Article

Compound and Conditioned Likelihood Ratio Behavior within a Probabilistic Genotyping Context

Kyle Duke *, Daniela Cuenca ©, Steven Myers and Jeanette Wallin ©

California Department of Justice, Richmond, CA 94804, USA
* Correspondence: kyle.duke@doj.ca.gov

Abstract: In cases where multiple questioned individuals are separately supported as contributors to a mixed DNA profile, guidance documents recommend performing a comparison to see if there is support for their joint contribution. Anecdotal observations suggest the summed log of the individual likelihood ratios (LR), termed the simple LR product, should be roughly equivalent to or less than the log(LR) for the joint likelihood ratio, termed the compound LR. To assist casework analysts in evaluating statistical weights applied to a case at hand, this study assessed how consistently compound LRs conform to an additive behavior when compared to the simple LR product counterparts. Two-, three-, and four-person DNA mixture data, of various mixture proportions and DNA inputs, were interpreted by STRmix® version 2.8 Probabilistic Genotyping Software. Relative magnitudes of LR increases were found to be dependent on both template level and mixture composition. The distribution of log(LR) differences between all compound/simple LR comparisons was ~−2.7 to ~28.3. This level of information gain was similar to that for compound LR comparisons, with and without interpretation conditioning (~−3.2 to ~27.7). In both scenarios, the probability density peaked at approximately 0.5, indicating the information gain from constrained genotype combinations has a comparable impact on the outcome of LR calculations whether the restriction is applied before or after interpretation.

Keywords: likelihood ratio; sub-source propositions; conditioning; compound LR; additive behavior

1. Introduction

Likelihood ratios (LRs) have been in use in forensic DNA typing for several decades now as a weight-of-evidence statistic for DNA mixture interpretation [1–7]. The construction of a sub-source LR necessarily requires a consideration of which propositions best represent the inclusionary (prosecution) and alternative (exclusionary/defense) theories about the source of the DNA evidence based on the relevant case information [8–11]. When the inclusionary proposition includes one individual as a contributor to the mixture and the alternative proposition replaces that individual with an unknown, unrelated contributor, the LR propositions for such a calculation, hereafter referred to as a “simple LR”, can be represented with the following shorthand notation (shown below for a two-person mixture), where “ID” represents a questioned individual and “U” represents an unknown, unrelated mixture contributor:

\[
\text{Inclusionary proposition} = \frac{ID1 + U}{U + U} \quad \text{(1)}
\]

This LR assesses how well individual ID1 explains the data. When there are multiple individuals in question and the presence of at least one of those individuals in the evidence can be reasonably postulated, an LR for the remaining individuals, hereafter referred to as a “conditioned LR”, may be calculated [12–15]. Basic properties of conditioned LRs in the forensic DNA context have been previously discussed in the literature. Generally, when the simple LRs for multiple individuals are greater than
1, conditioning on one or more individuals increases the magnitude of the LR for the remaining individuals, if all individuals are true contributors to the mixture [16]. The higher value from a conditioned LR is due to the reduction in ambiguity that results from a decrease in the possible genotype combinations in the mixture. This basic concept is represented below as:

\[
\frac{ID_1 + U}{U + U} \leq \frac{ID_1 + ID_2}{U + ID_2}
\]

where ID1 and ID2 represent two different individuals.

On the other hand, the most appropriate proposition pair in the context of a given case may place all questioned individuals in the inclusionary proposition versus an equal number of unknown, unrelated contributors in the alternative proposition. The propositions for this calculation, hereafter referred to as a “compound LR”, are represented below (with the same two-person mixture as an example):

Inclusionary proposition
Alternative proposition

\[
\frac{ID_1 + ID_2}{U + U}
\]

A criticism by the recent NISTIR 8351-DRAFT [17] speaks to the impact of specific propositions chosen for the calculated LR value: “This fact should encourage more effort to standardize development of propositions . . . ” Numerous publications on proposition formulation and LR reporting [10–12,15,18] provide valuable guidance. While simple, conditioned, and compound likelihood ratios are described in detail, the relative quantitative properties are not well represented in the literature. In particular, it has anecdotally been observed that when simple LRs (>1) for different individuals in a given case are multiplied together (i.e., their exponents are added), the product, hereafter referred to as the “simple LR product”, is typically less than or equal to the corresponding compound LR for the combination of those same individuals. This additive behavior is represented below, using the same two-person mixture from Equations (1) and (3):

Simple LR product for ID1 and ID2

\[
\frac{ID_1 + U}{U + U} \times \frac{ID_2 + U}{U + U} \leq \frac{ID_1 + ID_2}{U + U}
\]

A related question is whether a compound LR increases when adding a conditioned contributor, as represented below with a three-person mixture:

\[
\frac{ID_1 + ID_2 + U}{U + U + U} \leq \frac{ID_1 + ID_2 + ID_3}{U + U + ID_3}
\]

Exploring the robustness of these quantitative relationships is helpful both to individual casework analysts making decisions about setting propositions for compound LRs as well as standardization bodies that require empirical support for guidelines related to these LRs.

