Internal Validation of STRmix™ V2.4 for Fusion

NYC OCME

STRmix™ internal validation

This document describes the internal validation of STRmix™ V2.4 for the Promega Fusion™ multiplex at the OCME NY Laboratory (hereafter OCME). Internal validation describes the activities OCME has undertaken in-house before the implementation of STRmix™ for Fusion™ into routine casework.

This document follows the internal validation guidelines of the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems [1]. This included the examination of known and non-probative evidence samples, and investigations into reproducibility and precision, sensitivity and stochastic studies, and mixture studies. All numerical designations within refer to specific recommendations of the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems.

The results of all experiments related to the internal validation of STRmix™ at the OCME Laboratory are retained within the laboratory's quality system.

STRmix™ has previously been subjected to developmental validation following the SWGDAM Guidelines. This work has been published [2] with a more detailed summary of the tests undertaken included in the STRmix™ User's Manual.

STRmix™ parameters

4.1. *The laboratory should test the system using representative data generated in-house with the amplification kit, detection instrumentation and analysis software used for casework. Additionally, some studies may be conducted by using artificially created or altered input files to further assess the capabilities and limitations of the software.*

4.1.3. *Variable DNA typing conditions (e.g., any variations in the amplification and/or electrophoresis parameters used by the laboratory to increase or decrease the detection of alleles and/or artifacts)*

All testing was undertaken using the Promega PowerPlex® Fusion Amplification Kit. Samples were run on two 3130xl instruments using standard run conditions and subsequently analyzed using GeneMarker® HID v.2.8.2. Samples were only run under one electrophoresis condition (3kv 5seconds) and amplified under one set of conditions (29 cycles). The GeneMarker® HID stutter filters (including the global filter or minimum heterozygote filter) were turned off (set to 0%) prior to generation of GeneMarker® HID export data for use within STRmix™.

Mixtures were prepared in-house using extracted buccal swabs from staff members. In addition, older proficiency tests and validation samples from the semi-automated differential extraction of semen samples by the Qiagen Qiacube and EZ1 instruments will be used. Various combinations of two-person, three-person, and four-person mixtures were prepared at different dilutions, as indicated below.

The parameters described in the document Estimation of STRmix™ parameters for OCME – Fusion were used for all interpretations undertaken in this report. All other run parameters have been optimised by the STRmix™ developers. Following deconvolution by STRmix™, various hypotheses were tested.

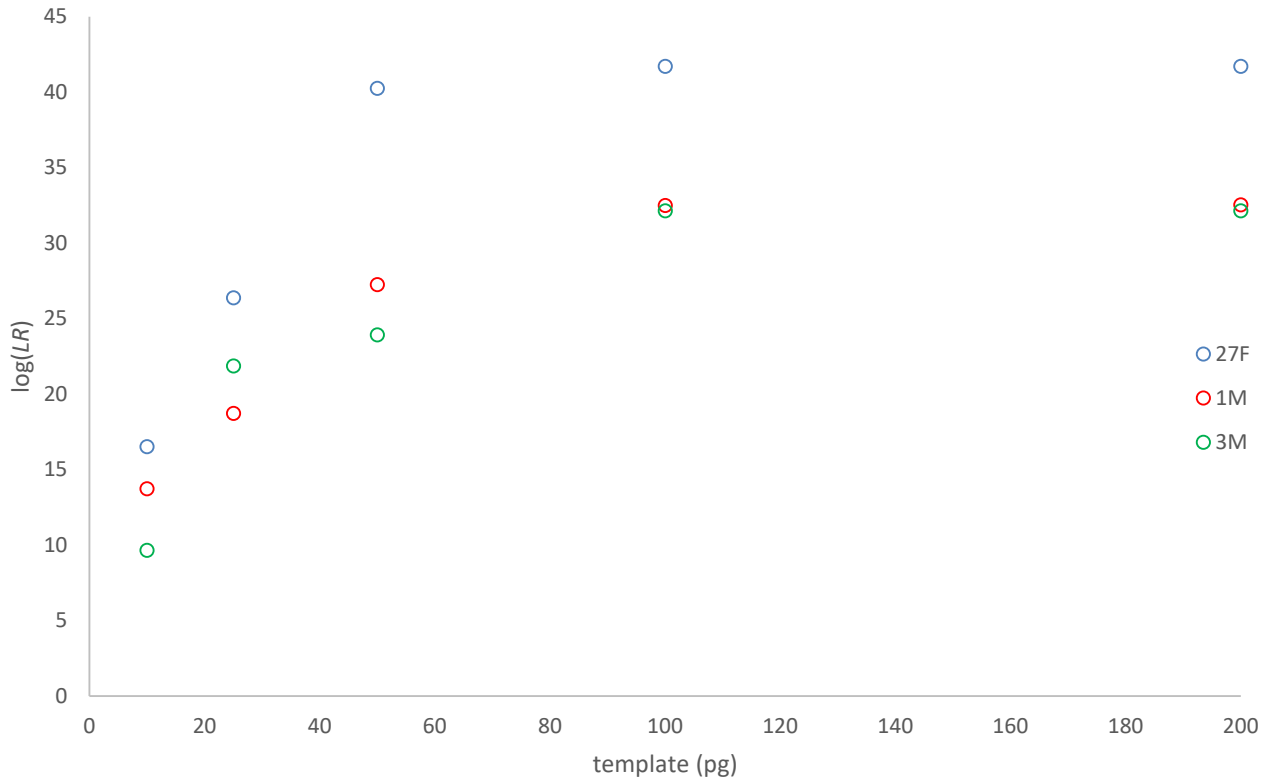
Experiments 1-3: Single Source Specimens

4.1.5 *Single-source specimens*

Experiment 1: A dilution series of single source profiles was constructed where the peak heights ranged from above the analytical threshold (AT) to below (ie had dropped out). Three single source samples were

amplified in PowerPlex® Fusion at the following input amounts: 200, 100, 50, 25, 10 pg, run on a 3130xl under standard conditions and analyzed with GeneMarker® HID version 2.8.2. The samples were analyzed with the laboratory's AT of 50 rfu. The samples were interpreted in STRmix™ and *LR*s calculated for the known contributors. The resultant *LR*s are plotted in Figure 1.

Figure 1: Plot of log(*LR*) versus input DNA (pg) for three single source dilution series



The known single source samples were run in STRmix™ with the known profile set as H_p and an unknown person set as H_d, i.e. *LR* = Comparison / Unknown. The resulting *LR*s show that STRmix™ is able to give expectedly high *LR*s (log(*LR*)>10E9) for known contributors to a single source sample.

Experiment 2: There is a small subset of profiles where the 'answer' is known or can be estimated easily [3]. These include single source profiles where the weight is one (or 100%) for each locus. The *LR* was calculated at each locus for five single source profiles (27F, 26F, 22F, 1M and 9M) and the individual locus *LR*s compared with the STRmix™ results. This was undertaken twice; once using an F_{ST} (or θ) value of 0 and once with F_{ST}=0.03. Setting θ to zero returns the product rule where:

$2p_ip_j$ for heterozygote loci

p_i^2 for homozygote loci

Where p_i is the allele frequency for allele *i*, p_j the allele frequency for allele *j*. When θ > 0, the Balding and Nichols [4] formulae (or equations 4.10 from NRCII [5]) are applied. For single source profiles:

$$\frac{2\left[\theta+(1-\theta)p_i\right]\left[\theta+(1-\theta)p_j\right]}{(1+\theta)(1+2\theta)}$$ for heterozygote loci [1]

$$\frac{\left[3\theta+(1-\theta)p_i\right]\left[2\theta+(1-\theta)p_i\right]}{(1+\theta)(1+2\theta)} \quad \text{for homozygote loci} \qquad [2]$$

Where p_i is the allele frequency for allele *i*, p_j the allele frequency for allele *j* and θ is the F_{ST} value. The allele frequency used within equations 1 and 2 are posterior mean frequencies. These are calculated using the following equation:

$$\frac{x_i+\frac{1}{k}}{N_a+1} \qquad [3]$$

Where x_i is the number of observations of allele *i* in a database, N_a is the number of alleles in that database and *k* is the number of allele designations with non-zero observations in the database.

The calculated and STRmix™ results for a single source profile for θ =0.00 and θ =0.03 are given in Table 1.

Table 1: 'By hand' (Excel) calculation of *LR* versus STRmix™ results for one of the five single source profiles (profile 27F) for the NIST Caucasian allele frequency dataset with varying F_{ST} values

Locus	Excel θ = 0	STRmix™ θ= 0	Excel θ = 0.03	STRmix™ θ= 0.03
D3S1358	11933.7793	11933.7793	61.493	61.493
D1S1656	28.7694	28.7694	21.849	21.849
D2S441	49.9506	49.9506	25.808	25.808
D10S1248	11.2967	11.2967	8.272	8.272
D13S317	35.7140	35.7140	22.022	22.022
Penta E	47.5313	47.5313	31.152	31.152
D16S539	10.1316	10.1316	7.583	7.583
D18S51	356.7607	356.7607	109.153	109.153
D2S1338	40.8770	40.8770	21.335	21.335
CSF1PO	7.7242	7.7242	6.080	6.080
Penta D	19.6534	19.6534	15.767	15.767
TH01	12.1815	12.1815	10.297	10.297
vWA	23.9850	23.9850	18.582	18.582
D21S11	15.5516	15.5516	13.103	13.103
D7S820	15.2550	15.2550	10.465	10.465
D5S818	226.9640	226.9640	37.877	37.877
TPOX	6431.1142	6431.1142	220.795	220.795
DYS391	-	-		
D8S1179	36.2402	36.2402	19.661	19.661
D12S391	91.3481	91.3481	45.599	45.599
D19S433	1794.6476	1794.6476	116.660	116.660
FGA	28349.4422	28349.4422	89.280	89.280
D22S1045	4.0777	4.0777	3.993	3.993
Total	5.126E+41	5.126E+41	2.725E+30	2.725E+30

The results in Table 1 show that STRmix™ is giving the expected answer based on the population genetic model being used. The NIST Caucasian allele frequencies [6] were used for the calculations.

An F_{ST} (θ) value of 0.03 will be for casework *LR* calculations using the NRC II 4.10 calculations. Studies have shown [5] that 0.01 F_{ST} values are adequately conservative for most populations with 0.03 offering an even greater value of conservatism and a lower *LR*.

Experiment 3: Off-scale peaks, saturation

4.1.4. *Allelic peak height, to include off-scale peaks*

Thirty-five single source profiles were amplified with above-optimal DNA input and analyzed in STRmix™ in order to review the impact of saturated data on profile interpretation. Ten different profiles were each amplified using 1.5 ng, 1 ng and 750 pg DNA and five profiles amplified using 2 ng DNA. The resultant profiles were interpreted in STRmix™. All interpretations resulted in intuitive genotypes where the weight = 1.0 for the known contributor genotype at each locus. A plot of the log(LR) for each profile is given in Figure 2. A plot of the per profile stutter variance (k^2) versus template (mean over the post burn-in accepts) for the 35 single source saturated profiles is given in Figure 3. As expected, the observed stutter variance (k^2) is higher for high template. This is because as alleles are more likely to be over the camera saturation limit (8000 rfu) their corresponding stutter peaks are larger than expected. In this case, the expected height of the stutter peak is calculated from the expected height of the allele and not the observed height which leads to slightly higher than expected variance between the observed and expected stutter peaks.

Figure 2: log(*LR*) for 35 profiles amplified after the addition of above-optimal DNA

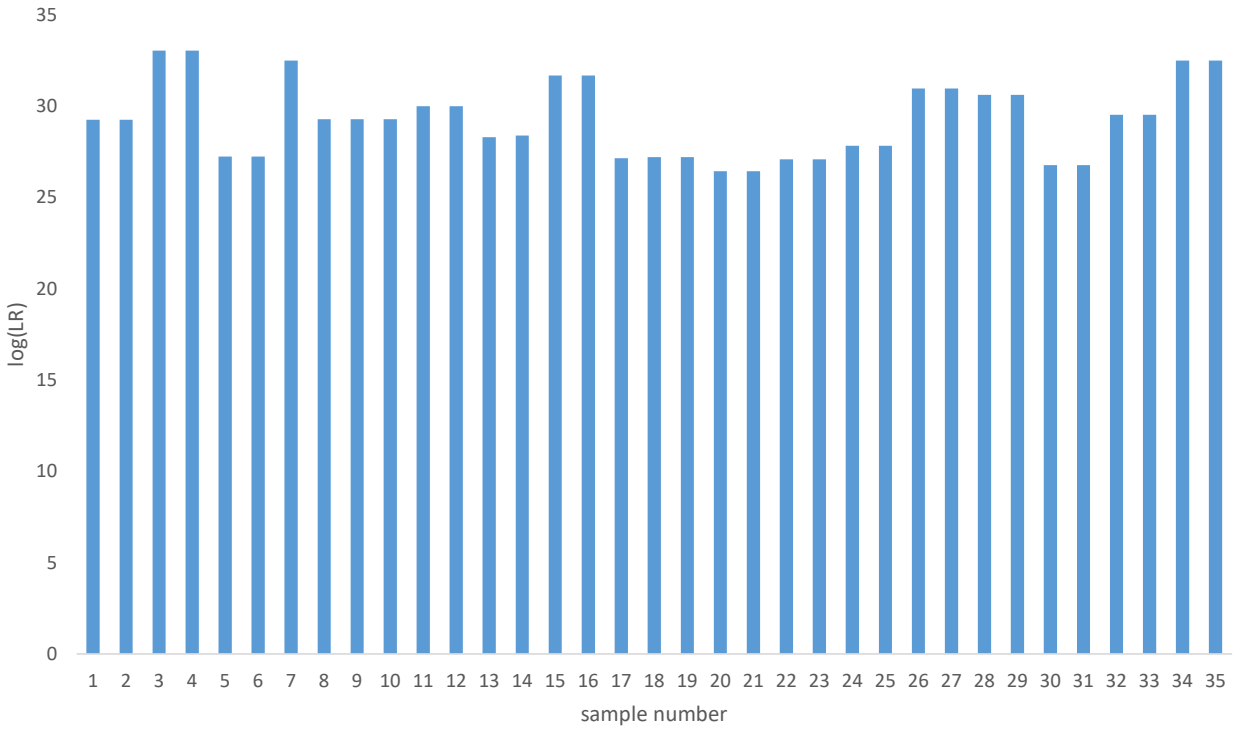
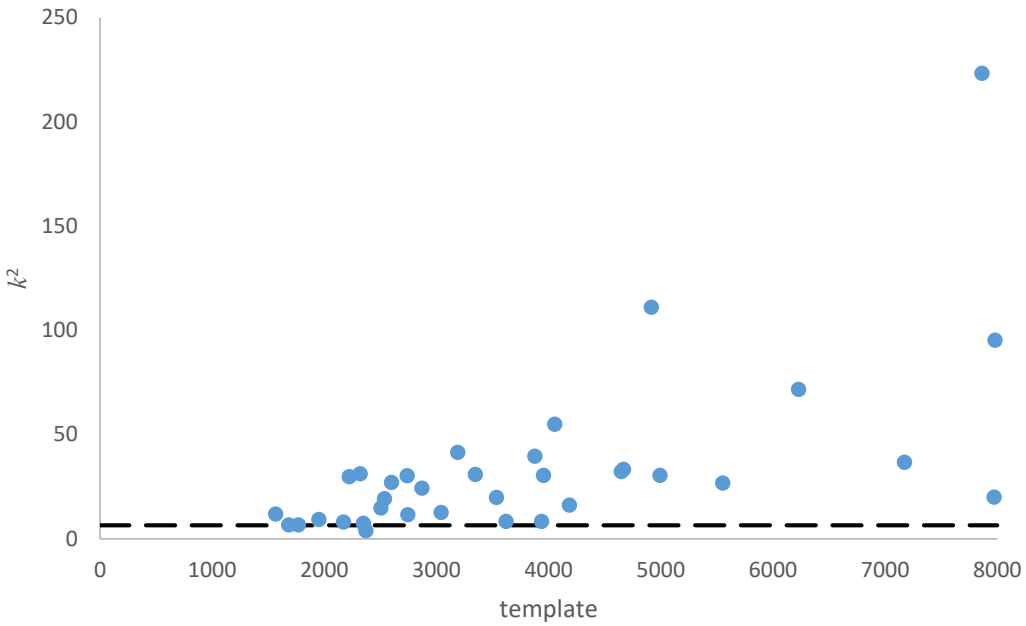


Figure 3: Per profile stutter variance (k^2) versus template (mean over the post burn-in accepts) for the 35 single source saturated profiles. The horizontal dashed line at 6.5 corresponds to the mode of the OCME stutter variance parameters, Γ(1.5007,12.9748)



Experiments 4-7: Specificity, Sensitivity, and Precision

4.1.1. *Specimens with known contributors, as well as case-type specimens that may include unknown contributors.*

4.1.2. *Hypothesis testing with contributors and non-contributors*

4.1.6. *Mixed specimens*

4.1.6.1. *Various contributor ratios (e.g., 1:1 through 1:20, 2:2:1, 4:2:1, 3:1:1)*

4.1.6.2. *Various total DNA template quantities.*

4.1.6.5. *Sharing of alleles among contributors.*

4.1.7. *Partial profiles, to include the following*

4.1.7.1. *Allele and locus drop-out*

4.1.13. *Sensitivity, specificity, and precision, as described for Developmental Validation*

With respect to DNA profile interpretation methods, sensitivity is defined as the ability of the software to reliably resolve the DNA profile of known contributors within a mixed DNA profile for a range of starting DNA templates. Specificity is defined as the ability of the software to reliably exclude non-contributors (H_d true) for a range of starting DNA templates. Specificity and sensitivity are tested by calculating the *LR* for a number of two, three and four-person profiles for both known contributors and known non-contributors.

Experiment 4: Various hypotheses were tested using two, three, and four-person mixtures. The contributor profiles contain homozygous and heterozygous loci, the amount of allele sharing between contributors varies, and the minor component of each mixture is expected to experience dropout as the DNA template amount decreases (4.1.6.5 and 4.1.7.1).

The following propositions were tested:

H_p: For each dilution series, the DNA originated from the person of interest and *N*-1 unknown individuals

H_d: For each dilution series, the DNA originated from *N* unknown contributors

where *N* is the apparent number of contributors (NOC) based on the observed profile. The true number of contributors to a crime sample is unknown and unknowable. In order to ensure the results are relevant to casework the apparent NOC was determined. The assignment of the number of contributors to a profile is complicated by allele sharing, artefacts such as stutter and pull up peaks below the analytical threshold. The effect of an incorrect assignment of the number of contributors to an interpretation is explored later in this document. The following process was followed when determining the apparent NOC:

1. The profile as reviewed as a whole, assessing the level of degradation, presence of low level peaks, noisy or clean baseline and general quality (template) of the profile

2. Likely stutter peaks were identified (both forward and back) by reference to OCME's per allele stutter ratio expectations (plots of SR ~Allele or within the stutter exceptions file)

3. The locus with the highest number of unambiguously allelic peaks, A, was identified. If A was an odd number, 1 was added. A/2 was the initial postulate of the number of contributors to the profile

4. Peak height imbalances were reviewed at the most informative locus (greatest number of alleles). Taking into account allele sharing or 'stacking', visually try to 'pair' alleles and assign to contributors. If there was too much imbalance between alleles this meant the likely presence of an additional contributor above that indicated by allele count alone.

5. If one or more contributors at this locus was either trace or a clear major, this pattern was checked to ensure it was represented at other loci.

6. The general pattern of contributors (number and proportion) was the applied to other loci in the profile. If it holds, that number of contributors was assigned to the profile otherwise consideration was given to the addition or subtraction of one. Consideration of genetic variants such as trisomies was also made if the general pattern did not hold across the profile.

The samples were deconvoluted in STRmix™ assuming the apparent NOC and compared to the known contributors and 10,000 known non-contributors using the 'Start and Search' function within STRmix™. The non-contributors were artificially generated using the published NIST Caucasian allele frequencies. The aim of this experiment is three-fold:

1- To test the limits of STRmix™; to determine the range of expected *LR*s for both true and non-contributors

2- To see the effects of DNA concentration on LR

3- To see if the weights of STRmix™ are intuitively correct.

The log(*LR*) for known contributors (H_p true) should be high and should trend downward to 0 as less information is present within the profile. Information includes amount of DNA from the contributor of interest, conditioning profiles (for example the victim's profile on intimate samples), replicates and decreasing numbers of contributors. For non-contributors (H_d true) the log(*LR*) should trend to 0 as less information is present within the profile.

A summary of the number of profiles interpreted with their known and apparent contributors is provided in Table 2.

Table 2: Number of profiles interpreted as part of the sensitivity and specificity experiment

		Apparent N			
		1	**2**	**3**	**4**
True N	**2**	2	214	2	0
	3	1	13	113	1
	4	1	8	11	35

The log(*LR*) values are plotted against the average peak height (*APH*) *per contributor* for the two, three, and four contributor mixtures in Figures 4 through 6, respectively. The APH per contributor value was used as this is the most comparable to the information an analyst will have with forensic casework and is therefore the most relevant explanatory variable to plot. A plot of log(*LR*) versus total input DNA (from quantitation) is also provided in Appendix 2.

The per contributor amount for H_d true contributors was taken as the lowest of the known contributors. The *APH* per known contributor was taken from the unmasked and unshared alleles. Where no DNA from the individual was detected within the profile, the APH was set to half the AT (25 rfu). The lowest contributor *APH* for each profile was used for the H_d true contributors.

