 0.6. Calculated using* Bin(2N, *0.6).*

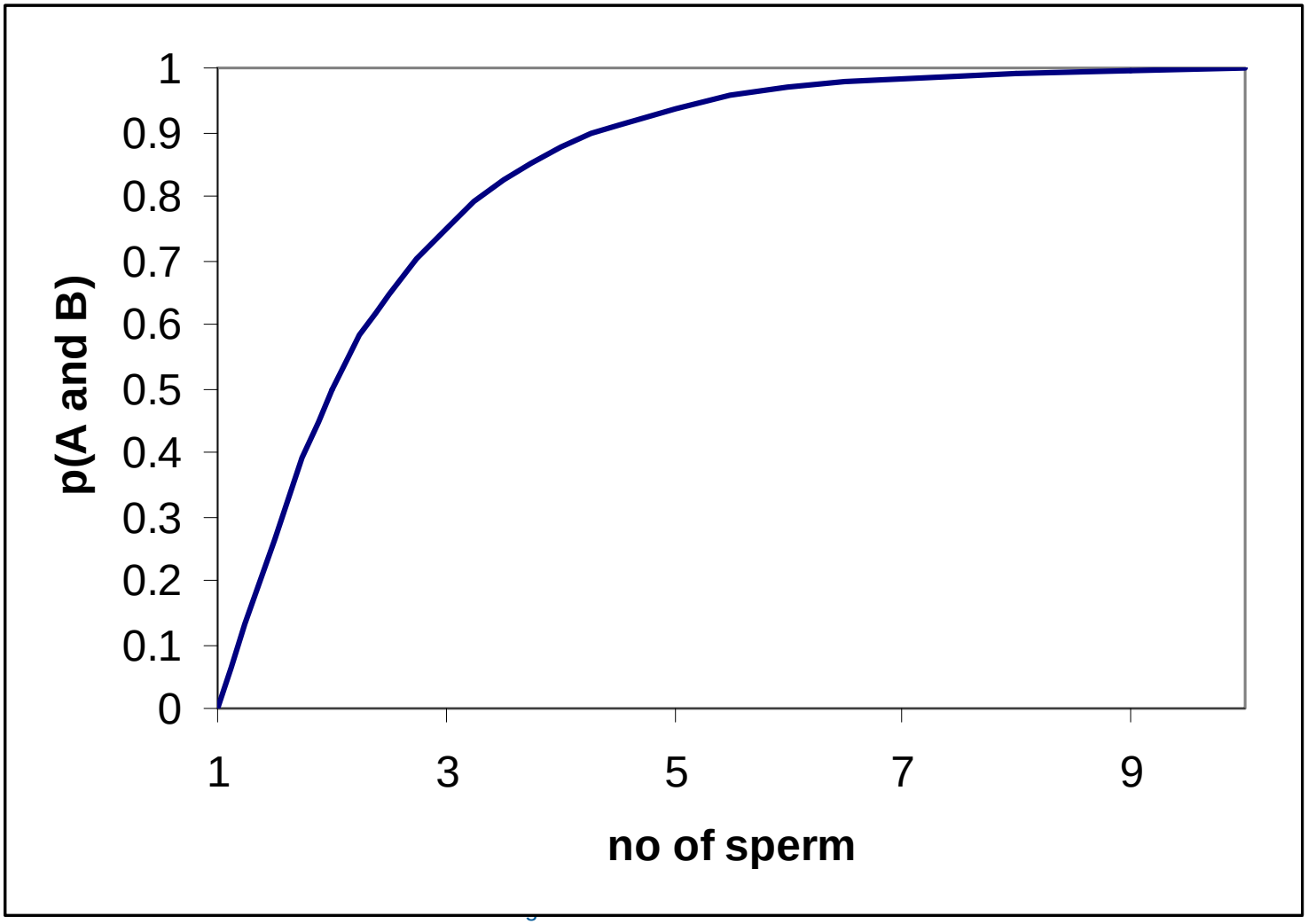
The Forensic Science Service®



Haploid v. diploid

- If we take a single diploid cell and analyse the DNA then we will have equal contribution of alleles
- How many haploid (sperm) cells are needed to obtain both alleles of a heterozygote?

The probability of observing both alleles A and B in a sample of n *sperm at a heterozygous locus*



Taking an aliquot for PCR

- Once we have extracted DNA from a sample, then we take an aliquot of the DNA for analysis – e.g. 20ul from a total of 66ul
- Again this results in imbalance of heterozygotes.
- Furthermore, if we repeat the experiment exactly, the numbers of DNA molecules recovered are different.

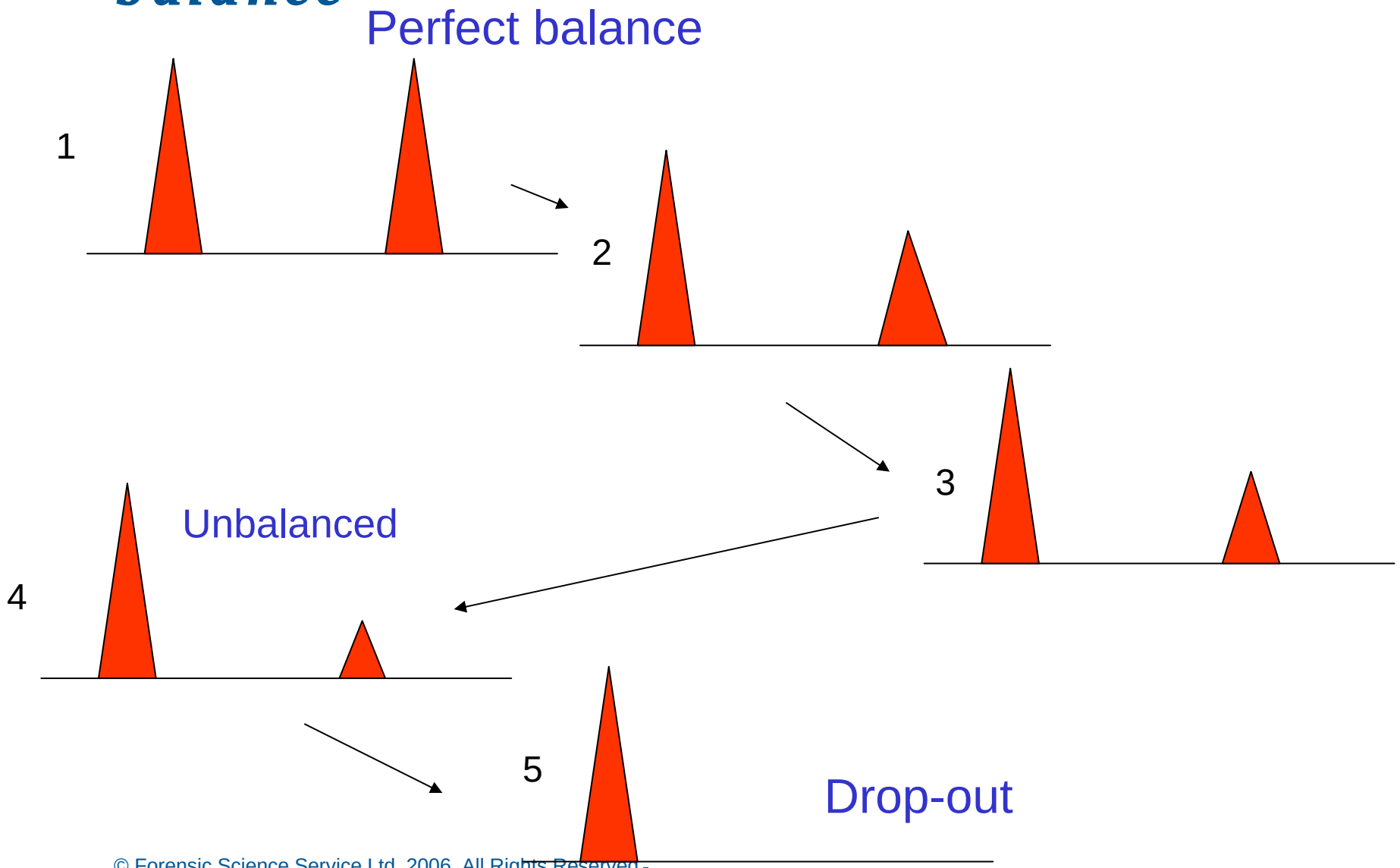
PCR

- PCR is not 100% efficient
- There is approximately 80% chance that a fragment will be amplified per cycle
- This will lead to some imbalance
- BUT we demonstrated that PCR efficiency does not have much effect compared to sampling.
- PCR is also simulated as a binomial model

