Science for Communities

Article:

Buckleton, J.S., Bright, J.A., Cheng, K., Budowle, B., & Coble, M.D. (2018). **NIST interlaboratory studies involving DNA mixtures (MIX13): A modern analysis**. *Forensic Science International: Genetics*, *37*, 172–179.

INSTITUTE OF ENVIRONMENTAL SCIENCE AND RESEARCH LIMITED
Kenepuru Science Centre: 34 Kenepuru Drive, Kenepuru, Porirua 5022 | PO Box 50348, Porirua 5240, New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770
www.esr.cri.nz

NIST interlaboratory studies involving DNA mixtures (MIX13): A modern analysis

John S. Buckleton[a,b*], Jo-Anne Bright[a], Kevin Cheng[a], Bruce Budowle[c], Michael D. Coble[c]

[a] *Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand*

[b] *University of Auckland, Department of Statistics, Auckland, New Zealand*

[c] *Center for Human Identification, Department of Microbiology, Immunology, and Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX, 76107, USA*

* Corresponding author at: Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand. Email address: john.buckleton@esr.cri.nz.

MIX13 was an interlaboratory exercise directed by NIST in 2013. The goal of the exercise was to evaluate the general state of interpretation methods in use at the time across the forensic community within the US and Canada and to measure the consistency in mixture interpretation. The findings were that there was a large variation in analysts' interpretations between and within laboratories.

Within this work, we sought to evaluate the same mock mixture cases analyzed in MIX13 but with a more current view of the state-of-the-science. Each of the five cases were analyzed using the Identifiler™ multiplex and interpreted with the combined probability of inclusion, CPI, and four different modern probabilistic genotyping systems. Cases 1–4 can be interpreted without difficulty by any of the four PG systems examined. Cases 1 and 4 could also be interpreted successfully with the CPI by assuming two donors.

Cases 2 and 3 cannot be interpreted successfully with the CPI because of potential of allele dropout. Case 3 demonstrated the need to consider relevant

background information before interpretation of the profile. This case does not show that there is some barrier to interpretation caused by relatedness beyond the increased allelic overlap that can occur. Had this profile been of better template it might have been interpreted using the CPI despite the (potential) relatedness of contributors.

Case 5 suffers from over-engineering. It is unclear whether reference 5C, a non-donor, can be excluded by manual methods. Inclusion of reference 5C should be termed an adventitious match not a false inclusion. Beyond this statement this case does not contribute to the interlaboratory study of analyst/laboratory interpretation method performance, instead, it explores the limits of DNA analysis.

Taken collectively the analysis of these five cases demonstrates the benefits of changing from CPI to a PG system.

Keywords: Forensic DNA; DNA mixture; Mixture interpretation; Interlaboratory study; Collaborative exercise; MIX13

Introduction

Mixture interpretation is one of the greatest challenges to forensic DNA typing. There are two general approaches that have been used to assess the strength of mixture evidence: the combined probability of inclusion (CPI) and the likelihood ratio (*LR*). Each have their strengths and limitations, and it is incumbent on the user to appreciate how to use each method appropriately. The CPI is limited often to two person mixtures and to loci in which there is little or no evidence of allele dropout [1, 2]. Thus, the CPI does not make full use of the mixture data. Even with the simplified approach of the CPI, some laboratories have not implemented it correctly. Indeed, because allele dropout was not adequately addressed, there have been CPI calculations that have been overstated and that inappropriately used the reference profile of the person of interest (POI) [3]. Two collaborative studies, known as MIX05 and MIX13, orchestrated by NIST, distributed in 2005 and 2013 [4, 5] also showed the lack of proper implementation of the CPI. Bieber et al. [1] described how to properly apply the CPI.

A more cogent approach to mixture interpretation is the assignment of an *LR* as it makes better use of the data. The ISFG DNA Commission (2006) [6] recommended the use of *LR* based methods.

Initially, *LR* methods suffered from the same limitation as the CPI of not being able to accommodate allele dropout. However, with the advent of the semi-continuous model [7] more challenging

mixtures could be analyzed. Today, continuous models (see references [8-12] for general discussions on the use of PG systems) expressed with an *LR* are the methods of choice as they allow for interpretation of complex mixtures and partial profile data.

Discussion of MIX13 continues [13, 14]. This work and previous presentations confirm that the translation of mixture interpretation theory and practice to the bench analyst was ineffective as of 2013 and resulted in some laboratories inappropriately interpreting mixture data [1, 3, 16]. These findings are not contested by the forensic genetics community, but neither do they shed light on the current state of mixture interpretation as of 2018. Discussions of mixture interpretation practices of 5-10 years ago or more are good for historical purposes and lessons learned, but we believe this leaves a gap in that they provide little guidance for moving forward. To place into perspective how mixtures can be interpreted we apply the capabilities of today (i.e., using several PG software systems) to NIST MIX13. Our analyses point out how PG can improve interpretation of mixtures over the CPI and early *LR* approaches, some design challenges of the MIX13 study that do not demonstrate methodology issues and instead show limitations of any extreme contrived mixtures, and hopefully a balanced position on the current state of mixture interpretation.

Data

As part of NIST MIX13, five mock cases were prepared or collated by NIST and made available electronically. The profiles, references, and study design are available at https://strbase.nist.gov/interlab/MIX13.htm. Each case consisted of an electropherogram file (.fsa format) and typed reference profiles from individuals described as victims, consensual partners, or suspects. The mixed profile electropherograms were made available as both Life Technologies Identifiler™ Plus (or Identifiler™) or Promega PowerPlex® 16 (or PP16 HS) profiles. Brief case scenarios were provided to put each case into perspective (Table 1). Note that we make no comment on what is task-relevant information in constructing a case scenario; we simply are stating what NIST provided to the participants. Mixtures 1, 3, 4, and 5 were made at NIST. Mixture 2 was made at Boston University (BU). In this work we describe the interpretation of the Identifiler™ Plus and Identifiler™ profiles only. However, the findings generally should apply to the PP16 HS profiles with proper modeling.

