Research paper

The a posteriori probability of the number of contributors when conditioned on an assumed contributor

Catherine M. Grgicak [a,b,*], Ken R. Duffy [c], Desmond S. Lun [b,d,e]

[a] *Department of Chemistry, Rutgers University, Camden, NJ 08102, USA*
[b] *Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA*
[c] *Hamilton Institute, Maynooth University, Ireland*
[d] *Department of Computer Science, Rutgers University, Camden, NJ 08102, USA*
[e] *Department of Plant Biology, Rutgers University, New Brunswick, NJ 08901, USA*

ABSTRACT

Forensic DNA signal is notoriously challenging to assess, requiring computational tools to support its interpretation. Over-expressions of stutter, allele drop-out, allele drop-in, degradation, differential degradation, and the like, make forensic DNA profiles too complicated to evaluate by manual methods. In response, computational tools that make point estimates on the Number of Contributors (NOC) to a sample have been developed, as have Bayesian methods that evaluate an A Posteriori Probability (APP) distribution on the NOC. In cases where an overly narrow NOC range is assumed, the downstream strength of evidence may be incomplete insofar as the evidence is evaluated with an inadequate set of propositions.

In the current paper, we extend previous work on NOCIt, a Bayesian method that determines an APP on the NOC given an electropherogram, by reporting on an implementation where the user can add assumed contributors. NOCIt is a continuous system that incorporates models of peak height (including degradation and differential degradation), forward and reverse stutter, noise, and allelic drop-out, while being cognizant of allele frequencies in a reference population. When conditioned on a known contributor, we found that the mode of the APP distribution can shift to one greater when compared with the circumstance where no known contributor is assumed, and that occurred most often when the assumed contributor was the minor constituent to the mixture.

In a development of a result of Slooten and Caliebe (FSI:G, 2018) that, under suitable assumptions, establishes the NOC can be treated as a nuisance variable in the computation of a likelihood ratio between the prosecution and defense hypotheses, we show that this computation must not only use coincident models, but also coincident contextual information. The results reported here, therefore, illustrate the power of modern probabilistic systems to assess full weights-of-evidence, and to provide information on reasonable NOC ranges across multiple contexts.

1. Introduction

In the forensic setting, evidentiary data is excised from a substrate by a series of laboratory procedures designed to extract and purify DNA. The DNA extract is then, typically, partitioned into one fraction that undergoes amplification and one fraction that is stored. For the portion that is amplified, a set of tens of short tandem repeats (STRs) are subjected to polymerase chain reaction (PCR), resulting in billions of fluorescently tagged synthesized fragments. The fluorescence is sized and binned into STR categories, which are reported as the STR alleles at a given locus. Since each person is expected to carry at most two STR

alleles per locus, a perfect system would render only one (i.e., homozygous) or two (i.e., heterozygous) peaks per person. For samples with only one or two contributors and processed within a laboratory system with a very high signal-to-noise, limited artifacts, and no sampling effects, a sufficient method to estimate the number of contributors (NOC) is to apply analytical and stutter ratio thresholds, count the number of peaks, divide by two and round up. This is often referred to as the Maximum Allele Count (MAC) estimate.

Forensic samples, however, contain DNA from an unknown possibly large number of unidentified contributors, where the number of DNA copies from each are also unknown. Even if the laboratory system was

analytically perfect, counting the number of detected peaks to surmise the number of contributors (NOC) often leads to underestimation when the samples are complex [1,2]. As is the case in forensic operations, the signal is further obfuscated by allele drop-out [3], signal filters that promote allele drop-out [4], and large contributor numbers [5]. Thus, qualitatively or subjectively (i.e., by visual inspection) establishing if there are adequate imbalances to justify one or two fewer or greater contributors to the MAC estimate are difficult given the explosion in the genotype combinations that may explain the evidence well [6]. In the worst case, subjectively inferring the NOC that explains the evidence may have the unwanted consequence of excluding reasonable, though subjectively unrecognized, NOC assignments from consideration requiring a re-evaluation post suspect-review, which has associated issues [7].

Another source of profile infidelity is the tendency for STR loci to produce stutter products whose relative signal increases with respect to the alleles as the number of DNA copies decreases [8–10]. The tendency for low-copy samples to produce high relative stutter signal complicates the NOC interpretation because it counteracts the effects of drop-out in that it suggests an additional contributor need be considered as a plausible proposition [11,12].

Whatever the source of signal perturbance, there are two competing, though important, factors that make peak counting ill-suited to assigning the number of contributors: that of extraneous signal from noise and artifacts that bear more weight as the template mass decreases; and that of signal-loss which also bears more weight as the template mass decreases. The chance of each occurring simultaneously in a profile while considering the frequency of the STRs in the population can be taken into consideration by determining the probability that the number of contributors, N, to the evidence, E, is n given the defense's hypothesis: $P(N = n|E, H_d)$. In Bayesian systems such as NOCIt [11,13], this quantity corresponds to the a posteriori probability of n contributors, APP(n), given the evidence and the defense's hypothesis.

In complementary work, applying reasonable assumptions Slooten and Caliebe [14] demonstrated that a Likelihood Ratio (LR), which here we call fullLR, can be evaluated where the NOC is treated as a nuisance variable rather than the commonly used approximation that solely considers a single "best" estimate of the NOC. That is, for electropherogram, E, with N unknown contributors, H_p being the prosecution's hypothesis and H_d being the defense's hypothesis, if it is assumed that the a priori distribution of the number of contributors is the same under both hypotheses in the absence of any data, i.e., $P(N = n|H_p) = P(N = n|H_d)$ for all n, and that the posterior probability of the NOC given the defense's hypothesis is positive for all possible NOCs, $P(N = n|E, H_d) > 0$ for all n, then the LR that treats the NOC as a nuisance variable satisfies

$$fullLR(E) = \sum_n \left(\frac{P(E|H_p, N = n)}{P(E|H_d, N = n)} P(N = n|E, H_d) \right)$$
$$= \sum_n \text{LR}(E|N = n) P(N = n|E, H_d)$$

Their finding was reduced to practice in Ref. [11] where the fullLR was calculated using NOCIt's APP(n) coupled with the computational tool CEESIt that evaluates the likelihood ratio given there are n contributors, LR($E|N = n$) [15]. In that work it was demonstrated that the fullLR can be determined without application of an AT by using NOCIt's APP(n) to articulate a range of n that describe the data. Unsurprisingly, the n associated with APP(n) values near 0 do not need to be included to obtain a precise computation since they are implausible and it is sufficient to renormalize APP(n) accordingly.

Building on the results of [11], we continue by exploring the effects of case-related information. That is, we assess the effect on the APP when conditioning on an assumed contributor with a known genotype. This has operational implications since computing the fullLR conditioned on case information, I, requires the APP be conditioned on I also. Following the derivation from [11], we see that

$$\text{fullLR}(E|I) = \frac{P(E|I, H_p)}{P(E|I, H_d)}$$
$$= \sum_n \frac{P(E|I, H_p, N = n)P(N = n|I, H_p)}{P(E|I, H_d)} \frac{P(E, N = n|I, H_d)}{P(E|I, H_d, N = n)P(N = n|I, H_d)}$$
$$= \sum_n \frac{P(E|I, H_p, N = n)}{P(E|I, H_d, N = n)} \cdot \frac{P(N = n|I, H_p)}{P(N = n|I, H_d)} \cdot P(N = n|E, I, H_d)$$
$$= \sum_n \text{LR}(E|I, N = n) \text{APP}(n|I),$$

where: E is the evidence (i.e., the electropherogram); H_p and H_d are the prosecution and defense hypotheses, respectively; $\text{LR}(E|I, N = n) \triangleq P(E|I, H_p, N = n)/P(E|I, H_d, N = n)$ is the likelihood ratio given case information I (such as a specified contributor) and that there are n total contributors; and $\text{APP}(n|I) \triangleq P(N = n|E, I, H_d)$ is the APP conditioned on case information I. In this derivation, we assume that $P(N = n|I, H_p) = P(N = n|I, H_d)$ for all n, i.e., that the prior distribution on the number of contributors before seeing the data is the same under both the prosecution and defense hypotheses.

