Fall 12-2019

DNA Mixture statistics using a likelihood ratio software tool: effect of variations in drop-out rates and number of contributors

Felicia M. Lucero
CUNY John Jay College, felicia.lucero@jjay.cuny.edu

DNA Mixture statistics using a likelihood ratio software tool: effect of variations in drop-out rates and number of contributors

A Thesis Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Forensic Science

John Jay College of Criminal Justice

City University of New York

Felicia Lucero

December 2019

DNA Mixture statistics using a likelihood ratio software tool: effect of variations in drop-out rates and number of contributors

Felicia Lucero

This Thesis has been presented to and accepted by the Office of Graduate Studies, John Jay College of Criminal Justice in Partial Fulfillment of the Requirements for the Degree of Master of Science in Forensic Science.

Thesis Committee

Thesis Advisor: Mechthild Prinz, PhD

Second Reader: Nicholas D.K. Petraco, PhD

External Reader: Adele Mitchell, PhD

Table of Contents **Pages**

Acknowledgements

Thank you to my thesis advisor, Mechthild Prinz, PhD, for her patience and encouragement throughout this journey. I would also like to thank my committee members: Nicholas D.K. Petraco, PhD, and Adele Mitchell, PhD, for their advice, comments, and contributions to this project. Thank you to my friends and family for their continued support. Finally, I would like to thank my mother, Belinda Jimenez, for her love, motivation, and words of encouragement.

Abstract

Complex DNA mixtures can be very probative evidence, but comparisons to a person of interest can be affected by allelic drop-out and uncertainty regarding the number of individuals having contributed DNA to a sample. Scientific organizations such as the International Society of Forensic Genetics (Gill et al., 2006) recommend that likelihood ratios should be used to provide a statistical weight when a positive association is made between the DNA profile of a person of interest and an evidentiary DNA sample. To this effect the New York City Office of Chief Medical Examiner (OCME) developed a software program, Forensic Statistical Tool (FST), which calculates likelihood ratios for different scenarios taking into account empirically developed drop-out and drop in rates for different types of mixtures. The FST software was used to explore the effect of underestimation of a contributor's true drop-out rate and effect of the incorrect estimation of the number of contributors on LR calculations. It was found that underestimating the allelic dropout rate for a true contributor almost always led to an either equal or lower LR than when the original dropout rate was used. It was also found that when the number of contributors was misspecified, there was an increase or decrease in LR values for true contributors. Variation of resulting LRs was higher for more complex mixtures. Finally, LRs for comparisons to individuals, whose DNA was known to not be present in the test mixtures, were lower when using the lower drop-out rates than when using the true drop-out rates.

Introduction and Literature Review

Forensic DNA typing is a method used for isolating and characterizing variable regions of deoxyribonucleic acid (DNA) with the goal of identifying an individual, establishing familial relationships between individuals, or attributing biological evidence to a source. Approximately, 99.7% of our DNA is identical between individuals, the remaining 0.3% varies greatly between individuals which makes each individual unique and makes identification possible. The variation at 13 or more loci, or locations, within the 0.3% are generally analyzed in forensic laboratories and used for identification purposes (Butler, 2010).

There are two types of variation that are traditionally analyzed in order to establish uniqueness: sequence polymorphisms and length polymorphisms. Sequence polymorphism is the variation in the sequence of DNA at a particular locus. Length polymorphism is the variation in the length of a specific repeating sequence at a particular location. Forensic laboratories use length polymorphisms known as short tandem repeats (STR) when attempting to individualize a DNA sample. A short tandem repeat is a short tandemly repeating segment of DNA that occurs at a specific locus. During testing, the number of times that a STR is repeated is measured at each locus and assigned to that particular fragment of DNA. An alternate form of DNA present at each locus is called an allele and each allele is represented by the number of STR repeats. Since each individual inherits one set of chromosomes from each parent, they will either have two STR lengths present at each locus, if they are heterozygous, or show a single allele if they are homozygous (Butler, 2010).

In forensic DNA evidence analysis, the goal is to identify the individual(s) who contributed their DNA to a particular item of evidence. Once an evidence sample has been interpreted and genotypes have been assigned to the contributor(s) of the DNA sample, the DNA profile of a person of interest (POI) can be compared to the evidence sample. If the DNA alleles that the POI carries are also found in the evidence sample, this is considered a positive association. For example, consider a blood stain found on a knife at the scene of a homicide. Once DNA testing is completed, it may be found to be a single source sample, meaning only one person contributed DNA to that sample. A comparison can then be made between the single source blood stain found on the knife and the DNA profile of the victim. If the DNA profile generated from the single source blood stain is the same as the DNA profile of the victim, this is considered a "match". If the DNA profile of the victim matches that of the single source blood stain, the question is if this positive association is coincidental. As per the Scientific Working Group on DNA Analysis Methods (SWGDAM), organized by the FBI, any DNA analysis with a positive association between an evidence sample and an individual should be supplemented with a statistical weight calculation (Scientific Working Group on DNA Analysis Methods, 2017).

The type of statistic used to evaluate the strength of DNA comparisons is dependent on the type of DNA results generated from the item of evidence. For single source samples or when an individual's profile can be deduced from a DNA mixture, the Random Match Probability (RMP) approach can be used. RMP is the probability of seeing the same profile as that generated from the evidence sample and the POI in a randomly selected, unrelated individual. This statistic is calculated with the use of allele

frequencies that are taken from representative ethnic populations (Butler, 2015). Based on genetic inheritance rules, these individual allele frequencies can then be used to calculate a genotype frequency at each locus, and then, because the STR loci are independent of one another, the frequency of an entire DNA profile within a population can easily be estimated (National Research Council (US) Committee on DNA Forensic Science, 1996). If an individual's alleles are rare within its population, the probability that their alleles would match the crime scene sample by chance would be lower than if the individual carries common alleles. A lower RMP suggests that it will be rare to see that particular DNA profile in a randomly selected, unrelated individual.

Many biological evidence items collected at crime scenes are not from a single source. Sexual assault evidence routinely generates a mixture of the victim's and the perpetrator's DNA. The majority of casework in forensic DNA laboratories now consists of touched objects, which often show more than one DNA contributor (Mapes, Kloosterman, van Marion, & de Poot, 2016). For example, if a doorknob at a bank is tested, it will likely produce a mixture because many people touch the doorknob and thus leave some of their skin cells and DNA on the doorknob when they enter the bank. It tends to be more difficult to determine the genotypes of the individual contributors to a mixture than it is with single source samples. Some mixtures show a clear signal intensity or peak height difference in the detected alleles of a major and a minor contributor. In this case the expected genotypes of the individual contributors can be deconvoluted (or "deduced") (Clayton, Whitaker, Sparkes, & Gill, 1998). This approach is valid because the amount of amplification product generated from PCR is generally proportional to the relative amount of DNA template from each contributor (Perlin & Szabady, 2001).

Mixtures that contain approximately the same amount of DNA from each contributor, or mixtures with more than two contributors, do not allow for making decisions on the underlying genotypes and cannot generally be deconvoluted. These mixtures can still be compared to known references like a victim or a POI but it is not possible to make a direct match and apply the RMP statistic (Bille, Bright, & Buckleton, 2013). Depending on the presence or absence of the alleles of a person in an evidence sample, a comparison can result in an exclusion or an inclusion (positive association). Again, the question is if this association is fortuitous and as per SWGDAM, a statistic should be calculated (Scientific Working Group on DNA Analysis Methods, 2017).

Several issues complicate the interpretation of STR profiles and resulting statistical weight assessment in a DNA mixture. One important step in mixture interpretation is the estimation of the number of contributors. Since the true number of contributors can never be known, different methods have been suggested for estimating the number of contributors in a mixture: maximum allele count and maximum likelihood estimator. Maximum likelihood estimator uses allele frequencies that are present at each locus in a sample and searches for the number of contributors that will give the maximum likelihood of the observed data (Haned, Pène, Sauvage, & Pontier, 2011; Haned, Pène, Lobry, Dufour, & Pontier, 2011). Maximum allele count is the process by which the number of contributors is determined by the locus with the maximum number of alleles present. Perez, Mitchell, Ducasse, Tamariz, & Caragine (2011) conducted a study to examine the characteristics of two-, three-, and four-person mixtures beyond the maximum allele count. They looked at the total number of different alleles present in three types of purposeful mixtures and determined that there is overlap between the

distributions of the number of labeled alleles for the different types of mixtures. In general, two-person mixtures were best described by having a total allele count of 49 or less, three-person mixtures had 52-59, and four-person mixtures had 64 or more over the two or three replicated amplifications performed by OCME for high or low template samples, respectively. While the allele count averages were determined from purposeful mixtures developed for that study, similar results can be expected from crime scene samples. The variation in the number of alleles could be due to drop-out of the low-level contributors' alleles. The same study also looked at how the total number of alleles was affected by varying mixture ratios of the different contributors. The authors found that the number of alleles was relatively consistent in mixtures with similar mixture ratios. Finally, they analyzed the different mixtures that were generated from touched items and interpreted them using the guidelines that had been previously set. The guidelines proved to be more applicable to the purposeful mixtures created rather than those generated from the touched items. One reason for this could be uneven distributions due to different shedder status of the individual contributors or loss of material during the DNA recovery from the item (Perez, et al., 2011).

It is important to look at a mixture as a whole and consider as much information as possible, e.g. signal strength, stutter artifacts, peak height ratios and degradation effects (Butler, 2015). For example, lower amounts of DNA can lead to peak height imbalance between a pair of heterozygous alleles, which can make it difficult to deduce a contributor's genotype. Low amounts of DNA and DNA degradation can also cause allelic drop-out, where an allele is not detected at all, making it difficult to determine whether the alleles present are from a homozygous or heterozygous contributor (Balding,

2005). Another artifact, allelic drop-in occurs due to the amplification of DNA that does not originate from the assumed contributors (Butler, 2015; Gill et al., 2012). Stutter artifacts are a byproduct of STR amplification and usually appear next to a more intense allele peak. In complicated mixtures stutter peaks can mask a minor contributor and cannot be distinguished from true allele peaks. In low level samples, stochastic effects can also introduce elevated stutter, which are stutter peaks that are higher than expected based on the intensity of the allele peak. Not only can this lead to an increase in the estimation of the number of contributors, but it can also lead to an incorrect genotype assignment at that particular locus (Butler, 2015).

All of these artifacts can have an effect on mixture interpretation. The number of contributors can be underestimated if allele sharing occurs between the contributors of a mixture, i.e. a father and son's alleles are both present in an evidence sample. Allelic drop-out can occur and lead to an underestimate of the number of contributors, especially in samples with low signal strength indicating low amounts of DNA (Haned et al., 2011). Drop-in and stutter peaks on the other hand can lead an analyst to overestimate the number of contributors present in a sample. The estimation of the number of contributors is important when interpreting DNA mixtures, and several authors, e.g. Perez et al. (2011) developed guidelines to try and differentiate between two, three, and four-person mixtures; however, this task can be especially challenging with LT-DNA samples which have a greater risk of allelic drop-out than HT-DNA samples. One way to overcome the issue of drop-out is by increasing the number of amplification cycles, which can decrease the possibility of drop-out for low template samples as the probability of allelic drop-out is dependent on the number of amplification cycles (Mitchell et al., 2011). For this

reason, the New York City Office of Chief Medical Examiner (OCME) had adopted a dedicated low template (LT) DNA testing strategy. Samples with less than 100pg of template DNA per amplification were amplified in triplicate for 31 cycles. Samples that contained at least 100pg of template DNA per amplification were considered high template (HT) DNA and are either amplified once or in duplicate for 28 cycles (New York City Office of Chief Medical Examiner, 2012).