Method

Although the information on the mixtures have been made public, we made the subjective decisions blindly. We note that there are far fewer (but not nil) subjective decisions when using PG. However, since we cannot preclude bias in ascertaining our decisions all relevant decisions are disclosed so the readership can judge whether the decisions made are justified.

Profiles were analyzed in GeneMapper ID-X using an analytical threshold (AT) of 50 rfu for the NIST profiles and 30 rfu for the BU profile (case 2). For the Lab Retriever interpretation, profiles were analyzed using an AT of 30 rfu [17], except for Case 4 where the NIST AT of 50 rfu was used as requested in the study design.

These five mixtures were interpreted using four PG software systems: EuroForMix [12, 18, 19], Lab Retriever [20, 21], LRmix Studio [22, 23], and STRmix™ [8-10]. All of these PG methods produce *LR*s and require two propositions and an assignment of the number of contributors. EuroForMix and STRmix™ use continuous models, Lab Retriever and LRmix use semi-continuous models. While there

are several software now available, using all in this study would be a substantial resource burden. These four were chosen because of their ready access to the authors and provide insight on how PG may perform compared with the CPI. Since the mixture data are publicly available, we welcome users of the other software to carry out similar studies as described herein so we all can gain a better understanding of how various software perform.

For these cases the number of contributors and propositions are listed in Table 2. The number of contributors was assigned by human examination of the number and height of allelic peaks. We acknowledge the value of new technologies such as NOCIt [24] and PACE [25] that may reduce human input, but those software systems were not applied here so as not to add another variable to the comparisons.

Many of these software systems require some input parameters. The key settings or processes are described in Table 3.

These *LR*s should always be less than or equal to approximately 1/RMP for the person of interest. This is approximate largely because of microvariants in the way rare alleles are calculated. RMP is calculated using the Balding and Nichols formulae [27] using the same allele probabilities and F_{ST} values as used in the various software systems. There are two instances of a rare (or unobserved) allele from the FBI Caucasian allele frequencies (Case 3, D21S11, allele 35 and vWA allele 21). EuroForMix and Lab Retriever used 5/2N as the probability for this allele. LRmix used the (user configurable) frequency of 0.001. STRmix™ uses a posterior mean frequency for all alleles, $\left(x+\frac{1}{k}\right)/\left(N+1\right)$, where x is the number of observations and k the number of allele classes at the locus. For the RMP calculation 5/2N was used for the two instances mentioned above but other allele frequencies below 5/2N were not altered.

Table 1. Description of five case scenarios and profiles as part of NIST MIX13

Case	Target mixed DNA	References	Case scenario (abridged)
1	2-person mixture, 1:1 ratio	One true donor (reference 1A) supplied	Alleged male on female rape. Consciousness lost. Evidence is the sperm fraction from a vaginal swab. Reference 1A is the accused.
2	3-person mixture, 6:1½:1 ratio (300pg)	2A, 2B, 2C, (true donors) and 2D (non-donor)	Homicide of store employee. Video … shows two perpetrators enter the store, with one individual holding a gun … handgun found nearby connected to homicide by firearms examination to fired bullets. Evidence is a DNA profile generated from a swab from handle of the recovered handgun. Four suspects (References 2A, 2B, 2C, and 2D)
3	3-person mixture, 7:2:1 ratio	The reference profile of the victim and consensual partner were provided. Two suspect references 3A (true donor) and 3B non-donor.	Female victim had too much alcohol to drink … at some point in the middle of the night, she awoke with someone on top of her performing intercourse. Consciousness lost. Evidence is a DNA profile generated from the sperm fraction from a vaginal swab collected from the victim. DNA samples from the two men remaining in the house according to the boyfriend before he remembers passing out: his brother (Reference 3A) and one other unrelated male (Reference 3B). The victim and her boyfriend confirmed that they had consensual sex about 12 hours prior to the assault.
4	2-person mixture, 3.5:1 ratio	Reference 4A (true donor)	A female waiting … is attacked from behind and pushed to the ground… the perpetrator bites the victim on the back of her neck … DNA profile generated from saliva found when swabbing the bite mark on the victim.
5	4-person mixture, 1:1:1:1 ratio	References 5A and 5B (true donors), 5C (non-donor)	Several gang-related robberies … typically involved two or three perpetrators. A ski mask was recovered in a trash can one block away from the latest bank robbery and is submitted for DNA testing. Evidence is a mixed DNA profile developed from a ski mask recovered near a bank robbery scene.

Table 2. Summary of analysis and interpretation settings of PG software

Case	Target mixed DNA	Analytical threshold (rfu)	Assigned number of contributors	Propositions
1	2-person mixture, 1:1 ratio	50 (LRmix, EuroForMix, STRmix™), 30 Lab Retriever	2	H_1: Complainant and POI H_2: Complainant and an unknown unrelated individual
2	3-person mixture, 6:1½:1 ratio (300pg)	30	3	H_1: POI and two unknown unrelated individuals H_2: Three unknown unrelated individuals
3	3-person mixture, 7:2:1 ratio	50 (LRmix, EuroForMix, STRmix™), 30 Lab Retriever	3 (this assignment can only be made by assuming the presence of the consensual partner)	H_1: Complainant, consensual, and POI H_2: Complainant, consensual and an unknown unrelated individual
4	2-person mixture, 3.5:1 ratio	50	2	H_1: Complainant and POI H_2: Complainant and an unknown unrelated individual
5	4-person mixture, 1:1:1:1 ratio	50 (LRmix, EuroForMix, STRmix™), 30 Lab Retriever	3 (there is no legitimate way to diagnose this mixture as a four person mixture without knowledge of the reference profiles)	H_1: POI and two unknown unrelated individuals H_2: Three unknown unrelated individuals

