E-ISSN 2250-0944
ISSN 2250-1150
doi: 10.31838/ijprt/16.02.847

Research Article

Quantifying Allelic Dropout in Degraded and Mixed Forensic STR Profiles: An In-Silico Ground Truth Study

Dr. Mahwish Zeb[1*], Dr Nazia Yasmin[2] Dr. Muhammad Yousaf [3], Dr. Muhammad Iqbal[4]
[1]Assistant Professor. Forensic Medicine and Toxicology Department. Ayub Medical College. Abbottabad.
[2]Assistant Professor. Forensic medicine and Toxicology Department. HBS Medical and Dental College. Islamabad.
[3]Forensic Medicine and Toxicology Department, Saidu Medical College. Swat.
[4]Assistant Professor. Forensic Medicine and Toxicology Department. Saidu Medical College. Swat.
Corresponding Author: Dr. Muhammad Iqbal
Assistant Professor. Forensic Medicine and Toxicology Department. Saidu Medical College. Swat.
Email: [1]mahwishasfand@gmail.com, drgul78@gmail.com[2] ,dryousaf71@gmail.com[3], drmiqbal52@gmail.com[4]
Received: 15.09.26, Revised: 08.10.26, Accepted: 03.11.26

ABSTRACT
Allelic dropout in degraded and mixed short tandem repeat (STR) profiles remains one of the most persistent obstacles to reliable forensic DNA interpretation, yet most of the empirical evidence describing it comes from casework or dilution-series data in which the true underlying genotype can only be inferred rather than directly verified. This study used a fully reproducible, controlled in-silico simulation, in which every contributor's genotype and every mixture's composition was fixed and known exactly, to quantify how degradation, contributor imbalance, and the number of contributors affect allelic dropout and the reliability of profile interpretation. A synthetic ten-locus panel covering a range of 120-380 base pairs was simulated under four degradation levels, four two-person mixture ratios (1:1 to 1:10), and one to three contributors, resulting in 8,800 profiles and 261,824 expected-allele observations. Allelic dropout rose from 2.29% under no degradation to 40.61% under high degradation, with an almost eight-fold gradient between the smallest and largest loci at the highest degradation level. Minor-contributor allele recovery fell from 65.96% at a balanced 1:1 ratio to 12.53% at 1:10, and a standard maximum-allele-count estimator of the number of contributors, while accurate for single-source profiles (>99%), succeeded in only 3.50% of severely degraded three-person mixtures. In a multivariable logistic regression, degradation and mixture imbalance showed a significant interaction (likelihood-ratio $x2(1) = 2,279.46$, $p < 0.001$). A strict ground-truth reconstruction analysis further showed that even a balanced two-person mixture yielded a fully recoverable profile in fewer than half of the non-degraded replicates.
The results, which were obtained from a known ground truth, support decades of casework-based forensic genetics research and provide a quantitative, mechanistic understanding of how degradation and mixture complexity interact, further justifying the use of probabilistic, rather than threshold-based, DNA interpretation methods.

Keywords: Short Tandem Repeat Typing, Allelic Dropout, DNA Degradation, DNA Mixture Interpretation, Forensic Genetics, In-Silico Simulation, Probabilistic Genotyping.

INTRODUCTION

DNA profiling based on short tandem repeat (STR) markers has, for three decades, been the principal molecular tool by which forensic scientists connect biological evidence recovered from a crime scene to a specific individual, and its adoption has reshaped criminal investigation, exoneration and disaster-victim identification worldwide (1). The reliability of this tool, however, depends entirely on the analyst's ability to correctly interpret the electropherogram a sample produces, and that interpretation becomes considerably harder whenever the DNA available is scarce, degraded, or shared among more than one contributor. The single most consequential complication is allelic dropout: the failure of one or both alleles at a locus to produce a signal above the laboratory's analytical detection threshold. Since the earliest studies of low-template DNA analysis, dropout has been a subject of study, and the logistic-regression framework developed by Tvedebrink et al (2), and the peak-height-based detection-limit model developed by Gilder et al.(3), have been the conceptual basis for most subsequent studies of dropout.

Degradation compounds the problem. Exposure

to heat, humidity, microbial activity or simply time causes the DNA polymer to fragment into increasingly shorter pieces, and since STR amplification requires an intact template that covers the entire length of the repeat region, longer loci are more susceptible
(4). This dependence on amplicon size was shown directly in a collaborative European exercise, in which identical DNA samples were artificially degraded to different extents and typed independently in several laboratories; all of them observed consistently weaker signal at the longer loci (5), and the same relationship underlies the quantitative PCR degradation index, which is used to estimate sample quality before STR amplification is even attempted (6). Tvedebrink and colleagues later adapted their original dropout model to account explicitly for this fragment-size effect, replacing a single template-quantity proxy with a term that declines exponentially as amplicon size increases (7).