In order to comply with SWGDAM requirements, positive associations to a mixture must be reported with a statistical weight and several approaches have been developed, each taking into account a varying amount of information. Statistical approaches either follow a binary model, a semi-continuous model, or a continuous model (Scientific Working Group on DNA Analysis Methods, 2017). Random man not excluded (RMNE), also known as combined probability of inclusion/exclusion (CPI), is an example of a statistical calculation that uses the binary model. The binary model only considers the presence or absence of alleles and cannot account for allelic drop-out. RMNE is the probability that all of the alleles in the profile of a randomly chosen person would appear in the mixture by chance. RMNE only applies when the analyst can be sure that all of the mixture alleles have been detected and that allelic drop-out did not occur. In this situation, RMNE is generally a conservative approach. However, RMNE does not take into account number of contributors to a sample, peak heights, assumed contributors, and cannot accommodate the possibility of allelic drop-out or drop-in (Bille, Weitz, Coble, Buckleton, & Bright, 2014). These types of limitations can make it unreliable for use with LT-DNA samples. Variations of the binary model have been developed, which Kelly, Bright, Buckleton, & Curran, (2014) have called “the semi-binary” model. This

model can be used if drop-out has occurred by omitting the locus in the calculation. Another limitation to the binary model is that replicate amplifications cannot be taken into account. This can become problematic when dealing with LT-DNA samples because for this sample type replicate amplifications are advantageous to obtain the most amount of data that can support more accurate interpretations (Kelly, et al., 2014; Bille et al., 2014). These limitations caused forensic geneticists to move towards the semi-continuous and continuous models.

Likelihood ratio methods can incorporate additional information such as known contributors or a drop-out rate and are used for both semi-continuous and continuous mixture evaluation. A likelihood ratio is a statistical calculation used to provide support for one scenario over another. In forensic DNA analysis, a likelihood ratio is the comparison of two competing scenarios: the probability of generating the DNA mixture if the POI is a contributor to the evidence sample versus the probability of generating the mixture if an unknown, unrelated individual is a contributor to the evidence sample instead. Allele frequencies are used to calculate LRs, similar to RMP calculations performed for the comparison of a POI to a single source sample (Buckleton, 2005). The terms prosecution hypothesis and defense hypothesis were created to refer to the two competing scenarios. A prosecutor would typically argue that the POI did contribute their DNA to a particular sample; whereas, the defense would argue that the DNA evidence originated from an unknown individual. Although, LRs are not calculated with the goal of supporting either hypothesis, the terms are still used to distinguish between the two. In the numerator is the probability of the STR data conditional on a prosecution hypothesis (H_p). This is the probability of generating the evidence mixture if the POI is a contributor

to the evidence sample. In the denominator is the probability of the STR data conditional on a defense hypothesis (H_d). This is the probability of generating the same evidence mixture if an unknown, unrelated person contributed to the evidence sample, rather than the POI. A likelihood ratio greater than one favors the prosecution hypothesis, suggesting that the mixture is better explained if the POI is a contributor to the crime scene DNA sample, rather than an unknown, unrelated person; the higher the LR, the greater the support for the prosecution hypothesis (Buckleton, 2005). A likelihood ratio less than one favors the defense hypothesis, suggesting that an unknown, unrelated individual contributed to the sample, rather than the POI; the lower the LR, the greater the support for the defense hypothesis (Buckleton, 2005).

One thing to keep in mind is that the value of the LR is dependent on the amount of data that is generated from the evidence sample and how much of that data is available to estimate the model parameters (Brümmer, 2013). While the standard likelihood ratio does not account for drop-in and drop-out, these occurrences can be incorporated into the LR calculation. Semi-continuous models can also employ the use of additional information that is provided from replicate amplifications (Kelly et al., 2014). LRs require the specification of the number of contributors in order to perform the calculations. LR calculations can take real-world phenomena into account, making it more flexible and realistic than other methods (Gill et al., 2006). Examples of semi-continuous mixture software tools are Lab Retriever (Inman et al., 2015) and the LRmix module of Forensim (Haned, Benschop, Gill, & Sijen, 2015). The Forensic Statistical Tool (FST) is a semi-continuous model developed by the OCME which incorporates replicate amplifications and drop-out/drop-in rates into the LR calculations (Mitchell et

al., 2012). As explained below, FST uses empirically determined quantitation based drop-out values, and the number of amplification cycles to calculate LRs. Other programs take drop-out and drop-in into account, but do not incorporate empirically determined drop-out and drop-out rates. As per Mitchell et al., (2012) "LoComation and Forensim require the user to specify drop-out and drop-in probabilities. Forensim then calculates the LR for a range of drop-out rates and displays the results graphically". Also, FST only identifies the presence of allelic drop-in, as other programs model drop-in as a function of the allelic frequencies. FST defines drop-in as "stutter as well as extraneous peaks that are not in stutter position", where other programs exclude stutter from their drop-in definitions and model stutter separately (Mitchell et al., 2012).

Although, FST takes drop-in and drop-out into account when calculating LRs, FST does not take peak heights into consideration and thus, is not using all available information. Peak heights are necessary in order to determine the approximate amount of DNA that each individual contributed to a sample and can be used to perform mixture deconvolutions. Since FST does not take peak heights into account, it does not deconvolute mixtures that are generated from evidence samples. Prior to the use of FST, the analyst determines whether or not a mixed sample can be deconvoluted and that information is entered into FST in order to calculate a LR (Mitchell et al., 2012).

Mixture deconvolution is important, since deduced contributor genotypes are eligible to be entered in the FBI DNA database system, but it is a challenging and time-consuming task (Butler, 2015). For this reason, the Scientific Working Group on DNA Analysis Methods (2017) has approved the use of continuous model probabilistic genotyping software, which incorporates peak height information to help analysts

determine the DNA profiles of the individuals present in a sample, and can calculate a statistical weight when comparing a POI's profile to a DNA sample. These continuous probabilistic genotyping programs, for example TrueAllele® and STRmix™, take into account several types of biological events in order to complete this task. Not only do they consider drop-in and drop-out, but they can also consider peak heights, amplification efficiencies, degradation, contributor ratios, stutter, among other events, in order to deconvolute DNA samples and calculate a statistical weight (Bright et al., 2016; Perlin et al., 2011). In 2009, Perlin & Sinelnikov studied the efficiency of manual deconvolutions compared to computer-based deconvolutions. Using the Cybergenetics TrueAllele® Casework program they determined that computer-based deconvolutions were more efficient, especially when dealing with samples with a low template amount (Perlin & Sinelnikov, 2009).

Similarly, the OCME switched from the use of the semi-continuous FST to a fully continuous probabilistic genotyping software called STRmix™ in January 2017. The new system was validated together with a new STR multiplex kit, Promega's PowerPlex® Fusion 5C, that tests a total of 24 loci, including the 20 CODIS expanded core loci. Prior to 2017, the OCME was utilizing the Applied Biosystems Identifiler® kit which tested a total of 16 loci. The Federal Bureau of Investigation, who manages the Combined Index DNA System, CODIS, had made the decision to increase the number of CODIS core loci from 13 loci to 20 loci and mandated the increase in tested loci for all CODIS laboratories (Hares, 2015). Testing more STR loci provides even higher levels of discrimination and continuous models have the advantage of utilizing qualitative and quantitative information in order to provide more informative conclusions.

Prior to the availability of STRmix™, the OCME had realized the need to incorporate allelic drop-out and drop-in into the LR analysis of DNA mixtures and developed and validated a semi-continuous computer program called the Forensic Statistical Tool (FST). The system was validated for use with the Applied Biosystems Identifiler® amplification kit on single source and mixtures from two, three, or four contributors tested under HT-DNA and LT-DNA conditions. Although the OCME protocols were used for the purpose of this study, the OCME does not interpret nor does it use FST for comparison to four-person mixtures. To allow for drop-out and drop-in within a LR framework, it is necessary to estimate the probability of each of these phenomena. FST uses empirical estimates of drop-out and drop-in. Drop-out and drop-in rates were determined separately for HT-DNA and LT-DNA amplification conditions. Drop-out rates were determined for each locus, heterozygous and homozygous loci, and DNA template quantity. Estimations of probability of partial and complete heterozygous drop-out, and complete homozygous drop-out were determined separately. Drop-out rates for deducible and non-deducible mixtures were also estimated separately. These estimates were incorporated into the appropriate number of contributors LR structure (Mitchell et al., 2012). The user selects a scenario with an appropriate defense and prosecution's hypotheses, specifies the quantity of template DNA that was amplified for the sample, and specifies whether or not the sample was deducible. FST will interpolate the data input by the user and will determine the appropriate drop-out rate to use for evidence samples that are amplified with quantities that qualify for drop-out rate estimations.

FST uses the estimated drop-out rate for each of the following: number of contributors, template DNA quantity, each locus, and ratio of mixed samples minus one standard deviation. The standard deviation was subtracted to lower the applied drop-out rate and generate a more conservative estimate of the LR, meaning a lower LR for non-contributors which would typically favor the defense hypothesis. The goal of this thesis research was to demonstrate the validity of this approach. Specifically, 1) does underestimating the true drop-out rate of a true contributor always lead to a lower LR than if the actual drop-out rate is used? Also, 2) does using an artificially low drop-out rate reduce the chance of obtaining a false inclusion when the POI is not a contributor to the mixture?

FST uses the information that the user provides in order to generate a LR; however, one critical step is the formulation of the two hypotheses (scenarios) to be compared to each other. One component here is the number of individuals assumed to have contributed DNA to the mixture. As explained above, with loci being homozygous and even unrelated individuals having alleles in common, this is a parameter that cannot truly be known and best estimates must be used. It is important to understand the impact of either under or overestimating the number of contributors. Therefore, this study includes an evaluation of incorrect estimates of the number of contributors to determine whether conservative estimates would be obtained for true contributors if the specified number of contributors was different from the actual number of contributors.

For the drop-out rate study, LRs for 19 two-person purposeful mixtures, 24 three-person purposeful mixtures, and 20 four-person purposeful mixtures were calculated using the empirically derived drop-out rate and again using half of the empirically

derived drop-out rate. LRs were compared to determine the effect of an underestimation of the drop-out rates on the LRs obtained for true contributors, as well as for non-contributors of the samples.

In addition, 15 two-person, 15 three-person, 15 four-person mixtures generated from touched items were analyzed using FST in order to evaluate the impact that misspecification of number of contributors has on the LRs. Each sample was compared to their true contributors one at a time using FST under the two-person, three-person, and four-person scenarios. No assumed known profiles were used for any of the scenarios.

Materials and Methods

The analyses presented here relied on samples that were generated and processed as part of OCME's validation of FST. For this thesis, DNA profiles for individual contributors and STR typing results from mock casework samples and purposeful mixtures had already been generated. This work begins with the labeled alleles in the mixtures and the individual contributors' profiles.

The validation included purposeful mixtures of known quantities of DNA from two, three, or four contributors as well as mock casework samples handled by two, three, or four-persons. Some of the items used for the mock casework samples were cleaned before being handling and some were not. The purposeful mixtures were prepared with varying proportions and amounts of DNA of known concentrations from known contributors. Each of the mock casework samples were handled by two, three, or four known individuals. The samples were processed, and the resulting mixtures were generated in accordance with the OCME protocols (Mitchell et al., 2012). In processing

the samples, information about the number and identity of the contributors was masked; the analysts treated the samples as they would treat crime scene samples.