Table 3. Summary of interpretation parameters used for each PG software

Software	
EuroForMix v1.10 and v1.11.4	MLE method used.
Lab Retriever v2.2.1 LRmix Studio	The probability of dropout was set, separately for each donor where possible, based on the average peak height for each contributor. The average peak height was transformed to a probability of dropout using the same model as embedded in STRmix™. This approach controls this variable. Peaks in back stutter positions were assigned as stutter or ambiguous stutter/allelic by examination of each profile electropherogram. If a peak in a stutter position was approximately the height of the minor contributor and there were minor alleles unvisualised it was deemed ambiguous. Otherwise these peaks were treated as stutter. Ambiguous peaks were treated by creating a virtual profile containing only the ambiguous peaks and treating this profile as a known contributor, adjusting the assigned number of contributors accordingly.
STRmix™ v2.5.11	Allelic variance (α, β) 3.57, 0.98, Stutter variance (α, β) 6.97, 1.75, Locus amplification variance 0.03. The point *LR* is reported in this paper for better comparison with the other methods. Most labs, however, report a lower bound in casework.
All software	The probability of drop-in was set to 0. The allele probabilities from the FBI extended dataset [26] were used. $\theta = 0.01$

Results

Case 1

This case presented little trouble in the pre-PG era. All (of 108) participants in the initial survey correctly included reference 1A and provided a statistic. A fairly wide range was observed in the statistical values reported [14] (note that participating laboratories optionally used different analytical thresholds and allele frequencies to calculate their match statistic, likely to create differences in reported values) [4, 5]. All four PG tested also included reference 1A with as much as four orders of magnitude difference between software systems (see Table 4). The continuous model software systems reported the larger *LR*s and the semi-continuous software systems essentially reported the same *LR*.

EuroForMix v1.10.0 initially produced an *LR* greater than 1/RMP. Investigation of this with the developer, Øyvind Bleka, identified a bug which was promptly fixed and reported (http://www.euroformix.com/?q=changes).

Table 4. PG results for NIST MIX13 case 1

Software	*LR*
STRmix™	1.4×10^{20}
EuroForMix v1.10.0	2.7×10^{20}
EuroForMix v1.11.4	1.5×10^{20}
Lab Retriever	4.1×10^{15}
LRmix	3.6×10^{15}
1/RMP	1.9×10^{20}

Case 2

It was reported that most (72 of 108) of the laboratories correctly included reference 2A [4, 5]. Fewer participants included the remaining true donors: references 2B (39 of 108) and 2C (15 of 108). The non-contributor, reference 2D, was falsely included in the mixture by one laboratory. Most laboratories either excluded (73 of 108) or gave an inconclusive result (33 of 108) for the non-donor 2D.

Table 5. PG results for NIST MIX13 case 2

Software	*LR*			
	Ref 2A	Ref 2B	Ref 2C	Ref 2D
STRmix™	9.6×10^{16}	1.8×10^{7}	6.7×10^{5}	9.3×10^{-15}
EuroForMix v1.10.0	1.9×10^{17}	7.5×10^{7}	1.3×10^{6}	4.4×10^{-3}
EuroForMix v1.11.4	1.9×10^{17}	7.5×10^{7}	1.3×10^{6}	4.4×10^{-3}
Lab Retriever	6.5×10^{4}	5.4×10^{5}	840	8.7×10^{-11}
LRmix	8.0×10^{4}	7.3×10^{5}	1,100	2.4×10^{-9}
1/RMP	9.5×10^{17}	9.3×10^{17}	3.6×10^{18}	n/a

The four PG tested in this work all correctly included or excluded the references (see Table 5). Again the continuous model software systems gave larger and similar *LR*s for the included reference samples; while the semi-continuous model software systems yielded lower, and similar, *LR*s for the included reference samples. For the non-contributor reference sample the most notable difference in performance was that of EuroForMix which had a more modest *LR* favoring H_2.

Case 3

NIST created this three-person mixture from a female sample (complainant), and a pair of real brothers, brother #1 (consensual partner) and brother #2 (POI, reference 3A) in a 7:2:1 ratio. A non-contributor was provided as a POI reference (reference 3B).

Most laboratories (90 of 108) correctly excluded the non-donor reference 3B. A summary of the

alleles of the crime profile and three reference profiles for the Identifiler™ Plus kit is provided in Table 6.

This scenario has two points to consider:

1. Can the consensual partner be used to assign the number of contributors and,

2. Does the relatedness of a genotyped contributor have any effect?

The agreed facts are that the victim, and consensual partner, had sex 12 hours prior. The time to sampling, which can be crucial, was not provided in the scenario. The electropherogram can be explained as a two-person mixture if the victim but not the consensual partner is assumed.

Some practitioners in the US (and in the NIST study) generally have not considered to condition on a consensual partner or even at times on a victim (even with reasonable expectations of either or both of them being a contributor such as from an intimate sample) for mixture interpretation. Based on case circumstances it may be reasonable to assume certain individuals as known contributors to a mixture. Assuming three contributors to the profile, the STRmix™ solution for the mixture proportions is about 0.71 victim, 0.26 consensual partner, and 0.03 unknown contributor. If the consensual partner is not assumed as a contributor, then reference 03A would have been excluded by all software systems under the assumption of two contributors.