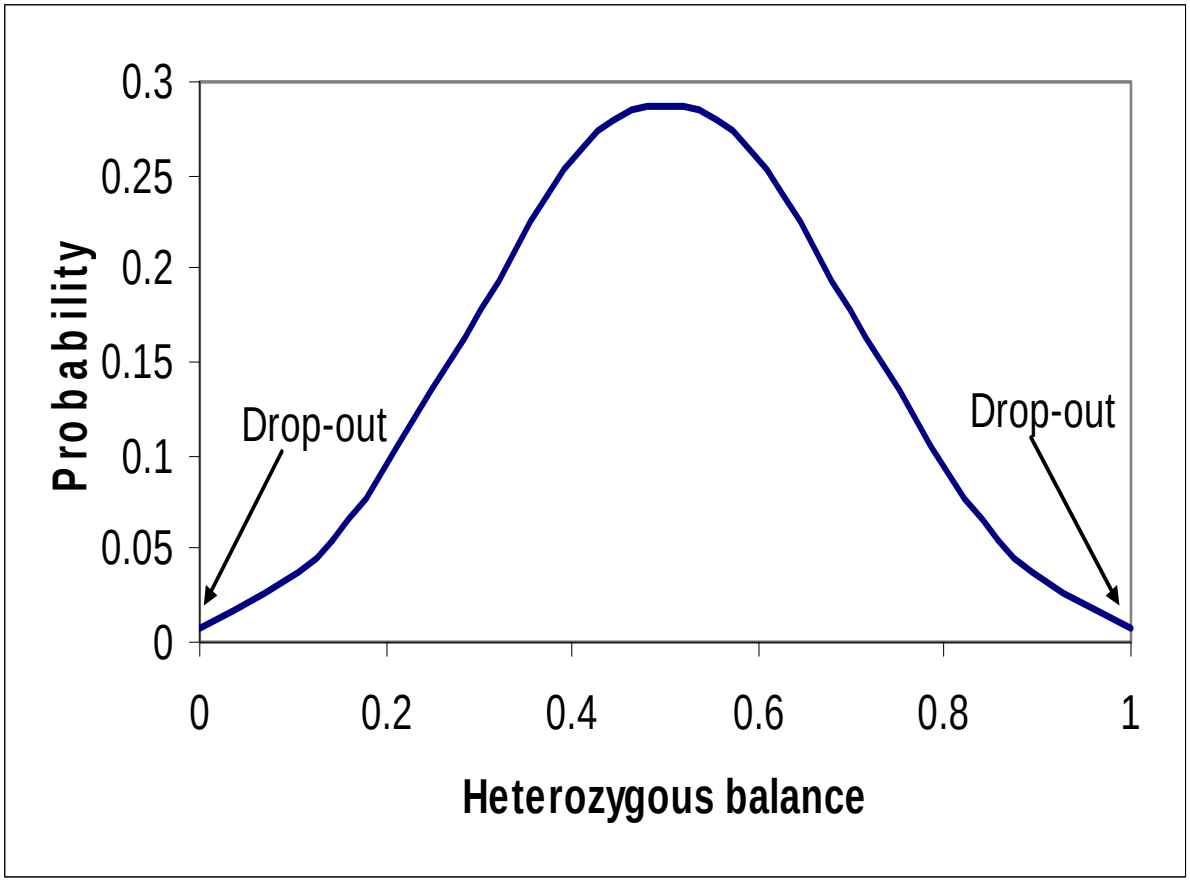
Heterozygous balance

- Imbalance is often observed in heterozygotes
- Why?
- Drop-out is sometimes observed – associated with low level DNA
- Why?

Illustration of heterozygous balance



How will heterozygous balance be affected if 7 alleles are randomly recovered?

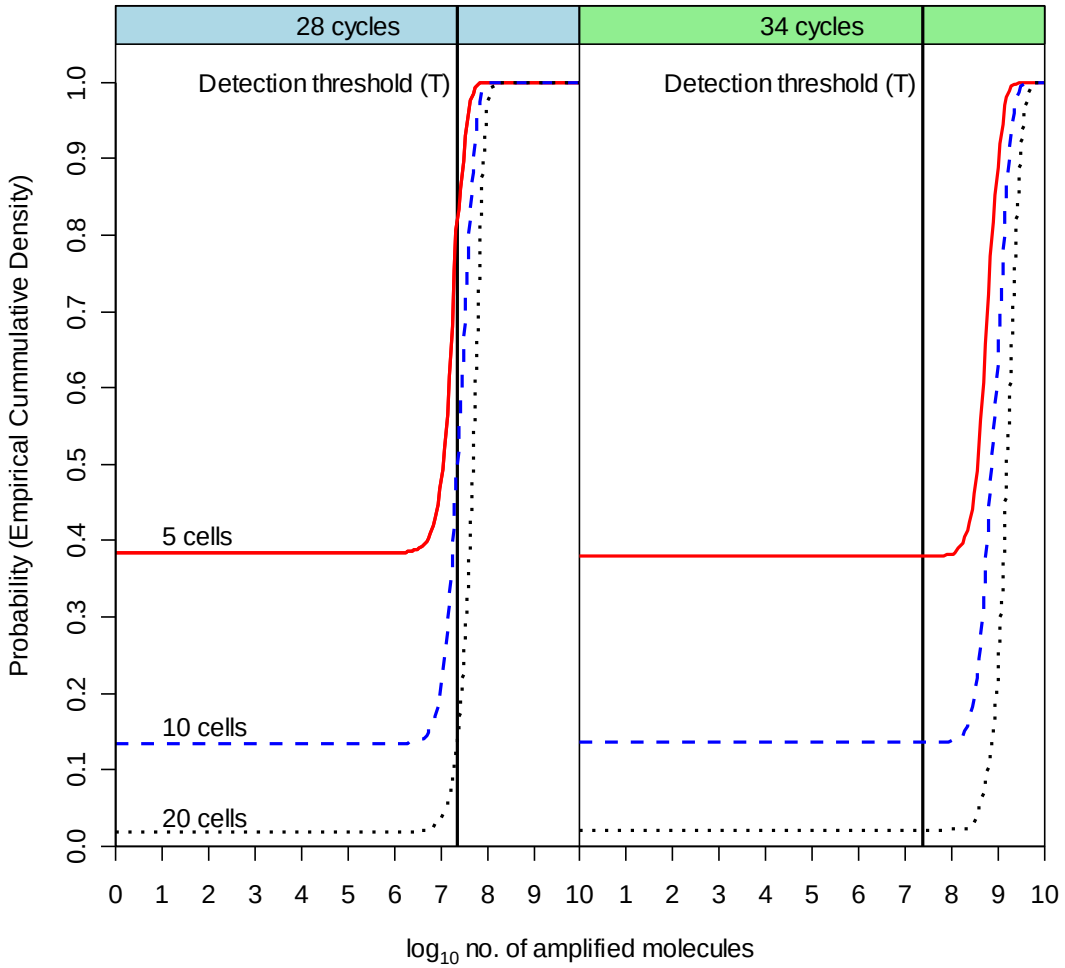


No of A's	No of B's	Hb
7	0	0
6	1	.14
5	2	.29
4	3	.42
3	4	.42
2	5	.29
1	6	.14
0	7	0

What causes drop-out? General principles of low level DNA:

- Dropout is a consequence of heterozygous imbalance (part of same phenomenon).
- There are 2 reasons for drop-out:
  - Stochastic ie no molecule is present in PCR reaction mix
  - Or: Insufficient molecules to trigger a signal
- BUT we have shown that if a molecule is present, and 34 cycles are used then there are always sufficient generated to exceed the threshold of detection i.e. it really is single molecule sensitive under ideal conditions.
- PCR inhibition and degradation will reduce sensitivity.