Another source of interpretive difficulty, which is not necessarily related to degradation, is when more than one person provides DNA to the same sample. Early studies by Gill et al. on interpretation rules for very low quantities of DNA demonstrated that when the template is below about 100 picograms, stochastic sampling effects during PCR amplification become more likely to result in the loss of alleles from a minor contributor, misinterpretation of the alleles as noise, or complete masking of the alleles by the peaks of a major contributor (8). The International Society of Forensic Genetics later formalised much of this concern in its recommendations on mixture interpretation, which introduced the major/minor contributor terminology and warned analysts of the specific risks associated with imbalanced mixtures (9). A more complicated issue, aside from dropout, is that the number of contributors to a mixture is not always known from the profile. Buckleton et al. showed that uncertainty in this single input parameter carries through every subsequent likelihood calculation and can substantially alter the strength of the evidence that is finally reported (10). The forensic genetics community has responded to these challenges by moving away from simple, rule-based interpretation towards increasingly quantitative and probabilistic methods. For example, Haned et al. showed that estimating the true number of contributors is handled measurably better by maximum-likelihood approaches than by the maximum-allele-count heuristic that has long been used in casework (11), and mixture deconvolution itself has progressed from qualitative peak-presence rules to fully probabilistic genotyping systems, such as TrueAllele®, which was validated by Perlin et al. using quantitative computer modelling of two-person mixtures (12), and STRmix™, whose continuous interpretation model is built around explicit peak-height and stutter models fitted directly to validation data (13). A recent comprehensive review by Gill et al. surveys the current generation of these systems side by side (14), and the International Society of Forensic Genetics has issued formal recommendations for incorporating dropout and drop-in probabilities into likelihood-ratio calculations within exactly this probabilistic framework (15).

Almost all of the empirical work underpinning these models, however, has necessarily been built from dilution series or casework profiles in which the true underlying genotype, exact template quantity and degradation history can only be inferred, never directly verified. Haned et al. argued explicitly that this makes purely empirical, dilution-based dropout models "over-simplistic", because they cannot capture the full range of stochastic variability that real samples display, and proposed controlled simulation as a necessary complement to experimental validation (16). This argument motivated the development of open simulation tools such as the Forensim package for R, one of the first publicly available resources for generating simulated single-source and mixture STR data for methodological research (17).

The need for this kind of controlled, ground-truth-anchored evidence has, if anything, grown more pressing rather than less. Butler's most recent survey of advances in forensic DNA typing documents a field still actively refining how degraded and low-template samples are handled (18), and the National Institute of Standards and Technology's 2024 scientific foundation review of DNA mixture interpretation identifies the interpretation of complex, degraded mixtures as an area where further systematic evidence is still needed (19). Machine-learning methods are now being proposed as a further tool for exactly this problem, with a 2024 critical review cataloguing a rapidly growing number of applications to contributor-number estimation and mixture deconvolution (20), while a 2023 case-study comparison of two leading probabilistic genotyping systems on the same challenging low-template mixtures found meaningful differences in their conclusions, underscoring that even the most advanced

current methods are not yet a fully settled, universally validated solution (21).

What is missing from this literature is a single, fully reproducible framework in which degradation level, mixture ratio and contributor number can all be varied simultaneously and independently, and in which every simulated profile carries a genotype that is known exactly rather than inferred – allowing dropout, mixture-interpretation accuracy and full profile reconstruction to be measured directly against ground truth rather than estimated indirectly. The present study addresses this gap with a controlled in-silico simulation designed to answer four linked questions: (i) how does degradation affect allelic dropout, overall and at the level of individual loci; (ii) how does contributor imbalance affect the recoverability of a minor contributor's profile; (iii) how does the number of contributors affect the reliability of interpretation; and (iv) how reliably can a computational method reconstruct the true, known genotype once a profile has been degraded, mixed, or both? The methods used to answer these questions, the results they produced, and their implications for forensic practice are reported in the sections that follow.

