



**The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:**

**Document Title:** Establishing Exclusion Criteria and the Significance of Inclusion for Complex Low-Template DNA Mixtures

**Author(s):** Catherine Grgicak, Ph.D.

**Document Number:** 252060

**Date Received:** October 2018

**Award Number:** 2012-DN-BX-K050

**This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**

**Boston University** School of Medicine  
Biomedical Forensic Sciences

72 E Concord St. Rm R806B  
Boston, Massachusetts 02118  
T 617-638-1950 F 617-638-1960  
[www.bumc.bu.edu/gms/biomedforensic](http://www.bumc.bu.edu/gms/biomedforensic)



---

## **Establishing Exclusion Criteria and the Significance of Inclusion for Complex Low-Template DNA Mixtures**

### **FINAL SUMMARY OVERVIEW**

June 24, 2015

Department of Justice, National Institute of Justice  
Award Number: 2012-DN-BX-K050  
(January 1, 2012 – March 31, 2015)

---

#### **Principal Investigator:**

Catherine Grgicak, PhD  
Assistant Professor  
Program in Biomedical Forensic Sciences  
Boston University School of Medicine  
Boston, MA 02118  
Phone: (617) 638-1968  
Fax: (617) 638-1960  
e-mail: [cgrgicak@bu.edu](mailto:cgrgicak@bu.edu)

## Contents

Purpose of the Project .....	3
Project Design .....	3
Methods.....	4
Data Analysis.....	6
List of Scholarly Products .....	9
Implications to Criminal Justice Policy and Practice.....	10
References.....	10

## Purpose of the Project

The purpose of this work was to address the need to increase the knowledge and understanding associated with complex forensic DNA interpretation and to develop a novel approach to interpret low-template DNA samples containing many contributors. The objectives were to: 1) increase the fundamental understanding of the relationship between DNA input and signal in order to provide recommendations for DNA validation experimental design; 2) provide algorithms and prototype software that combines the features of a fully continuous model to calculate the LR and its distribution, conditioned on the defense hypothesis; 3) provide interpretation approaches, which utilize novel algorithms and signal information produced by the laboratory to reduce a laboratory's need to continuously re-evaluate validation data.

## Project Design

### Phase 1. Sample Preparation & Signal Stability

Phase 1. of the project focused on the creation of single-source and mixture samples that were utilized during algorithm development. Single source samples, amplified with the Identifiler™ Plus (IDPlus) Amplification Kit, which already existed in the Boston University profile database, were utilized to characterize the electropherogram signal. These 616 single source sample files were amplified over a period of 4 weeks using multiple capillary lots, kit lots and pipettes. An additional 316 single source IDPlus samples were generated under this project and include the amplification of samples containing template masses of 0.5 and 1 ng. This dataset also included the amplification of a single sample (Sample 19) using target ranges of 0.25 to 0.008 ng. Additional mixture sets, shown in Table 1, were also prepared. Lastly, 50 PowerPlex16HS (PP16HS) single source and 80 PP16 HS mixtures were generated using a sub-set of the samples described above. Each was injected using 5, 10 and 20 sec injections.

Table 1. Samples prepared using the Identifiler™ Plus amplification kit.

No. of Contributors	Ratio of Contributors	Target Masses (ng)	Total # Amp Products
2	1:1; 1:2; 1:4; 1:9; 1:19; 1:49; 1:99	0.5, 0.25, 0.125, 0.063, 0.031, 0.016 and 0.008	57
3	1:1:1; 1:2:1; 1:4:1; 1:9:1; 1:2:2; 1:4:4; 1:9:9	Various (minor contributor ~ 2, 5, 10, 20 cells)	21

## Phase 2. MatchIt Algorithm Description and Assumptions

The Likelihood Ratio (LR) is defined as:

$$LR = \frac{\Pr(E|H_p, n_p)}{\Pr(E|H_d, n_d)}.$$

We have developed MatchIt to use the same number of contributors in both the numerator and denominator to calculate the LR. We note that for purposes of this work,  $n_p = n_d$  in all cases presented herein, and we use the known, and thus the true  $n$  to test the capabilities of MatchIt.

For this study, we use the following hypotheses for  $H_p$  and  $H_d$ :

$H_p$ : The evidence is a mixture of the genotype profile of a suspect ( $s$ ) and the profiles of  $n - 1$  other unknown, unrelated contributors, whom we term the ‘interference’ contributors.

$H_d$ : The evidence is from  $n$  unknown individuals unrelated to the suspect.

Our algorithm assumes a constant mixture ratio at all the loci. A constant mixture ratio model assumes that the mixture ratio is the same at all the markers. The STR loci used for forensic DNA analysis are assumed to be in linkage equilibrium and independent.