Presented here is a systematic assessment of compound and conditioned LR calculations aimed at establishing LR magnitude expectations. One of the primary objectives of this study was to determine how consistently compound and conditioned LRs conformed to an additive behavior across a variety of mixture proportions as well as a wide range of DNA inputs. Beyond the question of whether a compound or conditioned LR is equal to or greater in magnitude than its simple LR product/unconditioned counterpart, the study sought also to address the questions of how large the increase is and whether the relative magnitude of the increase can be predicted based on the properties of the mixture. Finally, the study assessed whether any parallels can be drawn between the magnitudes of compound and conditioned LRs. Such information can assist the casework analyst in
evaluating the statistical weights of the LRs calculated in a case at hand. For example, in sexual assault or touch DNA cases where the prosecution’s case posits multiple suspects as DNA mixture contributors, empirically supported expectations of compound and/or conditioned LR magnitude would allow an analyst to confidently provide information on whether those multiple suspects can explain the mixture data when considered together, instead of considering them in isolation. Per the authors of the NISTIR 8351-DRAFT, such information is currently insufficient in the literature.

2. Materials and Methods

2.1. Construction, Amplification, Capillary Electrophoresis, and Analysis of Ground Truth DNA Mixtures

Buccal cell DNA was collected with informed consent from healthy, unrelated laboratory volunteers and extracted using the PrepFiler\textsuperscript{TM} DNA Extraction Kit (Life Technologies, South San Francisco, CA, USA).

Because two-person mixtures are the easiest for STRmix\textsuperscript{®} to analyze, the composition of the two-person donor sets was not directed at achieving a given maximum allele count (see Table 1 and Supplementary Table S5). Meanwhile, the three- and four-person mixture donors were selected so that one set that had a maximum allele count consistent with N-1 contributors-, i.e., a maximum of four detected alleles per locus for the three-person mixtures and a maximum of six detected alleles per locus for the four-person mixtures-and one set that had a maximum allele count consistent with N contributors (5–6 alleles per locus for the three-person mixtures and 7–8 alleles per locus for the four-person mixtures).

Table 1. DNA donors for compound/conditioned LR study mixtures.

| Donor Group (Donor Number) | Max Number of Alleles | Donor ID | Sex   |
|---------------------------|----------------------|---------|-------|
| 2-person Group 1 (1)      | 4                    | F1      | Female |
| 2-person Group 1 (2)      | 4                    | F2      | Female |
| 2-person Group 2 (1)      | 4                    | F3      | Female |
| 2-person Group 2 (2)      | 4                    | M1      | Male   |
| 3-person Group 1 (1)      | 4                    | M2      | Male   |
| 3-person Group 1 (2)      | 4                    | F1      | Female |
| 3-person Group 1 (3)      | 4                    | F2      | Female |
| 3-person Group 2 (1)      | 6                    | F4      | Female |
| 3-person Group 2 (2)      | 6                    | F3      | Female |
| 3-person Group 2 (3)      | 6                    | M1      | Male   |
| 4-Person Group 1 (1)      | 6                    | M3      | Male   |
| 4-Person Group 1 (2)      | 6                    | F5      | Female |
| 4-Person Group 1 (3)      | 6                    | M4      | Male   |
| 4-Person Group 1 (4)      | 6                    | F6      | Female |
| 4-person Group 2 (1)      | 8                    | M5      | Male   |
| 4-person Group 2 (2)      | 8                    | F7      | Female |
| 4-person Group 2 (3)      | 8                    | F8      | Female |
| 4-person Group 2 (4)      | 8                    | F9      | Female |

Mixtures at a variety of DNA input amounts and mixture ratios were constructed by quantifying extraction yield for each donor with Quantifiler\textsuperscript{TM} Trio DNA Quantification Kit (Life Technologies) and combining the amounts of template indicated in Table 2. Each template series for a given mixture was made by diluting a single high-level mixture. Particular attention was focused on the inclusion of mixtures with high major to minor contributor ratios (e.g., 99:1, 100:100:4, 100:100:100:6), since these extreme ratios are occasionally encountered in casework and the PCAST review of forensic science [19,20] indicated that more empirical support is needed to demonstrate reliable software performance under these conditions: “... the studies involve few mixtures in which a sample is present at an extremely low ratio. By expanding these empirical studies, it should be possible to test validity and reliability across a broader range”.
Table 2. Sample composition of mixtures for compound/conditioned LRs. Donor mixture ratios are listed from left to right in accordance with the donor numbers listed in Table 1. The input amounts listed are for total DNA.