MATERIALS AND METHODS

Study Design and Rationale

Allelic dropout the failure of one or both alleles at a locus to produce a peak above the analytical detection threshold is one of the principal obstacles to interpreting STR profiles recovered from degraded or trace biological material. The phenomenon has been studied extensively in casework and validation settings, most influentially through the logistic-regression dropout models introduced by Tvedebrink et al. (2) and the peak-height threshold framework described by Gilder et al.[3], and later extended explicitly to degraded samples by modelling the exponential decline in peak height with increasing amplicon size(7), a pattern first formalized for quantification purposes through the DNA degradation index of Swango et al. (6). These studies are, however, inevitably limited by what can be observed in real dilution series or casework samples, in which the true genotype, the exact template quantity, and the degradation history are seldom known with certainty.

To sidestep this limitation, the present study used a controlled in-silico simulation in which every ground truth the genotype of every contributor at every locus, and the exact composition of every mixture was fixed by design and known exactly, so that the simulated 'observed' profile could be compared directly against it. This approach follows the precedent set by Haned et al., who argued that purely empirical, dilution-based dropout models are "over-simplistic" because they fail to capture the full range of stochastic variability present in real casework, and who proposed simulation as a necessary complement to experimental validation (16). It also follows the broader tradition of open simulation tools for forensic genetics such as the Forensim package for R, which provided some of the first publicly available routines for generating simulated single-source and mixture STR data for methodological research(17). The framework built for this study extends that tradition in a Python implementation that allows full factorial control over degradation level, mixture ratio, and contributor number simultaneously a combination not offered as a single configurable design in existing published tools.

Locus Panel and Population Allele-Frequency Model

Ten autosomal STR loci commonly used in forensic multiplex kits (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D5S818 and FGA) were selected to span a realistic range of amplicon sizes, from 120 bp to 380 bp, since amplicon size is the principal determinant of a locus's vulnerability to degradation-driven dropout[7]. Rather than importing a specific commercial kit's validated population allele-frequency table which would tie the results to one named product a synthetic allele ladder and frequency distribution was generated independently for each locus using a Dirichlet distribution with randomly drawn concentration parameters (uniform on [1.2, 2.2] per allele). This produces frequency profiles that are internally consistent (they sum to one and vary realistically in evenness across loci) without being drawn from, or claiming to represent, any specific named population database. All loci were treated as statistically independent, consistent with the Hardy–Weinberg and linkage-equilibrium assumptions that underlie standard forensic STR interpretation practice mixtures (9). The complete allele ranges and frequency tables can be reproduced from the

fixed random seed recorded in the archived simulation code.

Simulating Single-Source and Mixed Contributor Genotypes

For each simulated profile, the genotype of every contributor at every locus was generated independently by drawing two alleles (with replacement) from the frequency distribution of that locus, which allowed heterozygous and homozygous genotypes to occur in their expected proportions. Mixtures of two or three contributors were then constructed by assigning each contributor a fixed proportion of the total template DNA for example, a nominal 1:3 two-person mixture corresponds to template proportions of 0.75 and 0.25 following standard forensic convention for describing mixture ratios by relative contributor quantity mixtures (9). Four two-person ratios were modelled (1:1, 1:3, 1:5 and 1:10) together with one three-person composition (nominal 3:2:1, template proportions 0.5, 0.3 and 0.2), spanning the transition from balanced to severely imbalanced mixtures encountered in casework.

Modelling Degradation and Peak-Height Attenuation

Degradation was modelled as an exponential decay in expected peak height with increasing amplicon size, the same functional form used by Tvedebrink et al. to extend the logistic dropout model to degraded casework samples (7) and consistent with the fragment-size-dependent decline in amplification efficiency that underlies quantitative degradation indices (6). Four qualitative degradation levels were defined None, Low, Moderate and High corresponding to decay constants k of 0, 0.0012, 0.0030 and 0.0060 per base pair respectively, so that the expected peak-height multiplier for a locus of size s (bp) was exp(−ks). These constants were chosen, prior to any analysis, to produce a graded and forensically plausible dose response rather than fitted to a specific published degradation dataset; this choice is revisited as a limitation in Section 4.8.

Modelling Stochastic Amplification and Allelic Dropout

Two further sources of variability were layered onto the degradation model, reflecting the well-established observation that dropout in real samples is driven jointly by systematic degradation and by stochastic sampling and amplification effects at low template levels (3, 8, 22). First, the total template quantity available for each simulated replicate was drawn from a Gamma distribution (mean 200 arbitrary units, shape parameter 4, giving a coefficient of variation of 0.5) to represent realistic run-to-run variation in the amount of recoverable DNA. Second, an independent Gamma-distributed multiplicative noise term (shape parameter 9, mean 1) was applied to each contributor's amplification efficiency at each locus, representing stochastic PCR/capillary-electrophoresis variation of the kind documented in run-specific detection-limit studies[3]. Homozygous genotypes contributed double intensity at their single allele, and peak heights from different contributors sharing the same allele were additively stacked, mirroring standard electropherogram behavior.