The FST program was used to calculate likelihood ratios for each of the mock casework samples. Each sample was tested using different scenarios, depending on the apparent number of contributors and deducibility. The LRs for each of the mixtures were calculated using the following scenarios:

Apparent two-person LR calculation:

$$LR = \frac{comparison + unknown}{two\ unknowns}$$

Apparent three-person LR calculation:

$$LR = \frac{comparison + two\ unknowns}{three\ unknowns}$$

Apparent four-person LR calculation:

$$LR = \frac{comparison + three\ unknowns}{four\ unknowns}$$

Nineteen two-person purposeful mixtures, twenty-four three-person purposeful mixtures, and twenty four-person purposeful mixtures were evaluated to determine the effects of the underestimation of drop-out rates. First, each mixture was evaluated to determine whether or not the profile of the major contributor could be deduced. The DNA profile of a true contributor was then visually compared to the samples to determine if they could be included, excluded, or if no conclusions could be drawn. This was achieved using the OCME's 2012 version of the standard operating procedure for STR analysis (New York City Office of Chief Medical Examiner, 2012). The true contributor profiles were used as the comparison sample in each scenario, as a POI profile would be used in casework. FST tests were run on each sample twice, once using the empirically

derived drop-out rate of the true contributor being compared and once using half of the empirically derived drop-out rate of that contributor. For example, if a true contributor was heterozygous at five loci and two loci had an allele drop out, the true rate of single allele drop out at heterozygous loci for this contributor would be 0.40 and the half drop-out rate of single allele drop out at heterozygous loci for this contributor would be 0.20. Similar calculations were made for two-allele drop out at heterozygous loci and drop out at homozygous loci. In this analysis, the true contributor that was used to compute the drop-out rate was treated as the POI. Thus, this was analogous to running FST in a case where the POI actually did contribute to the evidence sample. FST was also run treating each of the profiles in a database of ten thousand simulated non-contributor profiles as the POI. The purposeful mixtures and mock casework samples were compared to the database twice, again using the drop-out rates that had been previously used for the true contributor testing of each sample. The samples were compared to the database to determine the distribution of LRs if a POI is not a contributor to the evidence sample. This was analogous to running FST in a case where the POI did not contribute to the evidence sample. The purpose was to compare LRs for non-contributors if the drop-out rate was specified correctly for one of the true contributors or if half of the actual drop-out rate was used.

Fifteen true two-person mixtures, fifteen true three-person mixtures, and fifteen true four-person mixtures were also evaluated to determine the effects of the misspecification of the number of contributors on the LR. First, each mixture was evaluated to determine whether or not the profile of the major contributor could be deduced. The DNA profile of a true contributor was then visually compared to the

samples to determine if they could be included, excluded, or if no conclusions could be drawn. This was achieved using the OCME's protocols. FST was then used to calculate a LR for each of the true contributors' profiles against its respective mixture using each of the three different scenarios shown above. The purpose was to determine the distribution of LRs if a POI is a contributor to the evidence sample and the number of contributors was either correctly determined or was over or underestimated.

Results

Effect of different drop-out rates when testing of true contributors as POI

The mixtures tested to evaluate the effect of two different drop-out rates are shown in Table 1. Comparisons of log LRs obtained with empirically derived drop-out rates (x-axis) and underestimated drop-out rates (y-axis) are shown below for high template and low template two, three and four-person samples (Figures 1A-3B). Each point represents one mixture, analyzed twice with FST. The identity line on each plot indicates where points fall when the LR is identical using the empirically derived drop-out rate and the underestimated drop-out rate. A point below the identity line indicates that the LR decreased when half of the empirically derived drop-out rate was used compared to the empirically derived drop-out rate. A point above the identity line indicates that the LR increased when half of the empirically derived drop-out rate was used comapared to the empirically derived drop-out rate.

Table 1: Mixture samples analyzed with two different drop-out values

Type of Mixture	**Number of Samples**
Two-person, 28 cycles	N=8
Two-person, 31 cycles	N=10
Three-person, 28 cycles	N=12
Three-person, 31 cycles	N=11
Four-person, 28 cycles	N=9
Four-person, 31 cycles	N=11

Two-person

The logarithm of each LR for two-person mixtures was taken and the results are depicted in Figures 1A and 1B with a summary of the findings in table 2.

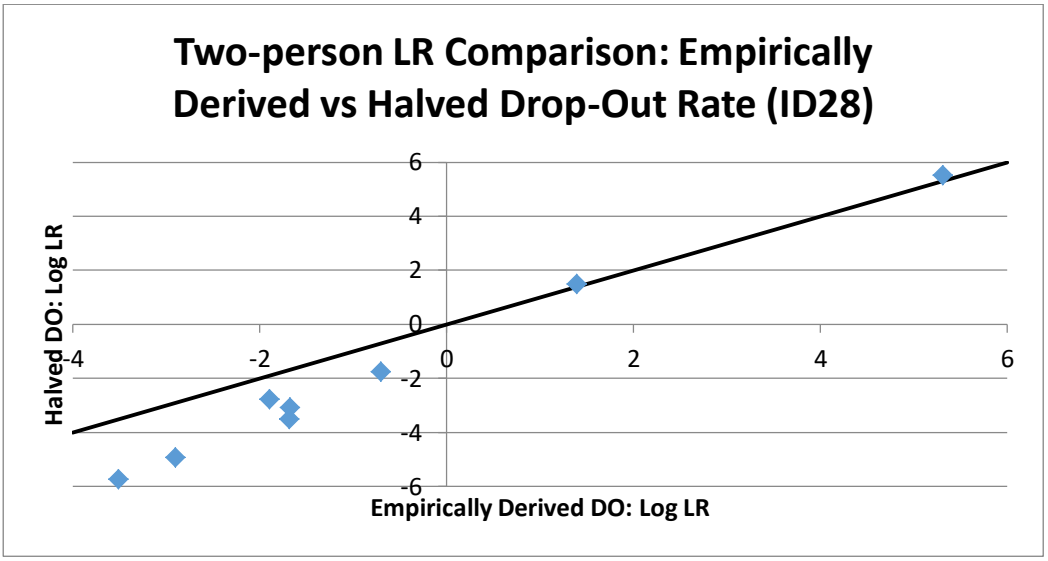

Fig. 1A Log LR plot for two-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 210pg/ul to 500pg/ul.

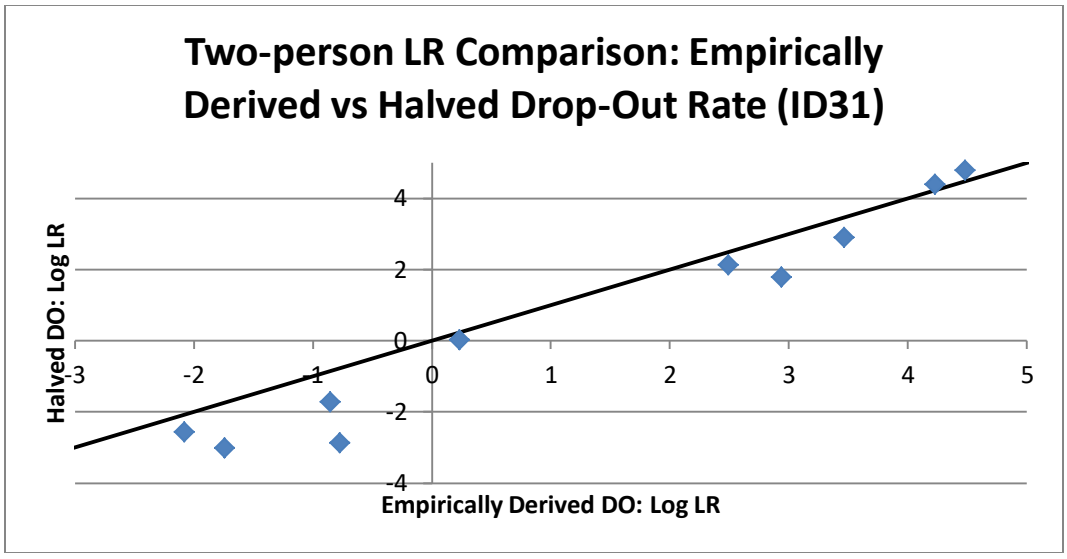

Fig. 1B Log LR plot for two-person deducible and non-deducible mixtures using the empirially derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation values ranged from 12pg/ul to 91pg/ul.

Table 2: Effect of using a lower than the empirically estimated drop-out rate for a true contributor to two-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Two-person, 28 cycles	0	2	7
Two-person, 31 cycles	0	5	5

As depicted in Figure 1A, seven of the nine two-person samples that were amplified using 28 cycles resulted in lower LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Two samples resulted in LRs that were slightly increased using the halved drop-out rates, relative to the empirically derived drop-out rates, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. As shown in Figure 1B, five of the ten samples that were amplified using 31 cycles resulted in lower

LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Five samples resulted in LRs that were slightly increased, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. Overall, using drop-out rates that were half of the empirically derived values resulted in LRs that were lower or approximately the same as those calculated using the empirically derived rates (Table 2).

Three-person

The logarithm of each LR for three-person mixtures was taken and the results are depicted in Figures 2A and 2B with a summary of the results in Table 3.

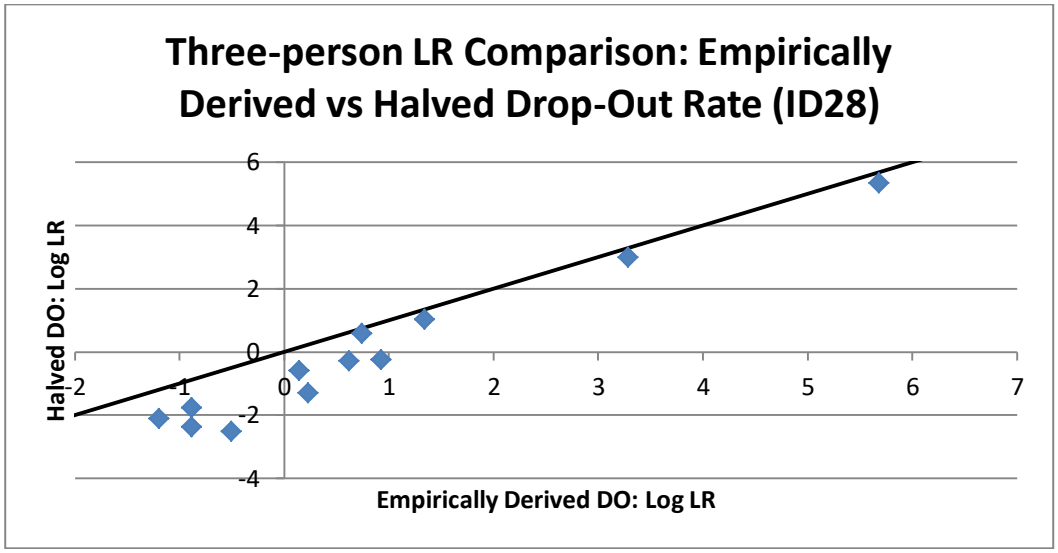

Fig. 2A Log LR plot for three-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 130pg/ul to 575pg/ul.