| Donor Group | Mixture Ratio | Input Amounts Tested | Replicates |
|-------------|---------------|----------------------|------------|
| 2-person Group 1 | 9:1 | 2 ng, 1 ng, 870 pg, 750 pg, 500 pg, 380 pg, 250 pg, 125 pg, 63 pg | 2 |
| 2-person Group 1 | 49:1 | 2.5 ng, 1.9 ng, 1.25 ng, 625 pg, 313 pg | 2 |
| 2-person Group 1 | 99:1 | 2.5 ng, 1.25 ng, 625 pg | 2 |
| 2-person Group 2 | 1:1 | 800 pg, 400 pg, 200 pg, 100 pg, 50 pg, 25 pg | 1 |
| 2-person Group 2 | 3:1 | 800 pg, 400 pg, 348 pg, 300 pg, 200 pg, 152 pg, 100 pg, 50 pg, 25 pg | 1 |
| 3-person Group 1 | 3:2:1 | 1.2 ng, 600 pg, 522 pg, 450 pg, 300 pg, 228 pg, 150 pg, 75 pg, 38 pg | 2 |
| 3-person Group 1 | 10:5:1 | 3.2 ng, 1.6 ng, 1.4 ng, 1.2 ng, 800 pg, 608 pg, 400 pg, 200 pg, 100 pg | 2 |
| 3-person Group 1 | 100:100:4 | 1.2 ng, 625 pg, 325 pg | 2 |
| 3-person Group 2 | 1:1:1 | 1.2 ng, 600 pg, 300 pg, 150 pg | 1 |
| 3-person Group 2 | 3:2:1 | 1.2 ng, 522 pg, 300 pg, 150 pg, 38 pg | 2 |
| 3-person Group 2 | 10:5:1 | 3.2 ng, 1.4 ng, 800 pg, 400 pg, 100 pg | 2 |
| 3-person Group 2 | 100:100:4 | 1.28 ng, 638 pg, 319 pg | 2 |
| 4-person Group 1 | 4:3:2:1 | 2 ng, 1 ng, 870 pg, 750 pg, 500 pg, 380 pg, 250 pg, 125 pg, 63 pg | 2 |
| 4-person Group 1 | 10:5:2:1 | 3.6 ng, 1.8 ng, 1.6 ng, 1.4 ng, 900 pg, 684 pg, 450 pg, 225 pg, 113 pg | 2 |
| 4-person Group 1 | 100:100:100:6 | 1.28 ng, 625 pg, 325 pg | 2 |
| 4-person Group 2 | 1:1:1:1 | 1.6 ng, 800 pg, 400 pg, 200 pg, 100 pg, 50 pg | 1 |
| 4-person Group 2 | 4:3:2:1 | 2 ng, 870 pg, 500 pg, 250 pg, 63 pg | 2 |
| 4-person Group 2 | 10:5:2:1 | 3.6 ng, 1.6 ng, 900 pg, 450 pg, 113 pg | 2 |
| 4-person Group 2 | 100:100:100:6 | 1.28 ng, 638 pg, 319 pg | 2 |

The mixtures were amplified with the GlobalFiler™ PCR Amplification Kit (Life Technologies), either singly or in duplicate as indicated, with a 28-cycle thermal cycling protocol on the ProFlex (Life Technologies) and electrophoresed on an Applied Biosystems™ 3500 Genetic Analyzer using POP-6 polymer (Life Technologies). Replicate amplifications were performed for any mixture with a minimum minor mixture proportion less than 25% in order to test a feature of STRmix unrelated to this study; however, since these replicates are independent data points, they were all included in this study, and each was independently interpreted with STRmix.

The capillary electrophoretic data from the amplified mixtures were subsequently analyzed in GeneMapper ID-X v1.6 (Life Technologies, South San Francisco, CA, USA) at channel-specific analytical thresholds of 51 RFU (blue), 71 RFU (green), 35 RFU (yellow), 41 RFU (red), and 61 RFU (purple) and analyzed with STRmix v2.8 (Institute of Environmental Science and Research, Auckland, New Zealand) using laboratory-validated settings, which include stutter models for $-1/+1$, $-2/+2$, $-0.5/+0.5$, and $-1.5$ STR repeats. Upon completion of interpretation, the targeted mixture proportions were confirmed to be consistent with the mixture proportions from the STRmix Interpretation Report for the highest template amount for each mixture from Table 2 (data not shown).

For this study, both unconditioned and conditioned analyses were performed, with the conditioned contributors consisting of each combination of single and multiple mixture donors up to a total number of N-1. The “MCMC Accepts” settings used were the default MCMC accepts of 10,000 burn-in/50,000 post burn-in per chain unless stated otherwise, where the accepts were increased by a factor of 20 to 200,000 burn-in and 1,000,000 post burn-in per chain.
2.2. Compound LR Calculations

Compound LR calculations were performed for all combinations of true contributors to the 197 mixtures in Table 2 using STRmix v2.8 [21]; this amounts to a total of 1775 compound LRs calculated (see Supplementary Tables S1–S4 for raw LR data). Table 3 is a summary of the proposition sets used for these calculations. Note that all combinations of two contributors were tested for the three-person mixtures, and all combinations of two and three contributors were tested for the four-person mixtures. One-sided 99.9% lower credible intervals, termed “Highest Posterior Density” (HPD), were calculated for all LRs unless otherwise indicated. NIST databases for African American, Caucasian, and Hispanic populations [22] were used for the calculations. Equal proportion population-stratified LRs were reported as the primary weight-of-evidence statistics.

Table 3. Propositions for simple LR product v. compound LR comparisons. Note that the designations “C1”, “C2”, etc. are generic and represent more than one combination of mixture contributors. For instance, under “Two true contributors”, C1 and C2 in proposition set (2) represent the three different combinations of two true contributors in a three-person mixture that could be paired together.

| Compound LR Comparison Propositions (2 True Contributors) | Simple LR product propositions | Compound LR propositions |
|------------------------------------------------------------|--------------------------------|--------------------------|
| C1 + Unk + Unk                                             | C2 + Unk + Unk                 | C1 + C2 + Unk + Unk      |
| Unk + Unk + Unk                                            | Unk + Unk + Unk                | Unk + Unk + Unk + Unk    |

| Compound LR comparison propositions (3 true contributors) | Simple LR product propositions | Compound LR propositions |
|----------------------------------------------------------|--------------------------------|--------------------------|
| C1 + Unk + Unk + Unk                                      | C2 + Unk + Unk + Unk + Unk     | C1 + C2 + C3 + Unk + Unk |
| * Unk + Unk + Unk                                         | Unk + Unk + Unk + Unk + Unk    | Unk + Unk + Unk + Unk    |
| * C1 + Unk + Unk                                           | Unk + Unk + Unk + Unk + Unk    | Unk + Unk + Unk + Unk    |

| Compound LR comparison propositions (4 true contributors) | Simple LR product propositions | Compound LR propositions |
|----------------------------------------------------------|--------------------------------|--------------------------|
| C1 + Unk + Unk + Unk                                      | C2 + Unk + Unk + Unk + Unk     | C1 + C2 + C3 + C4 + Unk  |
| * Unk + Unk + Unk                                         | Unk + Unk + Unk + Unk + Unk    | Unk + Unk + Unk + Unk    |
| * C1 + Unk + Unk                                           | Unk + Unk + Unk + Unk + Unk    | Unk + Unk + Unk + Unk    |
| * Unk + Unk + Unk                                          | Unk + Unk + Unk + Unk + Unk    | Unk + Unk + Unk + Unk    |

Simple LRs for all contributors to the mixtures in Table 2 were calculated, and the log(LR) were plotted to compare the magnitude of each compound LR to its corresponding simple LR product for each contributor combination.