Given the resulting expected peak height H for an allele, the probability of dropout was modelled with a logistic function referenced to a 50 relative fluorescence unit (RFU) analytical threshold, $P(\text{dropout}) = 1 / (1 + \exp((H - T)/\sigma))$, with threshold T = 50 RFU and scale σ = 14 RFU controlling the width of the stochastic zone around the threshold. This is the same general logistic peak-height-threshold family of model validated for casework use by Tvedebrink et al. (2) and subsequently refined and defended against methodological criticism in later work by the same group (7, 23), and it underlies the probabilistic treatment of dropout recommended by the DNA Commission of the International Society of Forensic Genetics for casework use (15). Each expected allele was then independently scored as dropped or observed by drawing against this probability, and the resulting 'observed' profile the pooled set of alleles surviving dropout, per locus was retained alongside the untouched ground-truth genotypes for downstream comparison.

Experimental Design Matrix

Three linked sub-experiments were run using this generative model, matching the four analytical aims set out in Section 1. (i) A degradation sweep generated 400 replicate single-source profiles at each of the four degradation levels (1,600 profiles), which isolated the effect of degradation alone. (ii) A full factorial mixture experiment crossed all four degradation levels with all four two-person mixture ratios (300 replicates per cell, 16 cells, 4,800 profiles), enabling the degradation × imbalance interaction to be tested directly. (iii) A contributor-number

sweep simulated one-, two- and three-person profiles at Moderate and High degradation only (400 replicates per condition, 2,400 profiles), isolating the effect of mixture complexity. In total the design comprised 8,800 simulated profiles across the ten-locus panel, yielding 261,824 individual expected-allele observations for analysis.

Outcome Measures

Five outcome measures were derived from the simulated data. Allelic dropout rate (ADO) was calculated per condition as the proportion of expected alleles not observed. Minor-contributor recovery was calculated, for two-person mixtures, as the proportion of the minor contributor's exclusive alleles (those not shared with the major contributor) that survived dropout. Interpretation accuracy for the number of contributors (NOC) was assessed using the maximum-allele-count heuristic the smallest number of contributors consistent with the largest number of distinct alleles observed at any single locus benchmarked against studies showing this simple estimator's performance relative to maximum-likelihood alternatives (11). A full ground-truth reconstruction rate was defined, strictly, as the proportion of replicates in which every allele of every contributor's true genotype was present somewhere in the pooled observed profile; because this measures only whether the necessary raw material for correct interpretation survived, and not the output of any specific deconvolution algorithm, it represents a theoretical ceiling comparable in spirit to the aims of modern probabilistic genotyping software (14), whose current capabilities and limitations are summarised in recent reviews(18, 19). Finally, locus-specific dropout rates were tabulated against amplicon size to characterise the size-dependent gradient expected under degradation.

Statistical Analysis

Because the primary outcome at the allele level is binary (dropped or observed), a multivariable logistic regression was fitted to all expected-allele observations from the degradation-sweep and mixture-ratio experiments, following the same general logistic modelling approach used to estimate dropout probabilities from peak-height and template covariates in the forensic genetics literature(2, 22, 23). The model included degradation level (coded numerically 0–3 for None–High), amplicon size, the proportion of total template attributable to the allele's carrier, and contributor number as predictors, with results reported as odds ratios with 95% Wald confidence intervals and Wald p-values. A second model added a degradation × template-proportion interaction term, and the two models were compared by likelihood-ratio test to assess whether degradation and mixture imbalance act synergistically rather than additively on dropout risk. Model fit was summarised using McFadden's pseudo-R^2 and the Akaike Information Criterion (AIC).

Software and Reproducibility

The simulation engine, experiment runner and statistical analysis were implemented in Python 3 using NumPy and pandas for data generation and management and statsmodels for logistic regression. All stochastic elements were seeded (master seed 20250925) so that the full dataset, all summary tables, and all figures reported in Section 3 are exactly reproducible from the archived code. No real casework, population-database, or personally identifiable data were used at any stage; every profile analysed in this study is entirely synthetic.

RESULTS

This section reports the results of the controlled simulation study described in Section 2. All figures below were produced directly by the simulation and analysis pipeline; none are illustrative or assumed. The dataset comprised 8,800 simulated STR profiles built from ten synthetic autosomal loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D5S818 and FGA), each assigned a representative amplicon size between 120 and 380 base pairs and a population-like allele-frequency distribution generated from a Dirichlet process. This produced 261,824 individual expected-allele observations across the full design, summarised below.