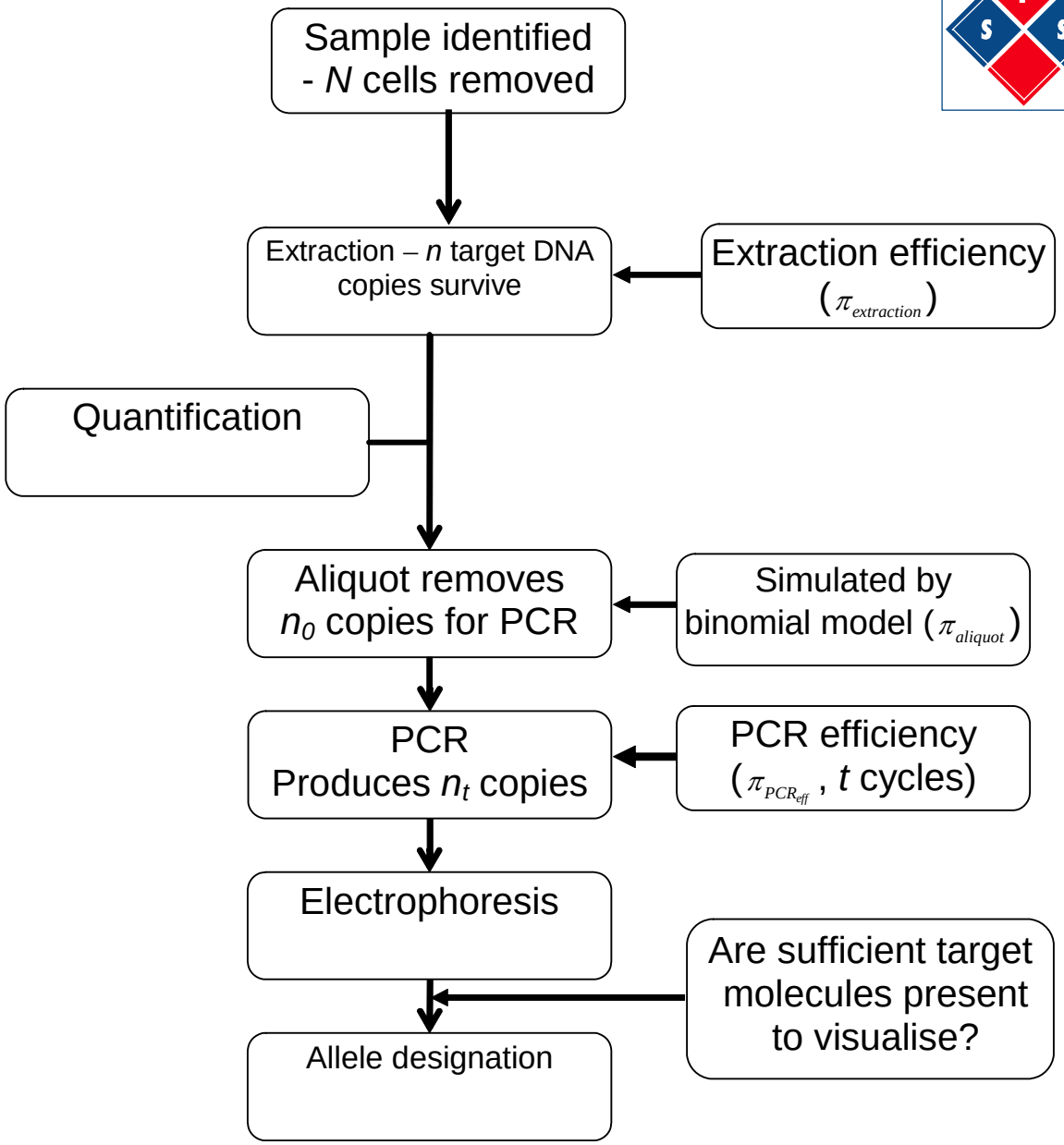
Cumulative probability density 28 v 34 cycles



probability density of 5,10,20 cells after extraction ($\pi_{Extraction} = 0.6$), selection of an aliquot ($\pi_{Aliquot} = 20/66$), and PCR ($\pi_{PCReff} = 0.8$) using 28 and 34 cycles respectively.

Bayes net (graphical model) – A computer simulation of PCR

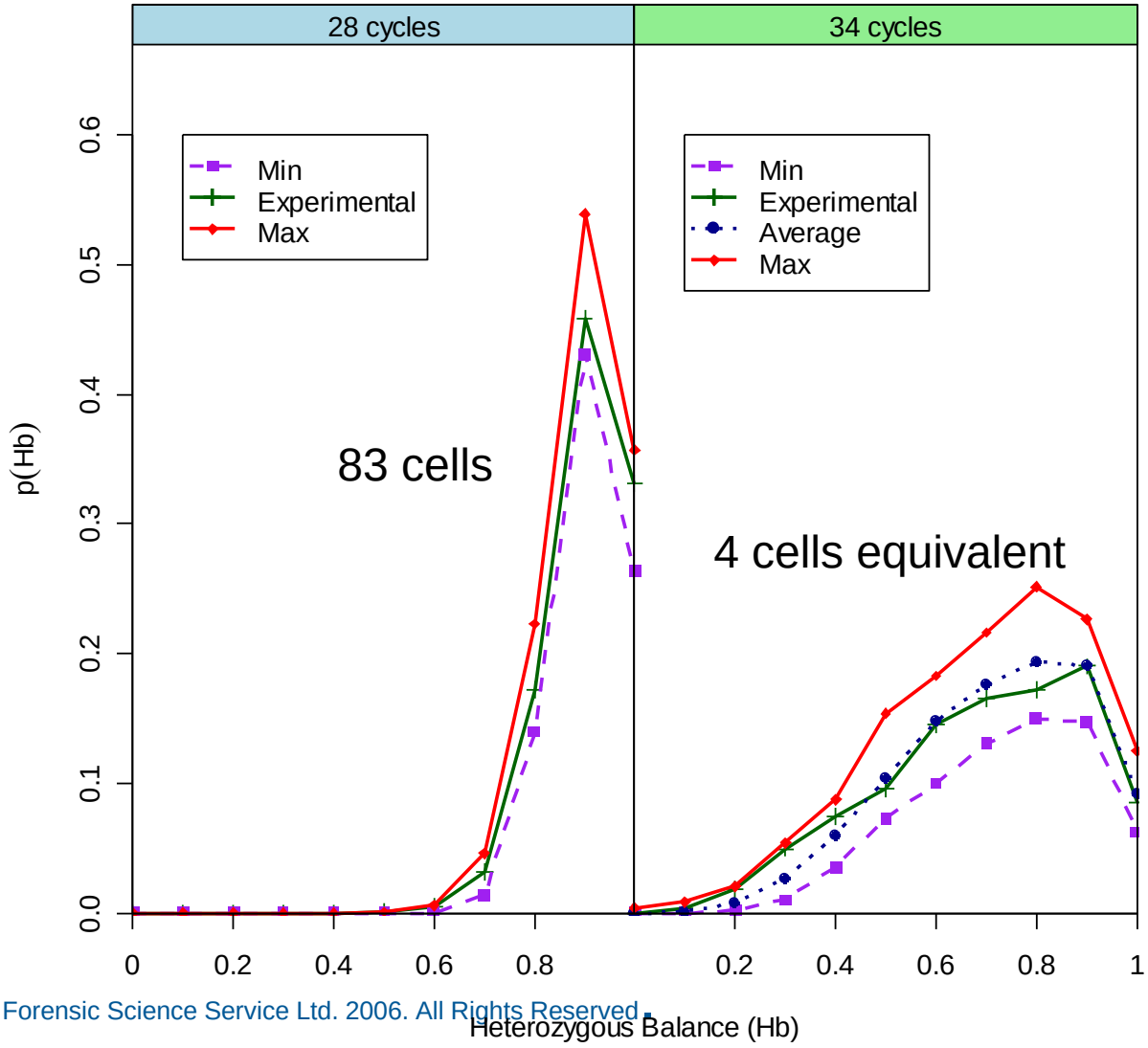
- We have built a simple computer model that randomly selects molecules (e.g from extracted products) and replicates molecules in PCR.
- i.e. The entire DNA process can be simulated at the molecular level by a series of simple binomial models using efficiency parameters
- Two models – one for haploid and one for diploid cells.