2.3. Conditioned LR Calculations

LR calculations were performed using STRmix v2.8 for all combinations of non-conditioned true contributors to the 197 mixtures in Table 2, given a particular conditioned interpretation; this amounts to a total of 4686 conditioned LRs calculated (see Supplementary Tables S1–S4 for raw LR data). Table 4 summarizes the proposition sets used for these calculations.
Corresponding unconditioned LR calculations were performed, and the log(LR) were plotted comparing the magnitude of each conditioned LR to its corresponding unconditioned LR for each contributor combination.

### 3. Results

#### 3.1. Compound LR Calculations: 2- and 3-Person Mixtures

The results of the compound LR calculations show that the additive behavior of compound LRs largely holds over the entire per-contributor DNA input range of ~5 pg–2.5 ng for 2- and 3-person mixtures (see Figure 1), with a majority above the black line of equivalence (~92% and ~88%, respectively). Furthermore, the majority (100% for 2-person and 94% for 3-person) of the compound LRs that are less than their simple LR product counterparts, were within one order of magnitude from the line of equivalence.

In terms of the trends observed for the various mixture proportions, the compound LRs with the largest difference are those with the most ambiguity in the genotype combinations (i.e., the mixtures with the most even mixture proportions). The 2-person mixtures are so well resolved that only the higher level 1:1 mixtures, with total DNA input levels of 100 pg, 200 pg, 400 pg and 800 pg, show a substantial increase in the compound LR compared to the simple LR product, whereas the 3-person compound LRs show significant increase over the simple LR product for most mixture proportions.

#### 3.2. Conditioned LR Calculations: 2- and 3-Person Mixtures

LRs for true contributors to the 2- and 3-person mixtures interpreted with conditioning were generally equal to or higher than their unconditioned counterparts (see Figure 2). This effect is expected due to the simplification imposed upon the interpretation by fixing one or more contributor genotypes. Comparing these plots to the compound LR plots, the same trends in terms of the mixture proportions most affected by conditioning are apparent; the mixtures with the most ambiguity in the genotype combinations (i.e., the mixtures with the most even mixture proportions) tend to increase the most with conditioning.
3. Results

3.1. Compound LR Calculations: 2- and 3-Person Mixtures

The results of the compound LR calculations show that the additive behavior of compound LRs largely holds over the entire per-contributor DNA input range of ~5 pg–2.5 ng for 2- and 3-person mixtures (see Figure 1), with a majority above the black line of equivalence (~92% and ~88%, respectively). Furthermore, the majority (100% for 2-person and 94% for 3-person) of the compound LRs that are less than their simple LR counterpart, were within one order of magnitude from the line of equivalence.

\[\text{(a)}\]

\[\text{(b)}\]

\[\text{(c)}\]

Figure 1. Plots of log(compound LR) v. corresponding log(simple LR product) for (a) 2-person mixtures with 2 true contributors, (b) 3-person mixtures with 2 true contributors, and (c) 3-person mixtures with 3 true contributors.

3.3. Compound LR Calculations: 4-Person Mixtures

The compound LR plots for the 4-person mixtures also show that the additive behavior generally holds over the entire per-contributor DNA input range of ~5 pg–2 ng, as well as the same trends in upward deviation from the line of equivalence with increasing ambiguity of genotype combinations. However, many false negative compound LRs (defined as LRs of 0) are also apparent, which increase in frequency with the number of true contributors placed in the numerator of the LR (see Figure 3).

A critical observation in understanding the root cause of these false negative compound LRs is where they fall on the x axis compared to the rest of the data set. The false negatives occur at relatively low log(simple LR product) values, but do not primarily occur near the origin. Because LR magnitude correlates with DNA mixture input, this simple LR product range corresponds to mixtures that are relatively low level but not the lowest level in the data set. For instance, among the Donor Group 1 4-person mixtures, the false negatives in Figure 3c had modest total DNA input ranges of ~250–638 pg (~6–250 pg on a per-contributor basis); the Donor Group 2 false negatives had a similar total DNA input range of ~200–625 pg (~5–350 pg on a per-contributor basis). In terms of the complexity of the interpretation (and the corresponding length of the plausible genotype combination list generated by STRmix), these mixtures ranked among the most complicated, because
the contributors’ input amounts are all in the stochastic range-high enough to allow full allelic detection of some contributors but low enough that allelic dropout is a realistic (and in some cases expected) possibility for all contributors.