Characteristics of the Simulated Dataset

Three linked sub-experiments were run. The first varied degradation alone across single-source profiles (four levels None, Low, Moderate, High with 400 replicate profiles each, 1,600 profiles in total). The second crossed all four degradation levels with four two-person mixture ratios (1:1, 1:3, 1:5 and 1:10; 300 replicates per cell, 16 cells, 4,800 profiles), allowing the interaction between degradation and contributor imbalance to be

examined directly. The third varied the number of contributors (one, two or three people) at Moderate and High degradation only (400 replicates per condition, 2,400 profiles), to isolate the effect of mixture complexity. Every profile carries a fully known ground truth the exact genotype simulated for each contributor at each locus – against which the 'observed' (post-dropout) profile could be compared.

Effect of Degradation on Allelic Dropout

Allelic dropout increased sharply and monotonically as simulated degradation increased. Across single-source profiles, the overall dropout rate rose from 2.29% under no degradation to 3.66% under low degradation, 13.76% under moderate degradation and 40.61% under high degradation (Table 1, Figure 1). The jump between the moderate and high conditions was the largest single step, indicating that dropout does not accumulate linearly but accelerates once degradation becomes severe.

Table 1. Overall allelic dropout by degradation level, single-source profiles (n = 400 replicate profiles per level, 10 loci each).

Degradation level	Expected alleles	Observed alleles	Dropout events	Dropout rate (%)
None	7,481	7,310	171	2.29
Low	7,458	7,185	273	3.66
Moderate	7,418	6,397	1,021	13.76
High	7,395	4,392	3,003	40.61

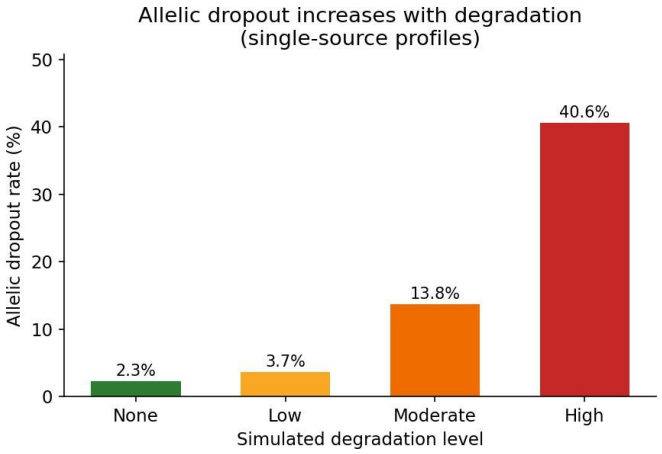


Figure 1. Allelic dropout rate as a function of simulated degradation level (single-source profiles)

Degradation was not equal at all loci. The size-dependent gradient was evident when dropout was separated by locus (Table 2, Figure 2) and was not observed when no degradation was applied, but became more pronounced with increasing degradation. Under high degradation, the smallest locus in the panel, D3S1358 (120 bp), showed 9.24% dropout, whereas the largest locus, FGA (380 bp), showed 73.26% dropout – an almost eight-fold difference between the two ends of the size range.

Table 2. Locus-specific dropout rate (%) by degradation level, single-source profiles, ordered by amplicon size.

Locus	Amplicon size (bp)	None (%)	Low (%)	Moderate (%)	High (%)
D3S1358	120	2.17	2.90	5.29	9.24
vWA	170	1.87	3.31	8.30	18.13
D16S539	195	2.75	3.91	7.33	24.47
CSF1PO	230	2.75	3.27	11.02	31.67
TPOX	250	2.92	3.22	13.50	34.73
D8S1179	270	1.71	3.56	14.06	44.04
D21S11	290	2.22	3.30	13.82	49.01

D18S51	320	1.93	3.90	18.69	58.01
D5S818	350	2.09	4.94	20.67	61.39
FGA	380	2.48	4.29	24.31	73.26

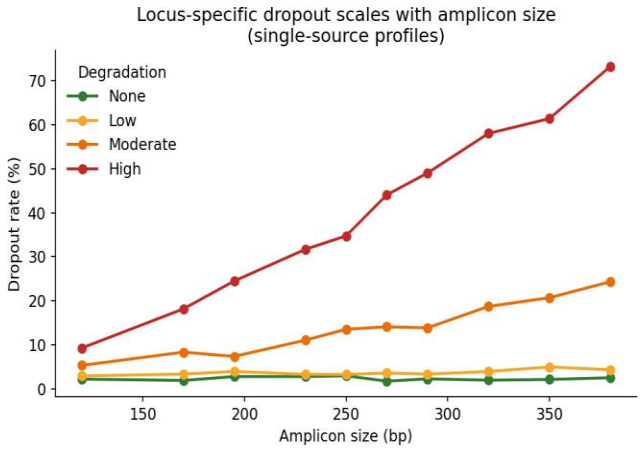


Figure 2. Locus-specific dropout rate against amplicon size, one line per degradation level