Figure 4a: Log(*LR*) versus *APH* per contributor for two person mixtures amplified by the OCME laboratory.

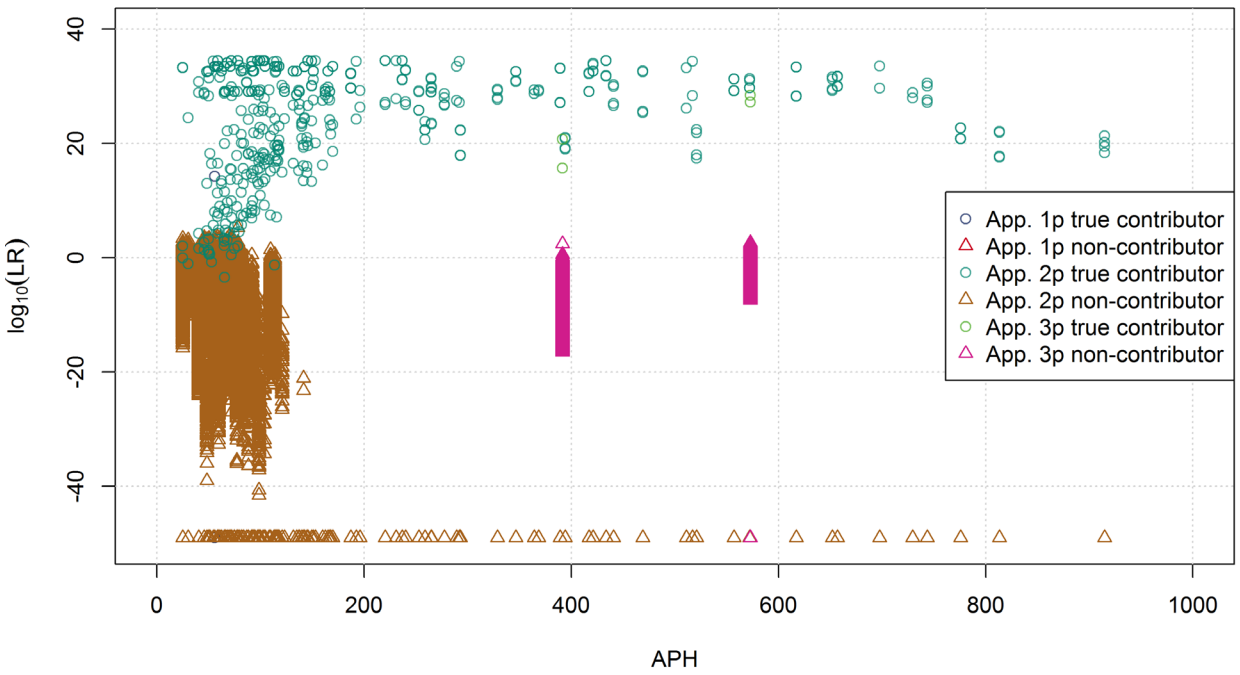


Figure 4b: Log(*LR*) versus *APH* per contributor (0 – 200 rfu x-axis) for two person mixtures amplified by the OCME laboratory.

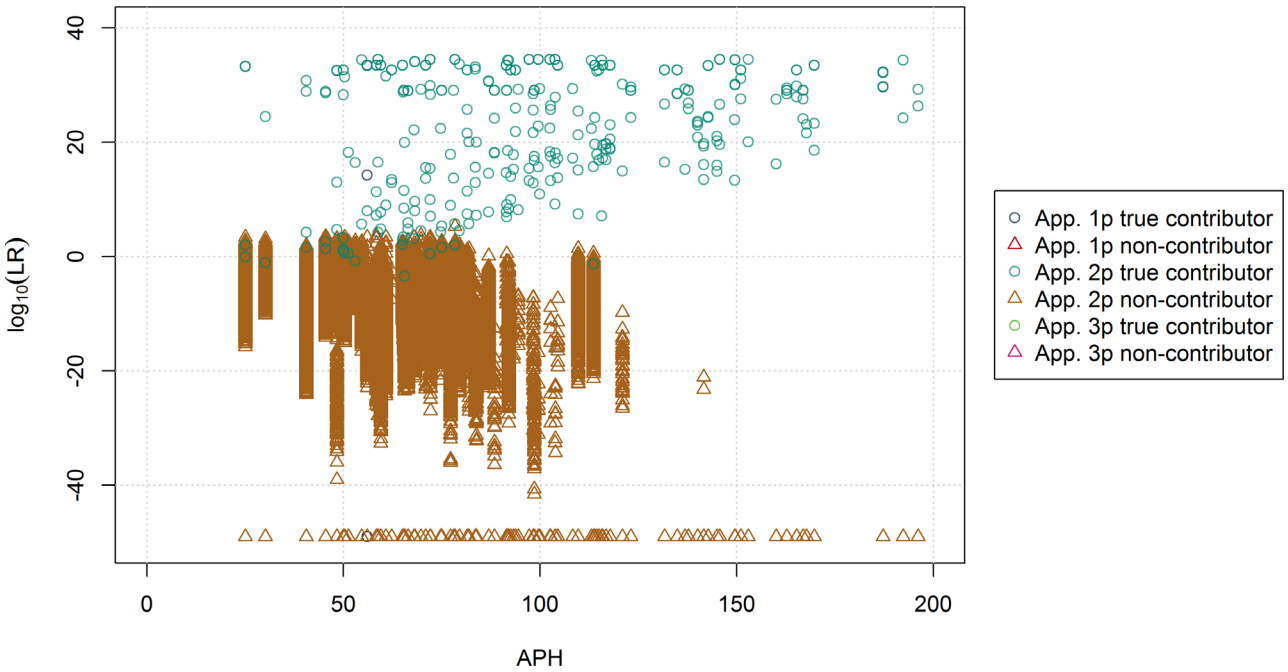


Figure 5a: Log(*LR*) versus APH per contributor for three person mixtures amplified by the OCME laboratory. An outlier (false exclusion) is highlighted

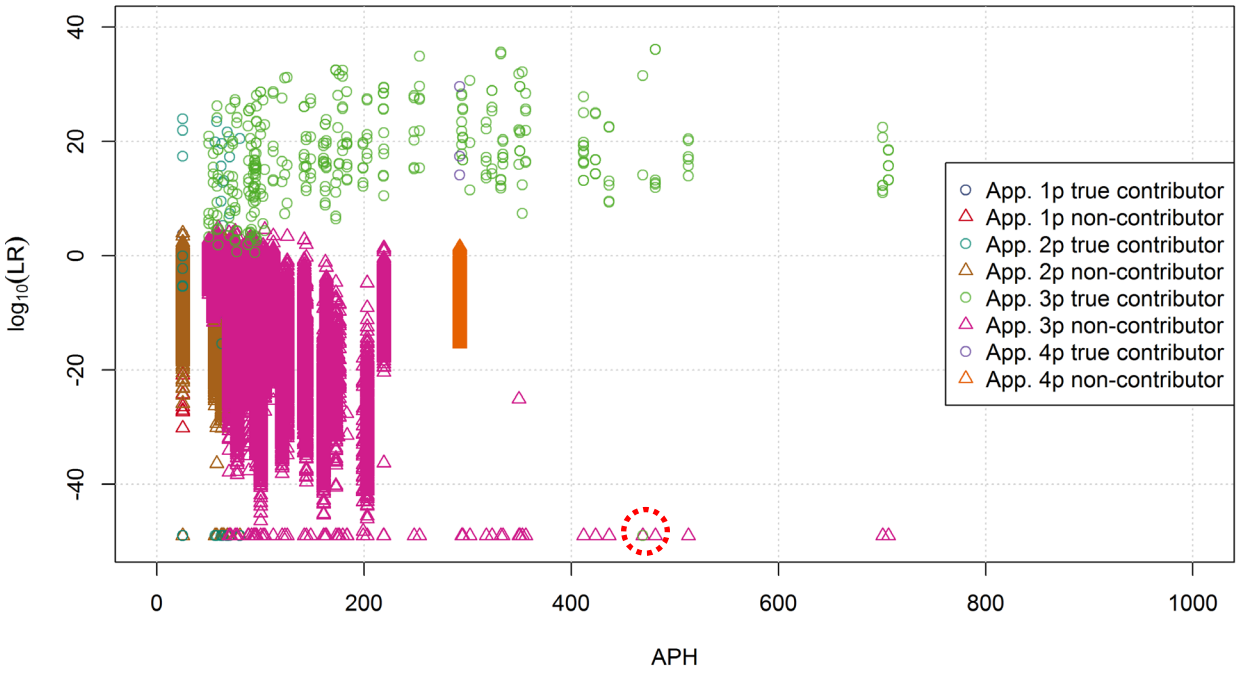


Figure 5b: Log(*LR*) for H_p true results versus APH per contributor (0 – 200 rfu x-axis logged scale) for three person mixtures amplified by the OCME laboratory.

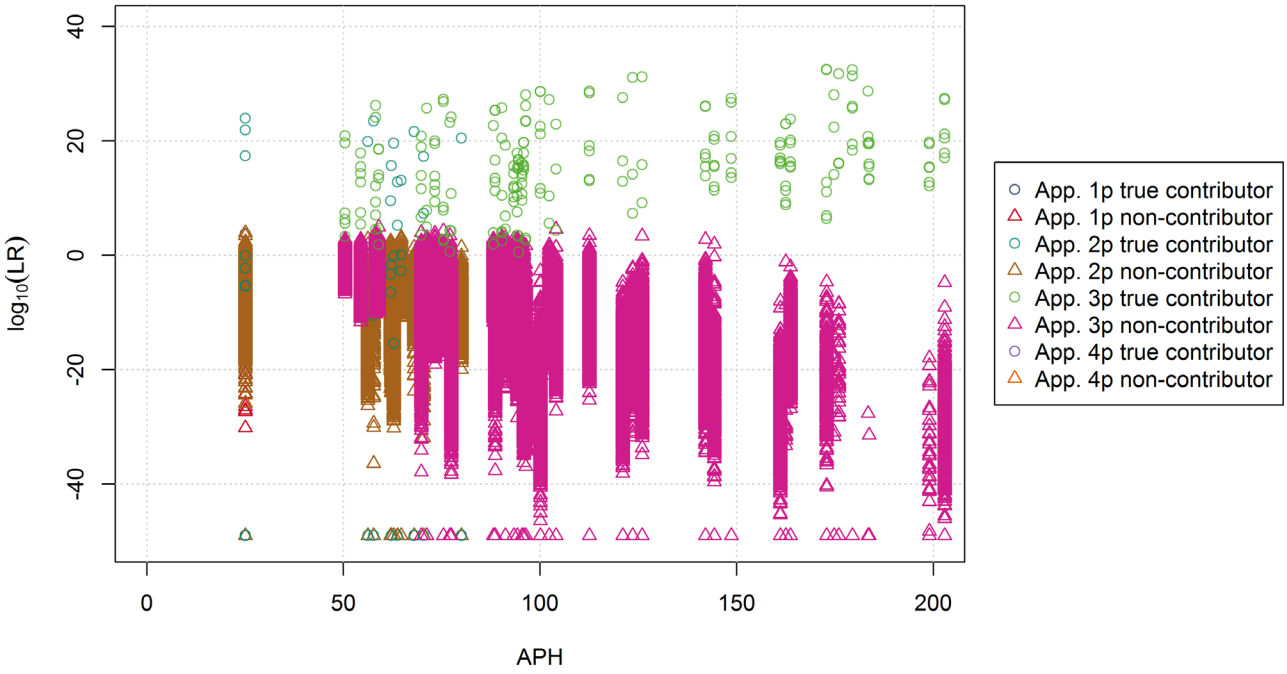


Figure 6a: Log(*LR*) versus APH per contributor for four person mixtures amplified by the OCME laboratory.

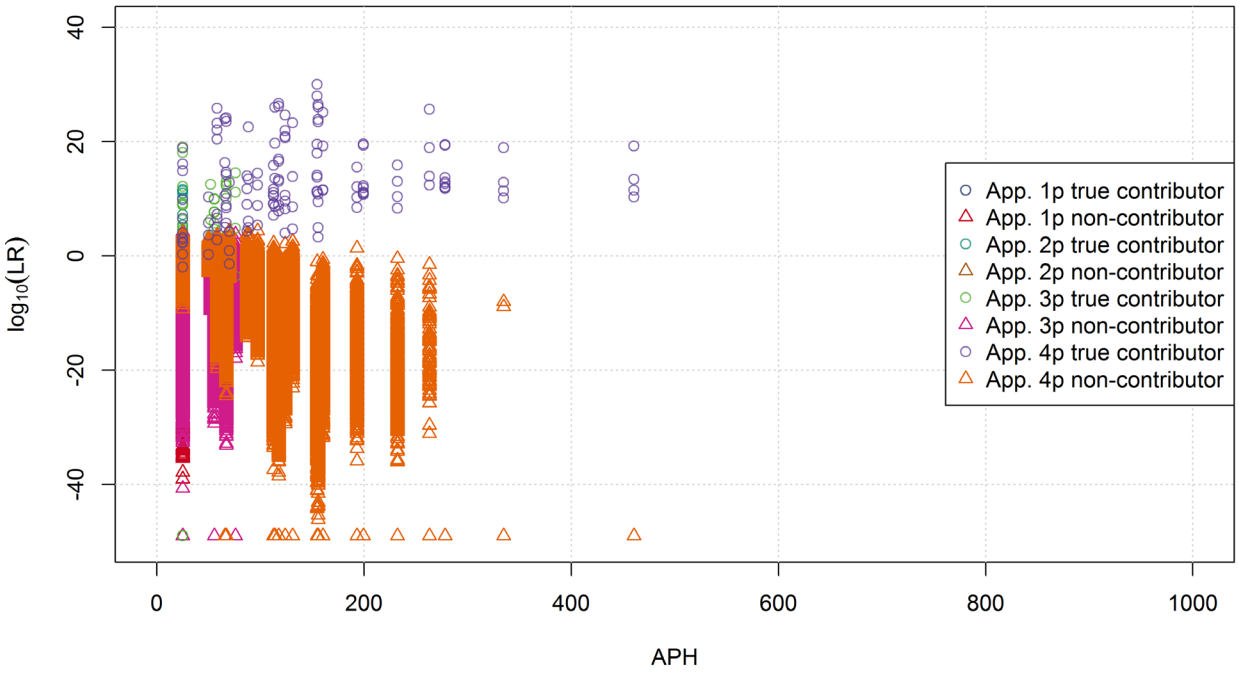
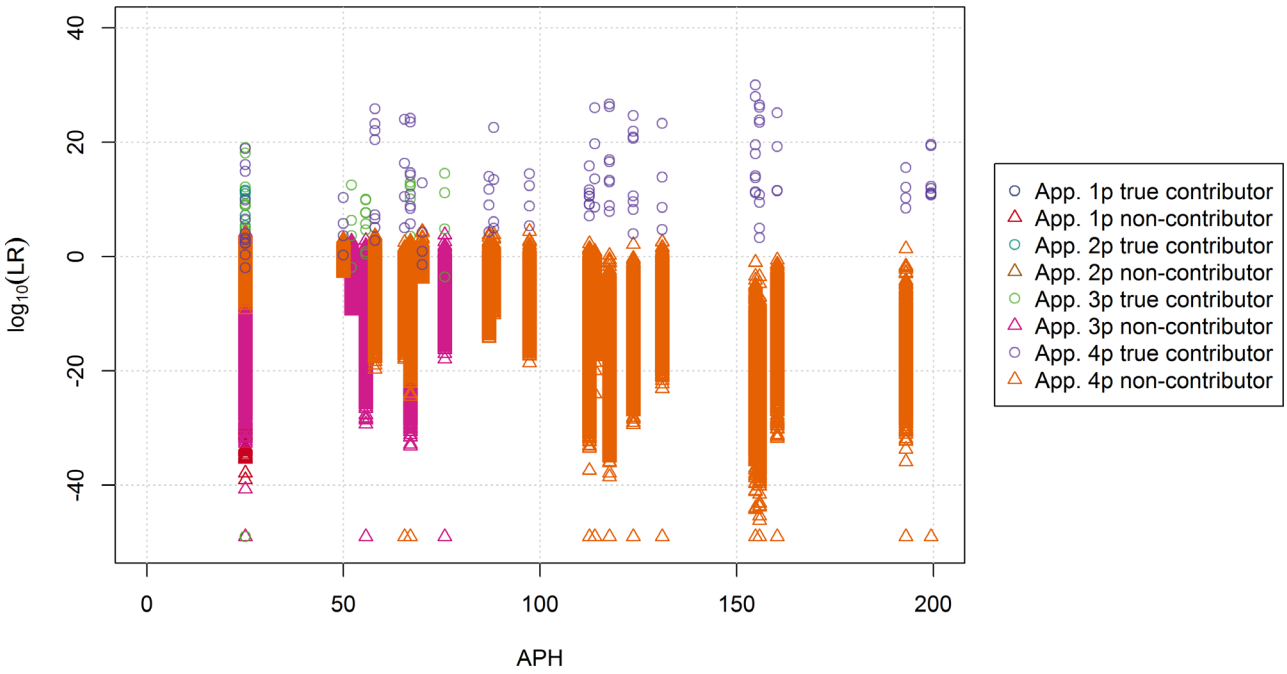


Figure 6b: Log(*LR*) versus APH per contributor (0 – 200 rfu x-axis logged scale) for four person mixtures amplified by the OCME laboratory.



Inspection of the plots in Figure 4 through 6 indicates that as expected at high template STRmix™ correctly and reliably resulted in high *LRs* for true contributors and low *LRs* for false contributors. A summary of the *LR*s for known contributors is provided in Table 3 and non-contributors in Table 4.

Table 3: Number of *LR*s less than 1 (exclusionary *LR*) and greater than 1 (inclusionary *LR*) for the known contributors

True N	# *LR*<1	# *LR*>1
2	7	435
3	26	358
4	31	189

The *LR*s for known contributors in support of exclusion are due to insufficient DNA from that contributor being present within the profile. This effect can be seen in Figures 4 through 6. There was one false exclusion in a three person mixture for a contributor with relatively high APH (sample 29-3p_CST8F_27M30_28M30_750pg_5-1-1). Inspection of the STRmix™ result indicated *LR* = 0 at both TH01 and D2S441 for the contributor 28M30. The relevant loci of the electropherogram are given in Figure 7. Inspection of TH01, indicates a 6.1 peak retained at analysis. This has been assigned to the minor contributor (28M30) at interpretation which is an exclusion. Inspection of D2S441 shows the likely incomplete separation of an 11.3/12 alleles. Reference 28M30 is an 11.3,12 at this locus. The profile was reinterpreted ignoring these loci resulting in a revised *LR* = 1.29E17. This result highlights the importance of ensuring that peaks within input files are correctly labelled prior to interpretation.

Figure 7: D2S441 and TH01 loci of sample 29-3p_CST8F_27M30_28M30_750pg_5-1-1

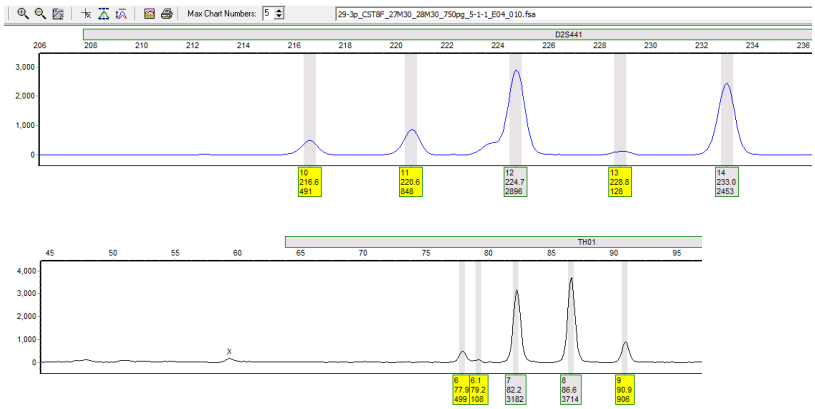
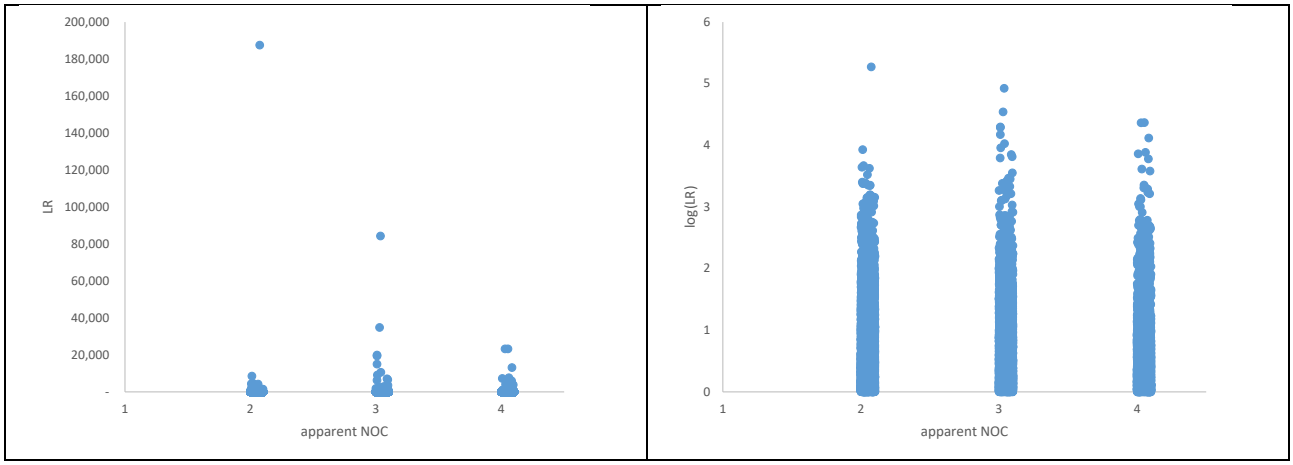


Table 4: Number of *LR*s less than 1 (exclusionary *LR*) and greater than 1 (inclusionary *LR*) for the non-contributors

True N	# *LR*<1	# *LR*>1
2	2,197,645	12,355
3	1,274,471	5529
4	540,012	9988

A plot of the *LR* > 1 for each apparent N is given in Figure 8 (logged and non-logged). There were two non-contributors that resulted in *LR* > 40,000.