Concomitant with the false negative compound LRs for the 4-person mixtures was an effect on the magnitude of the difference between the point estimate LR and HPD LR, which can be viewed as a secondary diagnostic of the issue. The 2- and 3-person plots of HPD LR v. point estimate LR (see Appendix A Figure A1) were as expected, given that the HPD is intended to be a lower bound on normal LR variation; a log-scale drop of 1–2 units from point estimate to HPD was observed for the compound LRs, the same as observed for the simple LRs also shown in the plots. However, very large differences between the point estimate LR and HPD LR (>5 log units, in some cases >10 log units) were observed in the 4-person HPD v. point estimate plots of the compound LRs, with the outlying data points increasing along with the number of contributors in the LR numerator (Appendix A Figure A2a).

Figure 2. Plots of 2- and 3-person log(LR)s based on conditioned interpretations v. corresponding log(LR)s based on unconditioned interpretations: (a) 2-person mixtures conditioned on 1 true contributor, (b) 3-person mixtures conditioned on 1 true contributor, and (c) 3-person mixtures conditioned on 2 true contributors.
compound LRs (defined as LRs of 0) are also apparent, which increase in frequency with the number of true contributors placed in the numerator of the LR (see Figure 3). Figure 3. Plots of log(compound LR) v. corresponding log(simple LR product) for 4-person mixtures, with (a) 2 true contributors, (b) 3 true contributors, and (c) 4 true contributors in the numerator proposition. Increasing numbers of false negative compound LRs (LRs of 0, plotted at $-30$ and circled in red) were observed with increasing numbers of true contributors in the LR numerator. Approximately 0.7% of the data points in (a), 7.4% of the data points in (b), and 31% of the data points in (c) are false negative LRs.

The increased complexity of the problematic 4-person mixture data, relative to the remaining 4-person data, explains both where the false negative compound LRs emerge in terms of x axis position and frequency, as well as why they are accompanied by large differences between the point estimate and HPD LRs. With a very complex interpretation and a long list of plausible genotype combinations comes a substantial diffusion of probability among the options. Because obtaining a non-zero LR requires that the interpretation places all of the true contributors together in the mixture, with a longer list of plausible options comes an increasingly remote opportunity for the true genotype combinations to be selected at all loci over the course of the interpretation. Hence, the sample effectively hits a “complexity ceiling”. In other cases, even if a non-zero LR is avoided, the same diffusion of probability leads to the very large drop in HPD, as very small genotype combination weights can lead to excessively low resampled weights from the HPD calculation process.

3.4. Optimization of Markov Chain Monte Carlo (MCMC) Accept Settings for 4-Person Compound LRs

If the complexity ceiling is the issue causing the false negative LRs and large differences between the point estimate LR and HPD LR, then more thoroughly exploring the probability space with increased MCMC accepts is a reasonable solution that has been suggested in
the relevant literature [23,24]. In order to preliminarily test a MCMC accept increase, we reanalyzed all of the problematic 4-person mixtures with the MCMC burn-in and post burn-in accepts set to 100,000 and 500,000 per chain, which is a 10-fold increase over the default settings of 10,000 burn-in and 50,000 post burn-in accepts per chain for STRmix v2.8. These altered settings produced non-zero LRs for every problematic mixture with the exception of a 250 pg 4:3:2:1 mixture. In this case, the LR with all four contributors in the LR numerator was 0, as was the LR with the three lower-level contributors in the numerator (at input levels of 75 pg, 50 pg, and 25 pg), indicating that fitting these low-level contributors together was the source of the issue. Upon performing 10 replicate analyses and 4-person compound LR calculations for this mixture, 3 of the 10 recalculations returned LRs of 0, indicating that further adjustment of the MCMC accepts was warranted.

To optimize the MCMC accepts to completely eliminate the observation of false negative compound LRs, the variations in burn-in and post burn-in accepts shown in Table 5 were tested on the 250 pg 4:3:2:1 sample. We observed that while increasing the post burn-in accepts 10-fold alone was more effective in preventing false negative LRs than increasing the burn-in accepts 10-fold alone, false negatives were still observed after both isolated setting adjustments. There was at least one false negative compound LR with a 20x increase in burn-in or post burn-in, when adjusted in isolation. However, no false negative compound LRs were observed with a 15x, 20x or 50x increase to both burn-in and post burn-in accepts. In order to achieve a balance between eliminating false negative compound LRs and avoiding unnecessary increases in runtime, we elected to move forward with the 20x MCMC accept increase (i.e., to 200,000 burn-in/1 million post burn-in per chain) for the remainder of the study.

Table 5. Effect of increasing MCMC burn-in and post burn-in accepts on the false negative rate for 4-person compound LRs calculated for a 250 pg 4:3:2:1 mixture.

| Fold Increase in Accepts | Burn-In or Post Burn-In | False Negative LRs in 10 Replicates | Average Decon Time |
|--------------------------|-------------------------|-------------------------------------|--------------------|
| None                     | -                       | 9/10                                | 1.36 min           |
| 10x Burn-in              | 10/10                   |                                     | 2.30 min           |
| 10x Post burn-in         | 4/10                    |                                     | 10.71 min          |
| 20x Burn-in              | 1/10                    |                                     | 21.07 min          |
| 24x Post burn-in         | 1/10                    |                                     | 25.44 min          |
| 15x Both                 | 0/10                    |                                     | 17.73 min          |
| 20x Both                 | 0/10                    |                                     | 23.90 min          |
| 50x Both                 | 0/10                    |                                     | 58.70 min          |

3.5. Repeated Interpretation/Compound LR Calculation at 20x MCMC Accepts

All of the 4-person interpretations and calculations from Figure 3 were performed with the adjusted 20x MCMC accepts setting in place (see Figure 4). No LRs of 0 were observed, and although some large differences between point estimate and HPD LR were still observed, their incidence and magnitude were greatly reduced (see Appendix A Figure A2).