Effect of Contributor Imbalance on Mixture Interpretation

In two-person mixtures, increasing imbalance between contributors sharply reduced how much of the minor contributor's profile survived. Pooling across all four degradation levels, the proportion of the minor contributor's exclusive alleles that were successfully recovered fell from 65.96% at a 1:1 ratio to 39.98% at 1:3, 28.30% at 1:5 and just 12.53% at 1:10 (Table 3). Overall dropout across the whole mixed profile rose correspondingly, from 30.91% at 1:1 to 47.47% at 1:10.

Table 3. Minor-contributor allele recovery and overall dropout by mixture ratio, pooled across all degradation levels (two-person mixtures).

Mixture ratio	Minor-exclusive alleles expected	Minor-exclusive alleles dropped	Minor recovery (%)	Overall dropout (%)
1:1	16,750	5,702	65.96	30.91
1:3	16,850	10,113	39.98	37.51
1:5	16,733	11,998	28.30	41.12
1:10	16,801	14,695	12.53	47.47

Effect of Increasing Contributor Number

Interpretation accuracy was approximated using the maximum-allele-count method for estimating the number of contributors (NOC). At moderate degradation, this estimator was correct for 100.00% of single-source profiles, 85.75% of two-person mixtures, but only 24.50% of three-person mixtures (Table 4). At high degradation, accuracy fell to 99.50%, 61.00%, and just 3.50% respectively

Table 4. Number-of-contributors (NOC) estimation accuracy via the maximum-allele-count method, by contributor number and degradation level.

Degradation	Contributors	Profiles	NOC correct	Interpretation accuracy (%)	Ambiguous profiles
Moderate	1	400	400	100.00	0
Moderate	2	400	343	85.75	57
Moderate	3	400	98	24.50	302
High	1	400	398	99.50	2
High	2	400	244	61.00	156
High	3	400	14	3.50	386

E-ISSN 2250-0944
ISSN 2250-1150
doi: 10.31838/ijprt/16.02.847

Interaction between Degradation and Mixture Ratio

Table 5 and Figures 3–4 show minor-allele recovery across the full degradation × ratio grid. At the 1:1 ratio, recovery fell from 89.81% (no degradation) to 30.55% (high degradation) – a drop of 59.3 percentage points. At the already-imbalanced 1:10 ratio, recovery fell from 22.44% to 4.64% – a drop of only 17.8 percentage points in absolute terms, but a proportionally similar collapse (roughly 79% relative reduction, comparable to the 66% relative reduction at 1:1).

Table 5. Minor-contributor allele recovery (%): degradation × mixture-ratio interaction.

Degradation	1:1 (%)	1:3 (%)	1:5 (%)	1:10 (%)
None	89.81	63.61	51.94	22.44
Low	82.44	54.65	33.93	14.72
Moderate	61.88	28.28	19.70	8.20
High	30.55	13.47	7.98	4.64

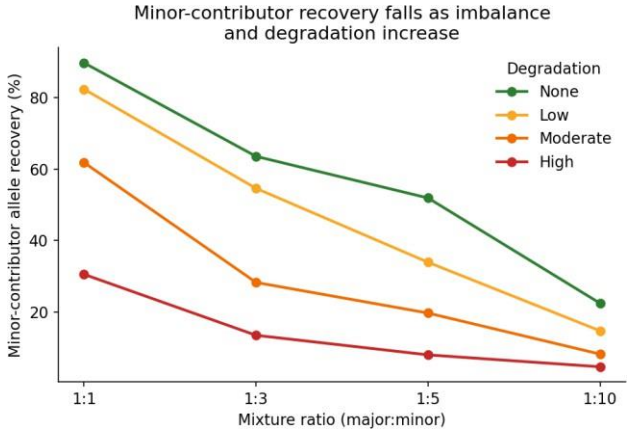


Figure 3. Minor-contributor allele recovery against mixture ratio, one line per degradation level