Figure 8: Plot of *LR* > 1 for the non-contributor comparisons per apparent number of contributor. Note that the x-axis has been jittered to give a sense of the data



The highest false inclusion (*LR* = 187,504) was for the two-person mixture 38-M1_C4_100_15-1_6M_22F_Newton. Non-contributor number 3544 shared all six obligate (non-shared and not in stutter positions) minor alleles within this profile. It shared a further 21 alleles with the major or minor or their corresponding stutter peaks. The second highest false inclusion (*LR* = 84,283) was for the three-person mixture Newton53-M2-C4-100-5-5-1-17M-27F-9M. Non- contributor number 9515 shares 25 alleles with the known contributors' alleles or corresponding stutter peaks. These false positives are therefore the nature of the DNA profile and not a failure of STRmix™.

The *LR*s from the database search are the point estimate and contain neither a θ correction nor any correction for MCMC or allele probability uncertainty. Table 5 below contains the statistics for all false inclusions greater than *LR*=10,000 with these corrections applied . *LR*s were calculated using the NIST Caucasian allele frequencies.

Table 5: H_d true database calculated LRs >10,000 with theta (θ), factor of N! and 99.0% 1-sided lower HPD interval applied.

Sample	Database profile	Database *LR*	Factor of N! *LR* θ = 3%	99.0% 1-sided lower HPD interval
38-M1_C4_100_15-1_6M_22F_Newton	3544	187,504	20,019	10,581
53-M2-C4-100-5-5-1-17M-27F-9M_Newton	9515	84,283	6751	3928
88-3p_23M30-CST_14M-18F30_250pg_5-2-1_Newton	7927	34,823	6	4
07-M1_250pg_4-3-2-1_4M-14M-7F-23F_Newton	7595	23,286	165	83
13-M2_112.5pg_10-5-2-1_1M-7F-8F-26F_Athena	4705	23,210	338	193
55-M2_C4_100_1-2-1_17M_27F_9M_Athena	9706	19,891	156	91

55-M2_C4_100_1-2-1_17M_27F_9M_Newton	9706	19,403	72	37
14-M3_50pg_1-3-3-1_3M-17F-18F-29F_Newton	6288	14,962	11	6
13-M2_112.5pg_10-5-2-1_1M-7F-8F-26F_Athena	2821	13115	452	164
20-M1_C4_100_5-1-1_12F_13M_6M_Newton	2641	10,634	206	116

Ten out of the 27,872 *LR*s > 1 (0.036%) were above 10,000. 99.7% of all H_d true *LR*s > 1 were less than *LR* = 1000 and 98.0% were less than *LR* = 100. Out of the > 4 million H_d true *LR*s calculated, approximately three out of four are zero.

The tests simulating the situation where the POI is not a donor are called 'H_d true' tests. Good [7] (quoting Turing) stated "*the expected factor for a wrong hypothesis in virtue of any experiment is 1.*" Following from this statement we can make two observations [8]:

1. The average *LR* for the H_d true tests should be about 1.

2. The probability of observing a likelihood ratio of x or larger from an unrelated non-donor is less than 1 in x.

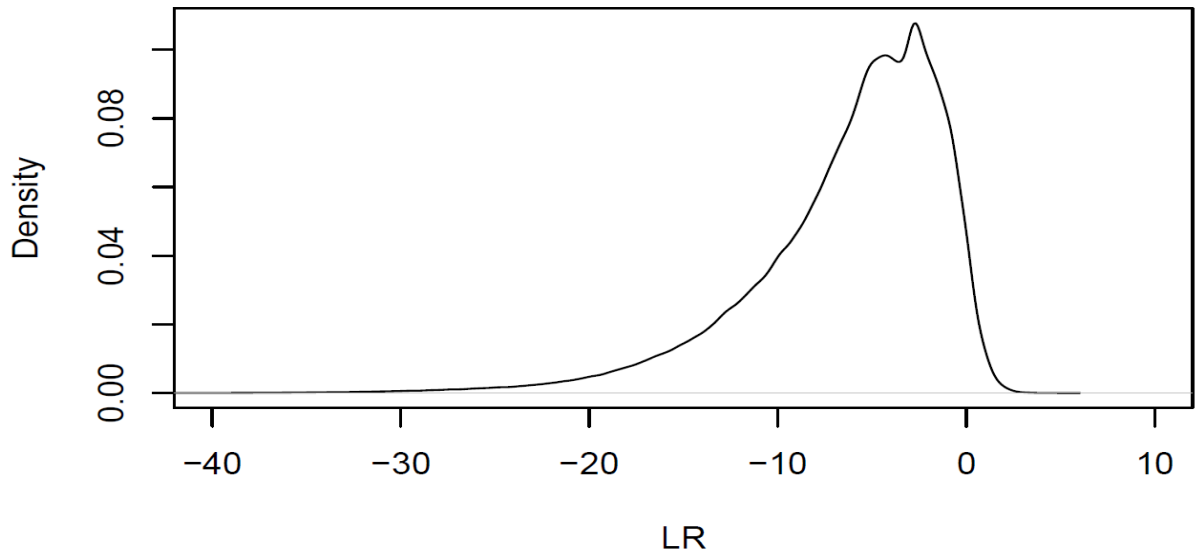
These two statements form the basis for assessing H_d true tests. In an experiment on 10,000 false donors (10,000 H_d true tests) we would expect at most about one *LR* ≥ 10,000, plausibly 10 above 1,000 and 100 above 100. This suggests that the tail that should be looked at is above about 10,000.

These occasional moderate *LR*s are no cause for concern, they are expected and occur when the false donor has the correct allele for the stain. The H_d true tests produce the point estimate *LR* with no θ correction. Hence the values are above those we would report.

The average *LR* for all H_d true comparisons equals 0.244. The observation of an average below 1 for this large number of comparisons suggests proper or even conservative performance.

The density estimates for the non-zero *LR*s is plotted below in Figure 9. This shows that the vast majority of non-zero *LR*s are small.

Figure 9: density estimates for the non-zero LRs.



For any given mixture there is a chance that a given non-donor will have sufficient matching alleles by chance to give a positive log(*LR*). This is the DNA false inclusion rate. The question is whether STRmix™ increases this rate by some error or misapplication. We have investigated this two ways:

1. The average *LR* for the false donors should be 1. In all tests done the average has been about 1 or less [2]. If STRmix™ was adding to the average *LR* by manufacturing falsely high LRs then this average would be markedly above 1.

2. We have investigated the two large *LR*s for false donors and they all have many alleles corresponding with the mixture.

We conclude that there is evidence that STRmix™ is not adding to the DNA false positive rate at all.

At low template or high contributor number STRmix™ correctly and reliably reported that the analysis of the sample tends towards an uninformative or inconclusive *LR*. As expected for the major donor, when the mixture has clear components (major/minor) the weightings for the genotype combinations are non-ambiguous and a strong inclusionary *LR* results. When the mixture proportions become more ambiguous, a decrease is seen in the *LR* where it is reasonable for increased uncertainty in assigning the alleles to a major or minor component.

These plots demonstrate the limits of the software/multiplex/laboratory combination, particularly the lower limits of DNA where an H_p true hypothesis results in a *LR* greater than 1 and the limit where false positives may arise (a *LR* greater than 1 where H_d is true). The results of these studies demonstrate the limits of the software by showing the range of *LR* values expected when H_p and H_d are true.

The primary diagnostics within STRmix™ used to diagnose the appropriateness of the interpretation are the genotypic weights and mixture proportions. Secondary diagnostics are the log(likelihood), Gelman-Rubin convergence diagnostic and the posterior mean of the allele and stutter variance parameters. A summary of the secondary diagnostics for the specificity and sensitivity runs is given in Appendix 1.

With over 3 million comparisons done to apparent 2- and 3-contributor samples for the H_d true experiment, a “false positive” rate of ~0.587% was calculated with a rate of ~0.001% for *LR*s greater than 1,000. This resulted in ~99.998% of all H_d true comparisons resulted in *LR*s <1,000. Because of this, the laboratory

"uninformative range" was set at 1,000. For any comparison, if the *LR* is 0.001 – 1,000, no support will be given to either hypothesis.

Table 6: Number and percentages of database calculated LRs in a given range for apparent 2 and 3 contributor mixtures.

	# of LRs <	% of LRs <	# of LRs >	% of LRs >
1	3588827	99.41349%	21173	0.58651%
100	3609534	99.98709%	466	0.01291%
1,000	3609946	99.99850%	54	0.00150%
10,000	3609993	99.99981%	7	0.00019%
100,000	3609999	99.99997%	1	0.00003%
1,000,000	3610000	100.00000%	0	0.00000%

Additional calculations were done for apparent 2 and 3 contributor mixtures with H_d true *LRs* greater than 1,000. Again, since the "start and search" function of STRmix™ returns point estimates with no θ or MCMC uncertainty corrections, the resulting *LRs* were much lower. Table 7 below shows that only two of the samples (38-M1_C4_100_15-1_6M_22F_Newton/Unknown 3544 and 53-M2_C4_100_5-5-1_17M_27F_9M_Newton/Unknown 9515) resulted in *LRs* still above 1,000. Based on the discussion above, this is due to the nature of the profile (minor profile sharing common alleles at lower amounts of input DNA) and not due to a failure of STRmix™.

Table 7: H_d true database calculated LRs >1000 re-calculated with 99.0% 1-sided lower HPD interval for apparent 2 and 3 contributor mixtures.

Sample	Database profile	Database LR	99.0% 1-sided lower HPD interval (Cau)
38-M1_C4_100_15-1_6M_22F_Newton.csv	3544	187504	9570
53-M2_C4_100_5-5-1_17M_27F_9M_Newton.csv	9515	84283	3710
88-3p_23M30-CST_14M-18F30_250pg_5-2-1_Newton.csv	7927	34823	3
55-M2_C4_100_1-2-1_17M_27F_9M_Athena.csv	9706	19891	109
55-M2_C4_100_1-2-1_17M_27F_9M_Newton.csv	9706	19403	36
14-M3_50pg_1-3-3-1_3M-17F-18F-29F_Newton.csv	6288	14962	5
20-M1_C4_100_5-1-1_12F_13M_6M_Newton.csv	2641	10634	114
56-M2_C4_100_3-2-1_17M_27F_9M_Newton.csv	4539	9126	31
37-3p_CST8F_27M30_28M30_37.5pg_5-1-1_Newton.csv	6542	8527	89
23-M1_C4_100_1-2-1_12F_13M_6M_Athena.csv	4977	7102	23
23-M1_C4_100_1-2-1_12F_13M_6M_Athena.csv	2641	6529	32
08-M1_125pg_4-3-2-1_4M-14M-7F-23F_Newton.csv	3925	6235	7
68-2p_CST_5F-CST_13M_37.5pg_4-1_Newton.csv	4456	4686	217
15-M3_37.5pg_1-3-3-1_3M-17F-18F-29F_Athena.csv	6288	4419	14
83-M2_C3_250_1-15_15M_5F_Athena.csv	7371	4245	297
20-M1_C4_100_5-1-1_12F_13M_6M_Athena.csv	2641	3576	77
85-M2_C4_100_20-1_15M_5F_Newton.csv	4212	3338	283
13-M1_C3_250_5-1-1_12F_13M_6M_Athena.csv	2641	2926	23
20-M1_C4_100_5-1-1_12F_13M_6M_Athena.csv	4977	2832	47

82-3p_23M30-CST_14M-18F30_100pg_5-1-1_Newton.csv	7587	2615	3
83-3p_23M30-CST_14M-18F30_75pg_5-1-1_Newton.csv	2911	2523	31
44-M1_C4_100_1-4_6M_22F_Athena.csv	8456	2444	62
85-M2_C4_100_20-1_15M_5F_Newton.csv	9156	2442	235
79-3p_23M30-CST_14M-18F30_250pg_5-1-1_Newton.csv	2592	2418	3
23-2p_24F30-29M30_250pg_2-1_Newton.csv	5156	2403	96
90-2p_24F30-29M30_37.5pg_4-1_Newton.csv	7280	2401	296
93-M2_C4_100_1-10_15M_5F_Newton.csv	3049	2257	186
94-M2_C4_100_1-15_15M_5F_Newton.csv	2555	2255	172
24-2p_24F30-29M30_150pg_2-1_Newton.csv	2114	2218	19
14-M3_50pg_1-3-3-1_3M-17F-18F-29F_Newton.csv	1890	2144	38
52-M2_C4_100_5-1-1_17M_27F_9M_Newton.csv	9706	1926	18
53-M2_C4_100_5-5-1_17M_27F_9M_Newton.csv	9808	1850	67
23-M1_C4_100_1-2-1_12F_13M_6M_Athena.csv	4171	1847	10
08-M3_50pg_1-1-1-1_3M-17F-18F-29F_Newton.csv	4450	1638	47
15-M3_37.5pg_1-3-3-1_3M-17F-18F-29F_Athena.csv	3901	1572	189
21-M3_50pg_1-3-5-1_3M-17F-18F-29F_Athena.csv	4282	1562	50
52-M2_C4_100_5-1-1_17M_27F_9M_Athena.csv	8929	1514	10
15-M3_37.5pg_1-3-3-1_3M-17F-18F-29F_Athena.csv	90	1453	106
83-M2_C3_250_1-15_15M_5F_Athena.csv	409	1432	131
85-M2_C4_100_20-1_15M_5F_Newton.csv	9762	1426	123
45-3p_CST8F_27M30_28M30_37.5pg_5-2-1_Newton.csv	3527	1405	23
45-M1_C4_100_1-10_6M_22F_Athena.csv	1607	1356	136
53-M2_C4_100_5-5-1_17M_27F_9M_Newton.csv	5969	1348	54
80-3p_23M30-CST_14M-18F30_150pg_5-1-1_Newton.csv	2139	1283	11
20-M1_C4_100_5-1-1_12F_13M_6M_Newton.csv	4977	1254	12
47-M1_C4_100_1-20_6M_22F_Newton.csv	8235	1239	141
71-2p_CST_5F-CST_13M_250pg_2-1_Newton.csv	9049	1152	28
23-2p_24F30-29M30_250pg_2-1_Newton.csv	2114	1131	43
21-M3_50pg_1-3-5-1_3M-17F-18F-29F_Athena.csv	2626	1123	27
22-M3_37.5pg_1-3-5-1_3M-17F-18F-29F_Athena.csv	3256	1111	40
89-3p_23M30-CST_14M-18F30_150pg_5-2-1_Newton.csv	7927	1075	0.2
45-M1_C4_100_1-10_6M_22F_Athena.csv	8206	1058	99
15-M3_37.5pg_1-3-3-1_3M-17F-18F-29F_Athena.csv	4031	1037	90
20-M3_100pg_1-3-5-1_3M-17F-18F-29F_Athena.csv	833	1010	17

The data used above contained various sets of samples, some of which were identical amplifications re-injected on two different instruments. Although these would not be categorized as replicates, the data demonstrates that there is no instrumentation variability on the results. This also demonstrates, as discussed above, that the H_d true *LRs*>1 are due to the DNA and not the STRmix™ software.

4.1.2.1. *The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses. For example, if there are two*

persons of interest, they may be evaluated as co-contributors and/alternatively, as each contributing with an unknown individual. The hypotheses used for evaluation of casework profiles can have a significant impact on the results obtained.

Experiment 5: In addition to the above experiments, STRmix™ was run with varying scenarios as follows: Mixtures will be assessed through STRmix™ with known contributors as "assumed" contributors. 2, 3, and 4 person mixtures of varying ratios will be run through STRmix™ with 1 or multiple contributors as "known assumed" contributors and these *LR*s will be compared with the *LR*s when the contributors are run as "unknown".

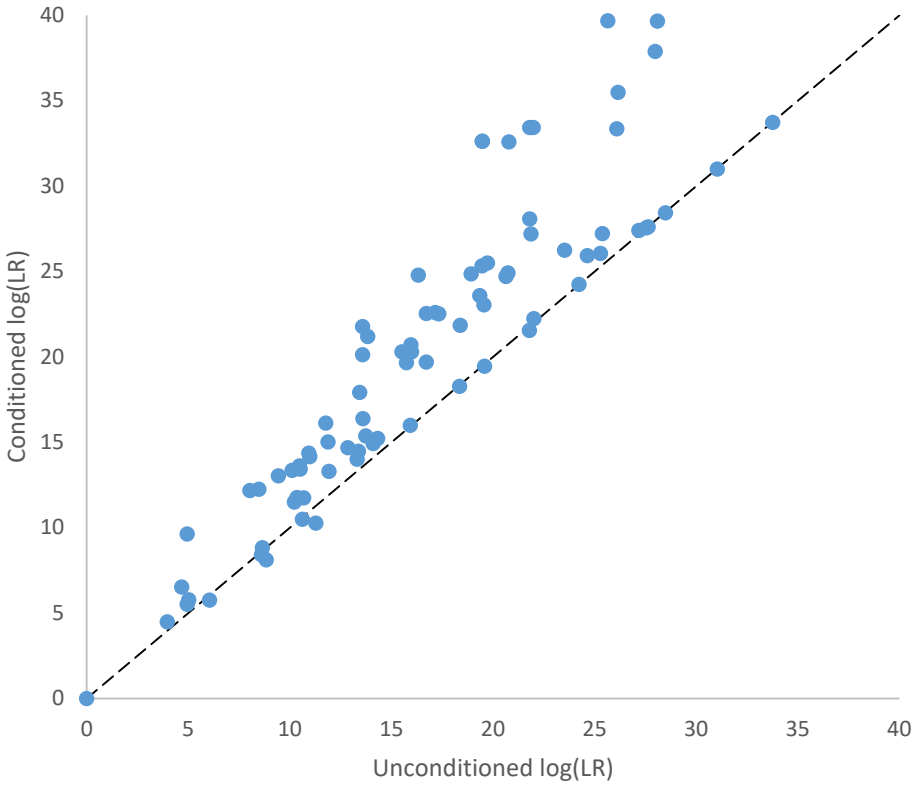
The effect of the addition of relevant information at interpretation has been shown to increase the *LR* for H_p true and reduce the *LR* for H_d true propositions [9]. Each of the two, three and four person profiles were re-interpreted in STRmix™ to test the effect of the assumption of the major contributor under both the prosecution and defence propositions. The following propositions were tested for *N* contributor mixtures:

H_p: The DNA originated from the major, the person of interest and *N*-2 unknown individual/s

H_d: The DNA originated from the major and *N*-1 unknown individuals

The log(*LR*) after conditioning on the major contributor was compared against the original log(*LR*) after having assumed no contributors. A plot of the log(*LR*) for each interpretation is provided in Figure 10.

Figure 10: Log(*LR*) after conditioning on the major contributor versus log(*LR*) with no conditioning



Values above the line at $x=y$ indicate that the *LR* increased when conditioning on, or assuming, the major contributor. This is the expected outcome. The addition of correct conditioning profiles (known major

contributor under both H_p and H_d) improves the performance of the tests. This is discussed further in Taylor [8].

Experiment 6: Precision testing (Reliability)

The MCMC process is used to generate the weights within STRmix™ for different genotype combinations. This is a sampling procedure and therefore the weights will vary slightly between each run. The variability in *LR*s between replicate interpretations has previously been explored [10]. The MCMC process was shown to be a small source of variability compared with other lab variables including the PCR and CE process. The variability due to the size of the allele frequency database and the MCMC process is taken into account within STRmix™ V2.4 using the highest posterior density (HPD) method [11-13] (a type of confidence interval).