3.6. Conditioned LR Calculations: 4-Person Mixtures

Observations of false negative LRs occurred for the conditioned 4-person interpretations with the default number of MCMC accepts, similarly to the compound 4-person LRs (see Figure 5). A prominent difference with the conditioned 4-person LR plot is that the false negatives occurred along both axes; that is, in some instances, unconditioned compound LRs were zero and the corresponding conditioned LRs were non-zero, whereas in other cases unconditioned compound LRs were non-zero and the corresponding conditioned LRs were zero.
Fold Increase in Accepts Burn-In or Post Burn-In False Negative LRs in 10 Replicates

| Condition     | Average Decon Time |
|---------------|--------------------|
| None          | 9/10 1.36 min      |
| Burn-in       | 10/10 2.30 min     |
| Post burn-in  | 4/10 10.71 min     |
| Both          | 1/10 21.07 min     |
| Both          | 2/10 25.44 min     |
| Both          | 5/10 58.70 min     |

3.5. Repeated Interpretation/Compound LR Calculation at 20x MCMC Accepts

All of the 4-person interpretations and calculations from Figure 3 were performed with the adjusted 20x MCMC accepts setting in place (see Figure 4). No LRs of 0 were observed, and although some large differences between point estimate and HPD LR were still observed, their incidence and magnitude were greatly reduced (see Appendix A Figure A2).

Figure 4. Four-person compound LR plots from Figure 3, recalculated based on interpretations run at 20x MCMC accepts, with: (a) 2 true contributors, (b) 3 true contributors, and (c) 4 true contributors in the numerator propositions.

Another difference in the conditioned 4-person LR plot is the reversal of the trend observed for increasing numbers of contributors in the LR numerator; whereas false negatives increased along with the number of contributors in the numerator, false negatives decreased with increased amounts of conditioning. This effect would be expected, given that increased conditioning correlates with decreased mixture complexity.

3.7. Conditioned LR Calculations: 4-Person Mixtures at 20x Increased MCMC Accepts

The interpretations and LR calculations from Figure 5 were performed again at 20x MCMC accepts. As with the compound LR plots, the increased accepts effectively addressed the false negative LRs observed with the default accepts; however, they were not completely eliminated (see Figure 6). Examining the EPG data for the conditioned interpretations with false negative LRs at 20x MCMC accepts (see Appendix A Figure A3) demonstrates a biological cause for these false negatives. In both cases, a very low level contributor produced unexpectedly high peak data at one locus, given the targeted mixture proportions, and the high peak data was produced for an allele exclusive to that low level contributor.
is that the false negatives occurred along both axes; that is, in some instances, unconditioned compound LRs were zero and the corresponding conditioned LRs were non-zero, whereas in other cases unconditioned compound LRs were non-zero and the corresponding conditioned LRs were zero.

Figure 5. Plots of 4-person log(LR)s based on conditioned interpretations v. corresponding log(LR)s based on unconditioned interpretations run at the default number of MCMC accepts. The number of conditioned true contributors are (a) 1, (b) 2, and (c) 3. False negative LRs (LRs of 0, plotted at $-30$ and circled in red) are observed along both the X and Y axes for all but the interpretations conditioned on three true contributors (c).

Under these amplification conditions, STRmix heavily favored sharing the low level contributor peak. Further exacerbating the drive toward an incorrect inference was the fact that in both cases, the low level contributor in question was conditioned upon. Since data at other loci were consistent with an overall low contribution to the mixture, sharing of the peak appeared to be the most consistent with the data, despite this not being the ground truth.

Since the reason for the false negative LRs from the conditioned interpretations is an aberrant PCR amplification, the issue would be unlikely to be effectively addressed by increasing MCMC accepts. In other words, it is the unexpected height of the low level contributor peaks, not the length of the genotype combination list that is causing genotype combinations inconsistent with ground truth to be favored by the software.
contributor produced unexpectedly high peak data at one locus, given the targeted mixture proportions, and the high peak data was produced for an allele exclusive to that low level contributor.

Figure 6. Four-person conditioned LR plots from Figure 5, recalculated based on interpretations run at 20x MCMC accepts. The number of conditioned true contributors are (a) 1, (b) 2, and (c) 3. A limited number of false negative LRs (LRs of 0, plotted at −30 and circled in red) are still observed with 20x accepts.

4. Discussion

The data indicate that compound and/or conditioned LRs can be expected to be greater than simple LR products in a vast majority of cases, whether the mixture in question is at a low or high level of DNA input. Furthermore, the typical amount of increase was dependent on both the DNA input level as well as the mixture composition.

The spread of the data around the equivalence line in each plot is small near the origin, then expands in the middle of the x-axis range before contracting again near the upper boundary of the x-axis range. This indicates that maximum information gain from compound LR calculation and/or conditioning occurred at input amounts in the middle of the studied input range, which on a per-contributor basis is ~1 ng for our dataset. Such a maximum is sensible, given that low-level mixtures contain less information in general, and high-level mixtures are more definitively analyzed and so have less information to gain from compound LR calculation or conditioning than mixtures at a moderate template levels.