In what follows, we evaluate the impact of setting I such that we condition on there being: no assumed contributor, I_{no}; one known contributor who happens to be the major donor, I_{maj}; and one known contributor who happens to be a minor donor to the mixture, I_{min}. We evaluate the influence of I on several APP features such as the breath of the distribution, shifts in the APP's mode, sensitivity of the number of plausible NOCs, and the probability assigned to incorrect NOC, $\sum_{n \neq TrueNOC} \text{APP}(n|I)$. The results make clear that it is necessary for the contexts, I, in $\text{LR}(E|I, N = n)$ and $\text{APP}(n|I)$ to be coincident. They also elucidate how the APP distribution concentrates more around the true number, shifts or broadens with contextual information, and whether contextual information has significant bearing on the range of potential NOC values that best describes the evidence.

2. Description of NOCIt's APP and the fullLR when conditioned on a known contributor

The way in which NOCIt calculates the APP conditioned on case information I is similar to the way in which NOCIt calculates the APP without conditioning on an assumed contributor, as previously described [11], which is available in the larger ValiDNA 4.1.5 software suite. The difference is that all probability calculations are now performed in the probability space conditioned on I. Note that N is the total number of contributors, which includes any known contributors assumed in the case information I so, if I assumes one known contributor, then P($N = 0$) = 0.

Briefly, conditioning on I, NOCIt computes:

$$P(E|I, N = n, H_d) = \iint_{(\theta,\lambda) \in (\Delta^{n-1} \times \Omega^{n-1})} P(E|I, \Theta = \theta, \Lambda = \lambda, N$$
$$= n, H\_d) f_\Theta(\theta) f_{\Lambda|\Theta}(\lambda|\theta) d\theta d\lambda$$

where Θ is the n-dimensional vector of DNA mixture proportions; $\Delta^{n-1} = \{(\theta_1, ..., \theta_n) \in \square^n \mid \sum_{i=1}^n \theta_i = 1, \theta_i > 0 \forall i$ is the $(n-1)$-simplex, which defines all possible DNA mixture proportions; Λ is the n-dimensional vector representing the change in mixture proportions for each contributor at a reference length of 200 bp from that at 0 bp owing to degradation; Ω^{n-1} defines all possible values for Λ; f_Θ is the probability density function of Θ; and $f_{\Lambda|\Theta}$ is the conditional probability density function of Λ given Θ. Given Θ and Λ, the electropherogram for each locus l, E_l, is assumed independent of the electropherogram at every other locus. Therefore,

$$P(E_l|I, \Theta = \theta, \Lambda = \lambda, N = n, H_d) = \prod_{l \in L} P(E_l|I, \Theta = \theta, \Lambda = \lambda, N = n, H_d),$$

where L is the set of loci. Lastly, we have

$$P(E_l|I, \Theta = \theta, \Lambda = \lambda, N = n, H_d)$$
$$= \sum_{g, d \in \Gamma^n \times \{0,1\}^{2 \times |L| \times n}} \left\{ \begin{array}{l} P(E_l|I, G = g, D = d, \Theta = \theta, \Lambda = \lambda, N = n) \\ \bullet P(D = d|I, \Theta = \theta, \Lambda = \lambda) \; P_{[f_0]}(G = g|I, H_d) \end{array} \right\},$$

where G is the genotypes of the contributors; $D \in \{0,1\}^{2 \times |L| \times n}$ is a $2 \times |L| \times n$ matrix that represents the drop-out configuration for the alleles of these contributors; and Γ^n is the space of all possible genotypes for n contributors.

If the case information I is that there is a known contributor with genotype g_1, then

$$P(E_l|I, \Theta = \theta, \Lambda = \lambda, N = n, H_d)$$
$$= \sum_{g, d \in \Gamma^n \times \{0,1\}^{2 \times |L| \times n}} \left\{ \begin{array}{l} P(E_l|, G = \{g, g_1\}, D = d, \Theta = \theta, \Lambda = \lambda, N = n) \\ \bullet P(D = d|\Theta = \theta, \Lambda = \lambda) P(G = \{g, g_1\}|I, H_d) \end{array} \right\},$$

where $P(G = \{g, H_d\}$ is the probability of drawing $n - 1$ random contributors with genotype g from the background population.

To calculate the term $P(E_l|G = \{g, D = d, \Theta = \theta, \Lambda = \lambda, N = n)$, we use a probabilistic model of the electropherogram at locus l developed from calibration data of electropherograms of samples with known genotypes as described in Ref. [11]. Apart from this modification to calculate the probability of the electropherogram at each locus $P(E_l|I, \Theta = \theta, \Lambda = \lambda, N = n, H_d)$ with conditioning on an assumed contributor, all other aspects of the NOCIt calculation described in Ref. [11] remain the same.

3. Materials and methods

We calibrated NOCIt with all but 100 of the publicly available 1-Person PROVEDIt samples amplified with the GlobalFiler® Kit and injected on the 3500 Genetic Analyzer for 25 s (i.e., 2611 single source samples). Though the entire PROVEDIt dataset was used, smaller sets of calibration data containing 50 serially diluted samples can adequately calibrate the models [13].

The 100 single-source electropherograms excluded from the calibration data were used as the 1-person test samples. In addition, all 666, 2- to 5- person GlobalFiler™/25s PROVEDIt mixtures were used to evaluate the changes to APP(n|I), as were 49 differentially degraded mixture samples. This is the same dataset as used in Ref. [11]. All information germane to the sample is contained in its name as detailed in Ref. [5]. Briefly, the DNA target masses of the test samples ranged from 0.75 to 0.0078 ng, and the mixture ratios ranged from equal parts from all contributors to a 1:9 ratio between any two contributors. Samples were analyzed with GeneMapper® ID-X at 1 RFU. Non-reproducible artifacts such as spikes and dissociated dyes were manually removed, while peaks associated with incomplete adenylation, pull-up, complex pull-up (i.e., raised baseline) were filtered using the CleanIt module. Details regarding artifact removal are in Ref. [5]. All data were imported into NOCIt as a CSV and the filtered datasets are available on lftdi.com.

The nominal NOC is taken to be the TrueNOC as it is reasonable to assume that each contributor's DNA is represented in the signal since: 1) the RFU signal from a single amplifiable molecule of DNA is resolved from noise at these laboratory and AT conditions [16]; 2) the smallest minor contributor of any mixture sample did not fall below two cell's worth of DNA (NB: some single-source samples were amplified at 0.0078 ng); and 3) all 25 s GlobalFiler™ single-source PROVEDIt samples amplified at 0.016 ng or higher rendered RFU signal at a minimum

of 12 known allele locations, whether they be pristine, degraded or inhibited [5]. Though it is likely that some signal at these locations may be attributed to noise, it is unlikely that all 12 can; thus, we expect some signal from each contributor for all samples, wherein the expectation was corroborated by affirming the LR for true contributors was greater than one for most samples in [12]. (Table 1).