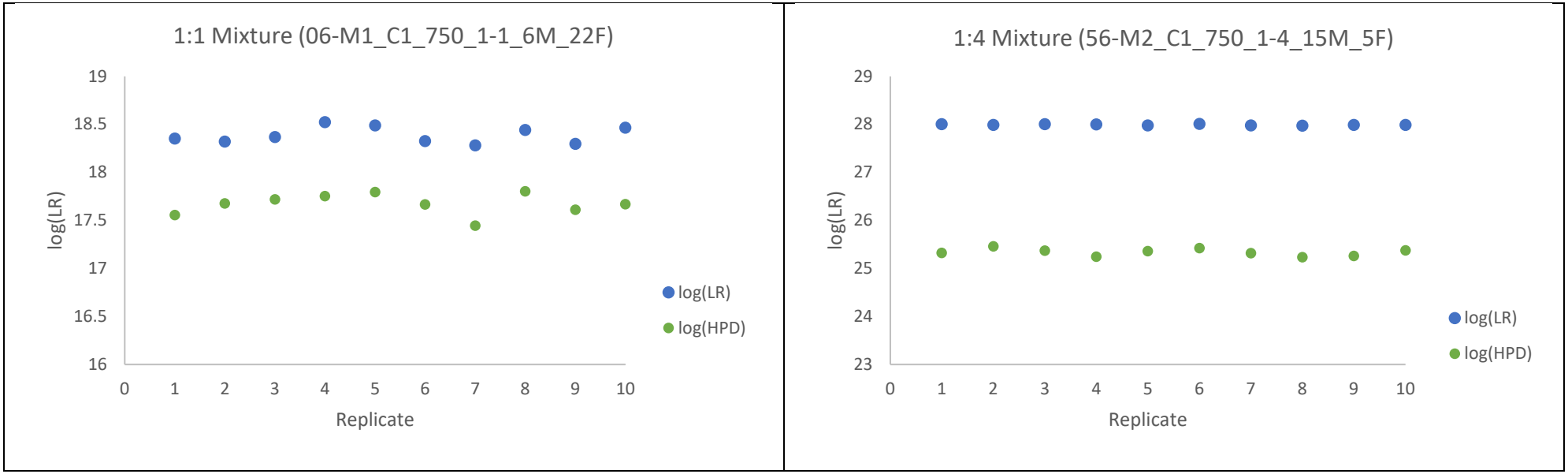
The extent of this variability was investigated by interpreting the following profiles:

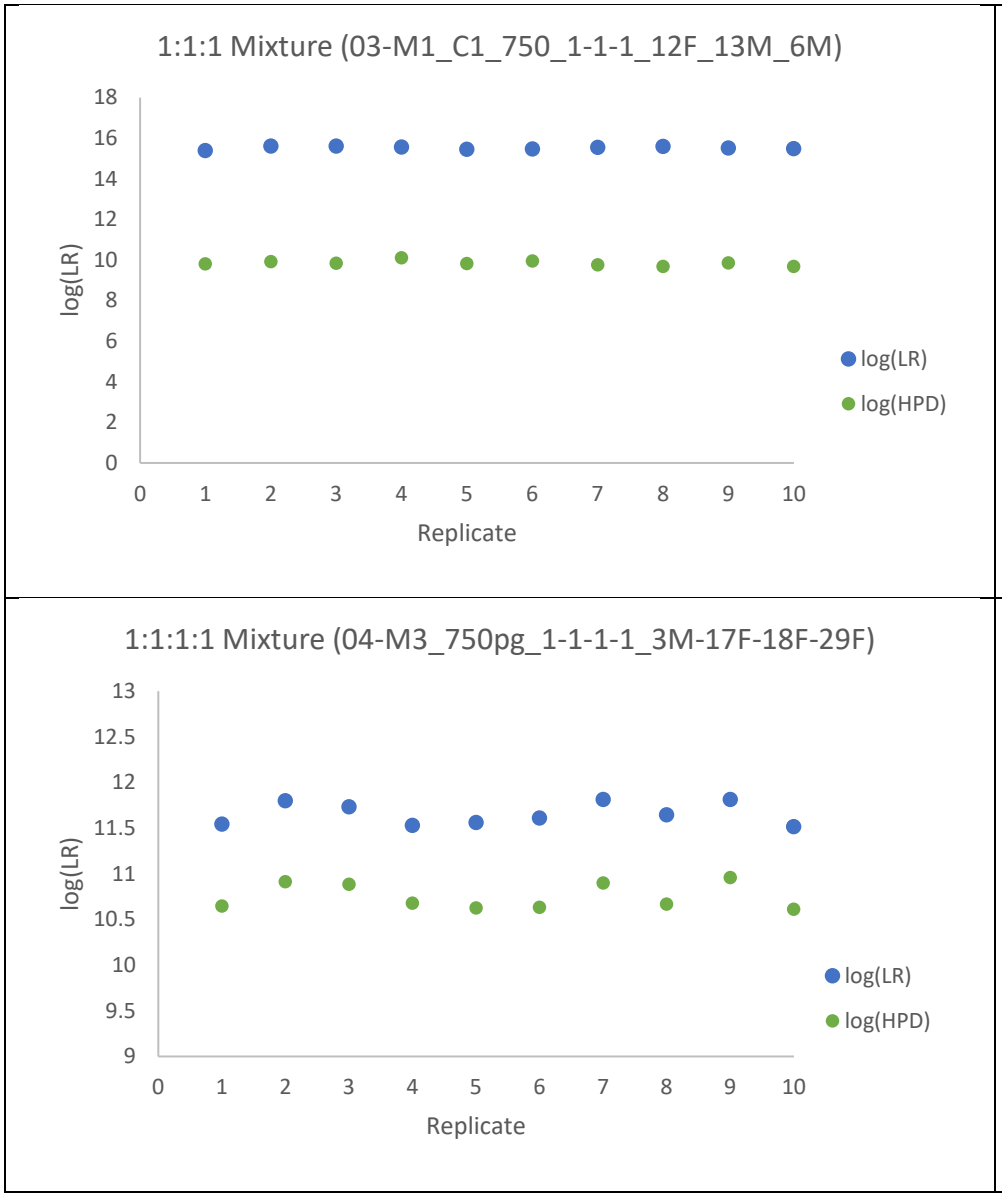
- Two of the two-person mixed DNA profiles (06-M1_C1_750_1-1_6M_22F and 56-M2_C1_750_1-4_15M_5F)
- Two of the three-person mixed DNA profiles (03-M1_C1_750_1-1-1_12F_13M_6M and 01-M1_C1_750_5-1-1_12F_13M_6M) and
- Two of the four-person mixed DNA profiles (04-M3_750pg_1-1-1-1_3M-17F-18F-29F and 12-M2_450pg_10-5-2-1_1M-7F-8F-26F)
- An additional interpretation was done using the 1:1:1 three-person mixture 03-M1_C1_750_1-1-1_12F_13M_6M. This additional interpretation was done with 1,000,000 total MCMC accepts (please note that the default setting for each of the other precision experiments was 500,000 total MCMC accepts)

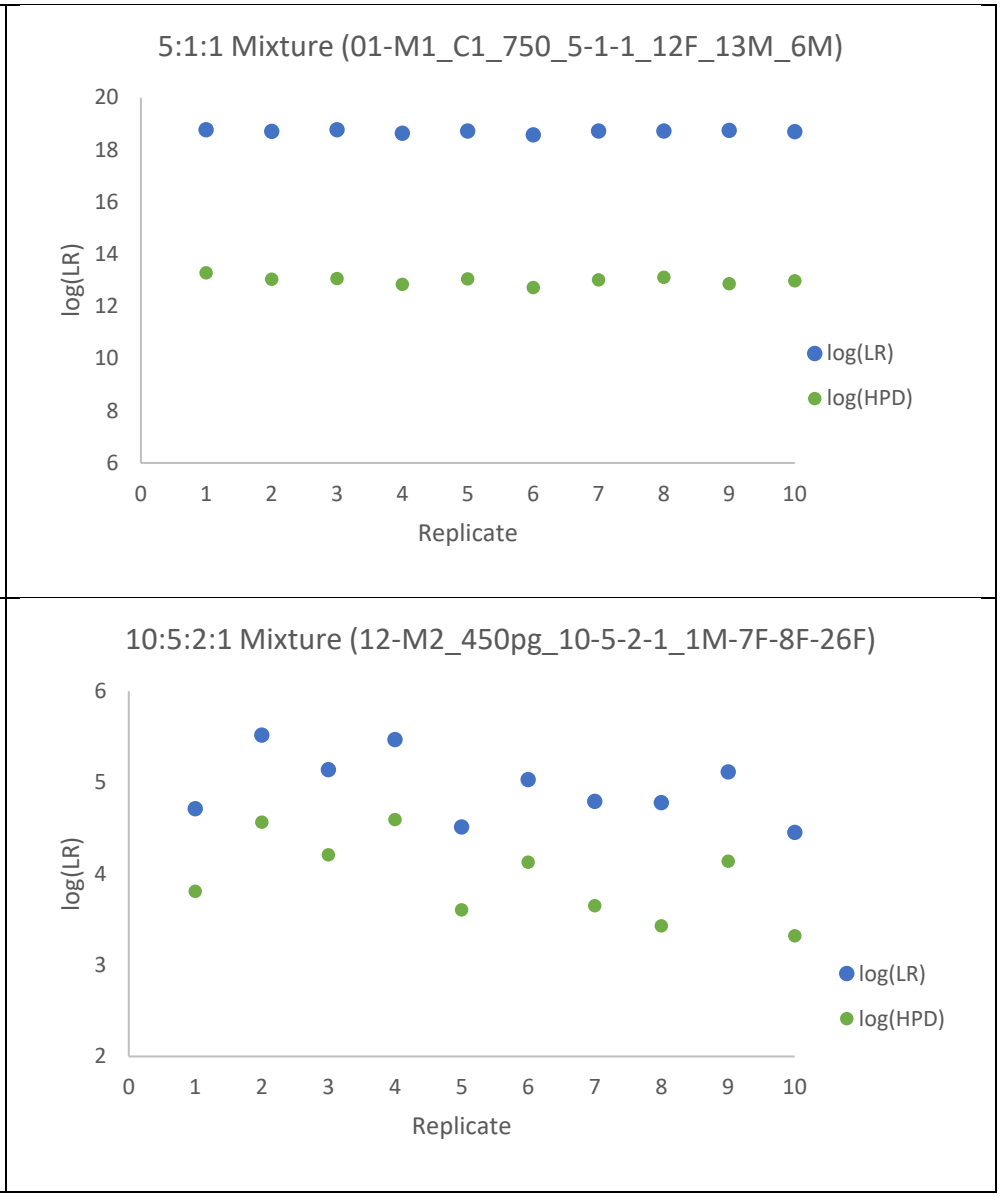
There was ambiguity in the genotype combinations of the minor contributor. In each instance the DNA profiles selected were interpreted ten times. A plot of log(*LR*) for each replicate is given in Figure 11. The blue dots indicate the *LR* values and the orange dots are the lower 99% bound of the HPD.

The results will demonstrate the variability in the *LR*s due to the MCMC process. They demonstrate that the variability is random and the values obtained for the various runs remain close; within one order of magnitude except for the four-person mixture. This is consistent with the model described by the developer. To further account for the variability of the allele frequency database and the MCMC process, STRmix™ V2.4 uses the highest posterior density (HPD) method, which is a type of confidence interval.

Figure 11: Plot of replicate log(*LR*) interpreted using STRmix™ for six different two, three and four person mixtures (minor contributors)

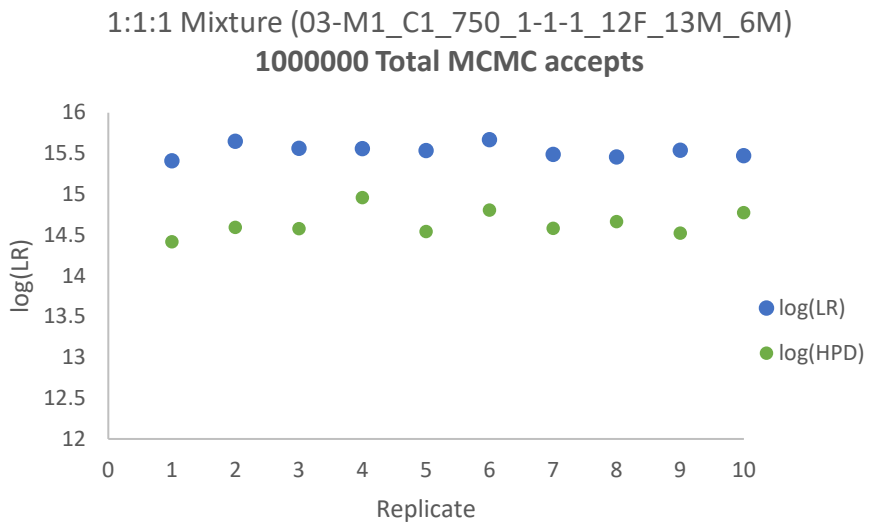



1:1:1 Mixture (03-M1_C1_750_1-1-1_12F_13M_6M)
log(LR)
Replicate
log(LR)
log(HPD)


5:1:1 Mixture (01-M1_C1_750_5-1-1_12F_13M_6M)
log(LR)
Replicate
log(LR)
log(HPD)

1:1:1:1 Mixture (04-M3_750pg_1-1-1-1_3M-17F-18F-29F)
log(LR)
Replicate
log(LR)
log(HPD)

10:5:2:1 Mixture (12-M2_450pg_10-5-2-1_1M-7F-8F-26F)
log(LR)
Replicate
log(LR)
log(HPD)


1:1:1 Mixture (03-M1_C1_750_1-1-1_12F_13M_6M)
1000000 Total MCMC accepts
log(LR)
Replicate
log(LR)
log(HPD)
16
15.5
15
14.5
14
13.5
13
12.5
12
0
1
2
3
4
5
6
7
8
9
10

Experiment 7: Case-type samples

4.1.1. *Specimens with known contributors, as well as case-type specimens that may include unknown contributors.*

Thirty two mock evidence samples were interpreted within STRmix™ and the Start and Search function used to compare to a database of known contributors and 10,000 non-contributors. The profiles were:

- 5 QiaCube Validation Mock Evidence Samples
- 5 QiaCube Validation Proficiency Test Samples
- 9 Pseudo Bottle Samples
- 13 Touched Item Samples.

Profiles were analysed on one or both 3130xl instruments. The .fsa files were interpreted in GeneMarker®HID and a stratified *LR* calculated. A summary of the results is given in Appendix 3. Where multiple reference profiles were submitted for the same profile they were both considered under the prosecution proposition. The alternate proposition for all interpretations was all unknown individuals.

The result of all comparisons was as expected. All H_p true samples (known contributors) gave high *LRs* while the H_d true comparisons (non-contributors) gave very low or 0 *LRs*. Sample 035136_40-Mock_11 gave the lowest *LR* value but that can be attributed to the nature of this low template sample (dropout, low APH, elevated stutter variance, etc.). This is showing that STRmix™ is performing as expected with mock samples by separating the known contributors from the known non-contributors.

Experiment 8: Inspection of the weights: the observation of the decrease in ability of deconvolution of mixtures and the simultaneous decrease in weights of individual genotypes in STRmix™ analyses.

4.2.1.3. *Generally, as the analyst's ability to deconvolute a complex mixture decreases, so do the weightings of individual genotypes within a set determined by the software.*

Two 2 person mixture series (both male:female mixtures) constructed in the following ratios: 15:1, 10:1, 4:1, 2:1, and 1:1 with total amounts of DNA of 500pg were amplified at standard conditions (PP Fusion, 29 cycles), run on a 3130xl, and analyzed in GeneMarker®HID v2.8.2, with an analytical threshold of 50rfu. First, the resulting mixtures were interpreted "blind" by two different analysts with the purpose of deconvolution of the major and minor contributors. The analysts were provided with electropherograms of the data with stutter filters both "on" and then "off", using the default stutter filter settings for PowerPlex® Fusion from the manufacturer, and a table of the average peak height ratios for single source samples run on two 3130xl instruments at DNA input amounts ranging from 0pg to 1ng. Second, the data was analysed in STRmix™ and component profiles were determined for both the major and the minor contributors containing all alleles reaching a ≥99% weighting. The profiles obtained from the manual deconvolutions were compared with those determined by STRmix™. The number of alleles deconvoluted by each analyst was graphed in comparison to the number of alleles deconvoluted by STRmix™, for both the major and minor, at each ratio interpreted. The results showed that as the ability to deconvolute a mixture decreased the called alleles for each contributor also decreased for both the two analysts and STRmix™.

Figure 12: Deconvoluted called alleles plotted against mixture contribution percentage for major component.

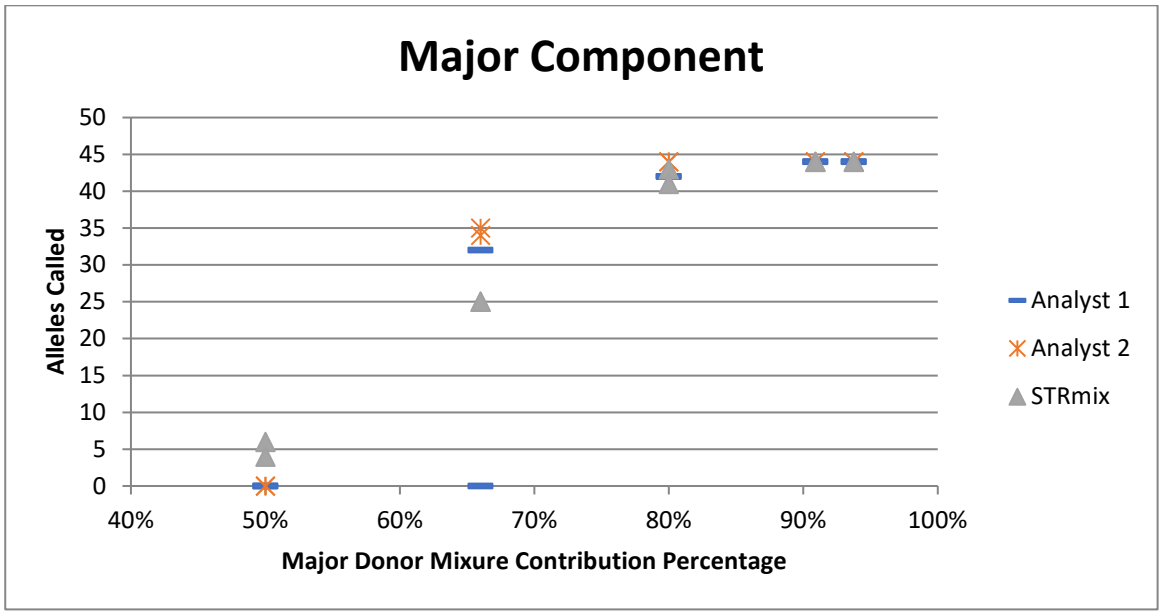
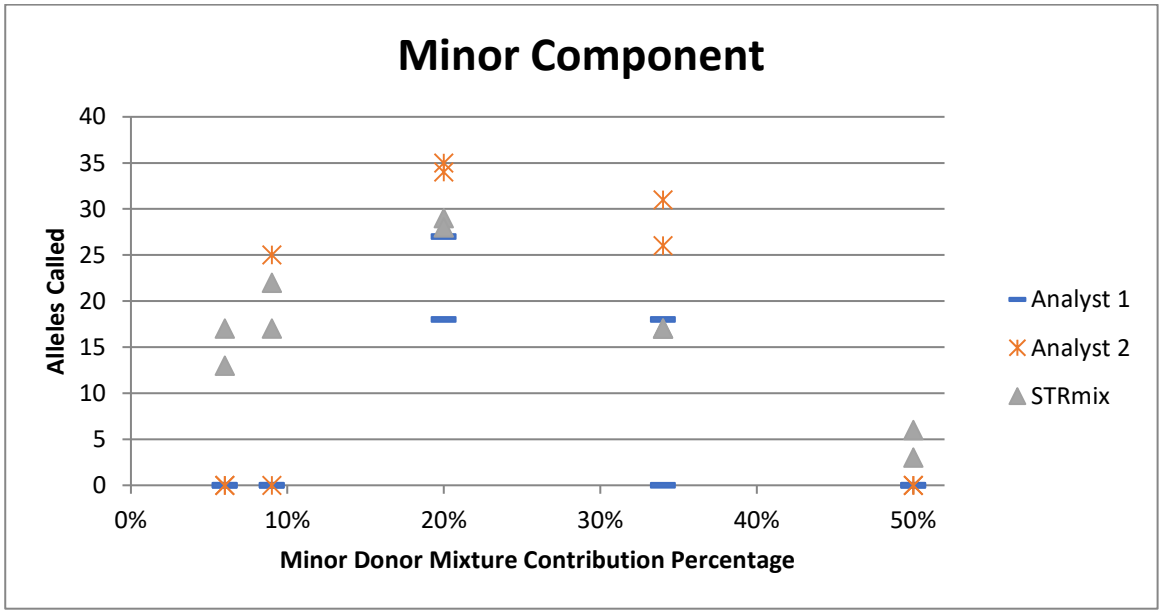


Figure 13: Deconvoluted called alleles plotted against mixture contribution percentage for minor component.