The  $p$ -value for the suspect is defined as the probability that a randomly picked person from the population would give rise to an LR at least as large as the one observed for the suspect.

$$p\text{-value}(s) = \Pr(LR(R) \geq LR(s)).$$

In this study, we use 1 billion or  $10^9$  random genotypes to compute the  $p$ -value.

Let  $\bar{R}$  be the genotype of an unknown contributor in the defense’s hypothesis, and  $\mathbf{U}^{n-1}$  the genotype of the  $n - 1$  other unknown contributors. The denominator of the LR can be written as:

$$\Pr(E|\mathbf{U}^n) = \sum_{\bar{r}} \Pr(E|\bar{R} = \bar{r}, \mathbf{U}^{n-1})\Pr(\bar{R} = \bar{r})$$

Since the number of possible values that  $\bar{R}$  can take is large and summing over all of them is computationally intensive, we utilize the random genotypes  $r^i$  that are sampled for the  $p$ -value computation to compute the denominator of the LR.

## Methods

The DNA was extracted from single source stains using standard organic extraction procedures. Absolute DNA quantification was performed using real-time PCR and the Quantifiler<sup>®</sup> Duo<sup>™</sup> Quantification kit according to the manufacturer's recommended protocol and one external calibration curve [1]. The extracted DNA was amplified using the manufacturer's recommended protocol (29 cycles) for AmpF $\ell$ STR<sup>®</sup> Identifiler<sup>®</sup> Plus Amplification Kit (Life Technologies, Inc) or PowerPlex 16 HS. Single source samples were amplified using 1, 0.5 0.25, 0.125, 0.063, 0.047, 0.031, 0.016 and 0.008 ng of DNA. Amplification controls were run and showed expected results. Fragment separation was accomplished on a 3130 Genetic Analyzer (Life Technologies, Inc.). Five, ten, and twenty second injections at 3 kV were performed on each of the samples. Fragment analysis was performed using GeneMapper IDX v1.1.1 (Life Technologies, Inc.) using an RFU threshold of 1. Artifacts were filtered as per the laboratory's standard operating procedure.

### Phase 1. Sample Preparation & Stability Studies

Data were exported, and sorted according the parameter being studied (capillary, injection, amplification, or kit lot), target mass, sample number, and locus. The data collected from samples analyzed on different capillary lots were identified as *Validation 1*; data collected from samples injected multiple times on one capillary were identified as *Validation 2*; data collected from samples amplified multiple times with one kit lot were identified as *Validation 3*; and data collected from samples amplified with different kit lots were identified as *Validation 4*. The between-replicate peak height variability was assessed by determining the variance in peak height for each allele  $a$ , for each validation  $v$ , which we term  $\text{var}(H_{a,v})$ . We utilize the following equation

$$\frac{\text{var}(H_{a,v})}{APH_{a,v}} = c_{a,v}^2$$

to determine  $c_{a,v}^2$ . Consequently, for laboratory processes with low variability we expect a small  $c_{a,v}^2$ .

Multiple regression was performed in order to identify the laboratory process most predictive for changes in  $c_{a,v}^2$ . A value of 0 was assigned if the same capillary was used between injections, and a value of 1 was assigned when different capillaries were utilized. Similarly, we assigned values of 1 when the group

contained multiple amplifications utilizing one kit and multiple amplifications utilizing multiple kits, respectively. Heterozygous balance within a locus was also determined. Since variability in  $Hb$  proportionally decreases with APH as per [2, 3],

$$\text{var}(Hb_{l,v}) = \frac{\sigma_{l,v}^2}{APH_{l,v}}$$

multiple regression of  $\sigma_{l,v}^2$  was performed to evaluate how laboratory changes combine to predict the variance of  $Hb_{l,v}$ . Impacts of the four laboratory alterations were also assessed by evaluating baseline noise between sets and the frequency of drop-out.

## Phase 2. Evaluation of MatchIt

The MatchIt algorithm was tested on 101 mock single-source, 2- and 3- person samples using the laboratory protocols described above, which were injected using 3 different injection times (5, 10 and 20 seconds) resulting in a total of over 300 electropherograms. The amplification targets ranged from 0.008 ng to 1 ng and the ratios ranged from 1:1 to 1:19 and 1:1:1 to 1:9:9 for the 2- and 3- person mixtures respectively.

## Data Analysis

### Phase 1. Stability Studies

The impact of four laboratory alterations on DNA signal were tested and compared by examining variability in peak heights. The results are shown in Tables 2.