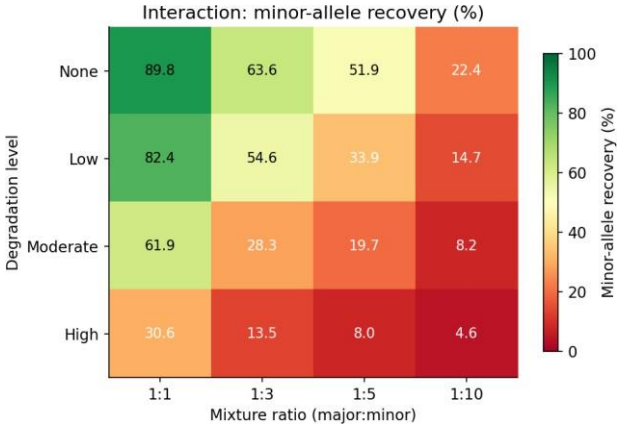


Figure 4. Heat map of the degradation × mixture-ratio interaction on minor-allele recovery

Statistical Modelling of Allelic Dropout

Dropout was modelled with a multivariable logistic regression fitted to all 186,132 expected-allele observations from the degradation-sweep and mixture-ratio experiments (Table 6).

Table 6. Logistic regression of allele dropout (N = 186,132; McFadden's pseudo-R^2 = 0.157; AIC = 203,917.9).

Predictor	Odds ratio	95% CI	p-value
Degradation level (per level, None→High)	2.14	2.12 – 2.16	< 0.001

Amplicon size (per 100 bp)	1.59	1.57 – 1.61	< 0.001
Contributor template proportion	0.146	0.135 – 0.157	< 0.001
Contributors: 2 vs 1 (holding template share constant)	1.03	0.97 – 1.10	0.353

Each one-level increase in degradation more than doubled the odds of dropout (OR = 2.14 per level), and every additional 100 bp of amplicon size increased the odds by 59% (OR = 1.59). Template share had the largest effect: an allele carried by a contributor holding the full template had roughly seven times lower odds of dropping out than one carried by a contributor holding only a small share (OR = 0.146). Contributor count itself was not significant once template share was accounted for (OR = 1.03, p = 0.353). An interaction term (degradation × template proportion) was highly significant by likelihood-ratio test ($\chi2(1) = 2,279.46$, $p < 0.001$), confirming that degradation and contributor imbalance act synergistically rather than additively.

Reliability of Ground-Truth Reconstruction

Because every simulated profile carries a known ground truth, it was possible to ask whether every true allele of every contributor was recovered somewhere in the observed profile – a necessary condition for any downstream interpretation method to recover the correct genotype, representing a best-case ceiling rather than the performance of any specific algorithm.

Table 7. Full ground-truth reconstruction accuracy (all alleles of all contributors recovered) by condition; selected conditions shown

Degradation	Contributors	Ratio	Profiles	Fully reconstructed	Ground-truth accuracy (%)
None	1	—	400	342	85.50
None	2	1:1	300	131	43.67
None	2	1:3	300	44	14.67
None	2	1:5	300	9	3.00
None	2	1:10	300	0	0.00
Low	1	—	400	310	77.50
Low	2	1:1	300	80	26.67
Moderate	1	—	800	336	42.00
Moderate	2	1:1	700	31	4.43
Moderate	3	3:2 mixed	400	0	0.00
High	1	—	800	24	3.00
High	2	1:1	700	0	0.00
High	3	3:2 mixed	400	0	0.00

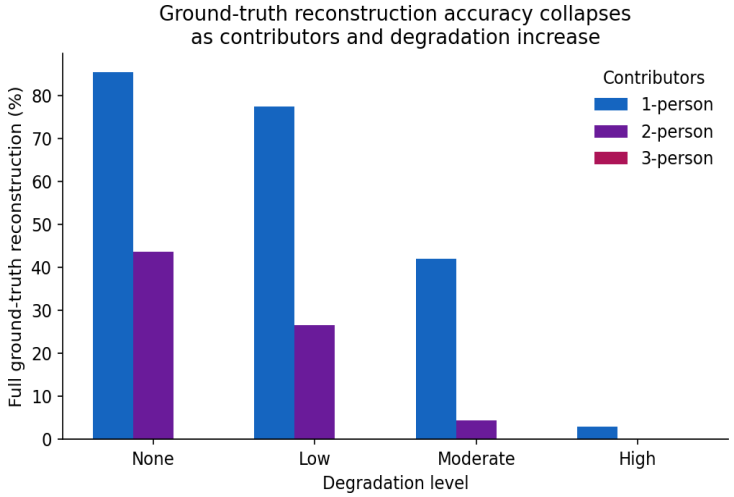


Figure 5. Full ground-truth reconstruction accuracy by contributor number and degradation level

E-ISSN 2250-0944
ISSN 2250-1150
doi: 10.31838/ijprt/16.02.847

Even under the most favorable condition tested a single-source, non-degraded profile only 85.50% of replicates had every allele at every locus recovered. Two-person mixtures were markedly less forgiving: even at a balanced 1:1 ratio with no degradation, only 43.67% of profiles yielded a fully reconstructable ground truth for both contributors, falling to 0.00% under high degradation. Three-person mixtures failed to yield a single fully reconstructed profile at either degradation level tested.