Experiment 9: The Effect of Replicates on STRmix™ *LR* values

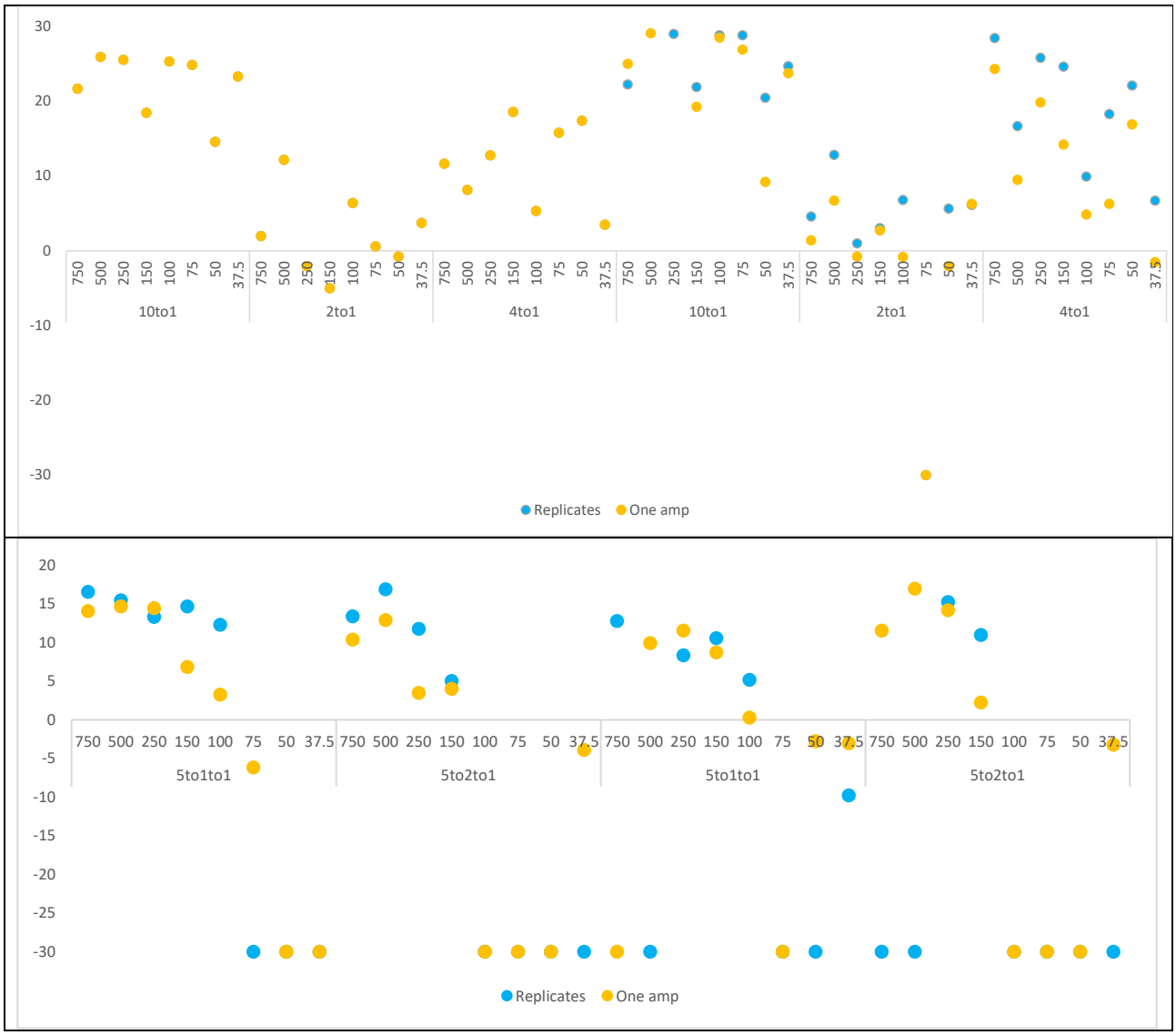
The Forensic Biology laboratory of the NYC OCME often performs repeat amplifications of DNA samples (replicates) in order to assist in deconvolutions and to confirm results seen in casework. The purpose of this experiment is to see how STRmix™ *LRs* change with the input of replicate data.

In addition, the following question will be posed: is there a threshold of DNA template input amount below which samples should automatically be replicated in order to obtain improved STRmix™ *LR* values?

Initial and replicate amplifications of known two- and three-person samples were interpreted together as replicate interpretations and an *LR* calculated for both the minor and major contributors. The profiles were interpreted assuming apparent NOC. In total, one profile was interpreted as a single source profile, 60 as two-person mixtures and 28 as three person mixtures. The *LRs* are compared to the *LRs* calculated from Experiment 4 using only one of the replicates for the major contributor (Figure 14) and for the minor contributor (Figure 15).

Figure 14: Log(*LR)* for major contributor for one amplification versus replicate amplifications (two-person mixtures top pane, three-person mixtures bottom pane)

Figure 15: Log(*LR)* for minor contributor for one amplification versus replicate amplifications



The one exclusion in the major three-person plot, as well as the majority of false exclusions for the minor contributor in the three-person mixtures is due to allelic drop-out. Due to low amounts of input DNA, the apparent NOC was often underestimated. Because the person of interest was most frequently the minor contributor in each of these profiles and is likely to be the first contributor to drop-out, we expect exclusionary *LRs*. More outright exclusions (*LR* =0) were seen for replicate interpretations than for single amplification interpretations as the second replicate is not adding any additional information for the minor contributor.

Three false exclusions were seen even though there was sufficient DNA information for the minor to have not dropped out. This can be attributed to incomplete separation of a 11.3/12 at D2S441. The 11.3 has not been resolved from the major 12 allele within both replicates, likely due to poor separation at CE. In the first replicate, more drop-out was modelled for the minor contributor and therefore a false exclusion did not result. With the addition of the second replicate a false exclusion resulted as less drop-out was modelled.

The interpretation of replicate amplifications has been shown to increase the *LR* for known contributors in the majority of situations where all peaks within a profile have been separated and called correctly. Because of this, OCME will attempt to run replicate amplifications with samples as needed.

Experiment 10: testing STRmix™ with hypotheses including different numbers of contributors, N, N-1 and N+1

4.1.6.3. *Various numbers of contributors. The number of contributors evaluated should be based on the laboratory's intended use of the software. A range of contributor numbers should be evaluated in order to define the limitations of the software.*

4.1.6.4. *If the number of contributors is input by the analyst, both correct and incorrect values (i.e., over- and under-estimating) should be tested.*

The true number of contributors to a profile is always unknown. Analysts are likely to add contributors in the presence of an artefact, high stutter, or forward stutter peaks. The assumption of one fewer contributor than that actually present may be made when contributors are at very low levels and dropping out (or visible below the analytical threshold), in constructed profiles where DNA is from individuals with similar profiles at the same concentrations, or family scenarios, such as DNA from a father, mother and their child where the child was the minor contributor.

The effect of the uncertainty in the number of contributors has previously been reported for a number of profiles with *N* and *N*+1 assumed contributors, where *N* is the number of contributors [14, 15]. When assuming *N*+1, many more low level adventitious matches were generated assuming the wrong number of contributors. The inclusion of an additional contributor beyond that present in the profile had the effect of lowering the *LR* for trace (low-level) contributors within the profile. STRmix™ adds the additional (unseen) profile at trace levels which interacts with the known trace contribution, diffusing the genotype weights and lowering the *LR*. There was no significant effect on the *LR* of the major or minor contributor within the profiles.

In order to assess the effect on the *LR* that under- and over-estimation of contributors has on STRmix™ results, the following profiles were interpreted.

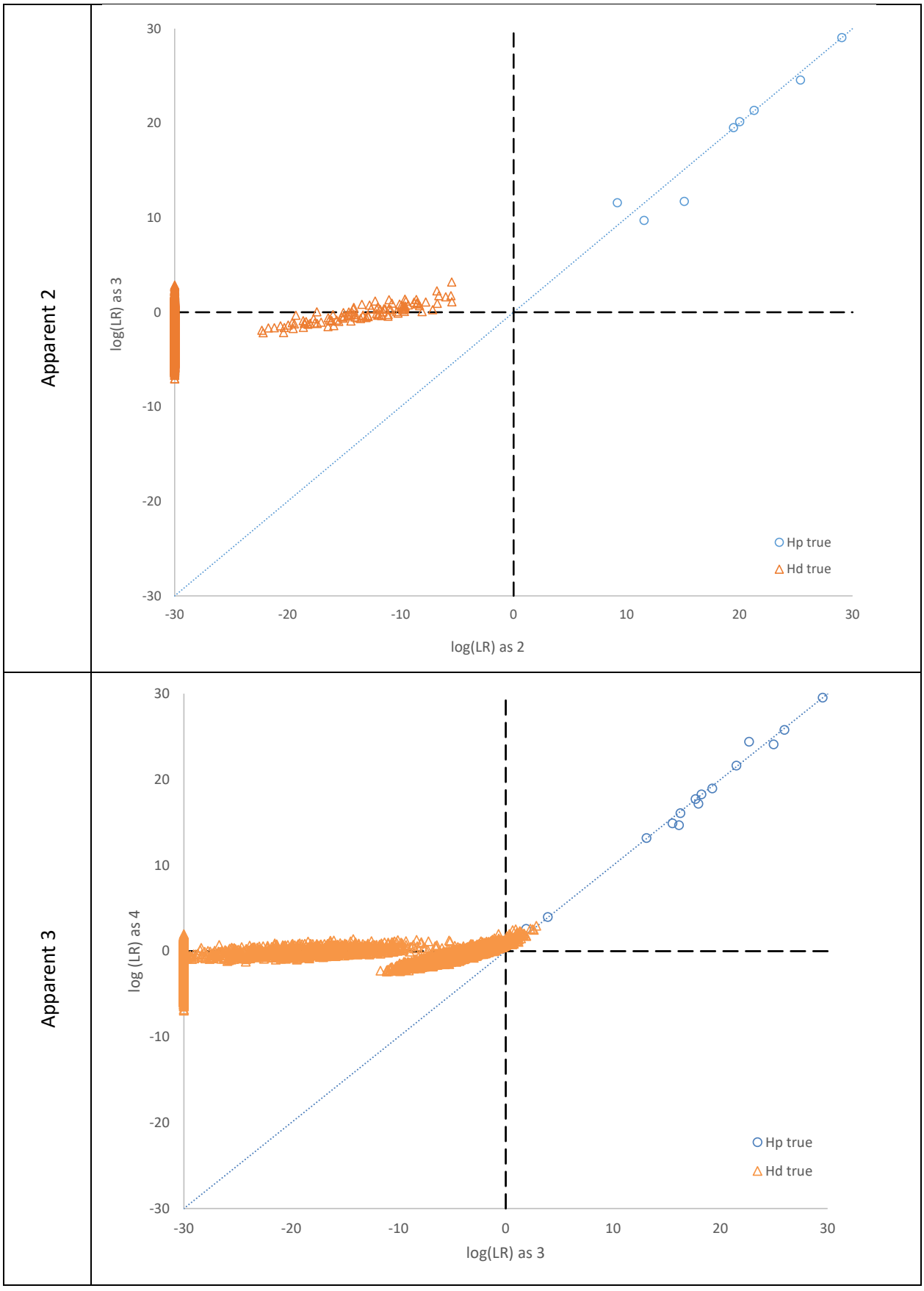
Table 9: Profiles with apparent and interpreted number of contributors

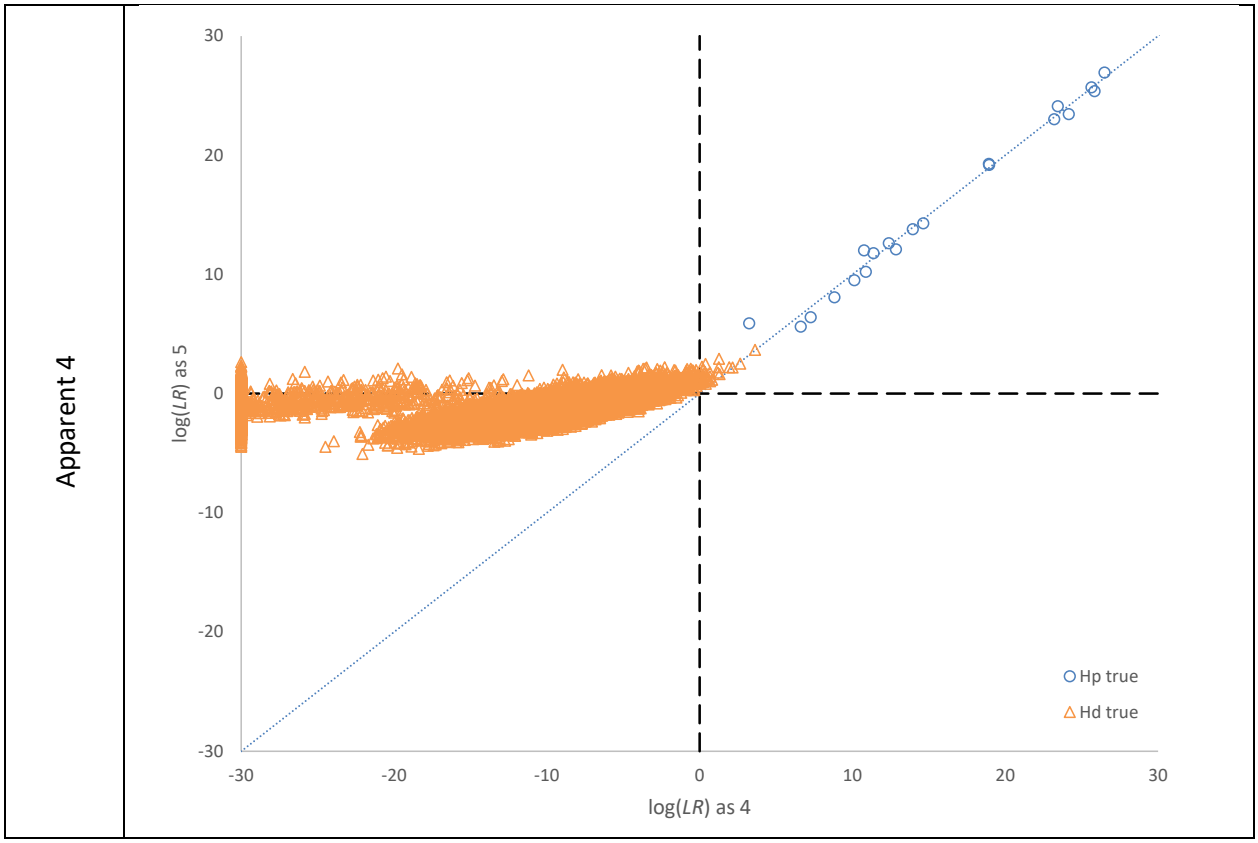
		Interpreted as N+1	Interpreted as N-1
Apparent N	2	01-M1_C1_750_20-1_6M_22F	94-M2_C4_100_1-15_15M_5F
		04-M1_C2_500_10-1_6M_22F	86-M2_C4_100_15-1_15M_5F
		75-M2_C3_250_15-1_15M_5F	47-M1_C4_100_1-20_6M_22F
		06-M1_C1_750_1-1_6M_22F	26-2p_24F30-29M30_75pg_2-1
		88-M2_C4_100_4-1_15M_5F	-
	3	34-3p_CST8F_27M30_28M30_100pg_5-1-1	25-M1_C4_100_5-2-1_12F_13M_6M
		43-M2_C2_500_3-2-1_17M_27F_9M	45-3p_CST8F_27M30_28M30_37.5pg_5-2-1
		01-M1_C1_750_5-1-1_12F_13M_6M	20-M1_C4_100_5-1-1_12F_13M_6M
		09-M1_C2_500_1-1-1_12F_13M_6M	44-3p_CST8F_27M30_28M30_50pg_5-2-1
		19-M1_C3_250_5-2-1_12F_13M_6M	37-3p_CST8F_27M30_28M30_37.5pg_5-1-1
	4	04-M1_1000pg_4-3-2-1_4M-14M-7F-23F	08-M1_62.5pg_4-3-2-1_4M-14M-7F-23F
		19-M3_250pg_1-3-5-1_3M-17F-18F-29F	08-M1-125pg-4-3-2-1-4M-14M-7F-23F
		16-M3_750pg_1-3-5-1_3M-17F-18F-29F	13-M2-225pg-10-5-2-1-1M-7F-8F-26F
		12-M3_250pg_1-3-3-1_3M-17F-18F-29F	21-M3_50pg_1-3-51_3M-17F-18F-29F
		10-M2_1800pg_10-5-2-1_1M-7F-8F-26F	09-M3_37.5pg_1-1-1-1_3M-17F-18F-29F

For over-estimation of contributors, (*N*+1) contributors:

The *LR* was calculated for both the known contributors and 10,000 non-contributors (as undertaken for Experiment 4: Sensitivity and Specificity). The *LR* for the known contributors and known non-contributors assuming N+1 is compared to the original *LR* assuming apparent N contributors within Figure 16.

Figure 16: Comparison of log(*LR*) for apparent N versus log(*LR*) for apparent N+1 contributors

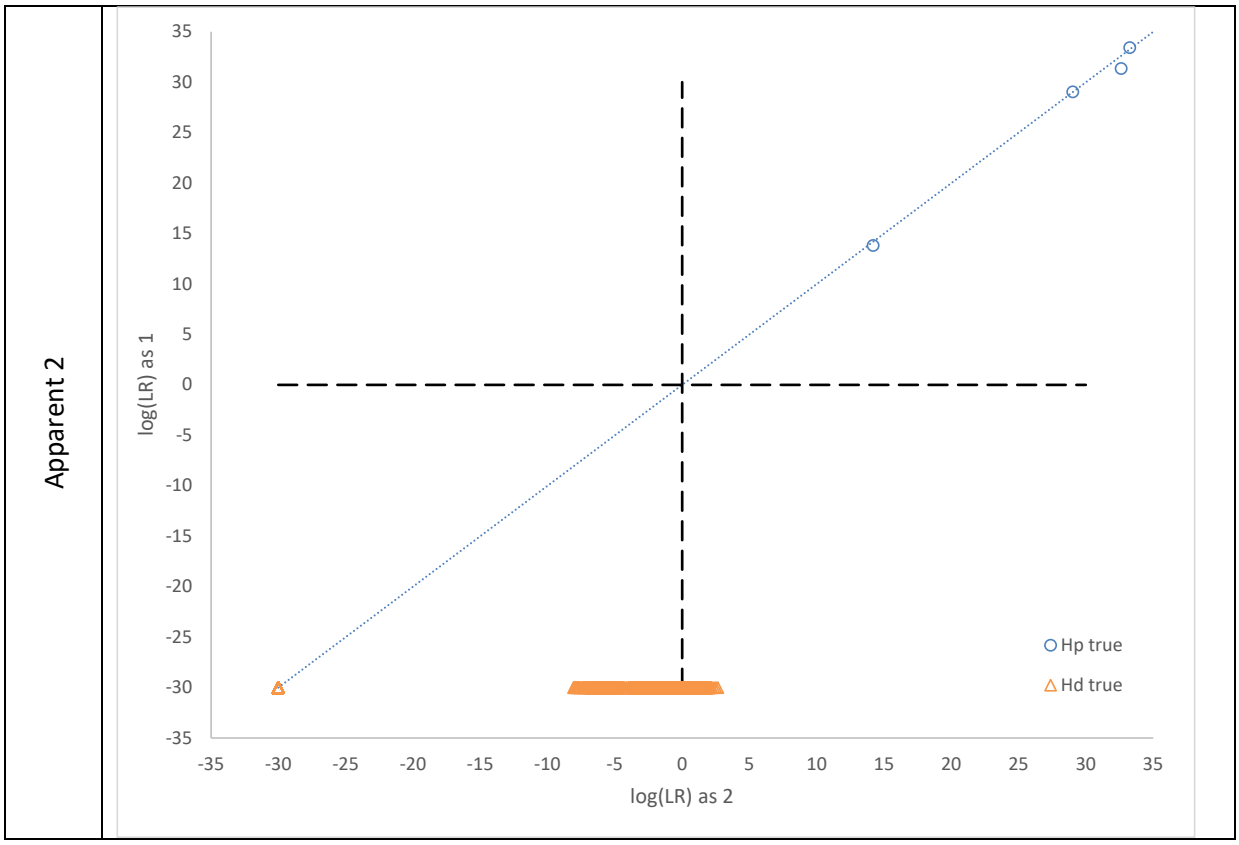




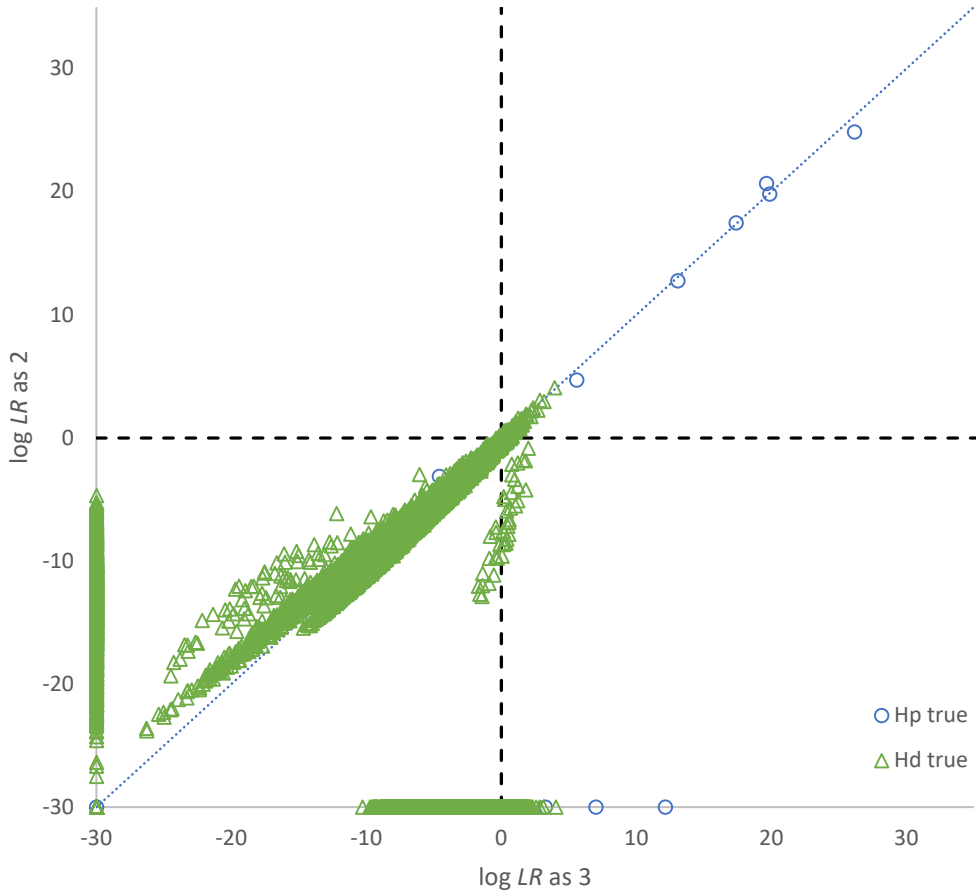
Subtraction of one contributor, (N-1) contributors:

The *LR* was calculated for both the known contributors and 10,000 non-contributors (as undertaken for Experiment 4: Sensitivity and Specificity). The *LR* for the known contributors and known non-contributors assuming N-1 is compared to the original *LR* assuming apparent N contributors within Figure 17.