Given the description of the case circumstances, there is a reasonable expectation of DNA from the consensual partner [28]. Therefore, it is sound to assume the presence of DNA from this donor. Given the assumption of two known contributors, the profile is better explained as a three-person mixture. This assumption seems reasonable and was performed by us in this manner when originally blindly analyzing this profile. In doing so, there are three alleles that are not masked by the victim and consensual partner or the stutters of their alleles (the unmasked peaks are in blue and italicized, and the unmasked but in a stutter position is in red and bold in Table 6; for illustrative purposes these alleles are highlighted for both mixture and reference 03A profiles). The decision about the mixture being composed of three contributors can be made without knowledge of the two suspects' genotypes which should not be examined prior to determining the number of contributors [2, 29].

There has been some confusion about relatedness regarding the application of the CPI and *LR* for mixtures. For example Butler et al. [14] state: Several of the laboratories in their responses recognized the issue of a related person in the mixture and responded with something like "due to the relatedness of the exemplars submitted for comparison, a statistical analysis cannot be provided at this time."

The relatedness of one of the persons of interest to the consensual partner poses no issues in determining inclusion or exclusion of this person. When it comes to providing a statistic for the inclusion, the "random person" referred to is not the POI and therefore there is no evidence of relatedness. Hence the relatedness of a genotyped individual to any person in the case may increase allelic overlap but does not affect the production of any statistic. All four PG software made correct inferences regarding inclusion and exclusion with similar *LR*s for reference 3A and reference 3B (see Table 7).

Table 6. Summary of the peaks (AT 50 rfu) and the three reference (true contributor) profiles for the Case 3 Identifiler™ Plus profile. ( ) indicates low signal peaks that may be stutter if an appropriate parent peak is present. There are three alleles in the mixture that are not associated with the victim (reference supplied), consensual partner (reference supplied), or explained by stutter.). Reference 3A has those three unmasked alleles (unmasked peaks are in blue and italicized, and the one unmasked peak in a stutter position is in red and bold).

Markers	Mixture	Victim	Consensual Partner	Suspect 03A (brother of consensual partner)
D8S1179	12,14,15	12,15	14,14	14,15
D21S11	28,(30.2),31.2,35	31.2,31.2	28,35	28,35
D7S820	(9),10,11	10,10	10,11	10,11
CSF1PO	10,11,12	10,11	10,12	12,12
D3S1358	(13),14,18	14,14	14,18	14,18
TH01	7,8,9.3	9.3,9.3	7,8	7,8
D13S317	11,12,13	11,12	12,13	12,13
D16S539	(**8**),9,10,(11),12	9,12	10,10	**8**,9
D2S1338	(16),(*17*),(19),20	20,20	16,20	16,*17*
D19S433	(13),14,(*14.2*)	14,14	14,14	14,*14.2*
vWA	(14),15,17,21	15,15	17,21	17,21
TPOX	6,8,9,11	9,11	6,8	8,9
D18S51	12,13,16	12,13	13,16	13,16
D5S818	10,11,12	11,12	10,12	10,12
FGA	20, *23*,26,27	20,26	26,27	*23*,27

Table 7. PG results for NIST MIX13 case 3

Software	*LR*	
	Ref 3A	Ref 3B
STRmix™	4.9×10^{7}	0
EuroForMix v1.10.0	6.8×10^{6}	0
EuroForMix v1.11.4	6.6×10^{6}	0
Lab Retriever	2.1×10^{7}	0
LRmix	2.1×10^{8}	0
1/RMP	9.8×10^{23}	n/a

A note on concerns about mixtures involving relatives

The confusion regarding how to address various scenarios regarding relatives and mixtures may have influenced the interpretation outcome on case 3 by some analysts partaking in the NIST MIX13 study. There is a general lack of clarity about exactly what issues arise from relatedness. Butler [30] at pg 336 -7 stated: "***Urban Legend #9: CPI works fine even if potential relatives are in the mixture.*** *The CPI statistic is based on a model using unrelated people … the model …may not hold up with related individuals in the mixture*." Perhaps the confusion comes from misapplication of the terminology "*in the mixture*." There is no issue using the CPI or an *LR* if the true contributors of the DNA in a mixture are related. The proper consideration arises if a postulated alternate donor is related to someone else that has supplied a reference profile. To appreciate the concept, two situations are discussed here:

1. The individuals postulated to be included in the mixture (H_1) are related and have been genotyped, and

2. One, or more, of the postulated unknown individuals (H_2) is related to the POI.

Situation 1 - Relatives in the mixture and genotyped: Consider a situation where an *N* donor mixture is proposed to be from relatives (H_1). The alternative proposition is that the mixture comes from *N* unknown persons. Even with the allelic overlap induced by their relatedness, under the contributor proposition it is straightforward to assign the probability of the mixture if the *N* relatives are the contributors.

Since the alternative proposition is that there are all unknown people (no relationships are known) contributing to the mixture, no new approaches to accommodate this scenario (and similar ones) are needed to determine the various genotypes that explain the mixture profile. Thus calculating the probability of observing the evidence given each hypothesis is straightforward.

Situation 2 - Postulated unknown individual (H_2) is related to the POI: Assigning an *LR* for this alternative proposition requires a different set of formulae than those used for Situation 1, described above [31]. There are many formulae required since the POI may be a homozygote aa or heterozygote ab, and the genotypes needed for the mixture calculations may be any or all of aa, ab, aQ, QQ where Q is neither a nor b. As an example we give some of the formulae in Table 8.

Table 8. Genotype and example genotype probabilities (for unrelated and siblings). *Q* is any allele other than a for a homozygote aa POI, and any allele other than a or b for the heterozygote POI ab. p_i is the probability of allele *i*. θ is the coancestry coefficient. Other formulae are possible if the POI and others are used in the conditioning.