NOCIt Settings were set to the standard settings detailed in [11], using a maximum contributor number setting of 6. NOCIt's APP($n|I$) results for I_{maj} and I_{min} are available in Supplement 1 and 2, respectively.

3.1. Shift in APP(n|I) with known contributor

To assess what impact changes to contextual information has on APP ($n|I$), we interrogate the distribution's similarities, changes to mode $n_{max|I} = \mathrm{argmax}_n \mathrm{APP}(n|I)$, and the shifts in the distribution's support. We note that in cases where there is an equal mixture proportion of contributors (e.g., 1:1:1) the results when conditioning on the first and second contributors are presented under the labels of 'minor' and 'major', respectively.

The similarity between distributions was assessed by employing the Bhattacharyya coefficient (BC) [17],

$$\mathrm{BC}(P, Q) = \sum \sqrt{P(n)Q(n)}$$

where P and Q are two distributions defined on the same range. The BC gives a number between 0 and 1, with 0 occurring when the two distributions have no overlap, and 1 occurring when the two distributions are equal. For each sample, the BC was evaluated for APP($n| I_{no}$) with each of APP($n| I_{maj}$) and APP($n| I_{min}$). When plotted as histograms, a large number of BCs near zero indicate a large number of APP distributions experience significant shifts under distinct contextual propositions.

We consider two supplementary statistics to assess changes in the a posteriori probability distribution on the NOC as a result of contextual information. As previously defined, let $n_{max|I}$ be the mode of the APP when conditioned on the contextual information I. The first statistic is the APP Range, $|\mathrm{APP}(n_{max|Ino}|I_{no}) - \mathrm{APP}(n_{max|Ino}|I)|$, which captures the absolute difference in the probability at the context-free mode by the introduction of the context I. We report the APP Range for the context I being both I_{maj} and I_{min}. For example, if no assumed contributor resulted in a mode of $n_{max|Ino} = 3$ with APP($n_{max|Ino}$) = 0.92 and the context of a known major or minor contributor resulted in APP($n_{max|Ino}|I$)=APP(3$|I$) = 0.24, then the APP Range would be $|0.92 - 0.24| = 0.68$. An APP Range greater than 0.5 necessarily means the mode of the APP has changed as a result of the introduction of the context.

The second statistic is the difference in the APP modes in the absence and presence of an assumed contributor, $n_{max|Imaj} - n_{max|Ino}$ and $n_{max|Imin} - n_{max|Ino}$. Negative values connote the distribution shifting to a smaller n in the context of an assumed contributor, while a positive value suggests the opposite.

We also explored if changing contextual information produces modifications to the size of the support of the APP, i.e., the number of non-negligible plausible NOCs, by directly reporting the number of APP ($n|I$) ≥ 0.001 for each sample.

Lastly, we conducted a multinomial logistic regression and report the P($n_{max|Imaj} - n_{max|Ino} = -1, 0, +1, +2$) for: β, the degree of electropherogram (EPG) sloping (continuous) (see [11] for a description of its determination); the TrueNOC (nominal); and whether the known contributor was one with "High', 'Some' and 'Low' levels of drop-out. This corresponded to the first bin representing the amplification of ca. 1–2 copies of amplifiable DNA, the 'Some' category representing the amplification of ca. 3–4 copies, while the last bin represents assumed contributors donating > 5 DNA copies. Since the limit of detection for

Table 1

Summary of the 815 GlobalFiler® sample set used to validate NOCIt.

| NOC	1	2	3	4	5
Number	100	193	170	186	166
Total Template Mass (ng)	0.5–0.0078	0.75–0.03	0.75–0.045	0.75–0.06	0.75–0.075
Contributor Ratio	N/A	1:1–1:9	1:1:1–1:9:9	1:1:1:1–1:9:9:1	1:1:1:1:1–1:9:9:9:1

our post-PCR method was 1 copy (i.e., all intact alleles that survived the pre-PCR steps are detected) [16], allele drop-out is solely the result of the ability to sample intact alleles, which can be accurately described as in [11]. Thus, the dropout rate for the intact assumed contributor in the 'High' drop-out bin ranges between 6% and 28%. The second bin contains samples with assumed contributors with drop-out rates between 0.3% and 6%, whereas the remaining bin includes samples whose assumed contributor has drop-out rates less than 0.3%.

3.2. APP(n|I) concentrations at n = TrueNOC

To assess if the APP shifts are toward or away from APP(TrueNOC|I)
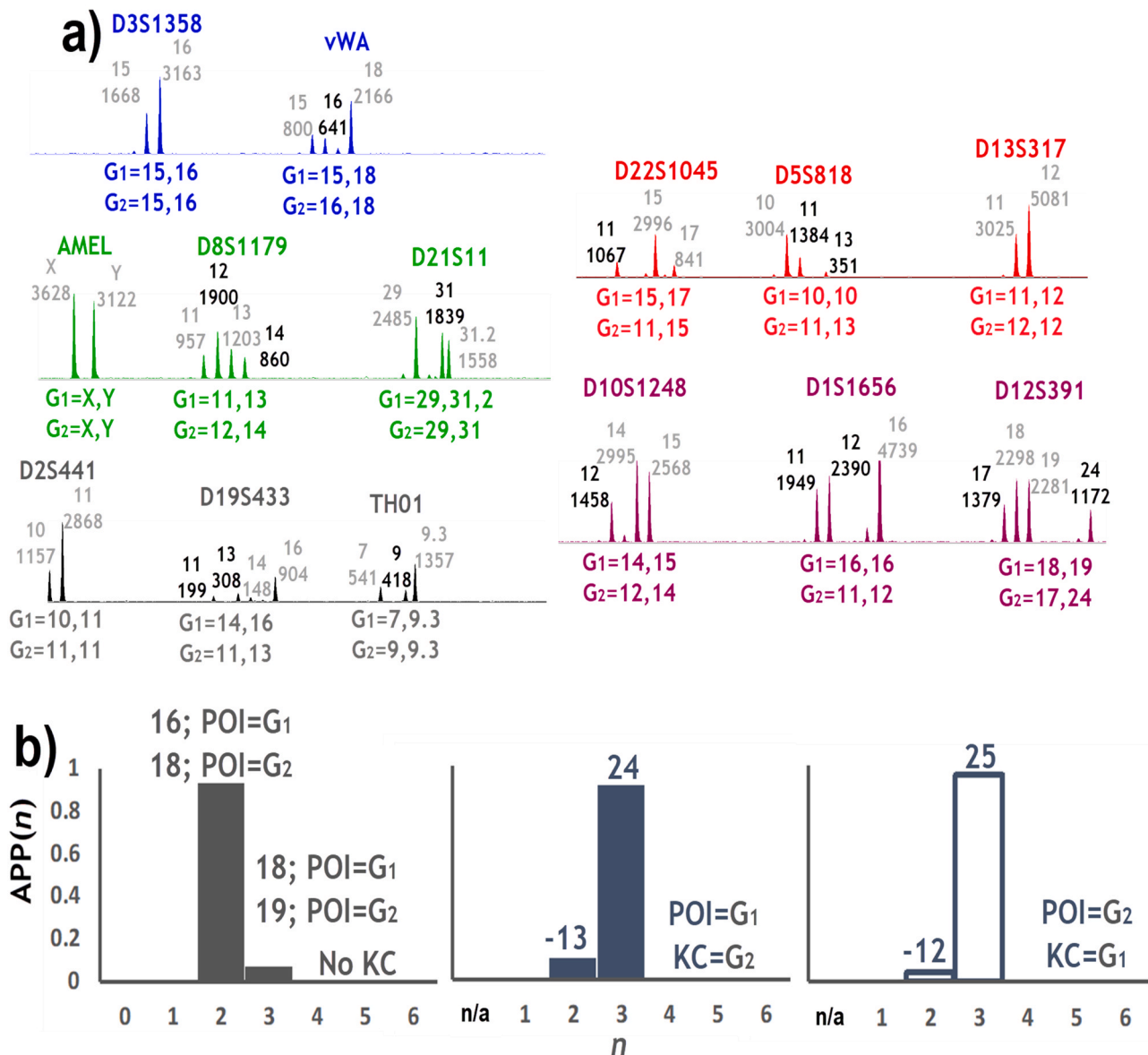
Fig. 1. (a) An electropherogram of 14 of the 22 STR GlobalFiler™ loci of PROVEDIt Sample RD14-0003-31_32–1;1-M2a-0.126GF-Q1.2_03.25s. hid, where this sample was mixed, extracted, quantified, amplified using 29-cycles, and injected for 25 s into a 3500 Genetic Analyzer [5]. (b) Bar charts of NOCIt's $APP(n|I)$ when no known contributor (KC) is assumed and when the known contributor is one or the other donor to the mixture. The values on top of the bars are the $logLR(E|n, I)$ for each POI under each proposition. Also provided are the known genotypes of each contributor.