Table 2. Coefficients and $p$ s after regression of $c_{a,v}^2$ , (constant associated with peak height variance) against variables categorically indicating whether different capillary lots, amplifications, or kits were utilized.					
Validation (v)	Laboratory Modification	Intercept	$\beta_v$	Error $\beta_v$	p
1	Capillary Lot		1.8	1.3	0.18
3	Amplification	0.31	33.4	1.4	1e-117
4	Kit Lot	(0.94)	-8.0	1.3	1e-09

Similarly, the coefficients ( $\beta_v$ ) of  $\sigma_{l,v}^2$  (constant associated with peak height ratio variance) against the categorical variables, indicating different capillary lots, amplifications and kit lots during validation, were 79 (p=0.53), 307 (p=0.01) and -128 (p=0.28), respectively, suggesting amplification kit lot did not have a substantial impact on peak height reproducibility or variation in peak height balance. The average peak

height at which the probability of drop-out is 0.05 was determined to be 169, 177, 182 and 186 for *Validations 1-4*, respectively, suggesting that changes in capillary and kit lot did not significantly impact levels of drop-out between validation sets. The largest difference in the means of the noise peaks between validation sets was 2, suggesting only minor changes in baseline noise occurs due to the introduction of various capillary or kit lots.

## Phase 2. Evaluation of MatchIt

### *LR and p-values for true contributors*

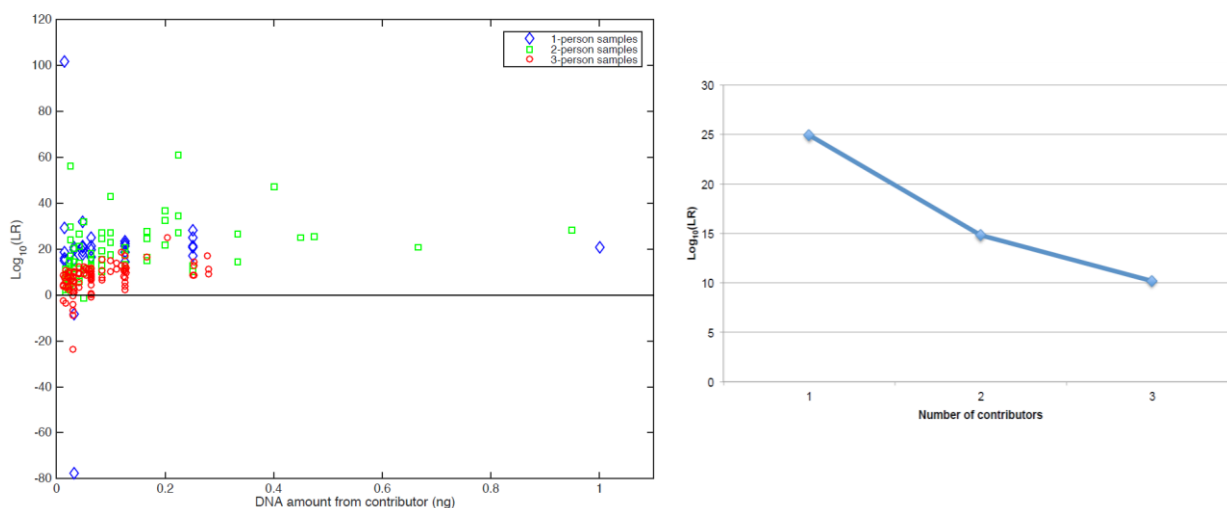


Figure 1. Left:  $\text{Log}(\text{LR})$  for 101 samples (10sec injection) 1-, 2- and 3-person mixtures when the known was an actual contributor. Right: The change in the  $\text{log}(\text{LR})$  as the complexity of the mixture increases.

Figure 1 (left panel) shows LR vs contributor template mass for all the contributors for 101, 1-, 2- and 3-person samples at the 10s injection in the testing set, when the known was one of the actual contributors. The LRs for the 5s and the 20s samples were similar to the 10s samples (data not shown). For each sample, the LR was computed for all the contributors. Hence, a single source sample would have 1 LR; a 2-person sample would have 2 LRs, etc. We observed that the amount of template DNA from the contributor impacted the LR from MatchIt, i.e. high LRs corresponded to high template DNA amounts. MatchIt computed the  $\text{log}(\text{LR})$  as less than 0 (or  $\text{LR} < 1$ ) for a contributor in 11 cases. When investigated, it was observed that the actual contributor was in low quantities, resulting in high levels of



dropout and stutter. Further, for a given single-source sample and template DNA mass, the LR was highest for the 1-person sample and decreased in the presence of an interference contributors (Figure 1, right panel).