Genotype of POI	Genotype for mixture calculation	Genotype probability	
		Unrelated	*siblings*
aa	aa	$\frac{(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}{(1+\theta)(1+2\theta)}$	$\frac{1}{4}+\frac{(2\theta+(1-\theta)p_a)}{2(1+\theta)}+\frac{(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}{4(1+\theta)(1+2\theta)}$
	aQ	$\frac{2(2\theta+(1-\theta)p_a)(1-\theta)p_Q}{(1+\theta)(1+2\theta)}$	$\frac{(1-\theta)p_Q}{2(1+\theta)}+\frac{(1-\theta)p_Q(2\theta+(1-\theta)p_a)}{2(1+\theta)(1+2\theta)}$
	Q,Q	$\frac{(1-\theta)p_Q(\theta+(1-\theta)p_Q)}{(1+\theta)(1+2\theta)}$	$\frac{(1-\theta)p_Q(\theta+(1-\theta)p_Q)}{4(1+\theta)(1+2\theta)}$
ab	aa	$\frac{(\theta+(1-\theta)p_a)(2\theta+(1-\theta)p_a)}{(1+\theta)(1+2\theta)}$	$\frac{(\theta+(1-\theta)p_a)}{4(1+\theta)}+\frac{(\theta+(1-\theta)p_a)(2\theta+(1-\theta)p_a)}{4(1+\theta)(1+2\theta)}$
	ab	$\frac{2(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{(1+\theta)(1+2\theta)}$	$\frac{1}{4}+\frac{(2\theta+(1-\theta)(p_a+p_b))}{4(1+\theta)}+\frac{(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{2(1+\theta)(1+2\theta)}$
	aQ	$\frac{2(\theta+(1-\theta)p_a)(1-\theta)p_Q}{(1+\theta)(1+2\theta)}$	$\frac{(1-\theta)p_Q}{4(1+\theta)}+\frac{(1-\theta)p_Q(\theta+(1-\theta)p_a)}{2(1+\theta)(1+2\theta)}$
	QQ	$\frac{(1-\theta)p_Q(\theta+(1-\theta)p_Q)}{(1+\theta)(1+2\theta)}$	$\frac{(1-\theta)p_Q(\theta+(1-\theta)p_Q)}{4(1+\theta)(1+2\theta)}$

Case 4

In the initial MIX13 survey nearly all laboratories correctly included reference 4A (106 of 108) and provided statistical weight to their conclusions [14]. All four PG software made correct inferences regarding inclusion with the continuous models providing larger *LR*s than the semi-continuous models (see Table 9).

Table 9. PG results for NIST MIX13 case 4. EuroForMix v1.10.0 (left number) and v1.11.4 (right number)

Software	*LR*
STRmix™	1.4×10^{20}
EuroForMix v1.10.0	9.8×10^{19}
EuroForMix v1.11.4	8.1×10^{19}
Lab Retriever	3.0×10^{16}
LRmix	5.8×10^{16}
1/RMP	1.5×10^{20}

Case 5

Case 5 is a constructed four-person mixture with genotypes selected using a software called Virtual Mixture Maker (David Duewer, NIST) in which there is an intended substantial allelic overlap among the contributors. The four "donors" were selected from the genotypes of 259 Caucasians from the NIST population dataset [32]. It was constructed so that there would be no more than four alleles at any locus. To put this extreme level of overlap in context, a mixture of DNA from four people can have up to 8 non-overlapping alleles detected, although Coble et al. [33] showed by simulation that on average four person mixtures tend to show 5 and 6 alleles per locus (the exact distribution may be calculated following Tvedebrink [34]). Hence, on allele count alone, Case 5 was intentionally designed so the Identifiler™ Plus profile would reasonably be assigned as a two-person mixture, even though the true state is a four-person mix in the ratio 1:1:1:1. Moreover, the peak heights presented as a best fit as a three-person mix in the ratio 2:1:1.

The only way to decide that this fabricated profile is a four-person mixture is to assume the presence of one of the POIs, i.e., reference 5A, which cannot be supported based on the case scenario. This assumption is disallowed by the SWGDAM 2010 [35] and 2017 [36] guidelines.

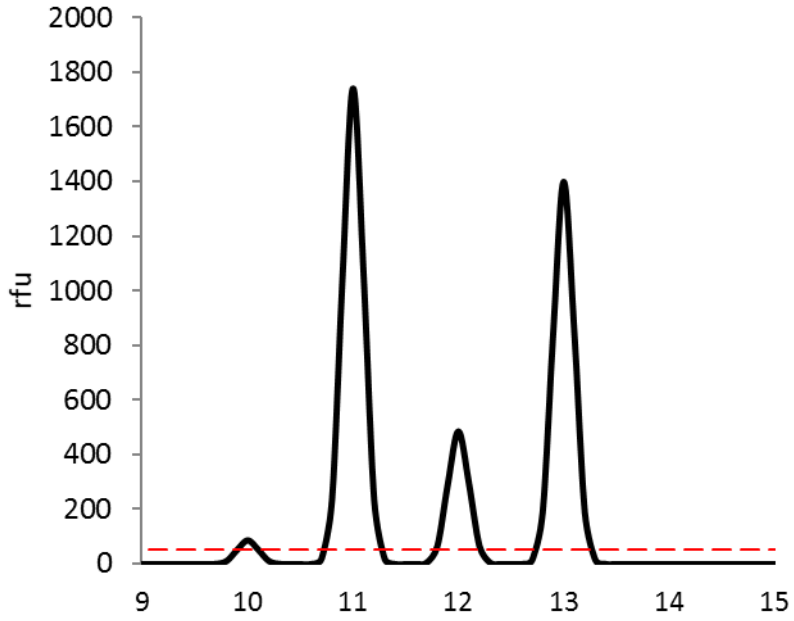
This particular profile may not be that useful for testing an analyst's ability to determine the true number of contributors in a mixture, but instead may be an extreme example of demonstrating the limitation of current CE-based STR typing technology. There are 183 million possible combinations of 4 people amongst 259 individuals that can be used to generate a mixture; only 7 of these fabricated mixed profiles showed a similar high level of allelic overlap. Furthermore, the mixture was in a perfect 1:1:1:1 ratio. Plausibly the chance of getting this particular combination from this set of 259 people in the 1:1:1:1 ratios adds to a probability of observing this mixture markedly less than 7 in 183 million.