Figure 17: Comparison of log(*LR*) for apparent N versus log(*LR*) for apparent N-1 contributors



Apparent 3



Apparent 4

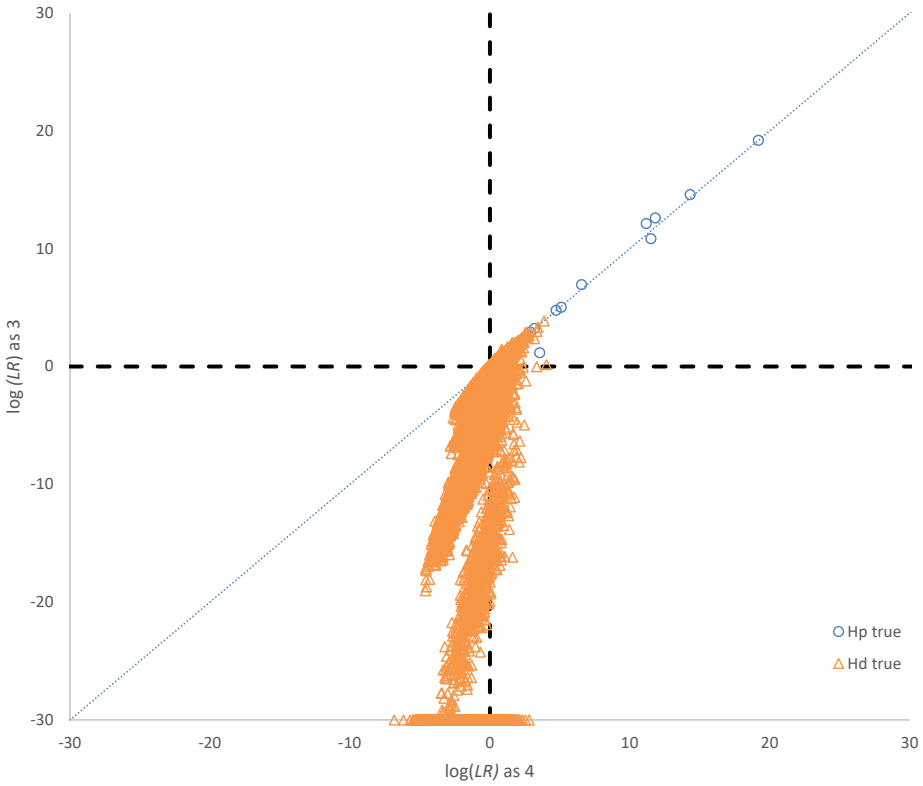


Figure 16 supports previous research finding that there is no significant effect on the *LR* of the major or clear minor contributors when the number of contributors is overestimated [14, 15]. The inclusion of an additional contributor beyond that present in the profile has the effect of decreasing the log(*LR*) for H_p true. This is because STRmix™ adds the additional (unseen) profile at low DNA amount (template) levels, diffusing the genotype probabilities. This allows more genotype combinations at loci, albeit with very low genotypic weight. Overestimating the number of contributors did not result in significant inclusionary *LR*s to non-contributors.

Figure 17 (apparent 3) shows that when you underestimate the number of contributors you can get false exclusions. In general there is no significant effect on the *LR* of the major or clear minor contributors to the mixture. Minor differences above and below the $x=y$ line are likely due to MCMC run variability. Underestimating the number of contributors results in higher *LR*s for H_p true comparisons as STRmix™ is not having to explain any additional trace components to the mixture as potentially allelic. Note that STRmix™ will not run if there are peaks present that cannot be explained using stutter modelling, or drop-in and can only be explained via an extra contributor being present in the mixture.

Experiment 11: The effect of allelic drop in in *LR* calculations

4.1.8. *Allele drop-in.*

Drop-in peaks will be artificially added *in silico* in a prepared single source sample. The log (*LR*) will be plotted to show how the presence of a drop-in peak affects the *LR*.

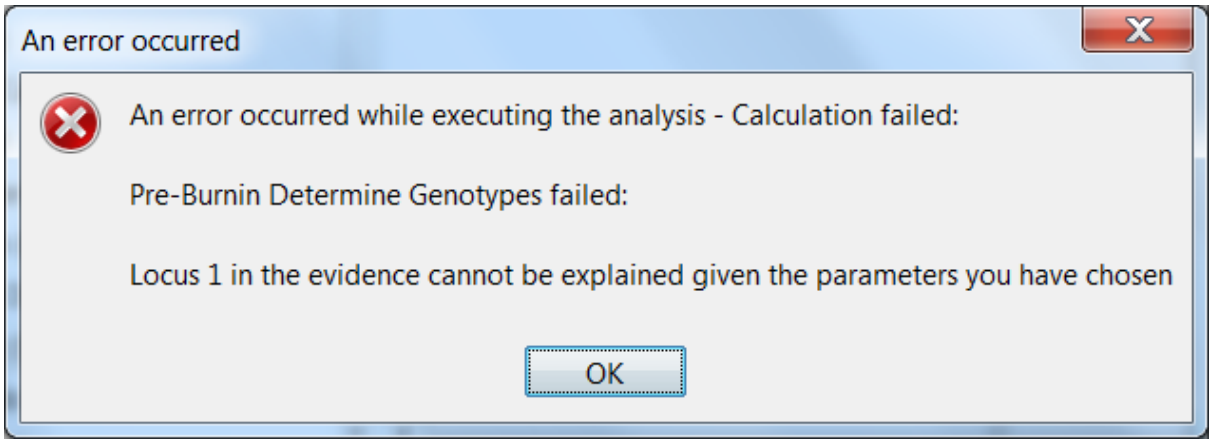
Observed drop-in rates at the OCME Laboratory have been modelled and the appropriate parameters are within STRmix™. OCME drop-in parameters for STRmix™ for the Fusion data:

Drop-in cap	100
Drop-in frequency	0.0024
Drop-in parameters	0,0

To test these settings four experiments were undertaken. In the first experiment, a realistically sized (height less than the maximum observed rfu) drop-in peak was artificially added to a *high template* single source STRmix™ input file (032416 29-Mock_1) at 60 rfu. The profile was interpreted as a single source profile. As expected, STRmix™ completely modelled the additional peak as drop-in because it could not pair with the high template alleles (>1000 rfu). The resulting *LR* was identical to the original profile *LR*.

In the second experiment, a realistically sized (height less than the maximum observed rfu) drop-in peak was artificially added to a *low template* single source STRmix™ input file (18-3M_25pg) at 70 rfu. The profile was interpreted as a single source profile. As expected STRmix™ modelled the additional peak as both drop-in and a true allele as it was of a similar height to the low template alleles at that heterozygote locus (<100 rfu). As expected, the resulting *LR* was less than the original profile *LR*.

In the third experiment, a drop-in allele was added to a heterozygote locus outside OCME's parameters (ie > maximum allowed height for drop-in) in a single source profile (sample 032416 29-Mock_1, drop-in at 120 rfu at locus 1). As expected, the interpretation could not be progressed as the profile could no longer be explained by one contributor:



Finally, a drop-in allele (below OCME drop-in cut-off) was added to a 0.97:0.03 proportion two person mixture (01-M1-C1-750-20-1-6M-22F-Athena) at D13S317, a locus with style 8,9:11 (where the 8,9 is the major). The drop-in allele was a 14 at 90 rfu which was added deliberately to possibly pair with the peaks of the minor contributor. A summary of the genotype combinations before and after the addition of the drop-in peak are given below in table 10.

Table 10: Genotype combinations before and after addition of a drop-in peak

Major	Minor	Weight	
		Before drop-in	After drop-in
8,9	11,11	0.5907	0.0080
8,9	9,11	0.1733	0.0021
8,9	8,11	0.1635	0.0027
8,9	11,Q	0.0722	7.42E-04
8,9	11,14	-	0.9863

As expected, STRmix™ considered the additional peak as both drop-in and allelic as can be seen by the genotype combination in the Table 10 above. This is due to the height of the drop-in allele being similar to the peak heights of the minor contributor.

4.1.9. *Forward and reverse stutter*.

Forward and reverse stutter were present in the various samples that were analyzed with STRmix™ for experiment 4 above.

4.1.10 *Intra-locus peak variance.*

The single source dilutions and some of the mixtures used in this study exhibited varying degrees of intra-locus peak variance. It can be seen from the plots within experiment 4 that STRmix™ is capable of analyzing electropherograms that exhibit intra-locus peak variance.

4.1.11 *Inter-locus peak variance.*

Inter-locus peak variance is routinely encountered in forensic casework samples. Past experience has found that inter-locus peak variance increases in samples that have an amount of DNA present that is less than the target level. The samples tested and reported in the experiments above contained varying

amounts of inter-locus peak variance. STRmix™ will be able to properly deconvolute the samples and return a valid *LR*.

4.1.12 *For probabilistic genotyping systems that require in-house parameters to be established, the internal validation tests should be performed using the same parameters.*

As indicated above, STRmix™ requires several parameters to be established in-house. All parameters were calculated and established prior to running the internal validation studies.

Experiment 12: Additional challenge testing

4.1.14. Additional challenge testing (e.g., the inclusion of non-allelic peaks such as bleedthrough and spikes in the typing results).

The input file for STRmix™ is an analysed file. Known artifacts must be edited out (unlabeled) of the input file prior to STRmix™ interpretation (refer above for a related discussion). Failure to remove a non-numeric peak call will cause STRmix™ to stop the interpretation. Inclusion of a numeric artefact (such as a pull up peak in an allelic position, forward or double back stutter peak) that is a similar height to peaks from the person of interest may cause a false exclusion. In addition, STRmix™ cannot model triallelic loci. These may cause a false exclusion at that locus and can be identified by reviewing STRmix™ results.

The effect of the inclusion of non-allelic peaks within a STRmix™ interpretation was trialled by variously editing DNA profiles. In the first experiment, an artefact peak falling outside an allelic bin was added to the fifth locus within the text input file of a single source profile (sample 032516-37-Mock-8) as an "OB" designation. The interpretation appeared to progress as usual but only the first four loci were read into the file. This is readily diagnosed by review of the input file in the results file:

Locus	Allele	Height	Size
1	16	501	124.8
1	17	5179	129
1	18	5854	133.4
2	11	332	164.6
2	12	5027	168.6
2	13	4578	172.7
3	10	500	216.5
3	11	8030	220.6
3	12	104	224.7

4	13	636	270
4	14	8046	274
5	10	0	0
5	11	0	0

In the second experiment, an artefact peak was artificially added that fell within an allelic bin. The presence of a non-allelic peak (or peaks) that has sized within an allelic bin position and is retained within the input file can cause a number of results depending on the crime profile and number of contributors. These include:

- An exclusionary *LR*. If the artefact is modelled as having originated from the person of interest (for example if the peak is of a similar height to the alleles corresponding to the person of interest in a mixed DNA profile) this may result in an exclusion.
- No effect. If the artefact peak was below the drop-in cap, the artefact may be modelled as a drop-in peak.
- Failure to interpret. If an artefact within an allelic bin is retained in a profile it may artificially increase the minimum number of contributors within the profile. For example an artefact at a heterozygous locus in a single source profile (not modelled as stutter or drop-in) will increase the minimum number of contributors by one. STRmix™ will not proceed assuming only one contributor.

Each of these expected outcomes was demonstrated by editing a single source input file and calculating a *LR* within STRmix™.

Experiment 13: Comparisons of manual interpretation guidelines with STRmix™ analyses

4.2 *Laboratories with existing interpretation procedures should compare the results of probabilistic genotyping and manual interpretation of the same data, notwithstanding the fact that probabilistic genotyping conclusions are inherently different from and not directly comparable to binary conclusions (e.g., exclusion or inclusion). Match statistics that are generated by these two approaches are based on different assumptions, thresholds and formulae. However, such a comparison should be conducted and evaluated for general consistency.*

4.2.1. *The Laboratory should determine whether the results produced by the probabilistic genotyping software are intuitive and consistent with expectations based on non-probabilistic mixture analysis methods.*

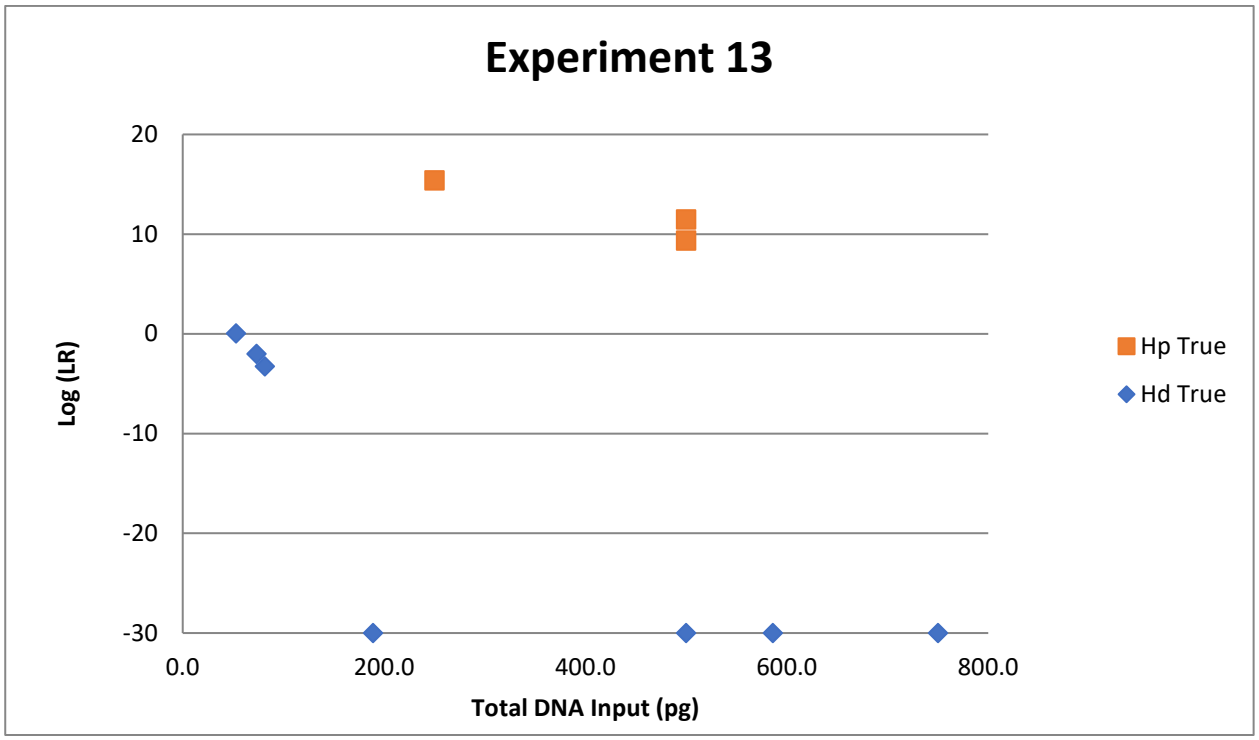
4.2.1.1. *Generally, known specimens that are included based on non-probabilistic analyses would be expected to also be included based on probabilistic genotyping.*

Analysts were asked to interpret the results based on their experience as well as data from the PowerPlex® Fusion validation (heterozygous ratio, stochastic and analytical thresholds, stutter %, etc.). 1 proficiency test sample, 4 mock evidence touched items from the PowerPlex® Fusion validation, and 5 mixtures from Mixture Set 2 were chosen for analysis (10 total samples). These samples were previously amplified with PowerPlex® Fusion for 29 cycles and analyzed with GeneMarker®HID v.2.8.2. Two analysts in the laboratory were provided with two sets of electropherograms (one with stutter filters on and one with stutter filters off) and one potential contributor profile for each sample. The two analysts were then asked whether they would deem the potential contributor profile as included, excluded, or inconclusive. The same 10 samples were then interpreted with STRmix™ v.2.4.05, and likelihood ratios (99.0% 1-sided HPD) of the potential contributor samples were calculated and the results were compared with the analysts' manual interpretations. The results of the experiment shown in Table 11 below that the two analysts' manual comparisons were consistent with the results obtained from STRmix™. Figure 18 shows true contributors obtained inclusionary log(*LR*) values while non-contributors obtained exclusionary (*LR*) values. As the input amount of DNA decreases the log(*LR*) values trend towards the uninformative range (0.001-1,000).

Table 11: Analysts interpretation of samples compared to STRmix™ LR output.

Sample	True Contributors	Potential Contributor Profile	Expected Answer	Analyst 1	Analyst 2	DNA Input (pg)	STRmix™ *LR*
5	24F30, 29M30 (2)	11M	excluded	excluded	excluded	500.0	0
14	24F30, 29M30 (2)	29M30	include	include	include	250.0	2.44E+15
22	24F30, 29M30 (2)	24F30	include	include	include	500.0	2.93E+11
51	12F, unknown (2)	10F	INC/ex	excluded	excluded	189.0	0
53	4M, unknown (2)	26	INC/ex	excluded	excluded	73.2	1.00E-02
54	20F, unknown (2)	4M	INC/ex	INC/excluded	excluded	52.8	1.10E+00
55	13F, unknown (2)	10M	INC/ex	INC/excluded	excluded	81.7	5.37E-04
61	CST_5F, CST_13M (2)	CST_13M	include	include	include	500.0	2.25E+09
77	23M30, CST_14M, 18F30 (3)	CST_4F	excluded	excluded	excluded	750.0	0
82	PT3	60	excluded	excluded	excluded	586.0	0

Figure 18: log(*LR*) values of true contributors and non-contributors plotted against total DNA input amount



Experiment 14: Partial profile testing

4.1.7. *Partial profiles to include the following:*

4.1.7.2. *DNA degradation*

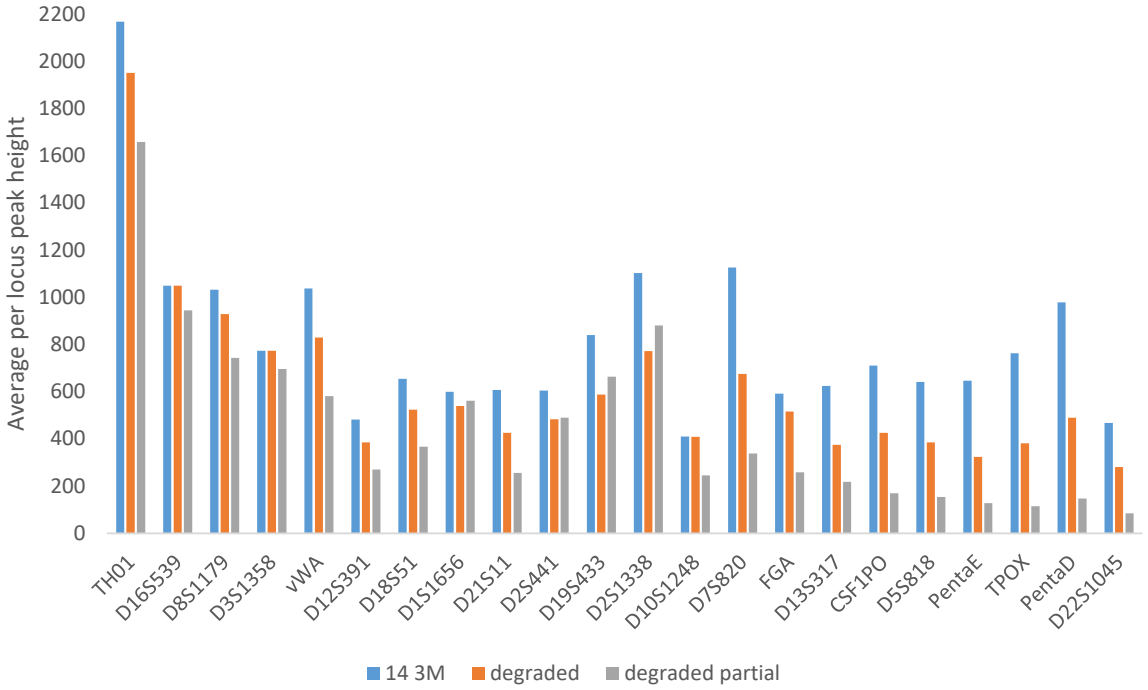
4.1.7.3. *Inhibition.*

A single source DNA profile from Experiment 2 was amended *in silico* to mimic samples with allele/locus drop-out, DNA degradation and partial inhibition. The samples were interpreted in STRmix™.