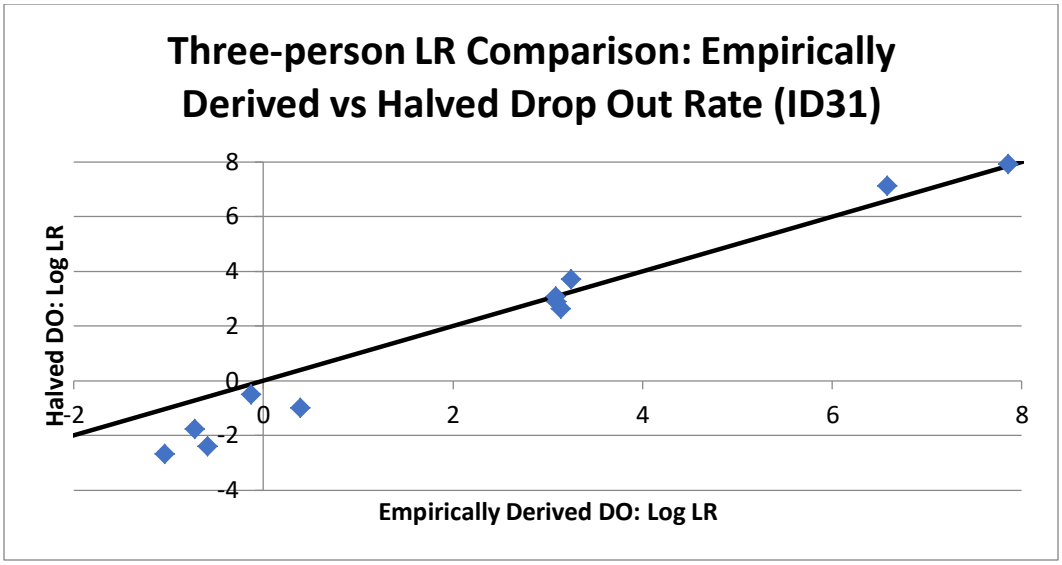

Fig. 2B Log LR plot for three-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation values ranged from 25pg/ul to 100pg/ul.

Table 3: Effect of using a lower than the empirically derived drop-out rate for a true contributor to three-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Three-person, 28 cycles	0	3	9
Three-person, 31 cycles	1	4	6

As depicted in Figure 2A, all of the three-person samples that were amplified using 28 cycles resulted in lower LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Nine of the twelve samples resulted in a decrease by at least one order of magniude, while three of the samples resulted in LRs that were within the same order of magnitude, yet slightly lower than the LRs calculated using the empirically derived drop-out rate. None of the

samples resulted in an increased LR when using the halved drop-out rate relative to the empirically derived drop-out rate. Four of the twelve samples that resulted in a decreased LR caused a shift from generating a LR greater than one to a LR less than one. As shown in Figure 2B, six of the eleven samples that were amplified using 31 cycles resulted in lower LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Two samples resulted in LRs that were slightly decreased, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. Two samples resulted in LRs that were slightly increased, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. One sample resulted in a LR that was one order of magnitude greater when calculated using the halved drop-out rate compared to using the empirically derived drop-out rate (Table 3). One of the eleven samples that resulted in a decreased LR caused a shift from generating a LR greater than one to a LR less than one. Overall, using drop-out rates that were half of the empirically derived rates resulted in LRs that were lower or approximately the same as those calculated using the empirically derived rates.

Four-person

The logarithm of each LR for four-person mixtures was taken and the results are depicted in Figures 3A and 3B with a summary of the results in Table 4.

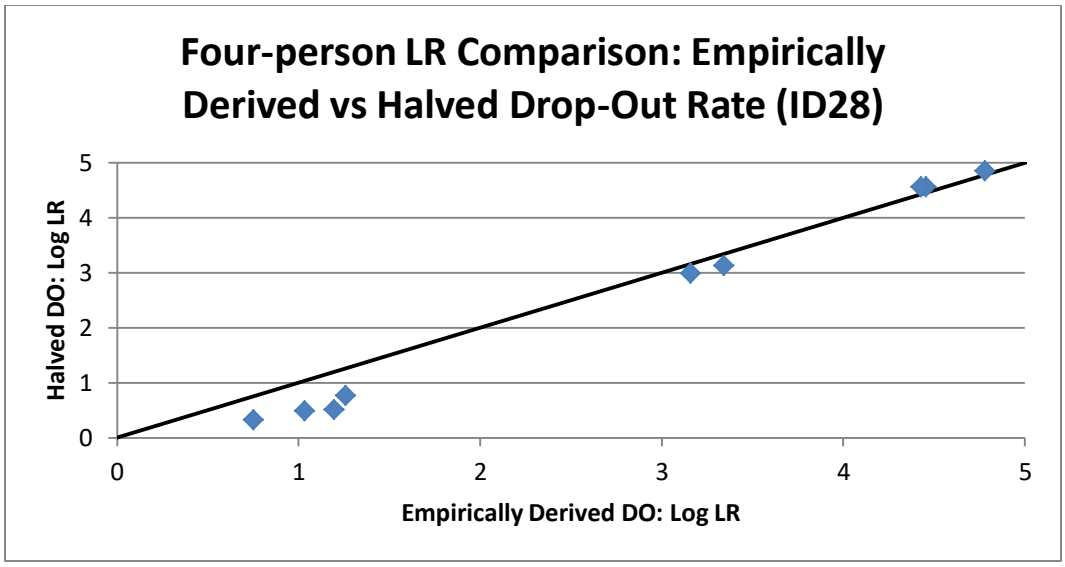

Fig. 3A Log LR plot for four-person non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 150pg/ul to 485pg/ul.

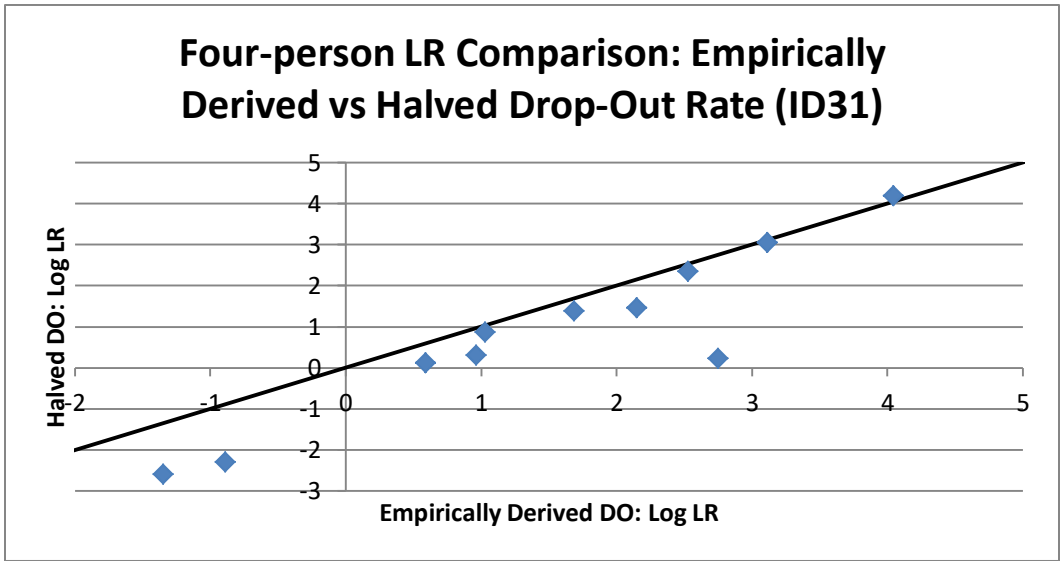

Fig. 3B Log LR plot for four-person non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation value ranged from 25pg/ul to 100pg/ul.

Table 4: Effect of using a lower than the empirically derived drop-out rate for a true contributor to four-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Four-person, 28 cycles	0	5	4
Four-person, 31 cycles	0	6	5

As depicted in Figure 3A, four of the nine four-person samples that were amplified using 28 cycles resulted in lower LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Three samples resulted in LRs that were slightly increased using the halved drop-out rates, relative to the empirically derived drop-out rates, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. Two samples resulted in LRs that were slightly decreased using the halved drop-out rates, relative to the empirically derived drop-out rates, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. As shown in Figure 3B, five of the eleven samples that were amplified using 31 cycles resulted in lower LRs when calculated using half of the empirically derived drop-out rate compared to the empirically derived drop-out rate. Five of the samples resulted in LRs that were slightly decreased, yet the order of magnitude was the same as the LRs calculated using the empirically derived drop-out rate. One sample resulted in a LR that was slightly increased, yet the order of magnitude was the same as the LR that was calculated using the empirically derived drop-out rate. Overall, using drop-out rates that were half of the empirically

derived values resulted in LRs that were lower or approximately the same as those calculated using the empirically derived rates (Table 4).

In general, the LRs decreased when half of the empirically derived drop-out rates were used compared to the empirically derived drop-out rates. Although, the majority of the LRs decreased, the change in LRs seemed to plateau as the number of contributors increased and the drop-out rate did not have as much of an effect on the LRs. This can be seen in tables 2-4. Using half of the empirically derived drop-out rates decreased the LR in approximately 63% of the two-person mixtures, 63% of the three-person mixtures, and 45% of the four-person mixtures. As the number of contributors increases, drop-out may not be as evident due to allele sharing between individuals. In that case, the LRs of the comparison of a POI may not be affected as much. In addition, five of the twenty-three three-person mixtures showed a change in support for inclusion to support for exclusion. This is consistent with the manual interpretation for these samples comparing each of these five true contributors to their respective samples. In all five cases, trained analysts found that no conclusions regarding inclusion or exclusion could be drawn. In inconclusive cases like this, it is expected that the LR would be mostly uninformative as well and close to one, which means that, if any value different from one is considered, slight variations in the calculation can cause a shift in support for one hypothesis over another. However, using half of the empirically derived drop-out rate is an approach to calculating LRs which usually lends more support for the defense hypothesis supporting the exclusion of a POI who is a true contributor to the sample.

Effect of different drop-out rates when testing of non-contributors as POI

The mixtures tested for effect of two different drop-out rates are shown in Table 1. Each mixture was compared to a database of 10,000 simulated non-contributors twice, using the same two drop-out rates that were previously used for the true contributor testing of each sample. Comparisons of log LRs obtained with empirically derived drop-out rates and underestimated drop-out rates are shown below for high template and low template two, three, and four-person samples (Figures 4-6).

Two-person

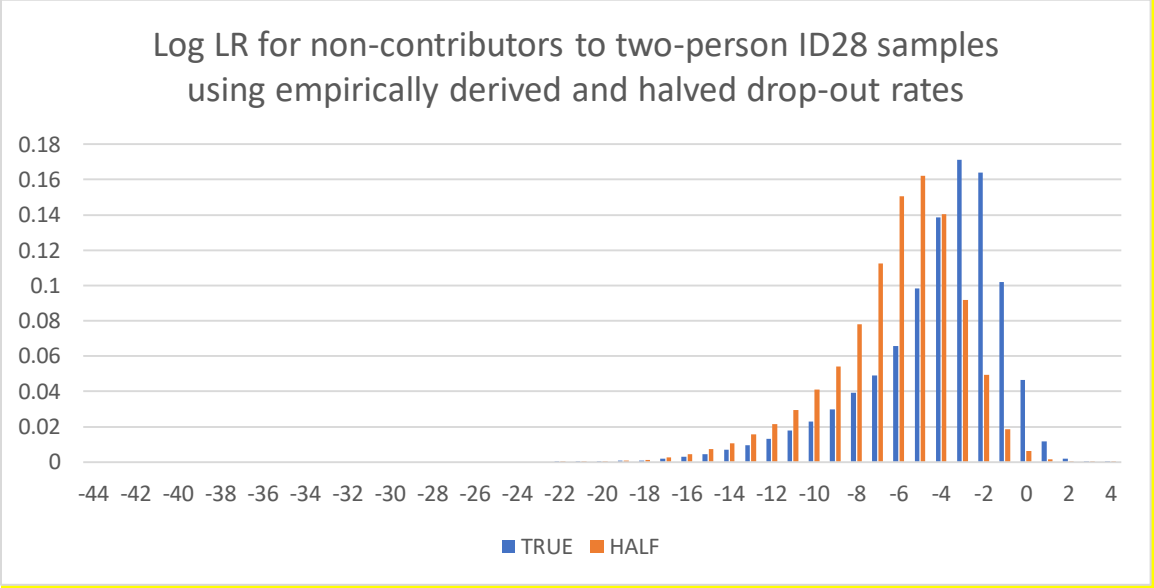

Fig. 4A Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to eight two-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 210pg/ul to 500pg/ul.