*High LRs correspond to low p-values*

Figure 2 shows how the  $p$ -values from MatchIt vary with LR for the samples.

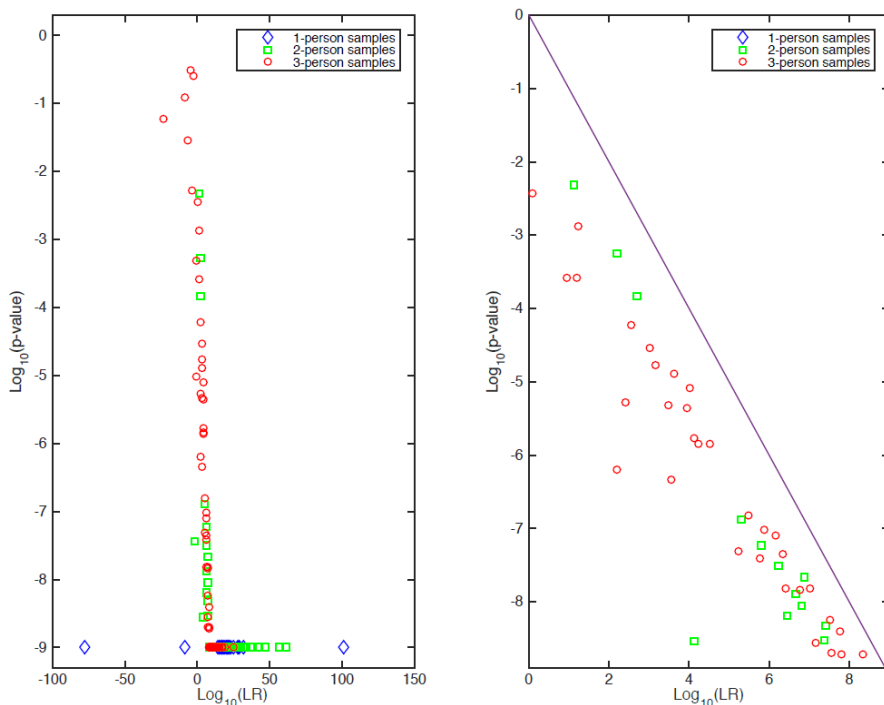


Figure 2. Left: The  $\text{Log}(p\text{-value})$  versus the  $\text{log}(\text{LR})$  for the 101 (10 sec injection) 1-, 2- and 3-person mixtures. Right: Zoomed in sample plot.

All the 1-person samples had the lowest possible  $p$ -value of  $10^{-9}$ . Even the 1-person samples that had a  $\text{LR} < 1$  had a  $p$ -value of  $10^{-9}$ . For the 2-person and 3-person samples, in all cases where the  $\text{log}(\text{LR})$  was greater than 8, the  $\text{log}(p\text{-value})$  from MatchIt was -9 (the lowest possible with 1 billion samples). Figure 2 (right panel) shows the same plot zoomed in on the  $\text{log}(\text{LR})$  values between 0 and 9. All the  $\text{log}(p\text{-value})$  points lie below the line representing  $-\text{log}(\text{LR})$ , as expected. The 5s and the 20s samples showed a trend similar to the 10s samples (data not shown).

*LRs for non-contributors are low*

MatchIt determines the LR distribution by sampling the genotypes of  $10^9$  non-contributors. Figure 3 (left

panel) is a representative example of the LR distribution for a given sample.