with changing context, we report the proportion of samples for which APP(TrueNOC|I) $\geq \alpha$, when $\alpha = 10^{-6}$, 0.001, 0.01, 0.1, 0.2, 0.5. As a reference, the true positive rates for the MAC method are also provided, reporting the minimum number of contributors that could explain the evidence. For the MAC method, the minimum NOC was calculated by applying an analytical threshold (AT) of 100 RFU, and filtering peaks in the stutter position that met the following two conditions: that the peak-in-question was one STR repeat shorter than a higher RFU peak to the right of it; and that the RFU intensity of said peak was less than the manufacturer's stutter filter thresholds [18]. We also report the $\sum_{n \neq \text{TrueNOC}} \text{APP}(N = n | I)$ for each sample for I_{no}, I_{maj} and I_{min}. Larger sums mean a shift in a posteriori belief away from the TrueNOC.

4. Results

We first provide illustrative examples demonstrating the way in which contextual information may impact the fullLR, the comparison of hypotheses that treats the NOC as a nuisance variable. Fig. 1a shows 14 of 22 STR loci of PROVEDIt Sample RD14-0003–31_32-1;1-M2a-0.126GF-Q1.2_03.25s.hid, which is a 2-person undamaged 1:1 mixture amplified from 0.126 ng of DNA, and injected for 25 s on a 3500 Genetic Analyzer. The values above the peaks are the STR alleles and their respective peak heights, where the dark text indicates alleles unique to the 2nd contributor. In loci D8S1179 and D5S818 there is obvious peak-height imbalances between the alleles associated with the 2nd contributor. When this sample is assessed with NOCIt we see a shift in the APP's mode from 2 to 3 under different contexts. On top of each bar in Fig. 1b is the logLR for the POI (person-of-interest) under the corresponding n assignment. This example shows the implications of assuming (or not assuming) the same context when ascertaining the range of n for downstream interpretation.

In instances where the casefile does not provide definitive indications of whether one or the other context is the only reasonable assumption, one could consider all plausible contexts by incorporating prior probabilities on I, $P(I = i)$, settled on before seeing the evidence, and using the law of total expectation [19]:

$$E[LR(E)] = \sum_i LR(E|I = i)P\left(I = i\right) \tag{1}$$

where LR$(E|I = i)$ is the previously described LR across all reasonable n within context i. Though mathematically straightforward, determining or settling on P$(I=i)$ is non-trivial for casework applications as it presumes some agreement between parties or justification about the probability that a given individual's genotype is in the data. In the absence any clear indications that a contributor should be assumed part of the mixture, one possibility is to apply $P(I = i) = 0.5$ across both contexts (i.e., that the assumed contributor is just as likely to be present in the data as not to be). In that case, the LR across all contexts is 24 and 25 for POI=G_1 and POI=G_2 (see Table 2).

In another scenario we show that assuming the presence of a specified contributor causes a shift to a lower n_{max}, corresponding to the TrueNOC. In this example, PROVEDIt Sample RD14-0003-49d3S10-0.0156GF-Q1.0_07.25s resulted in n_{max} shifting from 2 to 1 when Sample 49 was assumed present in the data. These exemplars can be taken to represent any electropherogram where the presence of a specified contributor in the data would be reasonable to assume, but not certain, such as an EPG of a sperm-fraction, where the victim's DNA is not unexpected, but also not assured. In Table 3 the summary of results for this sample is shown. Notably, with a known contributor correctly assumed, the fullLR, which treats the NOC as a nuisance variable, for a true non-contributor (PROVEDIt ID 01) transitions from 10^{-3} to 10^{-6}, which makes sense given most of the signal is explained by the assumed contributor (PROVEDIt Sample ID 49) when $n = 2$. In contrast, if only one n was assigned, then LR($E|n = 1$,I_{no}) is 10^{-40} but increases to 10^{-3} for LR($E|$ $n = 2$,I_{no}), demonstrating the marked impact n assignments can have on results with exclusionary influence. Notably, when all scenarios and all plausible n are considered, the overall LR approximates the largest LR of 10^{-3}. In all treatments a P($I=i$) = 0.5 is used as illustration but can be set to any value between 0 and 1.

Note that these examples are not presented as prescriptions suggesting what contexts ought to be considered, nor should they be interpreted as representative of most samples. Instead they serve to demonstrate that the range of n explaining the evidence depends on context and that coupling an n_{max} obtained when not assuming a specified contributor with an LR computation that does assume one could lead to unexpected results. To that end we present a large-scale evaluation of the impact on changing context on APP($n|I$) using the NOCIt system.

In Fig. 2 are stacked plots of APP($n|I$) for all test samples separated by the TrueNOC for I_{no}, I_{maj} and I_{min}. These plots demonstrate the distributions are consistently unimodal, regardless of context. We note that the samples are independently ordered according to their probabilities and, therefore, the mth sample of the first plot is not necessarily the same sample in the corresponding plots. In the case of TrueNOC= 1, there is no minor contributor and, therefore, NOCIt runs for TrueNOC= 1 and I_{min} are not applicable.

In addition to unimodality, another key result from Fig. 2 is that, in general, there is a tendency for the APP modes to shift from lower to higher n in the presence of an assumed contributor. This is more pronounced when the assumed contributor constituted the minor portion of the sample and when the TrueNOC is higher. A possible explanation of this phenomenon is that the addition of the minor assumed contributor does little to explain the EPG, and so a similar number of additional contributors is more likely to explain the data. Barring quantitative assessments of stutter, noise or drop-in, one can envision scenarios that explain such shifts. Take, for example, the scenario where there are two low-level peaks and the assumed contributor is homozygotic at that locus. If the first or second low-level peak is inconsistent with the assumed contributor, the estimate on n would increase.