How well does the model work?

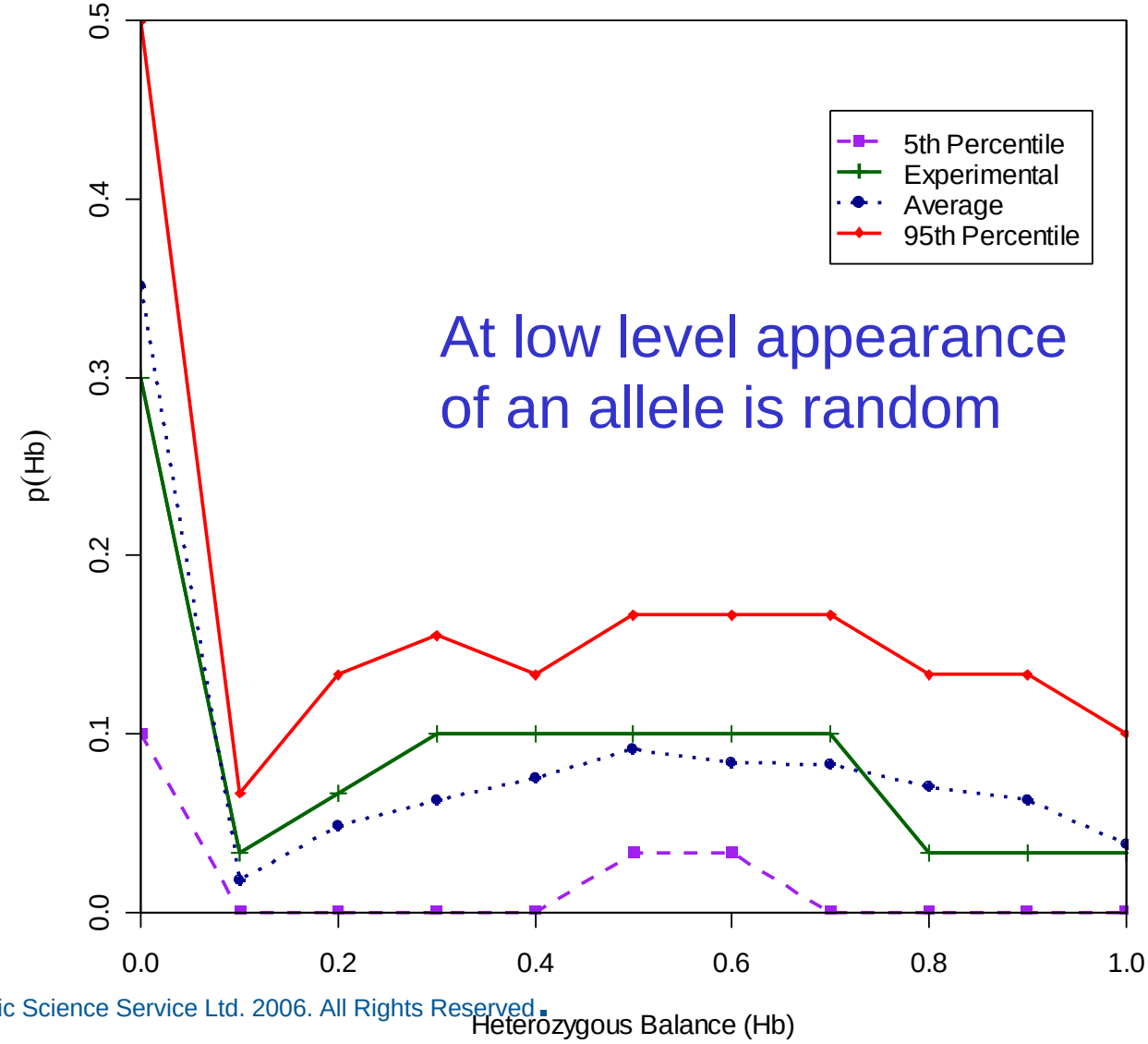
- How does it compare to real data?

Heterozgous balance



10 cells picked by LMD and subject to extraction

$\pi_{Extraction} = 0.46$, $\pi_{PCReff} = 0.8$, and $\pi_{Aliquot} = 20/66$



LoComatioN – a new probabilistic system to interpret low copy number profiles

Interpretation using LoComatioN

- Full probabilistic model
- Factors in contamination and dropout into the calculation.
- Enables the interpretation of complex mixtures up to 3 persons.
- Enables the probabilistic evaluation of multiple scenarios (impossible to do this previously

LoComatioN

The Forensic Science Service®

- Based on our new understanding of the biochemistry of low level DNA we have developed an expert system to calculate likelihood ratios
- The system is fully validated and about to be implemented into casework by the FSS
- It can be used for 34 cycles and 28 cycles low level DNA profiles.

A duplication experiment - the traditional method

	Amelo	D19	D3	D8	THO	VWA	D21	FGA	D16	D18	D2
1	- -	14 14	- -	15 15	- -	- -	28 32.2	20 -	- -	- 16	- -
2	X X	- -	18 18	15 15	- -	19 19	- -	- -	- 12	- -	- -
CONSENSUS				15,F							

Note that the 'F' designation means that allele drop-out cannot be discounted

LoComatioN

The Forensic Science Service®

- Doesn't depend upon a consensus profile
- It calculates probabilities across alleles across replicate PCR analyses
- The key is that we assign a probability of dropout and a probability of contamination to each allele
- It is therefore a much more efficient analytical tool because we don't have to make subjective decisions about whether or not to include a given allele in the DNA profile.

View Tools Help

se Options

Create New Case

Open Existing Case

Save Current Case

Close Current Case

se Browser

PROFILE INPUT

Manual Input

CASE OPTIONS

Profile Summary

Set Variables

Add hypotheses

se Details

eplicates: 1

uspects: 1

ictims: 1

Hd

Sus: nSus:

Vic: nVic:

Unk: nUnk:

Summary of profiles added for this case

Edit profiles | Close

Export to Word Report

Contributor Profiles

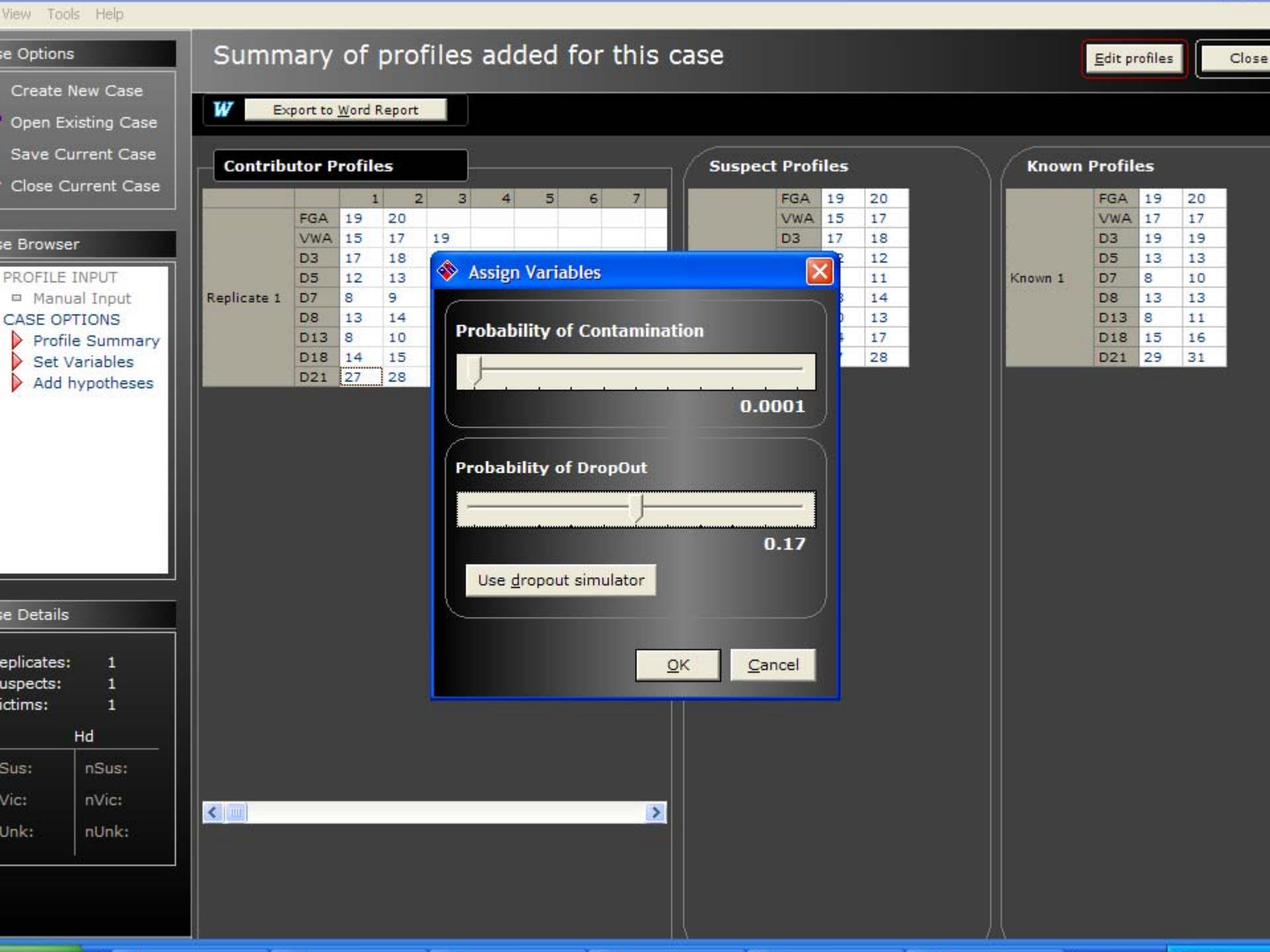
		1	2	3	4	5	6	7
Replicate 1	FGA	19	20					
	VWA	15	17	19				
	D3	17	18					
	D5	12	13					
	D7	8	9	10	11			
	D8	13	14					
	D13	8	10	11	13			
	D18	14	15	16	17			
	D21	27	28	29	31			