In the first experiment one profile (sample 14 3M) was artificially degraded. The average per locus peak heights before and after degradation are given in Figure 19. Both profiles were interpreted within STRmix™. Both interpretations resulted in intuitive genotypes where the weight = 1.0 for the known contributor genotype at each locus resulting in an identical *LR* (6.77E32). The mean of the post burn-in degradation parameter was 0.9669 rfu/bp for the original profile and 1.6012 rfu/bp after artificial degradation showing the STRmix™ is modeling the profile degradation expectedly.

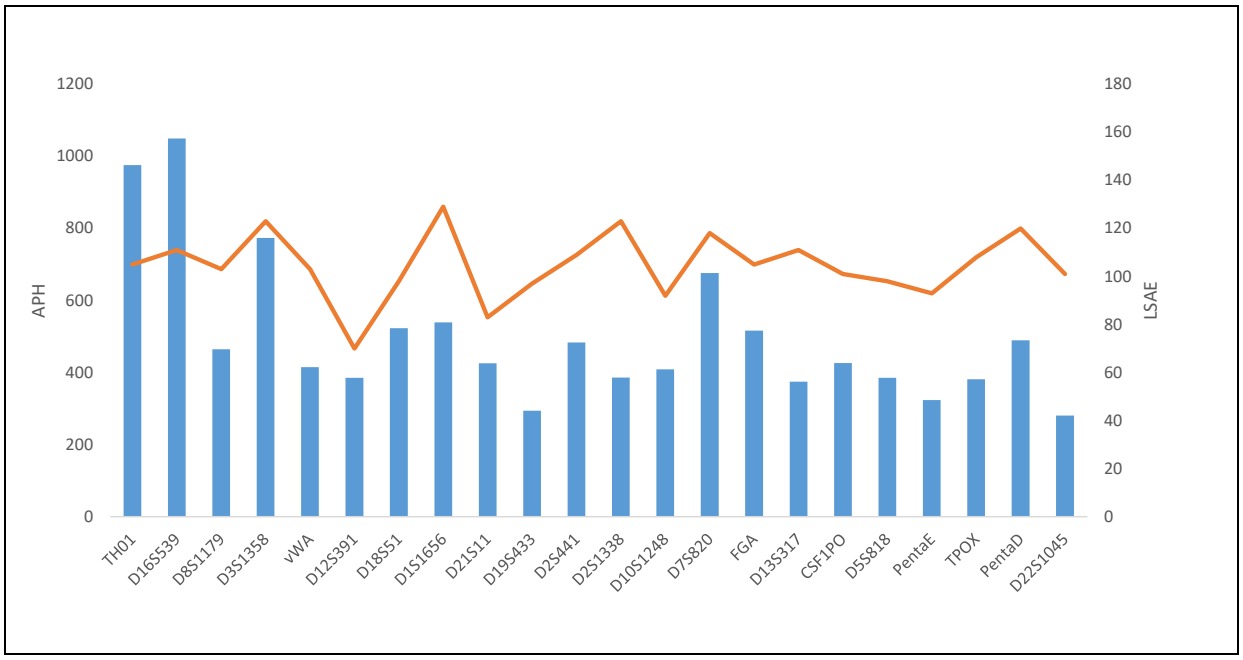
In the second experiment, the same profile (sample 14 3M) was artificially degraded further resulting in some allelic and locus drop-out. The average per locus peak heights before and after degradation are also given in Figure 19. The mean of the post burn-in degradation parameter was 1.713 rfu/bp and the *LR* reduced to 9.10E27 as a result of the dropped alleles.

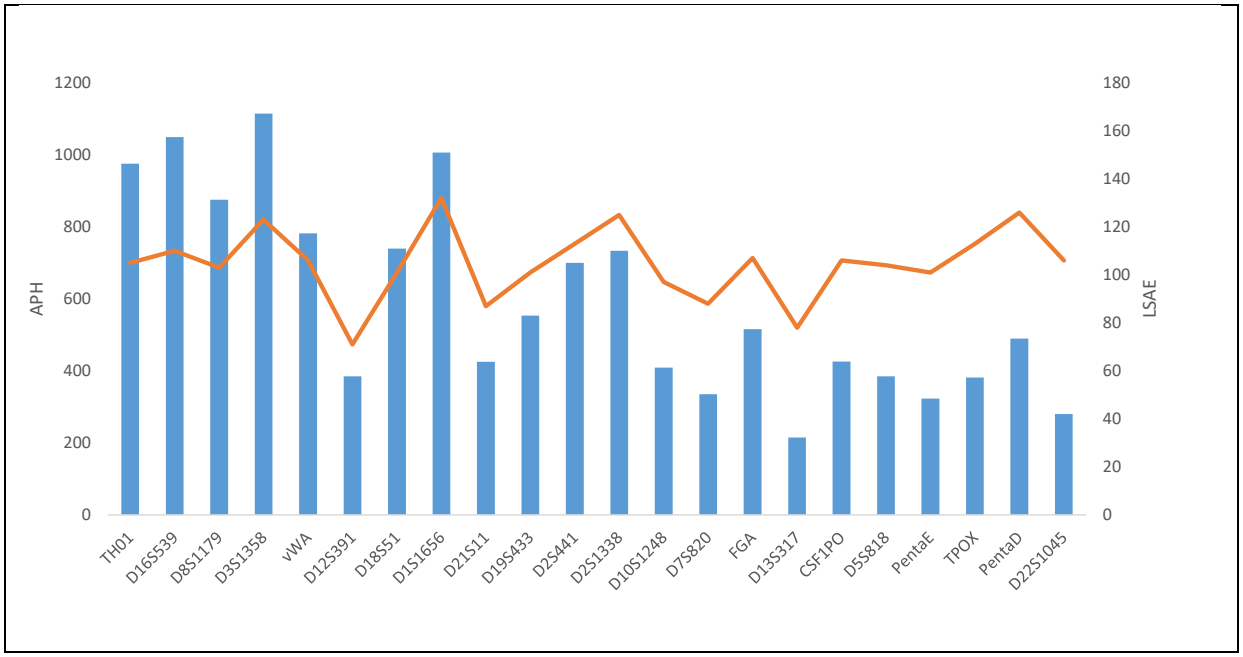
Figure 19: Per locus average peak height before and after artificial degradation



In the third experiment, the degraded profile was artificially inhibited. Inter locus peak variance is modelled in STRmix™ using locus specific amplification efficiencies (LSAE). The LSAE model reflects the observation that even after template DNA amount, degradation and variation in peak height within loci are modelled, the peak heights between loci are still more variable than predicted. The variance of this model is determined by directly modelling laboratory data. LSAE values for each STRmix™ interpretation appear within the results. We can demonstrate the relationship of LSAE values to average peak heights (APH) via a simple plot. The LSAE values should mimic the average peaks heights of the locus. This is demonstrated for one single source Fusion profile in Figure 20 where the degraded profile was inhibited at D7 and D13.

Figure 20: Plot of APH and LSAE value for each locus for a single source Fusion profile without inhibition (top pane) and with inhibition (bottom pane). The loci are sorted by molecular weight





Conclusion

This document describes the OCME laboratory's internal validation activities for Fusion profiles analysed using STRmix™ V2.4. It has been shown that it is suited for its intended use for the interpretation of profiles generated from crime scene samples.

References

[1] Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for the Validation of Probabilistic Genotyping Systems. 2015.

[2] Bright J-A, Taylor D, McGovern CE, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics. 2016;23:226-39.

[3] Bright J-A, Evett IW, Taylor D, Curran JM, Buckleton J. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics. 2015;14:125-31.

[4] Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International. 1994;64:125-40.

[5] National Research Council II. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press; 1996.

[6] Hill CR, Duewer DL, Kline MC, Coble MD, Butler JM. U.S. population data for 29 autosomal STR loci. Forensic Science International: Genetics. 2013;7:e82-e3.

[7] Good IJ. Good Thinking - The Foundations of Probability and Its Applications. Minneapolis: University of Minnesota Press; 1983.

[8] Taylor D, Buckleton J, Evett I. Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics. 2015;16:165-71.

[9] Taylor D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics. 2014;11:144-53.

[10] Bright J-A, Stevenson KE, Curran JM, Buckleton JS. The variability in likelihood ratios due to different mechanisms. Forensic Science International: Genetics. 2015;14:187-90.

[11] Taylor D, Bright J-A, Buckleton J, Curran J. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics. 2014;11:56-63.

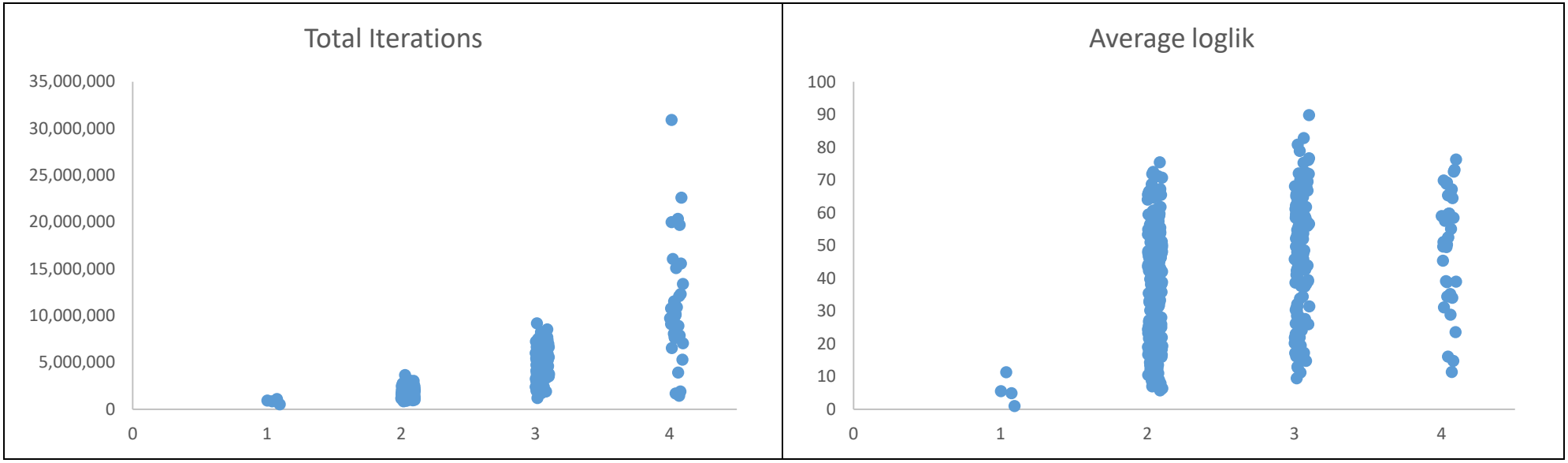
[12] Triggs CM, Curran JM. The sensitivity of the Bayesian HPD method to the choice of prior. Science & Justice. 2006;46:169-78.

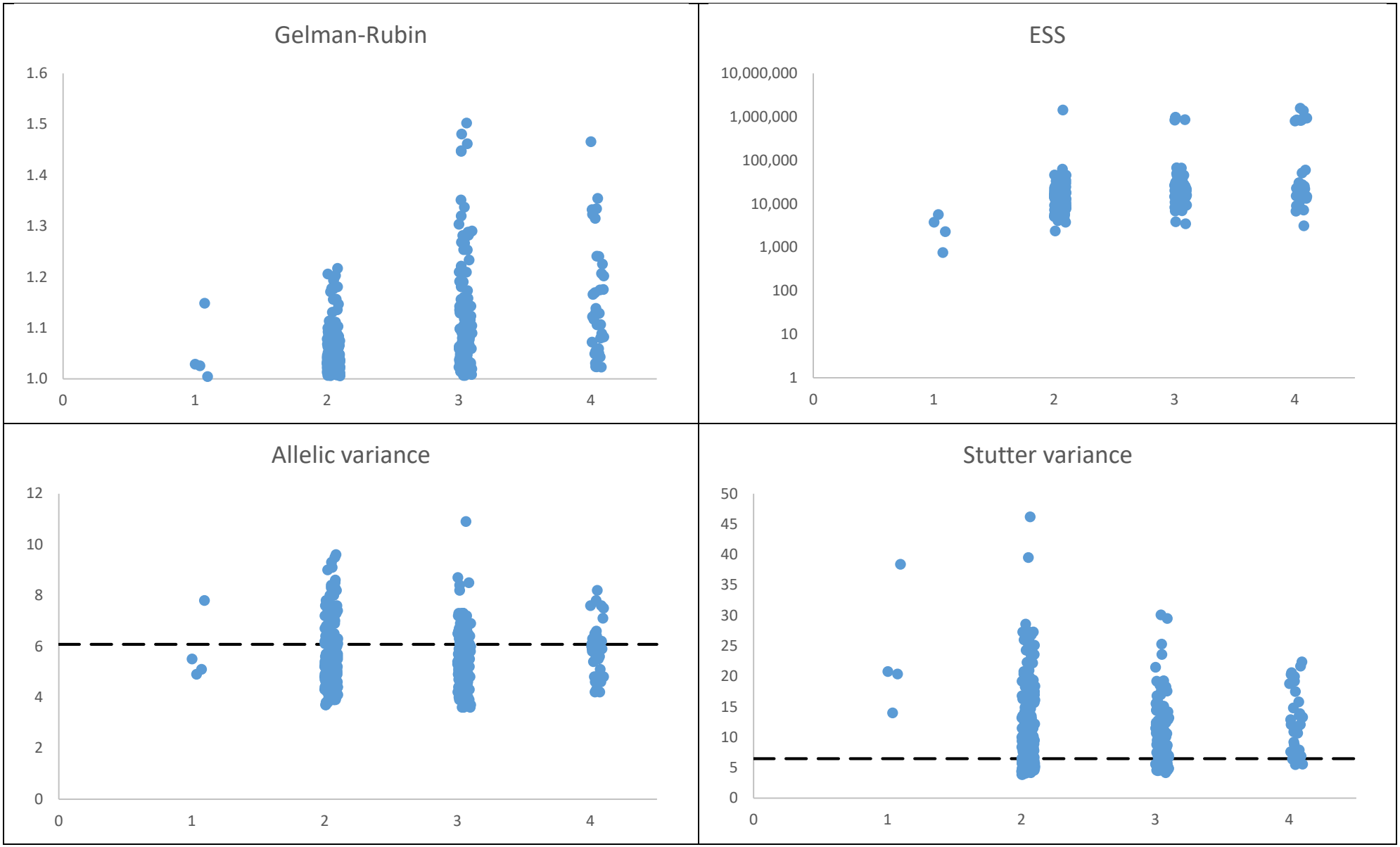
[13] Curran JM, Buckleton JS. An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations. Forensic Science International: Genetics. 2011;5:512-6.

[14] Bright J-A, Curran JM, Buckleton JS. The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. Forensic Science International: Genetics. 2014;12:208-14.

[15] Bright J-A, Taylor D, Curran J, Buckleton J. Searching mixed DNA profiles directly against profile databases. Forensic Science International: Genetics. 2014;9:102-10

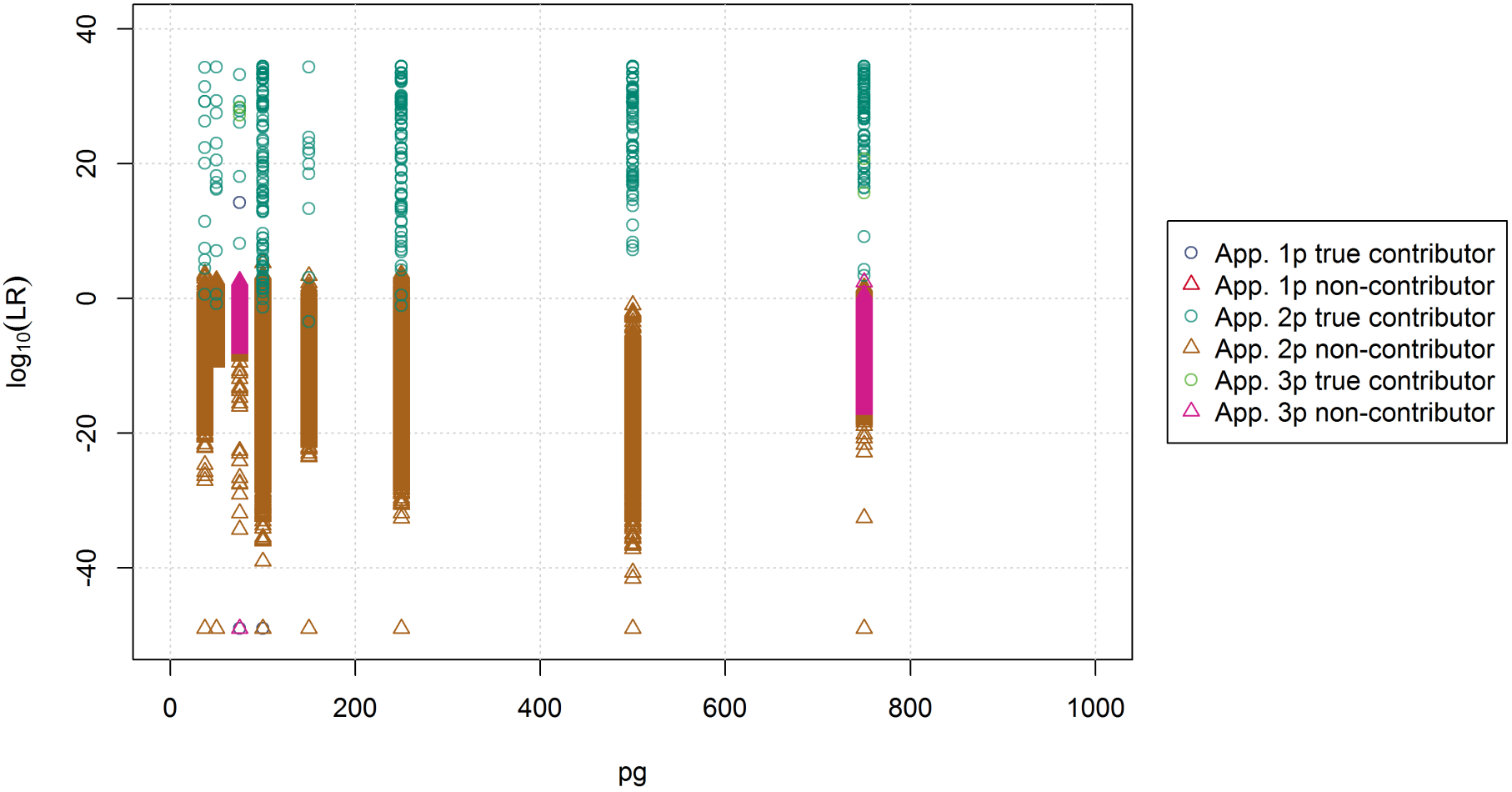
APPENDIX 1 Range of diagnostic values for runs undertaken as part of the sensitivity and specificity interpretations. The horizontal dashed lines in the allele and stutter variance plots represent the mode of their prior distributions



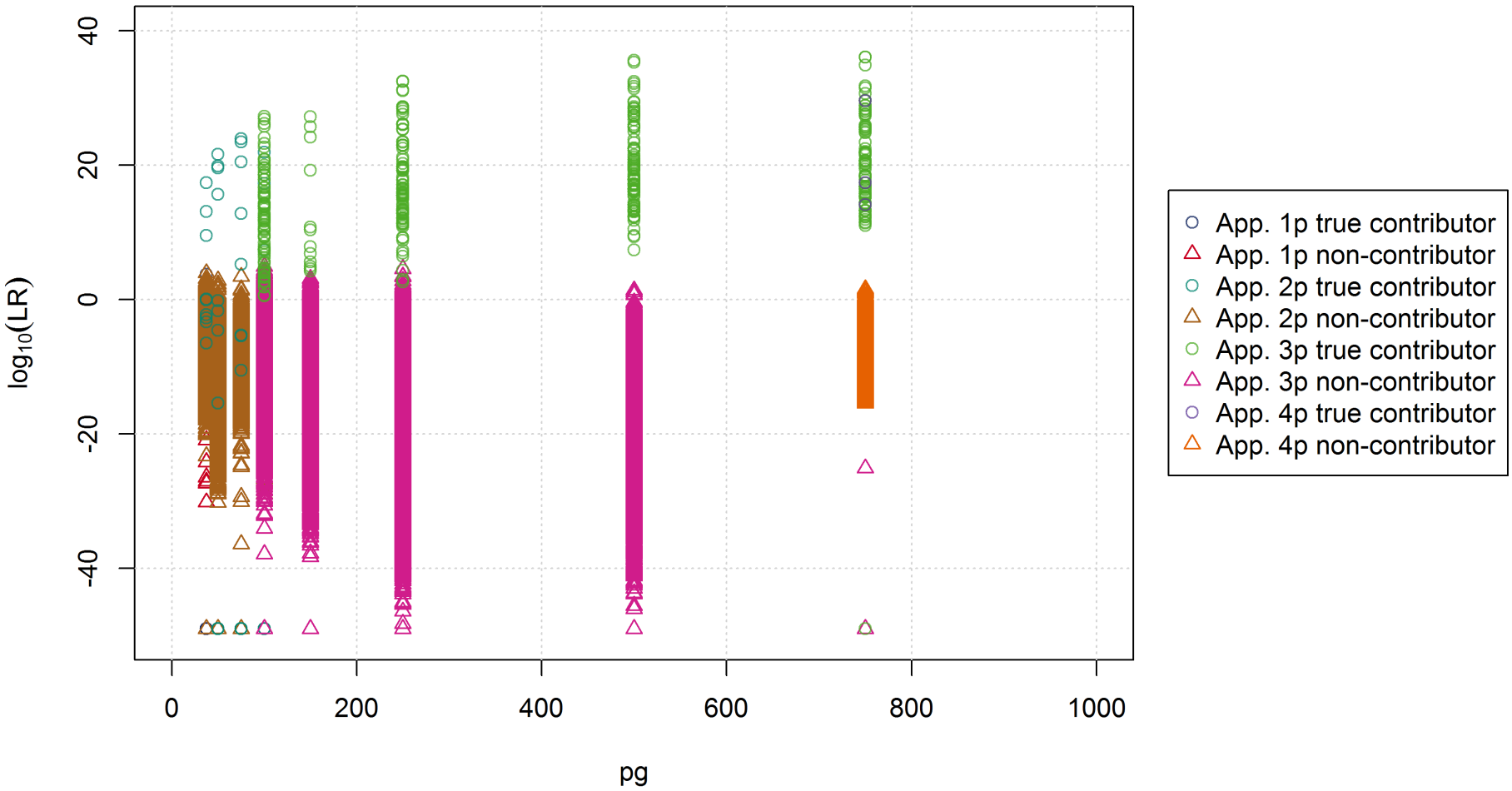

Gelman-Rubin
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0
1
2
3
4
ESS
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1
0
1
2
3
4
Allelic variance
12
10
8
6
4
2
0
0
1
2
3
4
Stutter variance
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4

APPENDIX 2 A plot of log(*LR*) versus total input DNA (from quantitation) for sensitivity and specificity pots from Experiment 4.