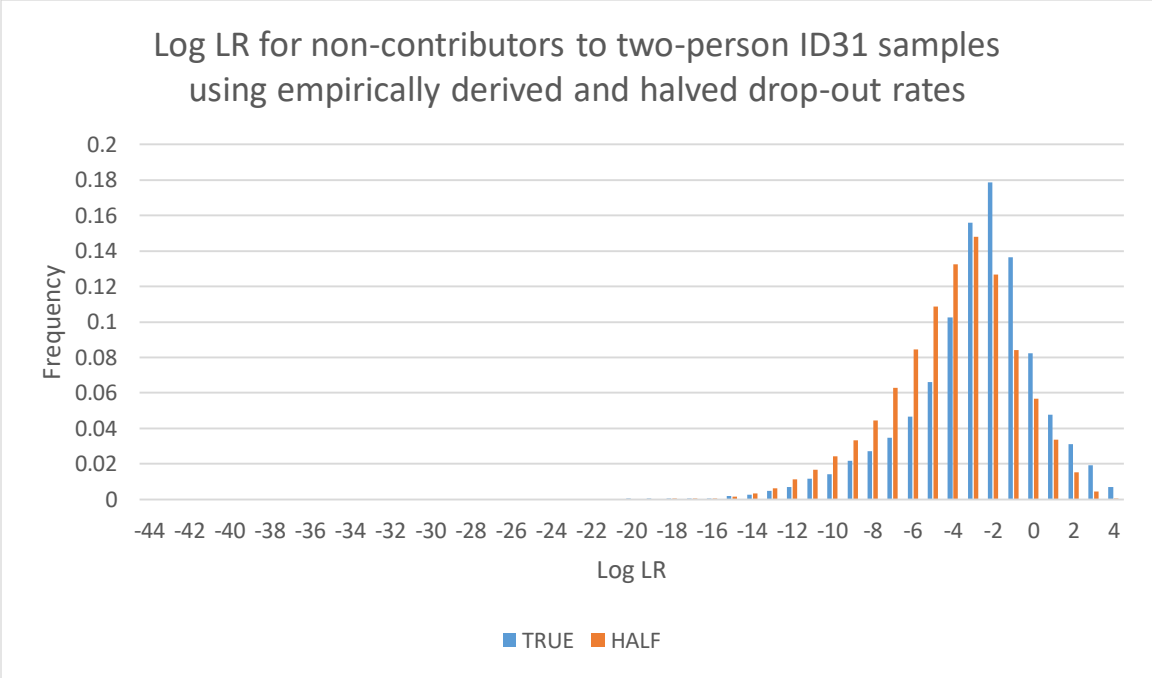

Fig. 4B Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to ten two-person deducible and non-deducible mixtures using the empirially derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation values ranged from 12pg/ul to 91pg/ul.

Figure 4A shows the distribution of log LRs for non-contributors to two-person high template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the eight two-person high template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with both "true" and "half" drop-out rates. Thus, the plot represents 80,000 non-contributor calculations for each drop-out rate.

Figure 4B shows the distribution of log LRs for non-contributors to two-person low template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the ten two-person low template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with both "true" and "half" dropout rates. Thus, the plot represents 100,000 non-contributor calculations for each drop-out rate.

Three-person

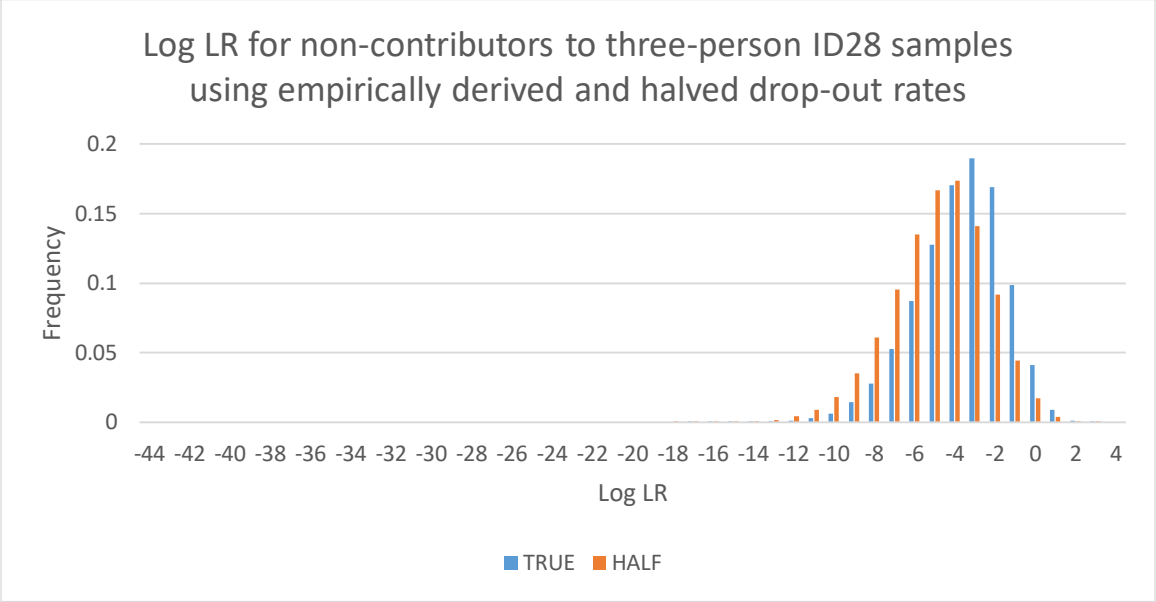

Fig. 5A Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to twelve three-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 130pg/ul to 575pg/ul.

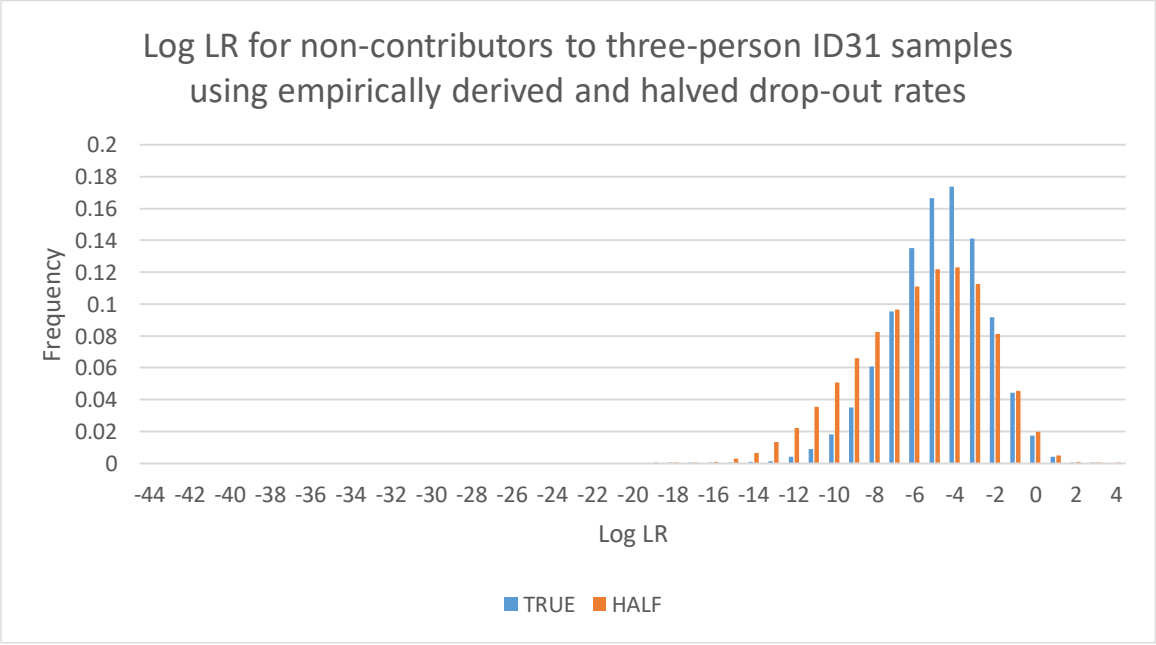

Fig. 5B Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to eleven three-person deducible and non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation values ranged from 25pg/ul to 100pg/ul.

Figure 5A shows the distribution of log LRs for non-contributors to three-person high template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the twelve three-person high template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with both "true" and "half" drop-out rates. Thus, the plot represents 120,000 non-contributor calculations for each drop-out rate.

Figure 5B shows the distribution of log LRs for non-contributors to three-person low template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the eleven three-person low template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with both "true" and "half" dropout rates. Thus, the plot represents 110,000 non-contributor calculations for each drop-out rate.

Four-person

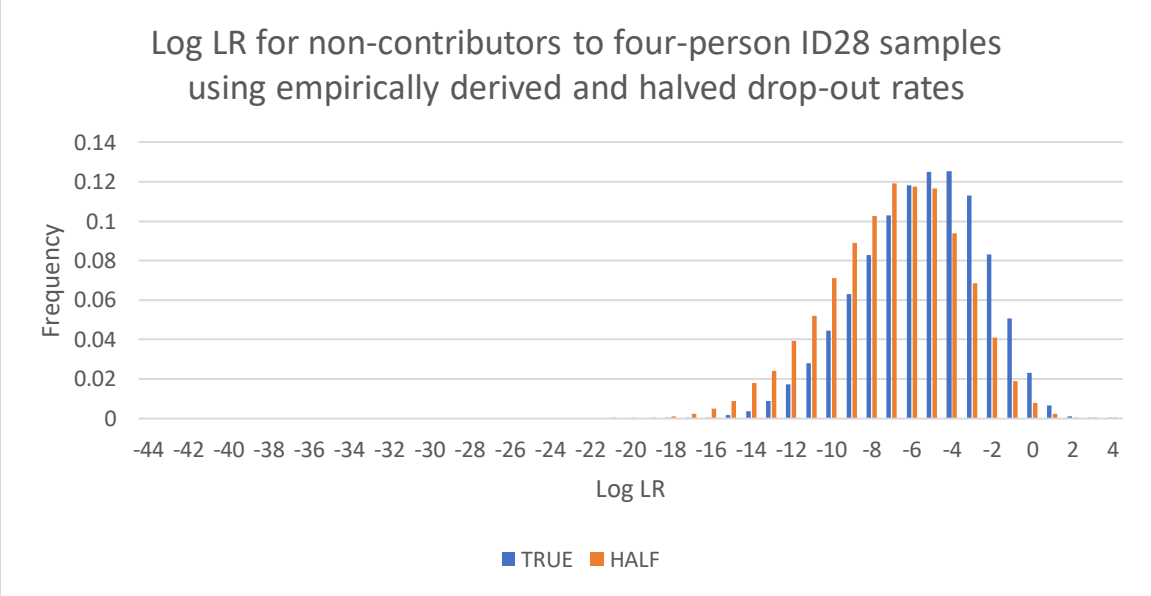

Fig. 6A Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to nine four-person non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 28 cycles. Quantitation values ranged from 150pg/ul to 485pg/ul.