4.1. Shift in APP($n|I$) with contextual information

We expand on the qualitative observations of Fig. 2 by examining the

Table 2

Computing the overall (across context and n) LR as per [14] and Eq. (1). The APP($N = n|E, I = i$) is from NOCIt and the LR($E|N = n, I = i$) is from CEESIt, two systems using the same underlying model, for the PROVEDIt sample RD14–0003–31_32–1;1-M2a-0.126GF-Q1.2_03.25s.hid. Here the POIs, G_1 and G_2, are PROVEDIt Sample ID 31 and 32, respectively.

| POI | KC | No KC | | | | One KC | | | |
		APP (n = 2|E,I_{no})	LR (E = n = 2,I_{no})	APP (n = 3|E,I_{no})	LR (E = n = 3,I_{no})	APP (n = 2|E,I)	LR (E = n = 2,I)	APP (n = 3|E,I)	LR (E| n = 3,I)
G_1	G_2	0.932	10^{16}	0.067	10^{18}	0.044	10^{-13}	0.956	10^{24}
			fullLR(E = I_{no}) (0.932•10^{16})+ (0.067•10^{18})= 10^{17}				fullLR(E = I) (0.044•10^{-13})+ (0.956•10^{24})= 10^{24}		
				Across context LR (10^{17}•0.5)+ (10^{24}•0.5)= 10^{24}					
G_2	G_1	0.932	10^{18}	0.067	10^{19}	0.094	10^{-12}	0.906	10^{25}
			fullLR(E = I_{no})= (0.932•10^{18})+ (0.067•10^{19})= 10^{18}				fullLR(E|I)= (0.094•10^{-12})+ (0.906•10^{25})= 10^{25}		
				Across context LR (10^{18}•0.5)+ (10^{25}•0.5)= 10^{25}					

Table 3

Computing the overall (across context and *n*) LR as per [14] and Eq. (1). The $APP(N = n|E, I = i)$ is from NOCIt and the $LR(E|N = n, I = i)$ is from CEESIt, two systems using the same underlying model, for the PROVEDIt sample RD14-0003-49d3S10-0.0156GF-Q1.0_07.25s. We test this sample using the POI PROVEDIt Sample ID 01, which is a true non-contributor to this sample.

		No KC				One KC			
POI	KC	APP (n = 1|E,I$_{no}$)	LR (E = n = 1,I$_{no}$)	APP (n = 2|E,I$_{no}$)	LR (E = n = 2,I$_{no}$)	APP (n = 1|E,I)	LR (E = n = 1,I)	APP (n = 2|E, I)	LR (E| n = 2, I)
01	49	0.484	10^{-40}*	0.516	10^{-3}	0.9999	N/A	0.0001	10^{-2}
			fullLR(E = I$_{no}$) $(0.484 \bullet 10^{-40}) + (0.516 \bullet 10^{-3}) = 10^{-3}$				fullLR(E = I) $(0.0001 \bullet 10^{-2}) = 10^{-6}$		
				Across context LR $(10^{-3} \bullet 0.5) + (10^{-6} \bullet 0.5) = 10^{-3}$					

* log(LR)< −40 are reported as − 40. N/A = not applicable, or zero, since the condition is that there is one person that constitutes the mixture and they are assumed present in the data so there cannot be another person, or POI, represented in the data.
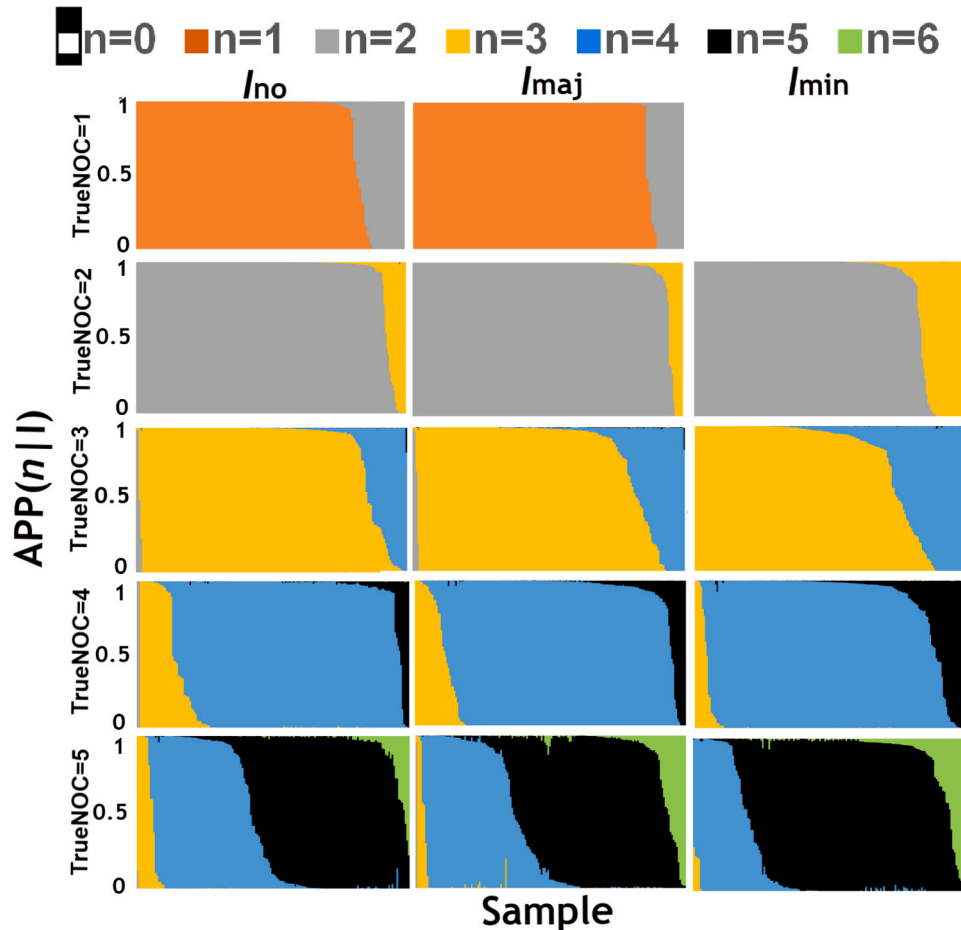

Fig. 2. Stacked plots of APP(n = I) for: I_{no}, when there is no assumed contributor; I_{maj}, when the assumed contributor is the major to the mixture; and I_{min}, when the assumed contributor constituents the smallest portion. Samples are ordered by probability rather than accession number such that the mth sample in the first row is not necessarily the same sample in the two corresponding stacked plots. Note that APP(0) > 0 is only possible when the context is I_{no}.

degree to which the APP shifts and to what extent the APP distribution broadens or narrows when we assume the presence of a specified contributor. Fig. 3 provides pie charts showing the number of APP(n |I) ≥ 0.001 for I_{no}, I_{maj} and I_{min} for the 715, 2- to 5- person mixture samples. Since there is no I_{min} for single-source samples, we exclude the 100 single-source data from Fig. 3. The pie charts indicate that, in general, the number of *n* associated with significant probabilities do not change. Thus, regardless of context, one to three *n* explain the mixtures well.