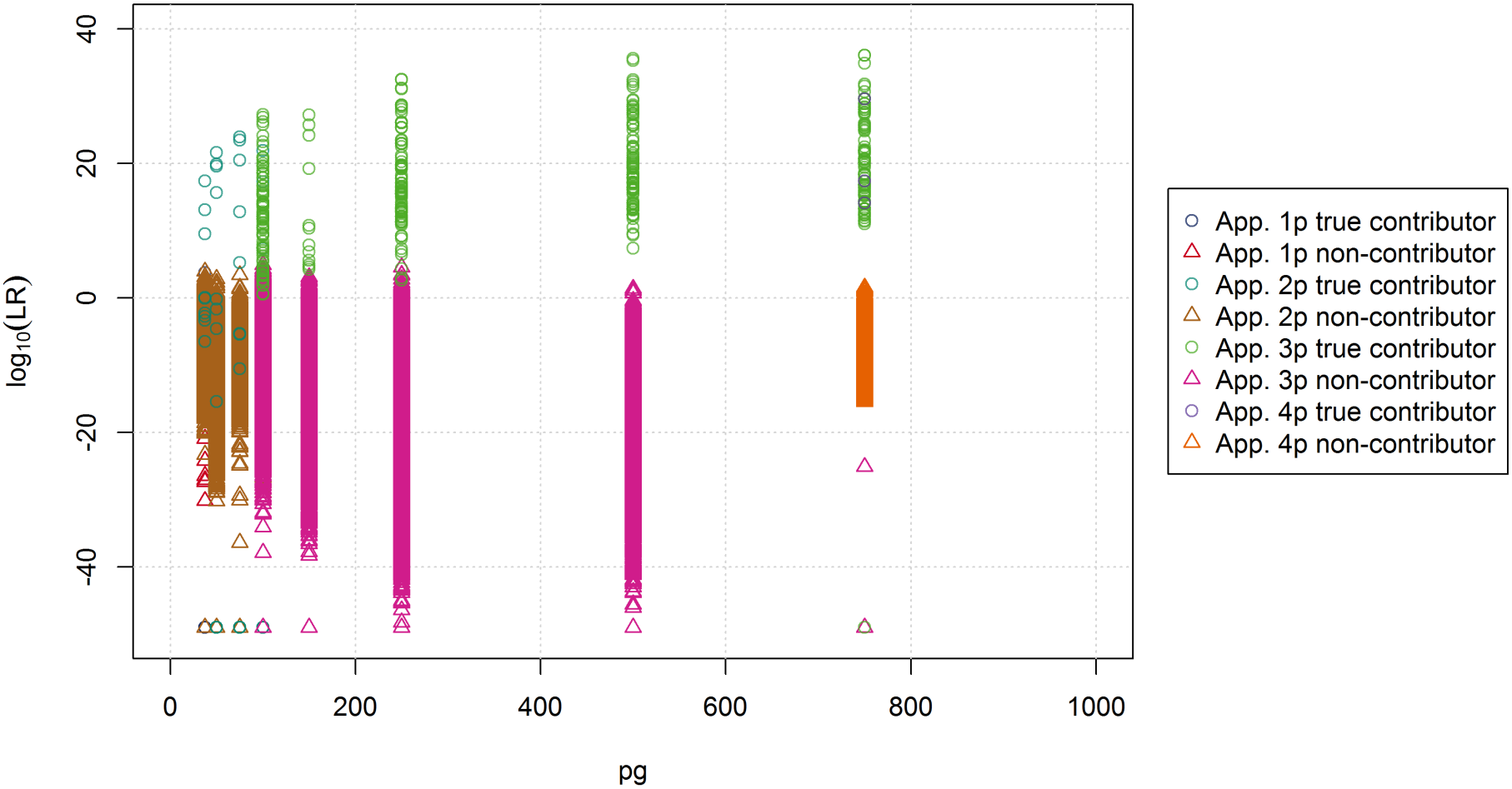
log(*LR*) versus total input DNA amount (pg) for known 2 person mixtures



log(*LR*) versus total input DNA amount (pg) for known 3 person mixtures



log(*LR*) versus total input DNA amount (pg) for known 4 person mixtures



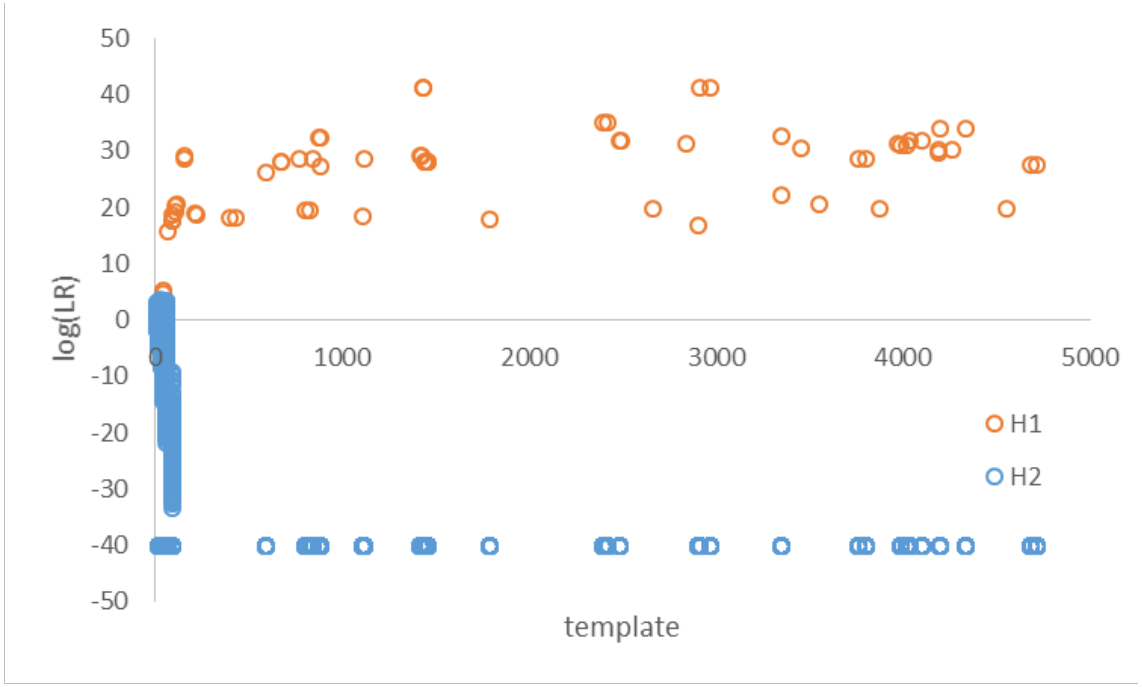
APPENDIX 3 Interpretation summary of 53 OCME mock profiles

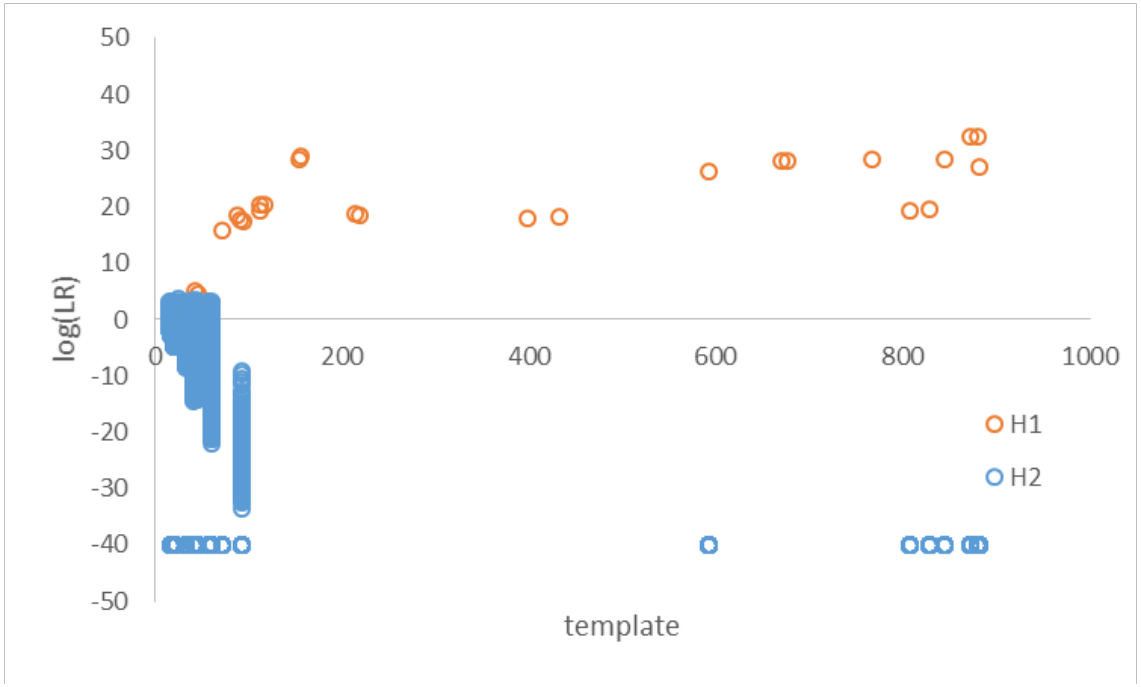
Fifty three mock crime scene profiles were interpreted. A summary of the number of contributors (NOC), total MCMC iterations, effective sample size, average log(likelihood), Gelman-Rubin and allele and variance constants for each sample is provided in the table below.

#	Sample name	NOC	Total iterations	ESS	Avg log(likelihood)	GR	allele variance	stutter variance	99.0% 1-sided lower HPD interval *LR* (NIST Cau)	Stratified *LR*	Number of *LR*=0
1	032516_29-Mock_1	1	565564	1977.6	42.6	1.02	5.8	22.1	2.79E34	4.76E32	10000
2	032516_31-Mock_3	1	655178	3796.4	55.7	1.02	5.5	7.3	1.31E29	2.63E28	10000
3	032516_32-Mock_4	1	522061	3270.1	26.7	1.12	9	12	6.00E29	1.29E30	10000
4	032516_33-Mock_5	1	653493	5791.7	48.8	1.01	5.2	21.1	1.16E27	4.42E27	10000
5	032516_34-Mock_6	1	558185	563080.0	44.2	1.05	5	7	1.09E27	3.77E26	10000
6	032516_35-Mock_7	2	1926211	19637.1	59.4	1.09	4.7	7.9	6.33E27	1.54E28	10000
7	032516_37-Mock_8	1	698542	7202.5	59.8	1.02	5.6	7.3	1.67E26	6.56E26	10000
8	032516_38-Mock_9	1	583987	2418.0	42.3	1.05	6.6	20.1	9.87E30	1.89E31	10000
9	032516_39-Mock_10	1	528655	1411.9	19.5	1.04	7.3	9.7	5.26E29	1.93E30	10000
10	032516_40-Mock_11	3	1103891	17876.2	5.7	1.04	4.3	18	1.76e4	2.93E04	0
11	032516_42-Mock_13	1	524727	4173.8	34.5	1.02	5.6	16.8	2.15E26	8.74E25	10000
12	032516_43-Mock_14	1	561160	2795.0	43.5	1.02	5.3	13	1.13E32	4.96E27	10000
13	032516_45-Mock_16	2	1261576	9993.9	13.9	1.03	4.2	15.9	9.27E16	1.44E17	6836
14	032516_46-Mock_17	2	1062068	4755.4	25.5	1.02	5.3	9.2	2.52E26	6.35E26	7608
15	032516_47-Mock_18	1	539546	4937.1	39.1	1.07	7.1	21.3	6.54E29	1.45E30	10000
16	032516_50-Mock_19	2	1045150	17319.0	10.2	1.07	5.4	27	4.07E15	6.96E15	1232
17	032516_51-Mock_20	2	2293121	13014.3	17.5	1.02	5.3	17.1	1.26E16	5.45E15	9862
18	032516_52-Mock_21	2	1273406	7655.6	23.6	1.03	6.5	15	6.59E33	1.24E32	7071
19	032516_53-Mock_22	3	1305909	8421.4	11.5	1.07	6.9	23.8	7.75E16	5.27E16	0
20	032516_54-Mock_23	2	1267027	18320.4	10.4	1.03	5.6	14.4	4.19E18	4.34E18	684
21	032516_55-Mock_24	2	1532511	32083.3	11.0	1.04	7.5	23.9	9.98E24	4.42E23	2218
22	032416_29-Mock_1	1	564013	6239.7	42.5	1.08	6	22	2.44E34	5.09E32	10000
23	032416_30-Mock_2	1	669677	6276.1	53.5	1.04	5.5	20.5	2.77E29	1.09E30	10000
24	032416_31-Mock_3	1	662039	6940.3	55.8	1.06	5.3	7.1	1.21E29	3.17E28	10000

25	032416_32-Mock_4	1	521108	3272.8	26.6	1.02	8.9	11.2	6.10E29	1.48E30	10000
26	032416_33-Mock_5	1	643875	14348.8	48.6	1.04	5.6	22.3	1.24E27	4.85E27	10000
27	032416_34-Mock_6	1	559674	7654.5	44.1	1.02	4.9	7	1.20E27	3.74E26	10000
28	032416_35-Mock_7	2	1885661	10492.3	59.4	1.01	4.2	8.2	8.61E27	1.71E28	10000
29	032416_37-Mock_8	1	701515	6555.4	59.9	1.02	5.4	7.3	2.13E26	7.70E26	10000
30	032416_38-Mock_9	1	576841	566576.0	42.0	1.05	7	19.7	1.01E31	1.35E31	10000
31	032416_39-Mock_10	1	535668	4447.3	19.4	1.03	7.5	9.7	4.57E29	1.77E30	10000
32	032416_40-Mock_11	3	1065743	7185.3	5.5	1.09	4.4	27.3	6.82E3	1.21E04	0
33	032416_42-Mock_13	1	524350	4443.6	34.5	1.03	5.8	16.3	1.79E26	9.40E25	10000
34	032416_43-Mock_14	1	563880	3706.4	43.7	1.04	5.1	13	1.57E32	7.95E27	10000
35	032416_45-Mock_16	3	1240834	16388.0	13.3	1.03	4.2	21.2	2.01E16	1.66E16	0
36	032416_46-Mock_17	2	1119340	6641.1	25.4	1.02	5.2	10.1	2.82E26	6.95E26	7608
37	032416_47-Mock_18	2	1167320	11528.0	43.6	1.03	6.6	11.5	3.58E29	5.15E29	0
38	032416_50-Mock_19	2	1041354	11850.4	10.0	1.01	5.2	18.7	3.26E15	5.09E15	1232
39	032416_51-Mock_20	3	2491670	20345.1	15.4	1.02	5.7	13.3	7.78E15	1.48E15	10000
40	032416_52-Mock_21	2	1375362	9923.0	23.3	1.01	7.1	11.9	1.29E34	1.59E32	8114
41	032416_53-Mock_22	3	1327831	7970.9	13.4	1.05	6.4	22.1	1.42E17	1.25E17	0
42	032416_54-Mock_23	2	1277524	5014.9	10.4	1.06	5.5	18.9	4.82E18	6.42E18	684
43	55-BUSS_SR	2	2004874	18557.7	33.4	1.08	5.8	8.2	2.78E40	1.06E41	10000
44	55-Mock_24	2	1536594	8475.4	10.6	1.06	7.4	23.3	2.44E24	1.58E23	2218
45	79-Proficiency_Sample_1_EC	2	1956937	10650.9	67.5	1.06	5.9	7	2.34E36	7.52E36	10000
46	80-Proficiency_Sample_2_EC	2	1388751	38695.3	41.5	1.08	5.1	14.1	1.10E35	1.60E35	10000
47	82-Proficiency_Sample_3_EC	2	1056936	6873.1	59.0	1.04	6.2	5.6	9.80E36	1.52E37	10000
48	83-Proficiency_Sample_4_EC	2	1190567	8794.4	68.0	1.02	5.2	6.9	8.22E35	6.53E35	10000
49	84-Proficiency_Sample_5_EC	2	1406639	3628.8	70.7	1.02	4.5	8.1	8.00E36	1.04E37	10000
50	85-BDSS_SR	2	1553204	16228.3	49.7	1.03	6.3	28.7	1.39E54	3.89E54	10000
51	86-RCSS_SR	2	1287810	12016.0	44.9	1.05	6.6	47.7	1.06E55	3.27E55	10000
52	87-BDSS_EC	2	1121200	10626.2	55.1	1.07	7.5	15.3	4.07E54	1.16E55	10000
53	88-RCSS_EC	2	1334449	3342.0	52.4	1.04	7.9	9.2	2.19E53	8.71E53	10000

Each interpretation was compared to the known donor and 10,000 non contributors. The non-contributors were created artificially using the Caucasian allele frequencies. A plot of stratified log(*LR*) versus average post burn-in template value (*t*) per contributor is given in the figure below. There are 530,063 data points within this plot. Exclusions (*LR*=0) are plotted as log(*LR*)=-40. The *t* per known contributor was taken from the STRmix™ output. The lowest contributor *t* for each *known* contributor was used for the H_d true contributors. A summary of the number of contributors (out of 10,000) that gave *LR*=0 is given in the table above. The second plot is a zoom of the x-axis (log scale) to better show the low level data.




50
40
30
20
10
0
-10
-20
-30
-40
-50
log(LR)
0
200
400
600
800
1000
template
H1
H2

Updates to summary as of 20 December 2019 (original version 18 November 2016) :

The written summary for the Internal Validation of STRmix™ V2.4 for Fusion NYC OCME was reviewed and updated in December 2019. These updates were made in order to correct transcriptional errors that were found, and to provide clarity in the labelling of tables, figures and end notes. A summary of the updates is listed below:

Page 12: Tables 3 & 4 - updated to list 'True' instead of 'Apparent' within the table header.

Page 13-14: Table 5- Sample 13-M2 112.5pg 10-5-2-1 in comparison to database profile #2821 was added to the table. The paragraph following the table was updated to indicate that there were ten LR's, and not nine. The percentage of LRs >1 did not change.

Page 14: The Good and Taylor parenthetical end note references were corrected. The Good reference was added as #7. This caused a shift of the numbers associated with the subsequent end note references throughout the document.

Page 15: The "false positive" rates listed on page 15 were updated from ~0.512% to ~0.587% and ~0.009% to ~0.001%. The percentage of H_d true comparisons was updated from ~99.9988% to a truncated value of ~99.998%. Table 6 was added. The paragraph following table 6 was updated to clarify that the calculations were performed on apparent 2 and 3 contributor mixtures.

Page 18: The first two paragraphs of Experiment 5 were reworded for clarity.

Page 22: The number of mock evidence samples and touched item samples were corrected for experiment 7.

Page 26: The graph in the lower half of Figure 15 was updated due to a transcriptional error from the underlying data. The two sentences following Figure 15 were removed, and the information in regards to the exclusion in the major three-person plot was incorporated into the following paragraph. For this sample, the apparent NOC was an underestimation of the true NOC, which was the cause of the LR=0 for the true contributor.

Throughout the document:

Figures which did not previously have headings were labelled with headings for clarity. This adjusted subsequent figure numbering.

Tables which did not previously have headings were labelled with headings for clarity. This adjusted subsequent table numbering.

The registered trademark symbol (®) and 'HID' was added to the mentions of the GeneMarker® HID software.

Additional grammatical, non-substantive fixes were made.

Updates to this validation summary do not change any standard operating procedures. All NYC OCME standard operating procedures for the use and interpretation of results from the STRmix™ software and PowerPlex® Fusion kit remain the same after review and updates to this document.