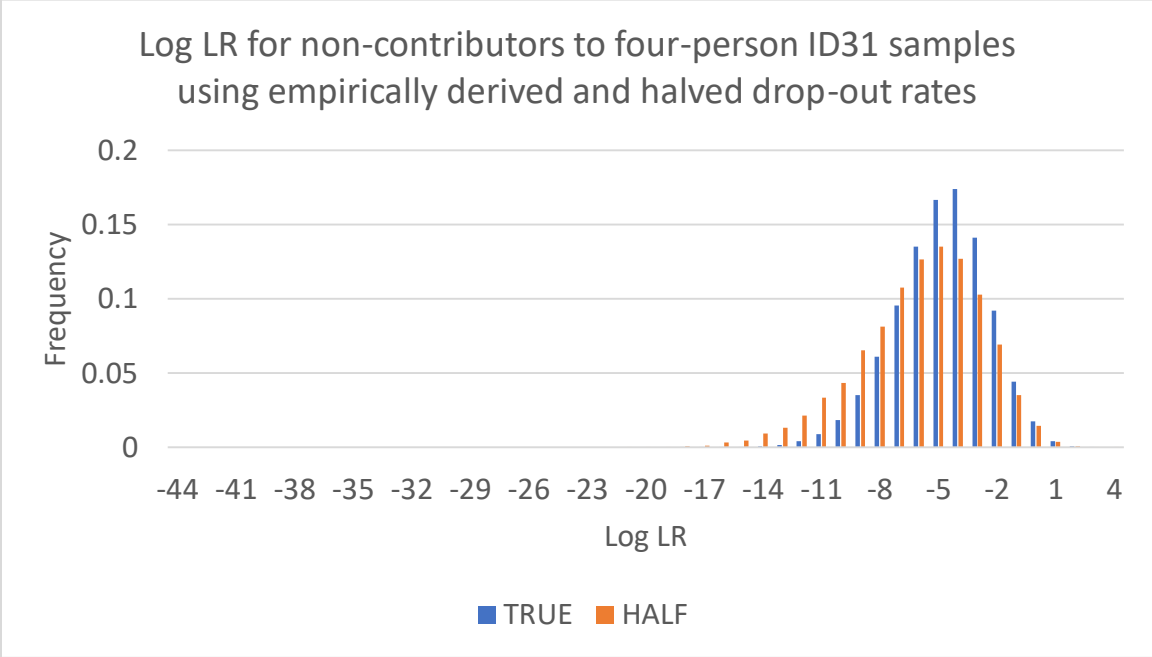

Fig. 6B Log LR distribution for the comparison of 10,000 simulated non-contributor profiles to eleven four-person non-deducible mixtures using the empirically derived drop-out rate versus the halved drop-out rate. Each sample was amplified using 31 cycles. Quantitation value ranged from 25pg/ul to 100pg/ul.

Figure 6A shows the distribution of log LRs for non-contributors to four-person high template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the nine four-person high template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with both "true" and "half" drop-out rates. Thus, the plot represents 90,000 non-contributor calculations for each drop-out rate.

Figure 6B shows the distribution of log LRs for non-contributors to four-person low template samples when the higher drop-out rates ("true") and lower drop-out rates ("half") were used. For each of the eleven four-person low template samples, 10,000 simulated non-contributor profiles were treated as the POI and LRs were computed with

both “true” and “half” dropout rates. Thus, the plot represents 110,000 non-contributor calculations for each drop-out rate.

Non-contributor testing assesses the risk of adventitious positive associations (LR>1) for unrelated individuals whose DNA is known not to be part of the tested mixture. Figures 4A – 6B show that using the lower drop-out rates shifts non-contributor LRs to the left, meaning that underestimation of drop-out rates yields lower LRs for non-contributors. This demonstrates again that underestimating the true drop-out rates for a mixture is an approach to calculating LRs which usually lends more support for the defense hypothesis supporting the exclusion of a POI who is a non-contributor. As expected, a low frequency of the calculated LRs were greater than one, which is expected as there will be some fortuitous matches through allele sharing between the true and non-contributors. For all mixture types tested, the number of non-controbutor LRs that are greater than one is lower for the reduced drop-out rate.

Effect of modifying the number of contributors

Table 5 provides a list of comparisons performed for a true contributor as the POI with correctly and incorrectly specified numbers of contributors. Comparisons of log LRs obtained when FST was run using the actual number of contributors (x-axis) and a misspecified number of contributors (y-axis) are shown below for high template and low template two, three, and four-person mixtures (Figures7A-9B). Each point represents one mixture calculation, analyzed multiple times for each of the actual number of contributors using the three different scenarios. The identity line on each plot indicates where points fall when the LR is identical using the actual number of contributors and the misspecified number of contributors. A point below the identity line indicates that the LR decreased

when the number of contributors were misspecified compared to the LR calculated using the actual number of contributors. A point above the identity line indicates that the LR increased when the number of contributors was misspecified compared to the LR calculated using the actual number of contributors.

Table 5: Mixture samples compared for different numbers of contributors.

Number of contributor comparison	**Number of Calculations**
Actual two-person vs three-person	N=30
Actual two-person vs four-person	N=30
Actual three-person vs two-person	N=45
Actual three-person vs four-person	N=45
Actual four-person vs two-person	N=60
Actual four-person vs three-person	N=60

Two-person

The logarithm of each LR for apparent two-person mixtures using true contributors to the mixtures as the POIs was taken and the results are depicted in Figures 7A and 7B with a summary of the results in Table 6.

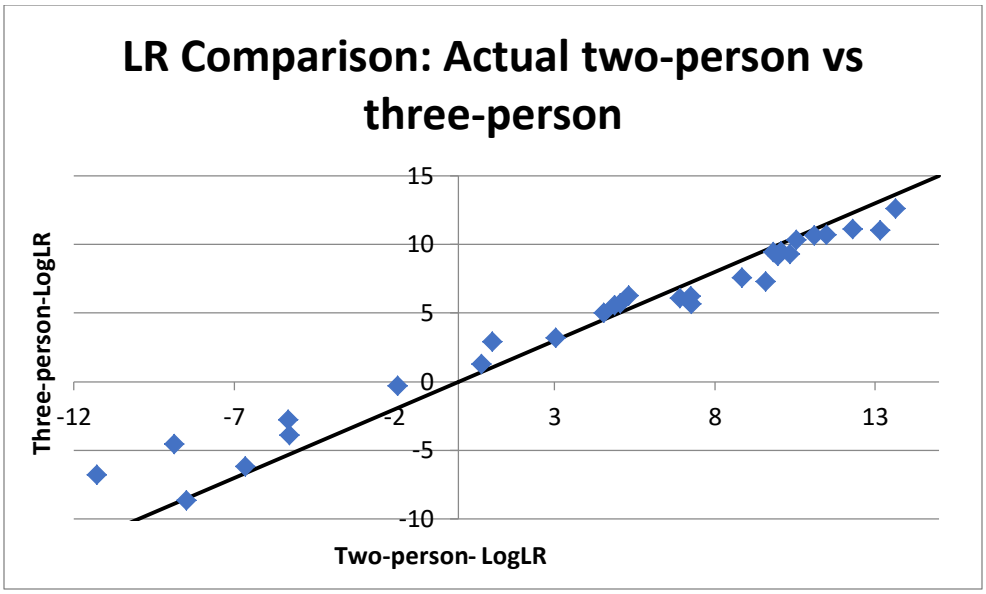

Fig. 7A Comparison of the Log LRs for actual two-person mixtures using the two-person model versus the three-person model.

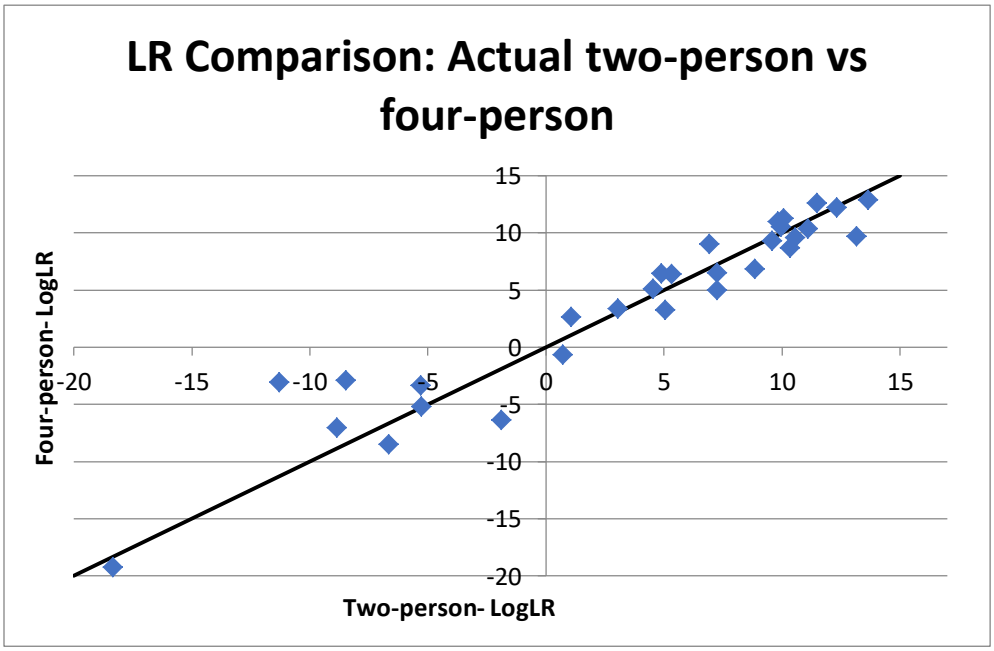

Fig. 7B Comparison of the Log LRs for actual two-person mixtures using the two-person model versus the four-person model.

Table 6: Effect of using a different number of contributors for actual two-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Actual two-person vs three-person	10	9	11
Actual two-person vs four-person	13	4	13

As depicted in Figure 7A, ten of the thirty actual two-person mixture calculations for true contributors resulted in a greater LR when run as a three-person mixture compared to a two-person mixture. Eleven of the thirty calculations resulted in a lower LR when run as a three-person mixture compared to a two-person mixture. Nine of the calculations resulted in LRs that were within the same order of magnitude. As shown in Figure 7B, thirteen of the thirty actual two-person mixture calculations resulted in a greater LR when run as a three-person mixture compared to a two-person mixture. Thirteen of the thirty calculations resulted in a lower LR when run as a three-person mixture compared to a two-person mixture. Four of the calculations resulted in LRs that were within the same order of magnitude (Table 6).

Three-person

The logarithm of each LR for apparent three-person mixtures using true contributors to the mixtures as the POIs was taken and the results are depicted in Fig. 8A and 8B with a summary of the results in Table 7.

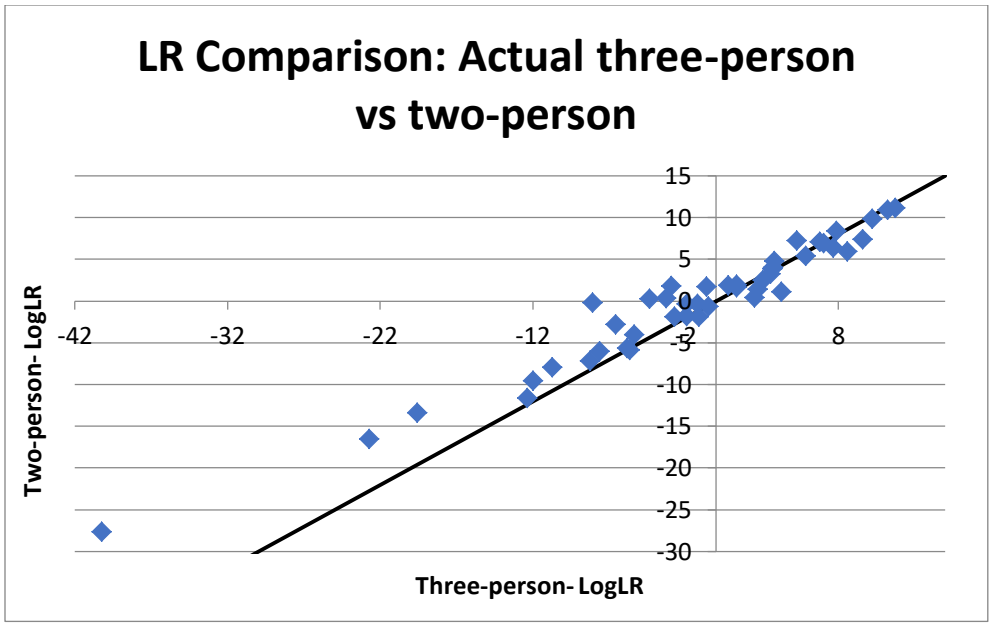

Fig.8A Comparison of the Log LRs for actual three-person mixtures using the three-person model versus the two-person model.