The impact of mixture composition on the magnitude of LR change for both the compound and conditioned LRs is most apparent when the constraints placed on the genotype combinations relevant to the LR calculations are maximized; that is, when N true contributors are placed in the numerator of an N-contributor mixture and when N-1 contributors are conditioned upon in the interpretation of an N-contributor mixture. Figure 1a,c and Figure 2a,c, for the 2- and 3-person mixtures, are the most striking in this
regard; mixtures with contributors in the same proportions have regular, distinct datapoint groupings, and the trends for the mixtures with more inherent ambiguity (e.g., 1:1, 1:1:1 and 100:100:4) are translated further upward than the trends for the more easily resolved mixture proportions. The corresponding Figures 4c and 6c, for the 4-person mixtures, have less clear demarcation of the datapoint groupings for the ambiguous mixture proportions, which is not surprising given their level of complexity.

Because not every compound LR was higher than its simple LR counterpart, our expectations about compound LR magnitude are most straightforwardly defined in terms of upper and lower bounds. The distribution of log(LR) differences between all compound LRs in this study and their corresponding simple LR products ranged from a minimum of ~−2.7 to a maximum of ~28.3, with the probability density peaking at approximately 0.5 (see Figure 7a).

![Figure 7](image_url)

**Figure 7.** Probability density distributions for the log difference between all compound LRs in the study and their corresponding simple LR products (a), as well as all conditioned LRs in the study and their unconditioned counterparts (b). An overlay of the two distributions (c) highlights their similarity.

Furthermore, in a similar fashion to the compound/LR product plots, the conditioned/unconditioned LR plots show that conditioning does not always lead to a higher LR. The distribution of log(LR) differences between conditioned LRs and their corresponding unconditioned counterparts ranged from a minimum of ~−3.2 to a maximum of ~27.7, with the probability density again peaking at approximately 0.5 (see Figure 7b).

The similarity of the log(LR) difference distributions for the compound/simple and conditioned/unconditioned comparisons (see Figure 7c) points to parallel mechanisms of action. In fact, because the weights produced by STRmix for the various genotype
combinations can be characterized as either unconditional or conditional probabilities, some connections can be drawn between the effects of compound LRs and conditioned LRs.

For instance, if we let Pr(A) be the weight assigned to a genotype combination including contributor A, we could also let Pr(A|B) be the weight assigned to the same genotype combination if the interpretation were conditioned on contributor B. Alternatively, if we place both contributor A and contributor B in the numerator of the LR for a compound LR calculation, we could let Pr(A,B) be the weight assigned to the combination of contributor A and contributor B together. However, we could also deconstruct Pr(A,B) into Pr(B) * Pr(A|B).

Since the same type of conditional probabilities come into play in calculating both compound LRs and LRs based on conditioned interpretations, this explains why the information gain from performing these types of calculations is similar. In fact, it could be said that calculating a compound LR is performing a type of conditioning, the difference being that the conditioning is applied at the point of LR calculation rather than at the point of interpretation.

5. Conclusions

The relative quantitative results of this LR study provide a basis for expecting additive behavior of LR exponents at all DNA input amounts. While not all of the compound LRs were greater than their simple LR product equivalents, a vast majority were equal or greater, and trends in the magnitude of the effect were seen in relation to both template DNA amount and mixture composition. We also determined that the substantial increases in analysis complexity inherent to higher-order mixture interpretations may necessitate increased MCMC accepts in order to avoid false negative compound LRs. In addition, the similarity of the log(LR) distributions for compound LRs and conditioned LRs relative to their simple LR product and unconditioned counterparts, respectively, derives from the similar information gain that occurs from either conditioning during the interpretation or the LR calculation of a compound LR.

Supplementary Materials: Full tables of the LRs calculated for the study, as well as a table of anonymized genotypes for the mixture donors that can be used to assess allele overlap, can be downloaded at: https://www.mdpi.com/article/10.3390/genes13112031/s1, Supplementary Table S1: 2PM LRs; Supplementary Table S2: 3PM LRs; Supplementary Table S3: 4PM LRs (default accepts); Supplementary Table S4: 4PM LRs (20x accepts); Supplementary Table S5: Mixture Donor Profiles.

Author Contributions: Conceptualization and methodology, K.D., J.W., S.M. and D.C.; writing—original draft preparation, K.D.; writing—review and editing, K.D., J.W., S.M. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the two anonymous peer reviewers for their valuable constructive feedback. The quality of the finished product was substantially enhanced as a result of their contributions.

Conflicts of Interest: The authors declare no conflict of interest.
Appendix A

Figure A1. Plots of log(HPD LR) v. log(point estimate LR) for 2-person (a) and 3-person (b) compound LR calculations.

Figure A2. Plots of log(HPD LR) v. log(point estimate LR) for 4-person compound LR calculations at the default and 20x default MCMC accepts. The frequency of substantial drops in HPD increased with increasing numbers of contributors in the LR numerator (a), in parallel with the incidence of compound LR false negatives, and these occurred less frequently with the 20x MCMC accept setting (b).
Figure A3. EPG data for conditioned mixtures with false negative LRs at 20x MCMC iterations. In both cases (see (a,b) below), unexpectedly high amplifications of peaks exclusive to one minor contributor at these loci led to incorrect inferences about peak sharing, which were exacerbated by conditioning. The black dotted line in each figure indicates the approximate expected height of the observed peak given exclusive attribution to the ground-truth contributor, based on per-contributor summary statistics in the STRmix Interpretation Report. This demonstrates visually why the software disfavored attributing the full height of each peak to its single true contributor.