Three reference samples were presented in the exercise called 5A, 5B, and 5C. References 5A and 5B were true donors and reference 5C was a non-donor. Reference 5C was constructed deliberately to share alleles among the four individuals in the mixture. Therefore, 5C, as is any profile, is unlikely to exist in any living person (the probability of this profile by the product rule using the FBI extended Caucasian database is approximately 5.66×10^{-15}). This case example is clearly over engineered. The evidence sample is unlikely to occur in reality (fewer than 7 in 183 million), and the chance of getting a non-donor that overlaps the profile is less than 1 in 38,000. But nonetheless, the case was analyzed with the PG software to illustrate performance capabilities.

It was reported that a total of 74 of 108 laboratories included reference 5C (along with true donors 5A and 5B) [4, 5].

It is worthwhile considering what evidence, if any, exists that could exclude reference 5C. In Table 10 the alleles of the donors used to construct this mixture and those of reference 5C are listed. Reference profile 5C fits well at every locus of the Identifiler™ Plus set except at D5S818 allele 12 (reproduced in Figure 1) where reference 5C is a homozygote 12 allele.

Figure 1. The D5S818 locus of MIX13 case 5



Assuming reference 5C is a contributor, the 12 peak is therefore expected at approximately twice the height of the unmasked heterozygote peaks (assuming no degradation or inhibition for this exercise). It is not trivial to identify the unmasked heterozygote peaks without knowledge of the true donors, which would not occur in the case scenario. With hindsight these alleles range between 314 and 662 rfu. Note that the 12 peak is in the back stutter position of the relatively large 13 peak and also in the forward stutter position of the relatively large 11 peak. The 12 peak is at 486 rfu which is within the range of the unmasked heterozygote peaks but less than twice the minimum value ($2 \times 314 = 628$ rfu). As this is the only locus where the 5C reference alleles suggest a poorer fit in an otherwise well-fitting profile, many examiners would still include the reference 5C as a possible contributor (or possibly render an inconclusive interpretation). When originally interpreting these profiles blind, we also included reference 5C as a potential contributor.

Table 10. Summary of Case 5 genotypes

		Loci														
		D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	vWA	TPOX	D18S51	D5S818	FGA	D2S1338	D19S433
True donors	Suspect 05A	10,15	30,31	8,10	12,12	16,17	7,7	11,13	11,12	15,16	8,9	13,15	11,13	22,24	17,23	13,14
	Suspect 05B	14,14	30,31.2	9,10	12,13	16,16	7,7	8,8	11,11	17,18	11,11	15,15	11,13	21,21	17,23	14,15
	"not tested"	10,10	31.2,32.2	10,10	11,11	16,17	6,9.3	11,12	11,13	16,17	8,8	13,15	11,13	21,21	18,20	12,14
	"not tested"	12,14	31,31	8,11	11,12	17,17	6,9.3	11,11	11,12	17,17	8,11	17,17	11,12	21,22	18,23	14,14
Constructed reference	Suspect 05C	10,14	31.2,32.2	10,10	11,12	17,17	6,7	8,11	11,13	15,17	8,8	15,17	12,12	21,24	18,18	14,14

Table 11. PG results for NIST MIX13 case 5

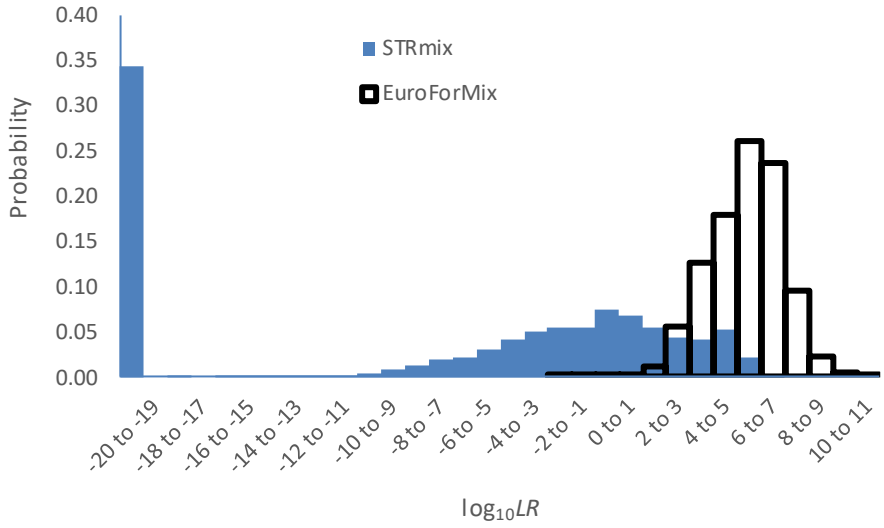
Software	*LR*		
	Ref 5A	Ref 5B	Ref 5C
STRmix™	2.8×10^{3}	2.1×10^{3}	1.2×10^{-8} (1 of 20 runs)
			0 (19 of 20 runs)
EuroForMix v1.10.0	1.0×10^{7}	3.9×10^{6}	5.0×10^{6} (3 of 10 runs)
			3.2×10^{3} (7 of 10 runs)
EuroForMix v1.11.4	1.8×10^{10} (1 of 10 runs)	7.9×10^{9} (3 of 10 runs)	5.0×10^{6} (3 of 10 runs)
	1.0×10^{7} (3 of 10 runs)	3.9×10^{6} (4 of 10 runs)	3.2×10^{3} (2 of 10 runs)
	0.9×10^{7} (4 of 10 runs)	3.4×10^{6} (3 of 10 runs)	2.6×10^{3} (5 of 10 runs)
Lab Retriever	5.8×10^{5}	5.6×10^{4}	4.9×10^{4}
LRmix	5.1×10^{5}	4.4×10^{4}	4.0×10^{4}
1/RMP	8.7×10^{16}	4.8×10^{18}	n/a