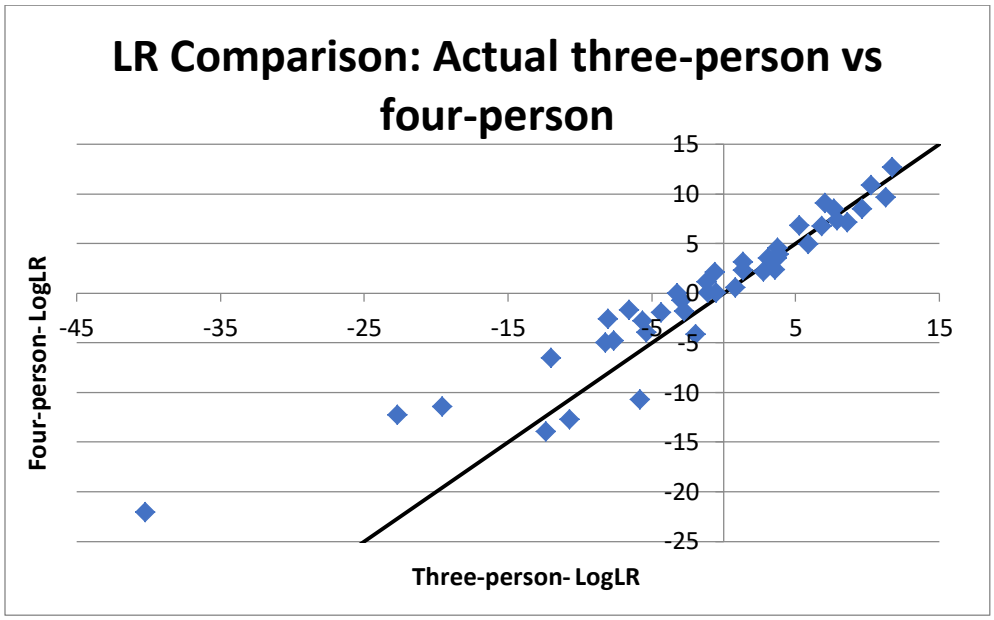

Fig. 8B Comparison of the Log LRs for actual three-person mixtures using the three-person model versus the four-person model.

Table 7: Effect of using a different number of contributors for actual three-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Actual three-person vs two-person	24	11	10
Actual three-person vs four-person	26	9	10

As depicted in Figure 8A, twenty-four of the forty-five actual three-person mixture calculations resulted in a greater LR when run as a two-person mixture compared to a three-person mixture. Ten of the forty-five calculations resulted in a lower LR when run as a two-person mixture compared to a three-person mixture. Eleven of the calculations resulted in LRs that were within the same order of magnitude. As shown in Figure 8B, twenty-six of the forty-five actual three-person mixture calculations resulted in a greater LR when run as a four-person mixture compared to a three-person mixture. Ten of the forty-five calculations resulted in a lower LR when run as a four-person mixture compared to a three-person mixture. Nine of the calculations resulted in LRs that were within the same order of magnitude (Table 7).

Four-person

The logarithm of each LR for apparent four-person mixtures using true contributors to the mixtures as the POIs was taken and the results are depicted in Figures 9A and 9B with a summary of the results in table 8.

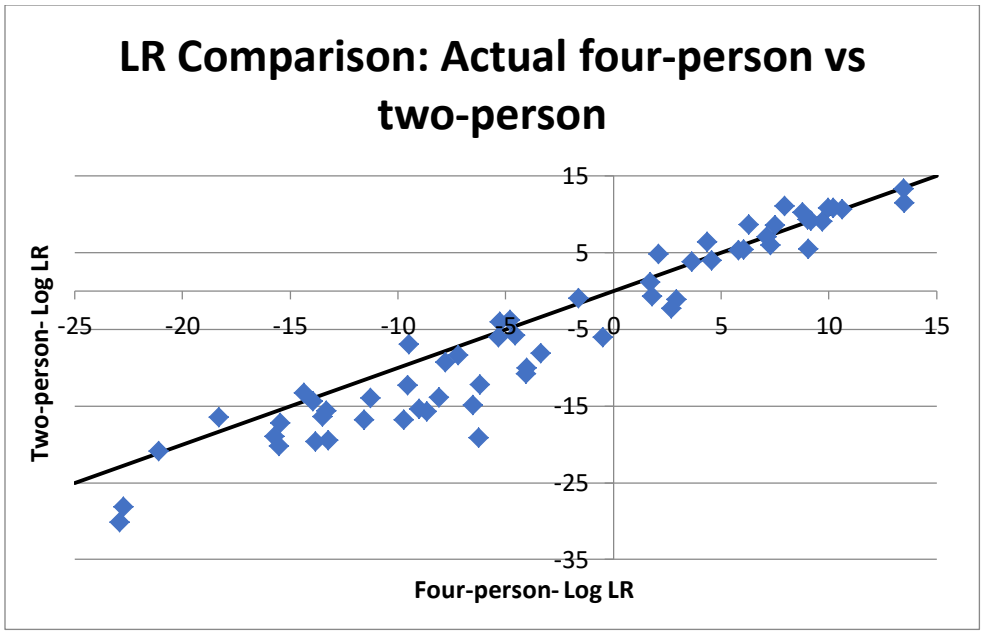

Fig. 9A Comparison of the Log LRs for actual four-person mixtures using the four-person model versus the two-person model.

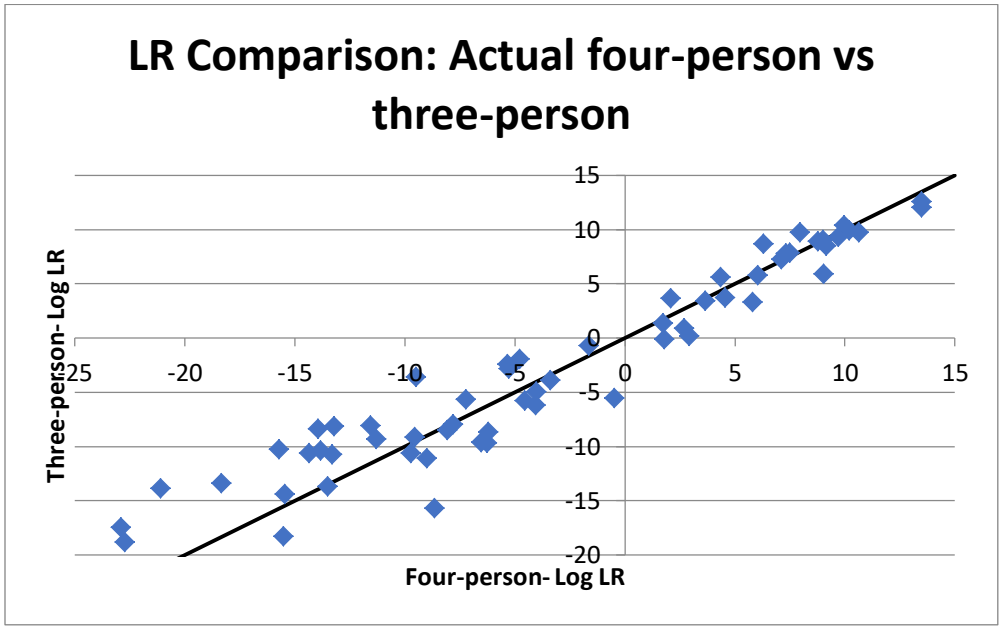

Fig. 9B Comparison of the Log LRs for actual four-person mixtures using the four-person model versus the three-person model.

Table 8: Effect of using a different number of contributors for actual four-person mixtures.

Type of Mixture	# of Samples with increased LR	# of Samples with the same order of magnitude	# of Samples with lower LR
Actual four-person vs two-person	14	10	36
Actual four-person vs three-person	24	13	23

As depicted in Figure 9A, fourteen of the sixty actual four-person mixture calculations resulted in a greater LR when run as a two-person mixture compared to a four-person mixture. Thirty-six of the sixty calculations resulted in a lower LR when run as a two-person mixture compared to a four-person mixture. Ten of the calculations resulted in LRs that were within the same order of magnitude. As shown in Figure 9B, twenty-four of the sixty actual four-person mixture calculations resulted in a greater LR when run as a three-person mixture compared to a four-person mixture. Twenty-three of the sixty calculations resulted in a lower LR when run as a three-person mixture compared to a four-person mixture. Thirteen of the calculations resulted in LRs that were within the same order of magnitude (Table 8).

For the most part, the LRs did not show a trend when the number of contributors was incorrectly estimated. Approximately 33% of the LRs decreased when the number of contributors was overestimated, compared to the 46% that increased. A similar results occurred when the number of contributors was underestimated. Approximately 42% of the LRs calculated with an understimate of the number of contributors resulted in lower LRs, compared to 38% of the LRs that increased. The majority of the LRs increased

when the number of contributors was incorrectly estimated for actual three-person mixtures, considering that over half of the LRs increased when the number of contributors were overestimated and underestimated. This could be due to the nature of the sample and the similarities that are shown between the characteristics of all three types of mixtures. In addition, approximately 1.7% of the two-person mixtures, 8.9% of the three-person mixtures, and 3.3% of the four-person mixture calculations showed a change in support for one hypothesis over another. The samples with the greatest change had a difference of five orders of magnitude in the LR calculations when the number of contributors was incorrectly estimated. For these cases, the LR calculations for two-person and four-person mixtures were greater than one when tested using the correct number of contributors, and the LRs were less than one when the number of contributors was incorrectly estimated. Again, the outcome was different for the three-person mixtures, here the 8.9% of the LR calculations that changed support were less than one when tested using the true three-person scenario, but greater than one when tested using the two-person and four-person scenarios. As with the drop-out rate study, the change in conclusions only occurred for samples inconclusive after visual comparisons. Using the OCME interpretation guidelines, experienced analysts found that no conclusions could be drawn for the comparison of the these true contributors to their respective samples.

Discussion and Conclusions

There are two characteristics of an evidentiary DNA samples that can never truly be known: the number of contributors and drop-out/drop-in rate of alleles; however, they are important factors in the calculation of a likelihood ratio for the comparison of a person of interest to an evidentiary DNA sample. Both of these factors can impact

likelihood ratio calculations and lead to either false exclusions or false inclusions of a POI. The NYC OCME developed a computer program, FST, to calculate likelihood ratios for the comparison of a POI to an evidentiary DNA sample when a positive association has been made. FST performs its LR calculations by using empirically determined drop-out rates adjusted by subtracting one standard deviation, thus using a lower, underestimated rate meant to be more conservative, in that a lower LR is obtained for non-contributors using the half drop-out rates, compared to the true drop-out rates. This study could show that LRs for true contributors are lower, and thus provide more support for the defense hypothesis, when the drop-out rate is underestimated than when the true drop-out rate is used. Haned et al. (2015) used the *likEvid* function of the R *Forensim* package to evaluate the effect of the variation in drop-out rates. Haned et al.'s study supports the findings in this thesis by also showing lower LR values for true contributors when using a lower drop-out rate. Haned et al. also tested higher drop-out rates and found these can artificially increase the LR leading to adventitious inclusions.