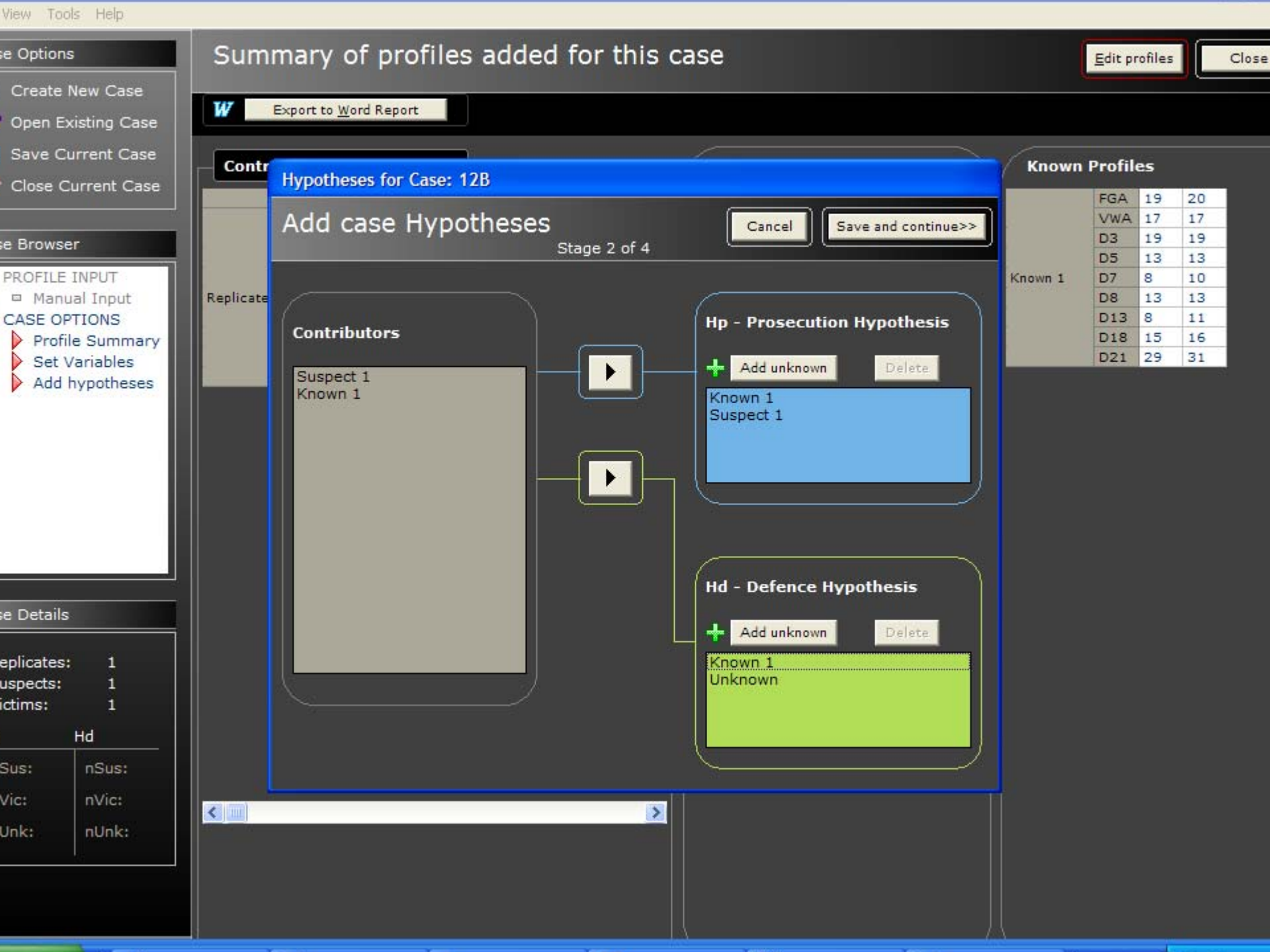
Suspect Profiles

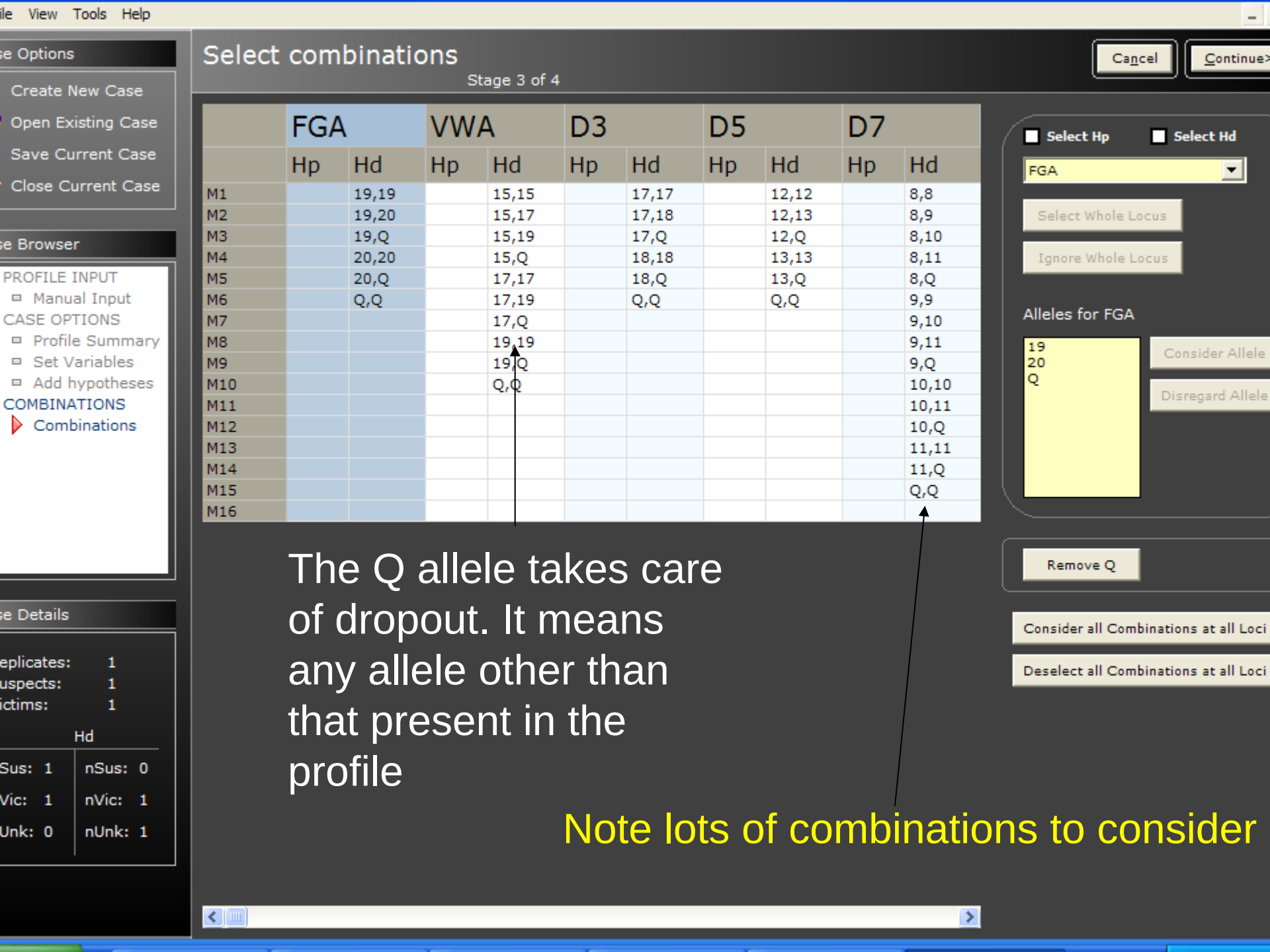
Suspect 1	FGA	19	20
	VWA	15	17
	D3	17	18
	D5	12	12
	D7	9	11
	D8	13	14
	D13	10	13
	D18	14	17
	D21	27	28

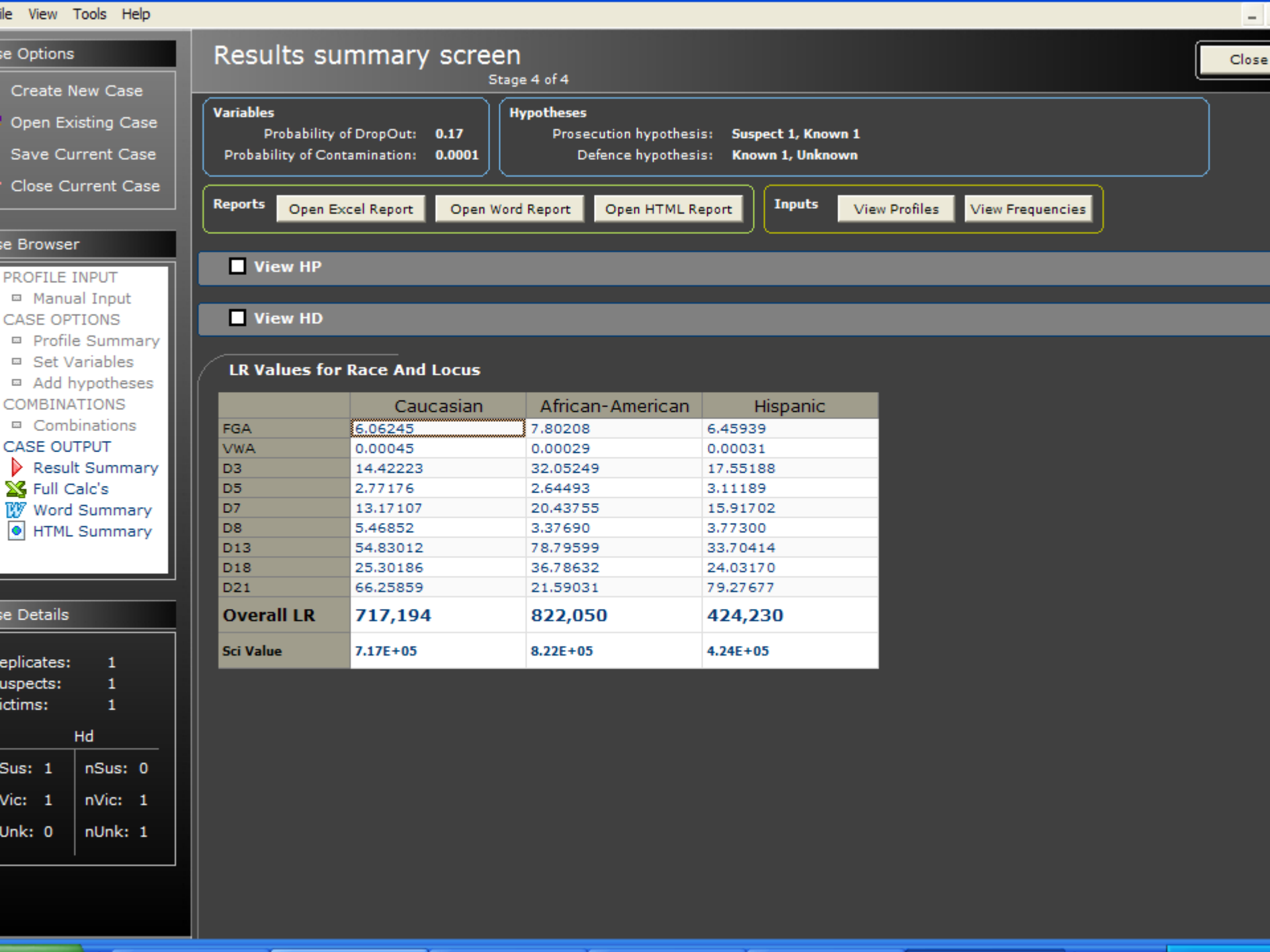
Known Profiles

Known 1	FGA	19	20
	VWA	17	17
	D3	19	19
	D5	13	13
	D7	8	10
	D8	13	13
	D13	8	11
	D18	15	16
	D21	29	31


View Tools Help
se Options
Create New Case
Open Existing Case
Save Current Case
Close Current Case
se Browser
PROFILE INPUT
Manual Input
CASE OPTIONS
Profile Summary
Set Variables
Add hypotheses
se Details
eplicates: 1
uspects: 1
ictims: 1
Hd
Sus: nSus:
Vic: nVic:
Unk: nUnk:
Summary of profiles added for this case
Edit profiles
Close
Export to Word Report
Contributor Profiles
Replicate 1
FGA 19 20
VWA 15 17 19
D3 17 18
D5 12 13
D7 8 9
D8 13 14
D13 8 10
D18 14 15
D21 27 28
Suspect Profiles
FGA 19 20
VWA 15 17
D3 17 18
Known Profiles
Known 1
FGA 19 20
VWA 17 17
D3 19 19
D5 13 13
D7 8 10
D8 13 13
D13 8 11
D18 15 16
D21 29 31
Assign Variables
Probability of Contamination
0.0001
Probability of DropOut
0.17
Use dropout simulator
OK
Cancel


View Tools Help
se Options
Create New Case
Open Existing Case
Save Current Case
Close Current Case
se Browser
PROFILE INPUT
Manual Input
CASE OPTIONS
Profile Summary
Set Variables
Add hypotheses
se Details
eplicates: 1
uspects: 1
ictims: 1
Hd
Sus: nSus:
Vic: nVic:
Unk: nUnk:
Summary of profiles added for this case
Edit profiles
Close
Export to Word Report
Contr
Replicate
Hypotheses for Case: 12B
Add case Hypotheses
Stage 2 of 4
Cancel
Save and continue>>
Contributors
Suspect 1
Known 1
Hp - Prosecution Hypothesis
Add unknown
Delete
Known 1
Suspect 1
Hd - Defence Hypothesis
Add unknown
Delete
Known 1
Unknown
Known Profiles
Known 1
FGA 19 20
VWA 17 17
D3 19 19
D5 13 13
D7 8 10
D8 13 13
D13 8 11
D18 15 16
D21 29 31