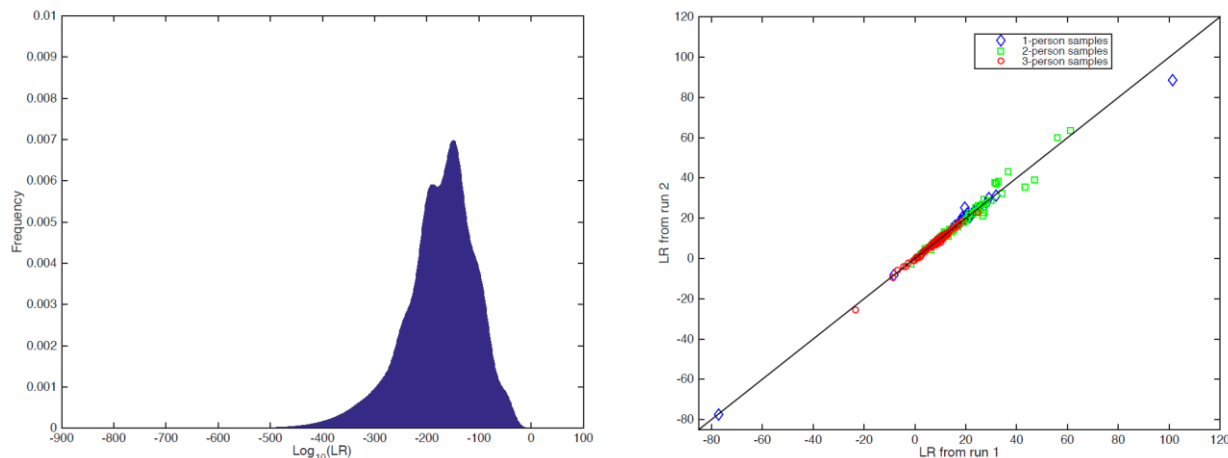


Figure 3. Left: Representative  $\log(\text{LR})$  distribution obtained from MatchIt for a 2-person mixture, with 0.25ng of template DNA and a 1:2 mixture ratio. Right: Repeatability results when 101 (10 sec) 1-, 2- and 3- person mixtures were run in MatchIt, in duplicate

For each sample, we calculated the fraction of the  $10^9$  genotypes that resulted in a  $\text{LR} > 1$  and multiplied it by  $\Pr(R \in \mathbf{R}_1)$  to calculate the Type I error Tippett statistic  $\Pr(\text{LR} > 1 | H_d)$ . The average Type I error statistic for the 1-person samples was  $8.56 \times 10^{-12}$  and it increased to 0.001 and 0.002 for the 2- and 3-person samples. The minimum LR observed for the random contributors was  $10^{-2581}$ , while the maximum was  $10^{31}$  (for a 1-person sample in which the true contributor to the sample also had a LR of  $10^{31}$ ).

#### *Repeatability and Runtime*

Figure 3 (left panel) shows the results of 2 different runs on the 101, 1-, 2- and 3-person samples at the 10s injection. There is little variation from run to run;  $R^2 = 0.9764$ , slope = 0.98, intercept = 0.14).

For single source samples, the average running time was 7.75 minutes, and it increased to 49.64 and 150.24 minutes for 2- and 3-person samples, respectively. The calculations were done using 8 cores on an Intel E3 3.4GHz processor.

#### List of Scholarly Products

1. Ullrich J. Moenich, Catherine M. Grgicak, Viveck Cadambe, Jason Yonglin Wu, Genevieve Wellner, Ken Duffy, Muriel Medard. A Signal Model for Forensic DNA Mixtures. Ansilomar Conference, Ansilomar, CA. November 2014.
2. Kayleigh Rowan, Genevieve Wellner, Catherine M. Grgicak. Exploring the Impacts of Ordinary Laboratory Alterations During Forensic DNA Processing on Peak Height Variation, Thresholds and Probability of Drop-out. Journal of Forensic Sciences (in press).

3. D. S. Lun, Harish Swaminathan, Abhishek Garg, Catherine M. Grgicak and M. Medard. Systems and Methods for Computational Interpretation of STR Profiles. WO Application PCT/US62/145,131, Filed: April 9, 2015

#### Implications to Criminal Justice Policy and Practice

Stability studies suggest that if single-source casework samples, run over a significant period of time, are not available the laboratory should evaluate intermediate levels of precision by incorporating multiple capillary lots during validation. Additionally, we have brought together a continuous method to compute the LR based on modeling of the peak heights in known data and to calculate a  $p$ -value from the LR distribution by simulation of genotypes based on allele frequencies. The method was tested on 306, 1-, 2- and 3-person electropherograms containing between 0.016 and 1 ng of DNA. The amount of template DNA from the contributor had an impact on the LR – small LRs arose from contributors with low template masses, indicating that high levels of dropout and stutter could decrease the probability of the evidence under the prosecution's hypothesis even for true contributors. Since we used  $10^9$  samples to calculate the  $p$ -value, the lowest possible  $p$ -value that can be achieved is  $10^{-9}$ , and this was obtained in all the cases where the LR was greater than  $10^8$ . The results from MatchIt were found to be repeatable after duplicate runs on all the samples in the testing set.

#### References

1. Grgicak, C. M.; Urban, Z. M.; Cotton, R. W., Investigation of Reproducibility and Error Associated with qPCR Methods using Quantifiler Duo DNA Quantification Kit. **2010**, 55, (5), 1331-1339.
2. Kelly, H.; Bright, J.-A.; Curran, J. M.; Buckleton, J., Modelling Heterozygote Balance in Forensic DNA Profiles. *Forensic Sci. Int: Genetics* **2012**, 6, 729-734.
3. Bright, J.-A.; Turkington, J.; Buckleton, J., Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Sci. Int. - Genetics* **2010**, 4, 111-114.