Though the range of reasonable *n* does not substantively change with conditioning, the histograms of Fig. 3 demonstrate that the APP Range

and BC do. The BC, which is a measure of dissimilarity across distributions, reveals that 12.4% and 21% of the paired comparisons rendered BCs less than 0.5 when APP($n|$ I_{no}) was compared to APP($n|I_{maj}$) and APP($n|I_{min}$) respectively, across the 715, 2– to 5- person samples. Similarly, the APP Range, which is a measure of change to the APP(n) support, illustrates that for 8.4% and 19% of the samples, the n_{max} changed between I_{no} and I_{maj}, and I_{no} and I_{min}, respectively. When we compare these APP Range values to those reported across three NOCIt replicates of I_{no} [11] using standard NOCIt conditions, wherein only 5.4% of the APP Ranges for the 715 mixture samples exceed 0.5, we
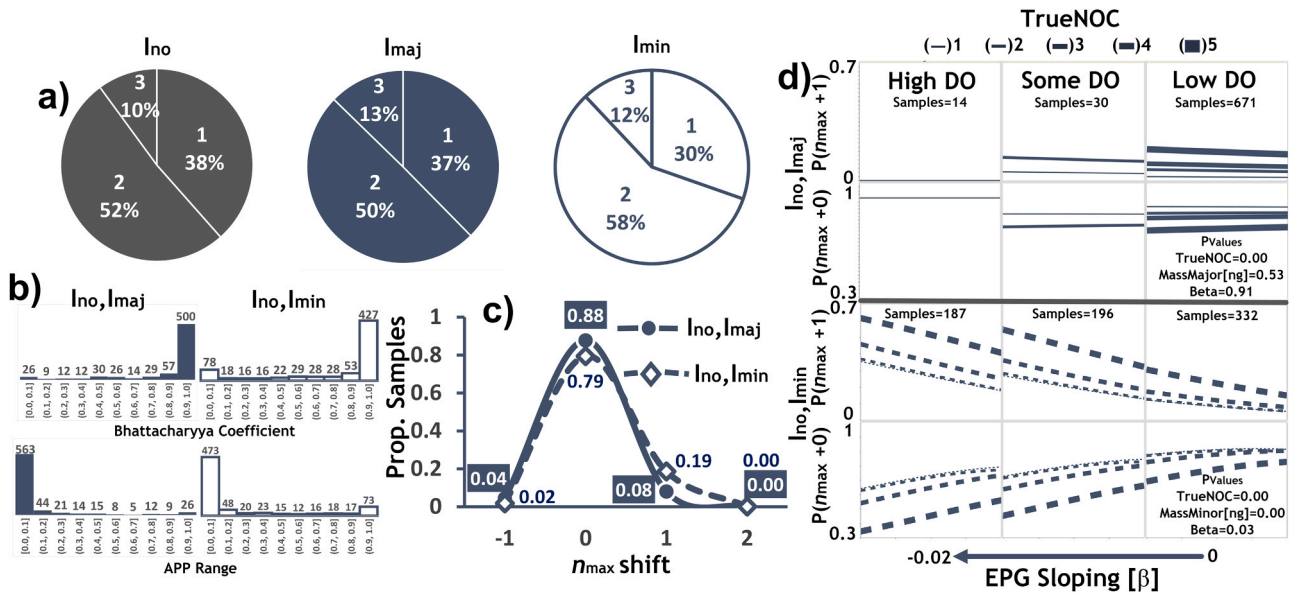
Fig. 3. (a) Pie charts depicting the percentage of samples resulting in one, two or three APP$(n|I) \geq 0.001$ when no specified donor is assumed, $I\!=\!I_{no}$; the assumed contributor is also the major donor to the mixture, $I\!=\!I_{maj}$; and the assumed contributor is the minor constituent, $I\!=\!I_{min}$. No sample exhibited greater than three APPs exceeding 0.001, regardless of context. (b) Histogram of the Bhattacharyya coefficient (BC) obtained between APP$(n|I)$ distributions when no and a major contributor is assumed, and the coefficient when no and a minor contributor is assumed. Also included are histograms of the APP Range with and without background contributor. The numbers on the top of the bars represent the number of samples falling into the corresponding category. (c) The proportion of 715 mixture samples for which $n_{max|I}$ changed when no contributor was assumed, and when a contributor was assumed and was either the major or minor component to the mixture. (d) Line graphs of the probabilities that the APP mode, $n_{max|I}$, shifts to one greater or does not shift when we assume a known contributor to be the major or the minor to the mixture. The probabilities are determined using multiple logistic regression for shifts of -1 to 2, wherein the probabilities for a shift of 0 and $+1$ are shown in the figure owing to the near 0 probabilities for all others. The features of TrueNOC, degree of electropherogram sloping [β], and the masses of the contributors binned into three categories representing high, some and low-levels of allele drop-out (DO) were the independent variables. The number of samples falling into the Low, Some and High DO are also depicted as are the p-values for each feature.

conclude that supplying contextual information impacts the APP and the number of contributors assignments for a significant number of samples, particularly when the assumed contributor is a minor component to the mixture.

To assess the direction of the distribution's shift, we record $n_{max|I}$ for each of the scenarios and report what proportion of n migrate to higher or lower values (Fig. 3c). When a change in $n_{max|I}$ occurred, it almost always occurred to a value that was one greater than originally obtained with $I\!=\!I_{no}$. To further explore, we conduct a multiple logistic regression to determine the probability that n_{max} shifts by $-1, 0, 1$ or 2 when we change I_{no} to I_{maj} or I_{min}. Notably, when the assumed is the major contributor, few samples are in the High to Some dropout (DO) category since, in most cases, the major will donate a large number of DNA copies. The exception are samples for which the total template mass is low and the contributions from each contributor are equal or near equal. Regardless of dropout category, we see that when the assumed contributor is the major one there is, overall, a low probability that contextual information will shift the APP distribution to higher n. This probability increases as the TrueNOC increases and reaches its largest value of 0.2, when the TrueNOC= 5 and the EPG sloping is severe. These trends, however, are significantly more pronounced when the assumed contributor is the minor one, wherein the probability of shifting to higher n is dependent on all three features and reaches its pinnacle at ca. 0.7 when the mass of the minor is low (i.e., high levels of dropout), the EPG sloping is severe, and the True NOC is 5.

4.2. APP(n) concentration at n = TrueNOC

Defining sensitivity as the proportion of samples for which APP (TrueNOC|I) $\geq \alpha$ for given $\alpha > 0$, in Fig. 4 the sensitivity of NOCIt is plotted, and separated by TrueNOC and background information. As in Ref. [11], the sensitivity of NOCIt when no contributor is assumed is high (i.e., $> 82\%$) at all α, except when TrueNOC = 5. Notably, NOCIt improves upon MAC's estimations across all sample complexities for all contextual scenarios. When we employ I_{maj}, the sensitivity tracks well with those of I_{no}. Notably, for TrueNOC= 1 if a contributor is known, then they are necessarily the major so sensitives for I_{min} are not offered. Interestingly, the sensitivity of APP($TrueNOC|I_{min}$) is noticeably lower for TrueNOC= 2 and to a lesser extent for TrueNOC= 3, but substantively improves for TrueNOC= 5. This implies different n might best explain the evidence under distinct propositions, requiring the integration of contextual or known contributors during NOC assessments, which may be difficult to accomplish intuitively in a consistent manner or would require substantiation in the literature.