File View Tools Help
Case Options
Create New Case
Open Existing Case
Save Current Case
Close Current Case
Case Browser
PROFILE INPUT
Manual Input
CASE OPTIONS
Profile Summary
Set Variables
Add hypotheses
COMBINATIONS
Combinations
Case Details
Replicates: 1
Suspects: 1
Victims: 1
Hd
Sus: 1
nSus: 0
Vic: 1
nVic: 1
Unk: 0
nUnk: 1
Select combinations
Stage 3 of 4
Cancel
Continue
FGA VWA D3 D5 D7
Hp Hd Hp Hd Hp Hd Hp Hd Hp Hd
M1 19,19 15,15 17,17 12,12 8,8
M2 19,20 15,17 17,18 12,13 8,9
M3 19,Q 15,19 17,Q 12,Q 8,10
M4 20,20 15,Q 18,18 13,13 8,11
M5 20,Q 17,17 18,Q 13,Q 8,Q
M6 Q,Q 17,19 Q,Q Q,Q 9,9
M7 17,Q 9,10
M8 19,19 9,11
M9 19,Q 9,Q
M10 Q,Q 10,10
M11 10,11
M12 10,Q
M13 11,11
M14 11,Q
M15 Q,Q
M16
Select Hp
Select Hd
FGA
Select Whole Locus
Ignore Whole Locus
Alleles for FGA
19
20
Q
Consider Allele
Disregard Allele
Remove Q
Consider all Combinations at all Loci
Deselect all Combinations at all Loci
The Q allele takes care of dropout. It means any allele other than that present in the profile
Note lots of combinations to consider

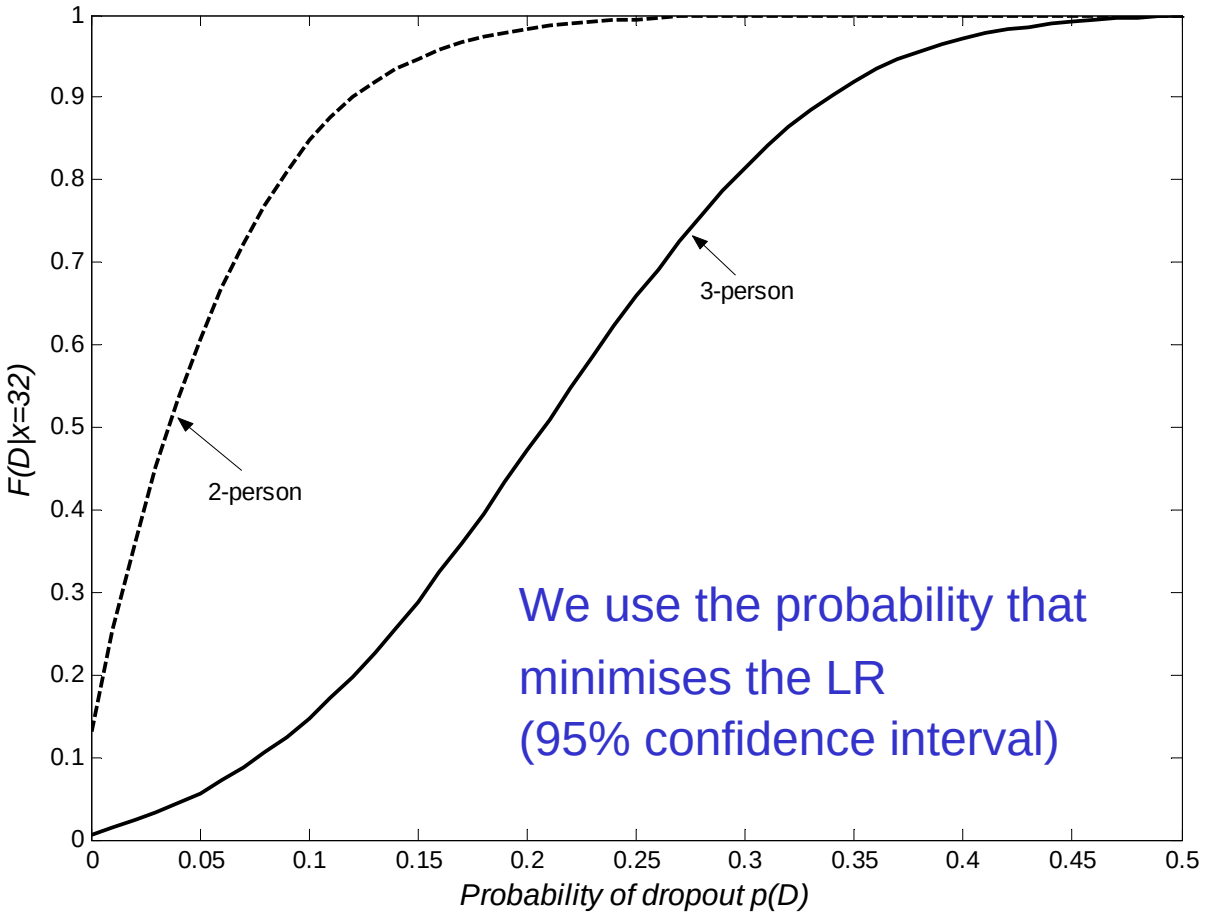

ile View Tools Help
se Options
Create New Case
Open Existing Case
Save Current Case
Close Current Case
se Browser
PROFILE INPUT
Manual Input
CASE OPTIONS
Profile Summary
Set Variables
Add hypotheses
COMBINATIONS
Combinations
CASE OUTPUT
Result Summary
Full Calc's
Word Summary
HTML Summary
se Details
eplicates: 1
uspects: 1
ictims: 1
Hd
Sus: 1
nSus: 0
Vic: 1
nVic: 1
Unk: 0
nUnk: 1
Results summary screen
Stage 4 of 4
Close
Variables
Probability of DropOut: 0.17
Probability of Contamination: 0.0001
Hypotheses
Prosecution hypothesis: Suspect 1, Known 1
Defence hypothesis: Known 1, Unknown
Reports
Open Excel Report
Open Word Report
Open HTML Report
Inputs
View Profiles
View Frequencies
View HP
View HD
LR Values for Race And Locus
Caucasian
African-American
Hispanic
FGA 6.06245 7.80208 6.45939
VWA 0.00045 0.00029 0.00031
D3 14.42223 32.05249 17.55188
D5 2.77176 2.64493 3.11189
D7 13.17107 20.43755 15.91702
D8 5.46852 3.37690 3.77300
D13 54.83012 78.79599 33.70414
D18 25.30186 36.78632 24.03170
D21 66.25859 21.59031 79.27677
Overall LR 717,194 822,050 424,230
Sci Value 7.17E+05 8.22E+05 4.24E+05

How do we calculate probability of dropout?

- Under the assumption that the probability of dropout is equivalent across all loci we simulate a mixture of 2 or more individuals
  - Use a random number generator to simulate dropout
  - Then count the number of surviving alleles
    - Example: Suppose that on average when there is no dropout p(D)=0 we observe c. 15 alleles
    - If p(D)= 0.5, we will observe on average 7.5 alleles
    - But if we carry out a simulation (1000x) then we may find that there will be a distribution between 4-12 alleles
    - We calculate 95% confidence interval

Calculation of the probability of drop-out from a sample with 32 alleles (SGMplus)

The Forensic Science Service®



We use the probability that minimises the LR (95% confidence interval)

A case-work example

- Two female victims were assaulted by a man wielding a hammer in their flat. Both sustained minor injuries
- The assailant fled the scene of crime.
- The hammer was recovered 100 yards from the crime scene
- Low level DNA from the head of the hammer revealed a mixture of DNA
- The suspect denied the offence and denied that the hammer was his.

The case-preassessment – we have to formulate both the prosecution and alternative defence hypotheses

- Prosecution hypothesis: The hammer was used by the suspect to commit the offence
- Defence hypothesis: The hammer was not used by the suspect to commit the offence
- Is there evidence to suggest that the suspect handled the hammer?
- Is there evidence to suggest that the victims were hit by the hammer?
- OK s0 formulating hypotheses can be the hardest part – there may be multiple possibilities suggested by both prosecution and the defence
- How do we explore this?