The interpretation of the assigned *LR* values in Table 11 is more challenging than in the other cases in MIX13. While all software systems correctly included 5A and 5B, STRmix™ yielded the lowest *LR*. STRmix™ appears to be affected by treating the sample as a three donor mixture, higher *LR*s were obtained when we treated the mixture as a four person mixture. However by this stage we were not blind. The difference in STRmix™ performance may contribute to excluding 5C, in contrast to the other continuous model software. The semi-continuous models performed similarly and included 5C.

The ground truth is that references 5A and 5B are true donors and 5C is constructed to be so close to the true donors (see Table 10) that it is unclear whether it should be counted as included or excluded. An inclusion of such a profile (reference 5C) is better termed an adventitious match. It is worth pointing out that it is subtle aspects of peak height that led STRmix™ to exclude. EuroForMix calculates an *LR* greater than 1, suggesting evidence for the inclusion of reference 5C. The two semi-continuous systems, LRmix and Lab Retriever, would not have excluded reference 5C, since they do not use peak height information. An *LR* > 1 using any semi-continuous model is therefore the expected result in this scenario. These findings may point to another important aspect of using any of the PG software and that is visual inspection for consistency. Whether the software included or excluded 5C, it incumbent upon the user to review the original data and not rely solely on the output.

Within Figure 2, we plot the log(*LR*) values for 10,000 individuals whose profiles were generated by sampling, with replacement, from the alleles of all four of the true donors to Case 5. *LR*s were assigned by STRmix™ (runtime of one minute) using the database search function [37] and EuroForMix (runtime of approximately four and a half days; For the purposes of this study, when calculating *LR*s for the 10,000 database individuals in EuroForMix, the F_{ST} was set to 0 due to runtime issues. It is assumed that *LR*s are marginally higher than if they were calculated using FST = 0.01.). STRmix™ and EuroForMix generated *LR*s less than 1 for 59% and 0.02% respectively of these non-contributors. Fabricating a profile in this way from only the alleles of the true donors creates genotypes with a reasonable expectation of inclusion. The results correctly demonstrate that adventitious associations will be made in some or many cases where all the correct alleles are present.

Figure 2. The results of 10,000 false donors tested against the Case 5 profile using STRmix™ and EuroForMix v1.10.0. The non-contributors have been created by sampling, with replacement, from the alleles of the true donors. 41% of *LR*s are greater than 1 ($\log_{10}LR$ >0) and 59% lower than 1 for STRmix™. 99.98% of *LR*s are greater than 1 for EuroForMix.



It was suggested by the organizers of the study that the goal of this case was to see if participants could diagnose it as too complicated [14]. The profile and one of the POI profiles appear to have been engineered to present insurmountable problems. The profile is a good fit to a good template three-person mixture with high allelic overlap. The high allelic overlap is not, in itself, an indication that interpretation is impossible.

The problem with Case 5 does not lie in the use of any one interpretation strategy. Without the use of continuous model PG systems, and even with the use of one, EuroForMix, we cannot think of any straightforward way to exclude reference 5C.

Conclusions

NIST MIX13 is of historic interest. However, interpretation methods have been strengthened or changed substantially, and there are newer multiplexes since 2013. Moreover, the outcomes of Cases 3 and 5 should be considered within the context on how they were constructed and because of this, the outcomes do not provide an accurate assessment of analyst performance but do provide an interesting example of the limitations of DNA profile interpretation.

Case 1 can be interpreted without difficulty by any of the four PG systems examined and could also have been analyzed successfully with the CPI. Cases 2 and 3 can be interpreted without difficulty by any of the four PG systems examined but the use of the CPI would be problematic due to the potential of allele dropout. Case 3 demonstrated the need to consider relevant background information, *I*, after initial assessment of the profile. By considering *I* it was possible to assume the presence of the victim and the consensual partner. This additional information allowed for an assignment of the number of contributors as three; without this information the profile would have been assigned (by some analysts) as coming from two donors. This case does not present some barrier to interpretation caused by relatedness of donors beyond the increased allelic overlap that can

occur. All four PG software properly included the true contributor and excluded the non-contributor with relatively similar *LR*s.

Case 4 can be interpreted without difficulty by any of the four PG systems examined and the CPI could be applied correctly, assuming two donors. Twelve loci would be suitable for the CPI statistic.

Case 5 suffers from over engineering. It is unclear whether reference 5C, a non-donor, can be (or even should be) excluded by manual methods. Inclusion of reference 5C should be termed an adventitious match not a false inclusion. Beyond assessing this statement of inclusion the case is not really suitable for this type of interlaboratory study as it instead explores the limits of DNA analysis not interpretation methods.