Also shown in Fig. 4 is the total APP($n|I$) for all n not equal to TrueNOC, where the median $\sum_{n \neq TrueNOC} APP(n)$ is 0.006, 0.008 and 0.014 for I_{no}, I_{maj} and I_{min}, respectively. In a well performing system these values would be near 0. The median associated with I_{min} being slightly larger than the other two contexts suggests that if an assigned contributor is the minor to the mixture, the APP's concentration at $n =$ TrueNOC is decreased but, typically, only by very small levels. The box plots also show that, regardless of context, the majority (i.e., 75%) of the samples resulted in at least 75% of the APP($n|I$) centered at $n =$ TrueNOC.
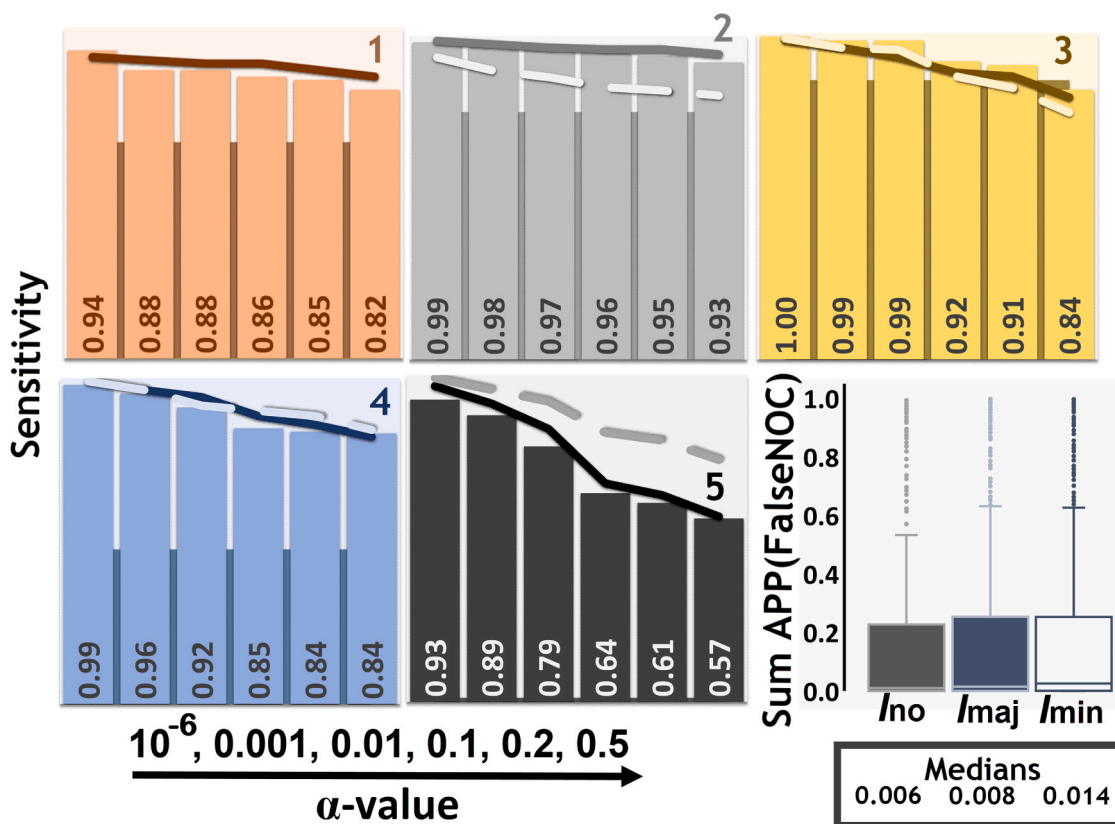
Fig. 4. Bar charts of sensitivity, i.e. correctly including TrueNOC for LR assignments, at α 10^{-6}, 0.001, 0.01, 0.1, 0.2 and 0.5, separated by TrueNOCs. When APP (TrueNOC) ≥ 0.5, then TrueNOC $= n_{\max}$. Included at the base of each bar is the sensitivity value at said α. Tracking along the bar are the sensitivities at each α (——) for I_{maj} and (— —) I_{min}. The MAC sensitivities are provided as the dark shaded regions between the bars. Note, in the case of TrueNOC= 5, MAC renders a sensitivity of 0. The box plot is the $\sum_{n \neq TrueNOC} APP(n|I)$ and the associated median for I_{no}, I_{maj} and I_{min} across the 715 mixture samples.

5. Discussion

Discussions and work regarding the methods by which n (or its range) are chosen are ongoing [14,20–22], where some take a single-step approach and evaluate the evidence across a pre-set range of n (typically 1–4) as described in Ref. [22] and discussed in Ref. [23], while others describe methods that are reliant upon user-specified n, requiring a pre-evaluation step. If the latter approach is utilized the pre-evaluation step may take many forms ranging from subjective evaluations, MAC, tools that aim to correctly classify a single n [12] or machine learning tools that provide probabilities that a given n explains the evidence [24,25], while others have discussed evaluating this within a Bayesian paradigm [11,13,26,27]. Notably, this problem extends beyond the forensic domain [28], which implies broad scientific interest in estimating the number of contributors to data.

Whatever the domain, the APP may be used to inform the scientist to the number of contributors that explain the evidence well, and the APP distribution can be used to do this in several ways. One approach, for example, is to select a range of n from the APP such that it surpasses some predefined α. This approach ensures all n associated with 'significant' APP(n) are considered and is the method taken in Ref. [11] and discussed here. An alternative use of the APP($n|I$) is to define a threshold β to exclude the largest set of $n \in N_e$ such that $\sum_{n \in N_e} APP(n) \leq \beta$. In other words, one ensures that the cumulative probability of all excluded n does not exceed β. This approach is discussed in Ref. [20] within the context of LRs. Other uses of the APP($n|I$) exist, such as selecting $n_{\max}$ as the single number of contributors that best explain the data. Though a straightforward approach, this work suggests that for complex scenarios and mixtures, determining the range of numbers of contributors that

effectively describe the evidence holds merit.

In most LR software, it is necessary to pick one or a very narrow range of n to evaluate the LR. The issue, therefore, is one of adequately choosing the n range for this purpose and by considering the relevant context while doing so. Notably, in current practice this is most often accomplished by the MAC method or by the analyst subjectively, by visual inspection, determining if there are adequate peak height imbalances to justify one (or two) additional or fewer contributors despite the explosion of genotype and peak height combinations that might explain the data [6] or publications that demonstrate MAC's insensitivity [1,3].

This work supplements the literature by exploring the impact of providing contextual information to the NOC inference. Here, we take the same 815 samples in Ref. [11] and run them under the assumption that all contributors are unrelated and there is one assumed contributor, wherein the assumed was either the major or smallest donor to the mixture. The results show that in a significant number of cases and in the presence of an assumed contributor, the range of n that explain the evidence shifts to higher n particularly when the assumed contributor constitutes a small portion of the signal. More generally, we demonstrate that when assigning n ranges the same contextual information as considered in the LR is needed. The certainty surrounding whether this can consistently be accomplished over time across analysts through visual inspection requires evaluation since some early literature suggests differences between analyst assignments [29,30], which can in turn differ from the range assigned with computational assistance as, for example, observed for PROVEDIt RD14–0003–44_45_46_47–1;1;4;1-M3a-0.105GF-Q0.8 in Refs. [29,11].

6. Conclusion

We demonstrate that changes to contextual information can impact the APP($n|I$), where the greatest impact was seen when an assumed contributor comprised the minor portion of the mixture. A unimodal APP($n|I$) is observed for all contexts. If the APP($n|I$) did exhibit a shift, it usually shifted to higher n suggesting that if contextual information is assumed in the LR, it also ought to be considered when determining the range of n. If calculating a 'fullLR', the APP($n|I$) and LR($n|I$) must be coincident. This, notably, has implications to forensic operations when contributors are assumed part of the signal.