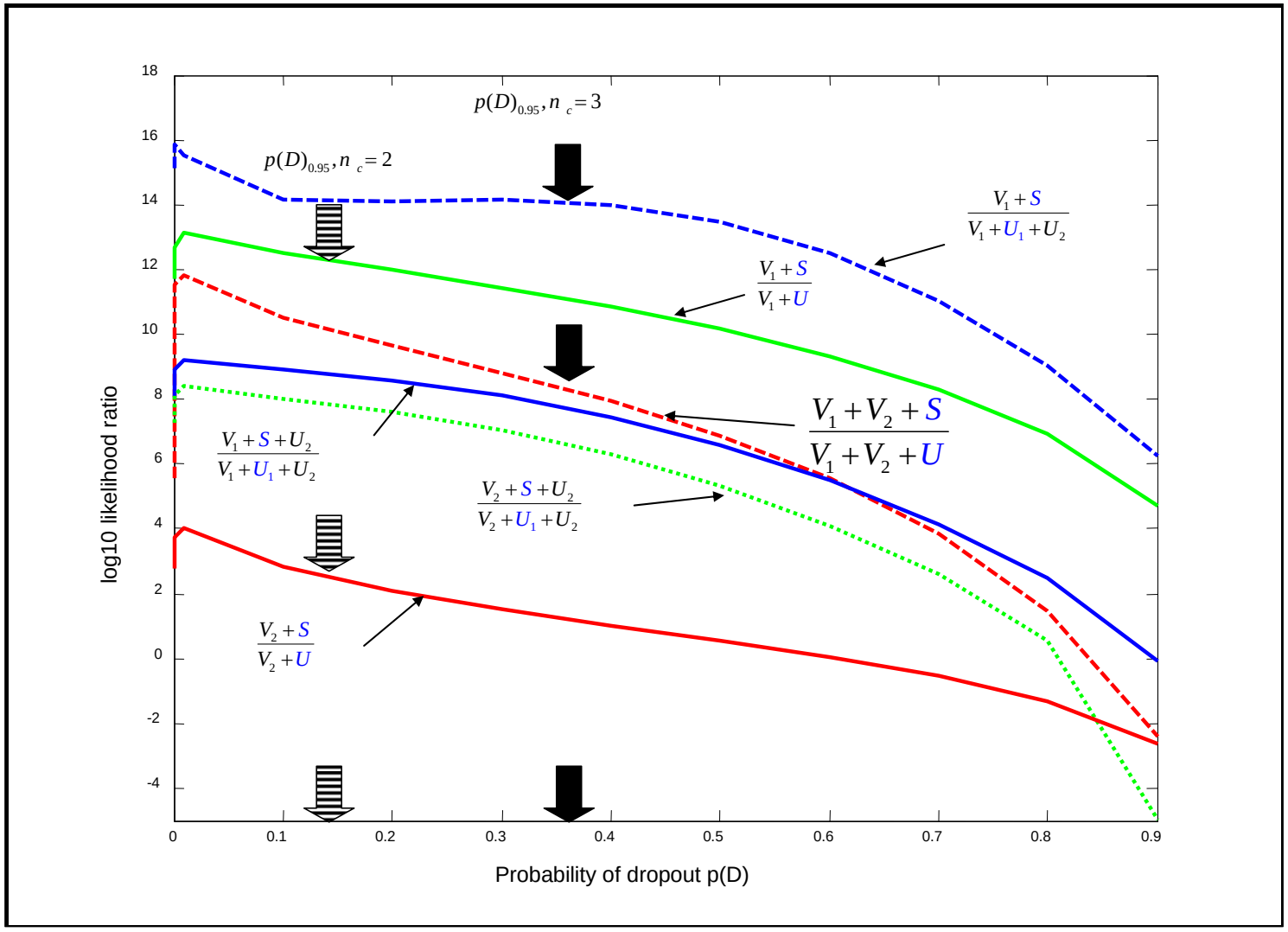
Case analysis – lets look a bit more closely at the data

	Allelic Results Observed At Each Loci Tested										
	Amelo	D3	VWA	D16	D2	D8	D21	D18	D19	THO	FGA
Sample (R_1)	X Y	14 16	15 16 19	11 13 14	20 23 24 25	11 12 13 15	28 31		12 14 15.2 17.2	6 8 9 9.3	22
Sample (R_2)	X Y	14 16	15 16 17 19	11 13 14	20 24 25	11 12 13 15	28 29 30 31 31.2	13 14 16 17	12 13 14 15.2 17.2	6 8 9 9.3	22 23 25
Victim 1	X X	16 16	15 16	13 13	20 20	11 15	29 30	17 17	12 14	6 8	22 25
Victim 2	X X	15 17	16 19	12 13	18 25	11 13	29 30	15 17	14 14	6 7	20 22
Suspect	X Y	14 16	15 19	11 14	24 25	12 13	28 31	14 17	15.2 17.2	9 9.3	22 23

Hypotheses

- The first set of hypotheses (based on casework circumstances):
  - *Hp*: Suspect + victim 1 +victim 2
  - *Hd*: unknown 1 +victim 1 + victim 2
- Both victims present and suspect?
  - What about VWA-17 allele; D18-13 allele?? Contamination??
- One victim and one unknown individual?
  - No need to invoke victim 2 in the profile because lots of alleles are shared.
- By inspection – the simplest explanation is
  - *Hp*: Suspect + victim 1 + unknown
  - *Hd*: Unknown 1 + victim 2 +unknown 2
- How would we evaluate these uncertainties??

Evaluation of multiple hypotheses



Evaluation of the hypotheses

- It doesn't matter if victim 2 contributed or not as the LRs are virtually unaffected i.e. the prosecution hypothesis is unaffected
- The consensus method gives a similar result in this case.
- The program allows extensive evaluation of any scenario.
- The prosecution hypothesis is seriously affected if it supposed that the DNA profile comprises victim 2 + unknown under *Hp* and *Hd* – this wouldn't make sense though.

Computer simulation and modelling

- We are using computer simulation to improve our understanding of the stochastic processes involved with low copy number DNA profiling
- We now understand the reasons for allele dropout and heterozygous balance using the PCRSIM model
- We have used this improvement in understanding to develop a powerful new probabilistic model to interpret low level DNA profiles *(LoComatioN)*.
- We anticipate that these new models will quickly supersede current methods of low level DNA analysis (at both 28 and 34 PCR cycles)

Important points

The Forensic Science Service®

- *LoCoMatioN* is not a black box – it does not give you **the** answer.
- It is an exploratory tool to assist the reporting officer to make a fair evaluation about a case
- It is used to evaluate multiple 'what-if' scenarios – these are complex calculations that can be completed within a few minutes (*in the example we show that some scenarios are not important to consider*).
- It is a very flexible tool; it will be essential that the user is well trained in low level DNA theory to use it properly.
- It is also up to the reporting officer to decide **how** to use it. There are no hard and fast rules.

Publications on low level DNA theory and background

- P. Gill, J. Whitaker, C. Flaxman, N. Brown,J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, Forensic Sci Int. 112 (2000) 17-40.
- P. Gill, Application of low copy number DNA profiling, Croat Med J. 42 (2001) 229-32.
- P. Gill,A. Kirkham, Development of a simulation model to assess the impact of contamination in casework using STRs, J Forensic Sci. 49 (2004) 485-91.
- P. Gill, J. Curran,K. Elliot, A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci, Nucleic Acids Res. 33 (2005) 632-43.

Publications on LoComatioN

- J. Curran, P. Gill,M.R. Bill, Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, Forens. Sci. Int. 148 (2004) 47-53.
- Peter Gill, Amanda Kirkham and James Curran (2007) *LoComatioN*: a software tool for the analysis of low-copy number DNA profiles. Forens. Sci. Int. 166, 128-138.

Acknowledgements

- Amanda Kirkham (validation)
- James Curran (theory and programming)
- Martin Bill (encouragement and support).