References
1. Evett, I.W. What is the probability that this blood came from that person? A meaningful question. J. Forensic Sci. Soc. 1983, 23, 35–39. [CrossRef]
2. Collins, A.; Morton, N.E. Likelihood ratios for DNA identification. Proc. Natl. Acad. Sci. USA 1994, 91, 6007–6011. [CrossRef] [PubMed]
3. Balding, D.J.; Nichols, R.A. DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. Forensic Sci. Int. 1994, 64, 125–140. [CrossRef] [PubMed]
4. National Research Council. The Evaluation of Forensic DNA Evidence, 2nd ed.; Committee on DNA Forensic Science: An Update; National Academy Press: Washington, DC, USA, 1996.
5. Weir, B.S.; Triggs, C.M.; Starling, L.; Stowell, L.I.; Walsh, K.A.J.; Buckleton, J.S. Interpreting DNA mixtures. J. Forensic Sci. 1997, 42, 213–222. [CrossRef] [PubMed]
6. Evett, I.; Weir, B. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists; Sinauer Associates: Sunderland, MA, USA, 1998.
7. Gill, P.; Brenner, C.H.; Buckleton, J.S.; Carracedo, A.; Krawczak, M.; Mayr, W.R.; Morling, N.; Prinz, M.; Schneider, P.M.; Weir, B.S. DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 2006, 160, 90–101. [CrossRef] [PubMed]

8. Cook, R.; Evett, I.W.; Jackson, G.; Jones, P.J.; Lambert, J.A. A hierarchy of propositions: Deciding which level to address in casework. *Sci. Justice* 1998, 38, 231–240. [CrossRef]

9. Evett, I.W.; Jackson, G.; Lambert, J.A. More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Sci. Justice* 2000, 40, 3–10. [CrossRef]

10. Buckleton, J.S.; Bright, J.-A.; Taylor, D.A.; Evett, I.; Hicks, T.; Jackson, G.; Curran, J.M. Helping formulate propositions in forensic DNA analysis. *Sci. Justice* 2014, 54, 258–261. [CrossRef] [PubMed]

11. Gill, P.; Hicks, T.; Butler, J.M.; Connolly, E.; Gusmao, L.; Kokshoorn, B.; Morling, N.; van Oorschot, R.A.H.; Parson, W.; Prinz, M.; et al. DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part I: Evaluation of DNA profiling comparisons given (sub-)source propositions. *Forensic Sci. Int. Genet.* 2018, 36, 189–202. [CrossRef] [PubMed]

12. Gittelson, S.; Kalafut, T.; Myers, S.; Taylor, D.; Hicks, T.; Taroni, F.; Evett, I.W.; Bright, J.A.; Buckleton, J. A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in an adversarial environment. *J. Forensic Sci.* 2016, 61, 186–195. [CrossRef] [PubMed]

13. Forensic Science Regulator, DNA Mixture Interpretation, FSR-G-222. Available online: https://www.gov.uk/government/publications/dna-mixture-interpretation-fsr-g-222 (accessed on 22 September 2022).

14. Hicks, T.; Kerr, Z.; Pugh, S.; Bright, J.A.; Curran, J.; Taylor, D.; Buckleton, J. Comparing multiple POI to DNA mixtures. *Forensic Sci. Int. Genet.* 2021, 52, 102481. [CrossRef] [PubMed]

15. Buckleton, J.; Taylor, D.; Bright, J.A.; Hicks, T.; Curran, J. When evaluating DNA evidence within a likelihood ratio framework, should the propositions be exhaustive? *Forensic Sci. Int. Genet.* 2021, 50, 102406. [CrossRef] [PubMed]

16. Taylor, D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Sci. Int. Genet.* 2014, 11, 144–153. [CrossRef] [PubMed]

17. Butler, J.M.; Iyer, H.; Press, R.; Taylor, M.K.; Vallone, P.M.; Willis, S. DNA Mixture Interpretation: A NIST Scientific Foundation Review. NISTIR 8531-DRAFT. 2021. Available online: https://www.nist.gov/forensic-science/dna-mixture-interpretation-nist-scientific-foundation-review (accessed on 22 September 2022).

18. Hicks, T.; Biedermann, A.; de Koeijer, J.A.; Taroni, F.; Champod, C.; Evett, I.W. The importance of distinguishing information from evidence/observations when formulating propositions. *Sci. Justice* 2015, 55, 520–525. [CrossRef] [PubMed]

19. President’s Council of Advisors on Science and Technology. Report to the President—Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. 2016. Available online: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed on 22 September 2022).

20. President’s Council of Advisors on Science and Technology. An Addendum to the PCAST Report on Forensic Science in Criminal Courts. 2017. Available online: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf (accessed on 22 September 2022).

21. Kelly, H.; Bright, J.A.; Coble, M.D.; Buckleton, J.S. A description of the likelihood ratios in the probabilistic genotyping software STRmix™. *WIREs Forensic Sci.* 2020, 2, e1377. [CrossRef]

22. Gettings, K.B.; Borsuk, L.A.; Steffen, C.R.; Kiesler, K.M.; Vallone, P.M. Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Sci. Int. Genet.* 2018, 37, 106–115. [CrossRef]

23. Bright, J.A.; Taylor, D.; McGovern, C.; Cooper, S.; Russell, L.; Abarno, D.; Buckleton, J. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. *Forensic Sci. Int. Genet.* 2016, 23, 226–239. [CrossRef] [PubMed]

24. Gelman, A.; Carlin, J.; Stern, H.; Rubin, D.; Dunson, D.; Vehtari, A. *Bayesian Data Analysis*, 3rd ed.; Electronic Edition for Non-Commercial Purposes Only; Chapman & Hall/CRC: Boca Raton, FL, USA, 2013. Available online: http://www.stat.columbia.edu/~gelman/book/BDA3.pdf (accessed on 27 October 2022).