PG software systems were able to address all scenarios (excluding the reference 5C which was intended to obfuscate) and presented *LR*s that are consistent with the quality of the profiles in the cases. This re-evaluation supports the utility of PG over that of the CPI and allows the community to assess the NIST MIX13 study in relation to the state-of-the-art of mixture analysis. This supports the ongoing transition from CPI to PG.

Acknowledgements

This work was supported in part by grant NIJ 2017-DN-BX-0136 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of their organizations.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.fsigen.2018.08.014.

References

[1] Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genetics. 2016;17:125.

[2] Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerreri RA, et al. Mixture Interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. Journal of Forensic Sciences. 2009;54:810-21.

[3] Budowle B, Bieber FR. Final Report on Review of Mixture Interpretation in Selected Casework of the DNA Section of the Forensic Science Laboratory Division, Department of Forensic Sciences, District of Columbia. http://dfs.dc.gov/page/usao-report-april-20152015.

[4] Coble MD. NIST inter-laboratory studies for DNA mixture interpretation. American Academy of the Forensic Sciences. Seattle, WA2014.

[5] Coble MD. MIX13: An interlaboratory study on the present state of DNA mixture interpretation in the U.S. 5th Annual Prescription for Criminal Justice Forensics. Fordham University School of Law2014.

[6] Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International. 2006;160:90-101.

[7] Gill P, Whitaker JP, Flaxman C, Brown N, Buckleton JS. An investigation of the rigor of interpretation rules for STR's derived from less than 100 pg of DNA. Forensic Science International. 2000;112:17-40.

[8] Bright J-A, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix™ – A multi laboratory response to PCAST. Forensic Science International: Genetics. 2018;34:11-24.

[9] Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. Internal validation of STRmix; for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2017;29:126-44.

[10] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2013;7:516-28.

[11] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL. Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences. 2011;56.

[12] Bleka Ø, Storvik G, Gill P. EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forensic Science International: Genetics. 2016;21:35-44.

[13] Augenstein S. Five Years On, Will MIX13 DNA Study Be 'Bombshell' Paper? 2018. https://www.forensicmag.com/news/2018/04/five-years-will-mix13-dna-study-be-bombshell-paper Accessed 21 August 2018.

[14] Butler JM, Kline MC, Coble MD. NIST Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned. Forensic Science International: Genetics. 2018;https://doi.org/10.1016/j.fsigen.2018.07.024.

[15] Augenstein S. Five Years On, Will MIX13 DNA Study Be 'Bombshell' Paper? Forensic Magazine2018.

[16] Texas Forensic Science Commission. Commission Releases Austin Police Department DNA Section Audit Report. 2016. http://www.fsc.texas.gov/blog/2016-07-12/commission-releases-austin-police-department-dna-section-audit-report Accessed 26 August 2016.

[17] Lab Retriever. Version 2.2.1 released Nov 21, 2014. https://scieg.org/wp-content/uploads/2017/07/Lab_retriever_2.2.1.rtf Accessed 31 January 2018.

[18] Bleka Ø, Eduardoff M, Santos C, Phillips C, Parson W, Gill P. Open source software EuroForMix can be used to analyse complex SNP mixtures. Forensic Science International: Genetics. 2017;31:105-10.

[19] Bleka Ø, Benschop CCG, Storvik G, Gill P. A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. Forensic Science International: Genetics. 2016;25:85-96.

[20] Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, et al. Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. BMC Bioinformatics. 2015;16:298.

[21] Lohmueller K, Rudin N. Calculating the weight of evidence in low-template forensic DNA casework. Journal of Forensic Sciences. 2013;58:234-59.

[22] Prieto L, Haned H, Mosquera A, Crespillo M, Alemañ M, Aler M, et al. Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. Forensic Science International: Genetics. 2014;9:47-54.

[23] Gill P, Haned H. A new methodological framework to interpret complex DNA profiles using likelihood ratios. Forensic Science International: Genetics. 2013;7:251-63.

[24] Swaminathan H, Grgicak CM, Medard M, Lun DS. NOCIt: A computational method to infer the number of contributors to DNA samples analyzed by STR genotyping. Forensic Science International: Genetics. 2015;16:172-80.

[25] Marciano MA, Adelman JD. PACE: Probabilistic Assessment for Contributor Estimation— A machine learning-based assessment of the number of contributors in DNA mixtures. Forensic Science International: Genetics. 2017;27:82-91.

[26] Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, et al. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International: Genetics. 2016;25:175-81.

[27] 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.

[28] Allard JE. The collection of data from findings in cases of sexual assault and the significance of spermatozoa on vaginal, anal and oral swabs. Science and Justice. 1997;37:99-108.

[29] Jeanguenat AM, Budowle B, Dror IE. Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. Science and Justice. 2017;57:415-20.

[30] Butler JM. Advanced topics in forensic DNA typing: Interpretation: Elsevier; 2014.

[31] Taylor D, Bright J-A, Buckleton J. Considering relatives when assessing the evidential strength of mixed DNA profiles. Forensic Science International: Genetics. 2014;13:259-63.

[32] 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.

[33] Coble MD, Bright J-A, Buckleton JS, Curran JM. Uncertainty in the number of contributors in the proposed new CODIS set. Forensic Science International: Genetics. 2015;19:207-11.

[34] Tvedebrink T. On the exact distribution of the numbers of alleles in DNA mixtures. Forensic Science International: Genetics Supplement Series. 2013;4:e278-e9.

[35] Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2010. http://www.forensicdna.com/assets/swgdam_2010.pdf Accessed 1 June 2018.

[36] Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2017.

https://media.wix.com/ugd/4344b0_2a08f65be531488caa8037ed55baf23d.pdf Accessed 17 March 2017.

[37] 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.