Similarly, LRs for non-contributors were lower when lower drop-out rates are used. This trend was observed across a wide range of DNA template amounts and mixture proportions. The calculated LRs were more conservative for both true contributors and non-contributors when drop-out rates were underestimated. More importantly, using lower drop-out rates gave lower LRs for non-contributors than higher drop-out rates, which is consistent with other research. Slooten (2017) studied the effect of varying the drop-out rate for true and non-contributors using the MixKin software and found that LRs for true contributors increased when using a higher drop-out rate compared to a lower drop-out rate. Slooten could show that this trend applied to non-

contributors as well. This effect was especially pronounced when known contributors were assumed in both the prosecutor and defense hypotheses. Slooten suggests that restricting the drop-out rates to lower values will lead to lower LRs, providing stronger support for exclusion of the non-contributors. This shift towards lower LRs for non-contributors was confirmed in this study. Only a low frequency of the LRs calculated were greater than one, indicating support for inclusion, when non-contributors were compared to the mixtures using the reduced drop-out rates. The number of LRs greater than one was higher for the actual drop-out rate. Although values above one for known non-contributors seem counterintuitive, these occurrences are expected due to allele sharing between individuals. In conclusion, underestimating the true drop-out rates for true contributors to a DNA mixture generally produces lower LRs that provide more support for the defense hypothesis which can reduce the chance of false inclusions of non-contributors.

Under or over estimating the number of contributors in a mixture can cause both, higher or lower LR values. Benschop, Haned, Jeurissen, Gill, & Sijen (2015) found that for many samples, assuming an incorrect number of contributors resulted in higher LRs that would have been in favor of the prosecution. In this study, as can be seen in table 7, this held true for the actual three-person mixtures, where the majority of LRs increased when the scenarios were changed to either two or four contributors. Regarding the actual two-person and four-person mixtures, the percentage of LR calculations that increased for under or overestimated contributors was similar to the percentage that decreased. A small percentage of the calculations lead to a change in the supported hypothesis when the number of contributors was incorrectly estimated. This is consistent with the visual

comparison that was made because the analysts found that no conclusions could be drawn for the comparison of the true contributors to their respective samples when the scenario chosen did not reflect the true number of contributors. Benschop et al. (2015) also generated LRs that were less than one when the number of contributors were underestimated using the LRmix model. They state that "these 'false exclusions' only occurred for mixtures from the 'extreme homozygote' datasets", indicating that they occur under high allele sharing conditions. Their study also used assumed known contributors in their scenarios, which caused a larger increase in LR values that provides more support for inclusion of a person of interest. However, the increase in LRs provides stronger support for the inclusion of the true contributors to the mixtures. This study did not test scenarios with assumed known contributors.

Marsden, Rudin, Inman, & Lohmueller (2016) used DNAMIX software to assess the frequency and under which conditions the incorrect estimate of the number of contributors could generate a LR greater than one for true contributors and less than one for non-contributors in complex mixtures. They found that 99.99% of the true contributors yielded a LR greater than one. The instances that generated a LR less than one resulted from samples that were from five-person mixtures. It was also found that 0.05% of the known non-contributors resulted in a LR greater than one. The samples which gave LRs greater than one were from higher order four and five-person mixtures. It can be more challenging to assign genotypes to the individual contributors in higher order mixtures due to the presence of a greater number of alleles present, making it more difficult to estimate mixture proportions, and to the greater probability of allele sharing. In those cases, the comparison of an individual to those mixtures could lead to an LR

closer to one, which doesn't necessarily provide strong support for either hypothesis. Nonetheless, the LRs for true contributors did either increase or decrease, however, overestimating the LR for a true contributor is less of a concern than overestimating LRs for non-contributors. The results presented here were consistent with those previously published by Benschop et al. (2015).

It is important to note that the second part of this project (effect of modifying the number of contributors) does not include a comparison of known non-contributors. Since non-contributors did not contribute their DNA to the evidence samples in question, it is expected that the LRs would be less than one. However, as has been shown for the drop-out rate study, adventitious LRs greater than one will occur due to allele sharing. Investigating the effect of treating known non-contributors as POIs and comparing them to mixtures using different number of contributor scenarios is something that could be a future project. This information could be useful to determine the risk of falsely including a non-contributor to a mixture if the number of contributors is incorrectly estimated.

As stated by Collins and Morton (1994), LR calculations for DNA identification increase efficiency and reliability and provide a more informative method to introduce and understand the evidence when presented to a jury. For this reason, statistical calculations in DNA cases have evolved and the use of likelihood ratios has been highly recommended for comparisons between individuals and evidence samples. The New York City Office of Chief Medical Examiner developed a program for this purpose, to provide a more informative interpretation of the results in a court of law[1].

[1] Overall, these results from this research can/should be viewed as a reference for laboratories when looking at these types of issues. OCME has validated protocols that dictate what their procedures and policies are. The opinions expressed in this thesis are mine, and not those of the OCME.

References

Balding, D. J. (2005). *Weight-of-evidence for Forensic DNA Profiles.* West Sussex, England: John Wiley & Sons Ltd.

Benschop, C. C. G., Haned, H., Jeurissen, L., Gill, P. D., & Sijen, T. (2015). The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures. *Forensic Science International: Genetics*, *19*, 92–99. https://doi.org/10.1016/j.fsigen.2015.07.003

Bille, T., Bright, J. A., & Buckleton, J. (2013). Application of random match probability calculations to mixed str profiles. *Journal of Forensic Sciences*, *58*(2), 474–485. https://doi.org/10.1111/1556-4029.12067

Bille, T. W., Weitz, S. M., Coble, M. D., Buckleton, J., & Bright, J. A. (2014). Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis*, *35*(21–22), 3125–3133. https://doi.org/10.1002/elps.201400110

Bright, J. A., Taylor, D., McGovern, C., Cooper, S., Russell, L., Abarno, D., & Buckleton, J. (2016). Developmental validation of STRmix^{TM}, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics*, *23*, 226–239. https://doi.org/10.1016/j.fsigen.2016.05.007

Brümmer, N. (2013). *Tutorial for Bayesian forensic likelihood ratio*. (October), 1–14. Retrieved from http://arxiv.org/abs/1304.3589

Buckleton, J. (2005) A Framework for Interpreting DNA Evidence. In J. Buckleton, C. M. Triggs, & S. J. Walsh (Eds.), *Forensic DNA Evidence Interpretation* (pp. 27-63). Boca Raton, FL: CRC Press.

Butler, J. M. (2010). *Fundamentals of Forensic DNA Typing.*New York: Elsevier

Butler, J. M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. New York: Elsevier.

Clayton, T. M., Whitaker, J. P., Sparkes, R., & Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International*, *91*(1), 55–70. https://doi.org/10.1016/S0379-0738(97)00175-8

Collins, A., & Morton, N. E. (1994). Likelihood ratios for DNA identification. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(13), 6007–6011. https://doi.org/10.1073/pnas.91.13.6007

Gill, P., Brenner, C. H., Buckleton, J. S., Carracedo, A., Krawczak, M., Mayr, W. R., … Weir, B. S. (2006). DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International*, *160*(2–3), 90–101. https://doi.org/10.1016/j.forsciint.2006.04.009

Gill, P., Gusmão, L., Haned, H., Mayr, W. R., Morling, N., Parson, W., … Weir, B. S. (2012). DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics*, *6*(6), 679–688. https://doi.org/10.1016/j.fsigen.2012.06.002

Haned, H., Pène, L., Sauvage, F., & Pontier, D. (2011). The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture. *Forensic Science International: Genetics*, *5*(4), 281–284. https://doi.org/10.1016/j.fsigen.2010.04.005

Haned, H., Benschop, C. C. G., Gill, P. D., & Sijen, T. (2015). Complex DNA mixture analysis in a forensic context: Evaluating the probative value using a likelihood ratio model. *Forensic Science International: Genetics*, *16*, 17–25. https://doi.org/10.1016/j.fsigen.2014.11.014

Haned, H., Pène, L., Lobry, J. R., Dufour, A. B., & Pontier, D. (2011). Estimating the Number of Contributors to Forensic DNA Mixtures: Does Maximum Likelihood Perform Better Than Maximum Allele Count? *Journal of Forensic Sciences*, *56*(1), 23–28. https://doi.org/10.1111/j.1556-4029.2010.01550.x

Hares, D. R. (2015). Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics*, *17*, 33–34. https://doi.org/10.1016/j.fsigen.2015.03.006

Inman, K., Rudin, N., Cheng, K., Robinson, C., Kirschner, A., Inman-Semerau, L., & Lohmueller, K. E. (2015). Lab Retriever: A software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. *BMC Bioinformatics*, *16*(1), 1–10. https://doi.org/10.1186/s12859-015-0740-8

Kelly, H., Bright, J. A., Buckleton, J. S., & Curran, J. M. (2014). A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science and Justice*, *54*(1), 66–70. https://doi.org/10.1016/j.scijus.2013.07.003

Mapes, A. A., Kloosterman, A. D., van Marion, V., & de Poot, C. J. (2016). Knowledge on DNA Success Rates to Optimize the DNA Analysis Process: From Crime Scene to Laboratory. *Journal of Forensic Sciences*, *61*(4), 1055–1061. https://doi.org/10.1111/1556-4029.13102

Marsden, C. D., Rudin, N., Inman, K., & Lohmueller, K. E. (2016). An assessment of the information content of likelihood ratios derived from complex mixtures. *Forensic Science International: Genetics*, *22*, 64–72. https://doi.org/10.1016/j.fsigen.2016.01.008

Mitchell, A. A., Tamariz, J., O'Connell, K., Ducasse, N., Budimlija, Z., Prinz, M., & Caragine, T. (2012). Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics*, *6*(6), 749–761. https://doi.org/10.1016/j.fsigen.2012.08.007

Mitchell, A. A., Tamariz, J., O'Connell, K., Ducasse, N., Prinz, M., & Caragine, T. (2011). Likelihood ratio statistics for DNA mixtures allowing for drop-out and drop-in. *Forensic Science International: Genetics Supplement Series*, *3*(1), e240–e241. https://doi.org/10.1016/j.fsigss.2011.08.119

National Research Council (US) Committee on DNA Forensic Science: An Update. (1996). The Evaluation of Forensic DNA Evidence. Washington (DC): National Academies Press (US); Available from: https://www.ncbi.nlm.nih.gov/books/NBK232610/ doi: 10.17226/5141

New York City Office of Chief Medical Examiner (2012) Standard Operating Procedure for STR Analysis. Unpublished.

Perez, J., Mitchell, A. A., Ducasse, N., Tamariz, J., & Caragine, T. (2011). Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts. *Croatian Medical Journal*, *52*(3), 314–326. https://doi.org/10.3325/cmj.2011.52.314

Perlin, M. W., Legler, M. M., Spencer, C. E., Smith, J. L., Allan, W. P., Belrose, J. L., & Duceman, B. W. (2011). Validating trueallele ® dna mixture interpretation. *Journal of Forensic Sciences*, *56*(6), 1430–1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Perlin, M. W., & Sinelnikov, A. (2009). An information gap in DNA evidence interpretation. *PLoS ONE, 4*(12). https://doi.org/10.1371/journal.pone.0008327

Perlin, M. W., & Szabady, B. (2001).Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples. *Journal of Forensic Sciences*, *46*(6), 1372-1378.

Scientific Working Group on DNA Analysis Methods. (2017). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 1–90.

Slooten, K. (2017). Accurate assessment of the weight of evidence for DNA mixtures by integrating the likelihood ratio. *Forensic Science International: Genetics*, *27*, 1–16. https://doi.org/10.1016/j.fsigen.2016.11.001