Funding Disclaimer and Acknowledgments

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2021.102563.

References

[1] M.D. Coble, J.-A. Bright, J.S. Buckleton, J.M. Curran, Uncertainty in the number of contributors in the proposed new CODIS set, Forensic Sci. Int. Genet. 19 (2015) 207–211.

[2] G.M. Dembinski, C. Sobieralski, C.J. Picard, Estimation of the number of contributors of theoretical mixture profiles based on allele counting: does increasing the number of loci increase success rate of estimates? Forensic Sci. Int. Genet. 33 (2018) 24–32.

[3] S. Norsworthy, D.S. Lun, C.M. Grgicak, Determining the number of contributors to DNA mixtures in the low-template regime: exploring the impacts of sampling and detection effects, Leg. Med. 32 (2018) 1–8.

[4] C.A. Rakay, J. Bregu, C.M. Grgicak, Maximizing allele detection: effects of analytical threshold and DNA levels on rates of allele and locus drop-out, Forensic Sci. Int. Genet. 6 (6) (2012) 723–728.

[5] L.E. Alfonse, A.D. Garrett, D.S. Lun, K.R. Duffy, C.M. Grgicak, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Sci. Int. Genet. 32 (2018) 62–70.

[6] P.C. Lynch, R.W. Cotton, Determination of the possible number of genotypes which can contribute to DNA mixtures: non-computer assisted deconvolution should not be attempted for greater than two person mixtures, Forensic Sci. Int. Genet. 37 (2018) 235–240.

[7] I.E. Dror, W.C. Thompson, C.A. Meissner, I. Kornfield, D. Krane, M. Saks, M. Risinger, Letter to the editor— context management toolbox: A Linear Sequential Unmasking (LSU) approach for minimizing cognitive bias in forensic decision making, J. Forensic Sci. 60 (4) (2015) 1111–1112.

[8] K.R. Duffy, N. Gurram, K.C. Peters, G. Wellner, C.M. Grgicak, Exploring STR signal in the single- and multicopy number regimes: deductions from an in silico model of the entire DNA laboratory process, Electrophoresis (2016).

[9] R. Cowell, A unifying framework for the modelling and analysis of STR DNA samples arising in forensic casework, arXiv Appl. (2018).

[10] R.G. Cowell, Computation of marginal distributions of peak-heights in electropherograms for analysing single source and mixture STR DNA samples, Forensic Sci. Int. Genet. 35 (2018) 164–168.

[11] C.M. Grgicak, S. Karkar, X. Yearwood-Garcia, L.E. Alfonse, K.R. Duffy, D.S. Lun, A large-scale validation of NOCIt's a posteriori probability of the number of contributors and its integration into forensic interpretation pipelines, Forensic Sci. Int. Genet. 47 (2020), 102296.

[12] M. Kruijver, H. Kelly, K. Cheng, M.H. Lin, J. Morawitz, L. Russell, J. Buckleton, J. A. Bright, Estimating the number of contributors to a DNA profile using decision trees, Forensic Sci. Int. Genet. 50 (2021), 102407.

[13] L.E. Alfonse, G. Tejada, H. Swaminathan, D.S. Lun, C.M. Grgicak, Inferring the number of contributors to complex DNA mixtures using three methods: exploring the limits of low-template DNA interpretation," (in English), J. Forensic Sci. Artic. 62 (2) (2017) 308–316.

[14] K. Slooten, A. Caliebe, Contributors are a nuisance (parameter) for DNA mixture evidence evaluation, Forensic Sci. Int. Genet. 37 (2018) 116–125.

[15] H. Swaminathan, M.O. Qureshi, C.M. Grgicak, K. Duffy, D.S. Lun, Four model variants within a continuous forensic DNA mixture interpretation framework: effects on evidential inference and reporting, PLoS One 13 (11) (2018), 0207599.

[16] K.C. Peters, H. Swaminathan, J. Sheehan, K.R. Duffy, D.S. Lun, C.M. Grgicak, Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: standardizing the forensic validation process, Forensic Sci. Int. Genet. 31 (2017) 160–170.

[17] A.K. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. 35 (1943) 99–109.

[18] ThermoFisher, "GlobalFiler™ PCR Amplification Kit User Guide," ed, 2016.

[19] D.P. Bertsekas, J.N. Tsitsiklis, Introduction to Probability, Athena Scientific, 2008.

[20] R. Meester, K. Slooten, An epistemic interpretation of the posterior likelihood ratio distribution, Law Probab. Risk 19 (2) (2020) 139–155.

[21] I.W. Evett, S. Pope, Is it to the advantage of a defendant to infer a greater number of contributors to a questioned sample than is necessary to explain the observed DNA profile? Sci. Justice 54 (5) (2014) 373–374.

[22] S. Manabe, C. Morimoto, Y. Hamano, S. Fujimoto, K. Tamaki, Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model, PLoS One 12 (11) (2017), 0188183.

[23] C. Brenner, "How to Decide How Many Contributors in a Mixture," presented at the American Academy of Forensic Sciences, Baltimore, MD, 2019.

[24] C.C.G. Benschop, J. van der Linden, J. Hoogenboom, R. Ypma, H. Haned, Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach, Forensic Sci. Int. Genet. 43 (2019), 102150.

[25] M.A. Marciano, J.D. Adelman, Developmental validation of PACETM: automated artifact identification and contributor estimation for use with GlobalFilerTM and PowerPlex® fusion 6c generated data, Forensic Sci. Int. Genet. 43 (2019), 102140.

[26] H. Swaminathan, C.M. Grgicak, M. Medard, D.S. Lun, NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping, Forensic Sci. Int. Genet. 16 (2015) 172–180.

[27] T. Egeland, I. Dalen, P.F. Mostad, Estimating the number of contributors to a DNA profile, Int. J. Leg. Med. 117 (5) (2003) 271–275.

[28] S.A. Sethi, W. Larson, K. Turnquist, D. Isermann, Estimating the number of contributors to DNA mixtures provides a novel tool for ecology, Methods Ecol. Evol. 10 (1) (2019) 109–119.

[29] J.-A. Bright, K. Cheng, Z. Kerr, C. McGovern, H. Kelly, T.R. Moretti, M.A. Smith, F. R. Bieber, B. Budowle, M.D. Coble, R. Alghafri, P.S. Allen, A. Barber, V. Beamer, C. Buettner, M. Russell, C. Gehrig, T. Hicks, J. Charak, K. Cheong-Wing, A. Ciecko, C.T. Davis, M. Donley, N. Pedersen, B. Gartside, D. Granger, M. Greer-Ritzheimer, E. Reisinger, J. Kennedy, E. Grammer, M. Kaplan, D. Hansen, H.J. Larsen, A. Laureano, C. Li, E. Lien, E. Lindberg, C. Kelly, B. Mallinder, S. Malsom, A. Yacovone-Margetts, A. McWhorter, S.M. Prajapati, T. Powell, G. Shutler, K. Stevenson, A.R. Stonehouse, L. Smith, J. Murakami, E. Halsing, D. Wright, L. Clark, D.A. Taylor, J. Buckleton, STRmix; collaborative exercise on DNA mixture interpretation, Forensic Sci. Int. Genet. 40 (2019) 1–8.

[30] E.R. Roman E. Aranda, K. Mereus, P. Spenser, R. Tontarsk, "DNA Mixture Interpretation Study: Inter- and Intra-laboratory Variation," in *International Symposium on forensic Science Error Management*, Gaithersburg, MD, 2015.