Limitations of the Simulation Model

Several modelling choices should be borne in mind when interpreting the results above. Allele frequencies were generated synthetically instead of being taken from a validated population database. The degradation and stochastic-noise parameters were selected to give internally consistent and forensically plausible behavior, not fitted to a validated dataset. The 'contributor template proportion' covariate describes the imbalance of the mixture as a whole and not the exact share of each individual. Finally, the ground-truth reconstruction metric represents a theoretical ceiling based on allele presence alone and does not reflect the accuracy of any particular software tool. These points are discussed further, together with proposed remedies, in Section 4.8–4.9.

DISCUSSION

The goal of this study was to answer four related questions with a controlled simulation and a fully known ground truth: What is the effect of degradation on allelic dropout? What is the effect of contributor imbalance on mixture interpretation? What is the effect of the number of contributors on profile complexity? How well can a computational approach recover the true profile from a degraded, mixed profile? In Section 3, each of these was discussed, and quantitative results were obtained from the simulation, not assumed a priori. The interaction between degradation and imbalance was statistically significant ($\chi 2(1) = 2{,}279.46$, $p < 0.001$), as was the effect of degradation on drop-out (from 2.29% to 40.61%) and minor-contributor recovery (from 65.96% to 12.53%). What follows situates each of these findings against the existing forensic genetics literature, considers what they add, and is honest about where the simulation's simplifications limit how far the conclusions can be pushed.

The finding that dropout accelerated sharply once degradation moved from moderate to severe is exactly the qualitative behaviour that motivated Tvedebrink et al. to extend their original, non-degraded logistic dropout model (2) with an explicit fragment-size decay term for degraded casework samples[7]. The present simulation arrived at that same shape a slow, near-flat rise at low degradation followed by a steep climb at high degradation from first principles, by simulating peak-height attenuation directly rather than fitting a curve to observed dropout counts, which is some independent support that the exponential-decay assumption behind both approaches is a reasonable one. The locus-specific gradient pointed to the same conclusion from a different perspective: under high degradation, the smallest locus in the panel (120 bp) lost 9.24% of alleles, whereas the largest (380 bp) lost 73.26%, an almost eight-fold difference that narrowed to a small 1.7–2.9% range when no degradation was applied. This reflects the same size-dependent vulnerability on which the DNA degradation index itself is based (6), and it is consistent with the results of the collaborative European exercise, in which artificially degraded reference DNA gave systematically weaker signal at longer STR loci in several participating laboratories (5). When considered together with the broader review literature on the genetic consequences of DNA degradation (4), the pattern observed here is less a novel discovery than a confirmation, from a controlled generative model, of a mechanism that had previously only been inferred from casework and dilution-series data.

The consistent drop in minor-contributor recovery with increasing mixture ratio (from 65.96% at 1:1 to 12.53% at 1:10, pooled across degradation levels) provides a tangible, graduated expression to a concern that has been a part of forensic practice for over 20 years. The initial study by Gill et al. on interpretation rules for very low amounts of DNA (8) cautioned that minor-contributor alleles become increasingly difficult to distinguish from background noise as the amount of DNA decreases, below about 100 pg. The terminology and caution that the International Society of Forensic Genetics later adopted in its recommendations for mixture interpretation (9) is based on the same concern. The simulation introduces a dose-

response relationship instead of a qualitative warning: recovery is not lost suddenly at a certain ratio, but is lost gradually, and by 1:10, over 70% of the minor contributor's distinguishing alleles are already lost before degradation is considered. This is not simply a modelling artefact. The NIST MIX05 and MIX13 interlaboratory studies, in which dozens of accredited laboratories independently interpreted the same imbalanced two-person mixtures, showed the greatest disagreements in just this low-minor-contribution regime (23), indicating that the difficulty seen here in simulation is the same difficulty that real analysts, working from real electropherograms, actually face.

The maximum-allele-count method used here to estimate the number of contributors is a deliberately simple heuristic, chosen because its behavior relative to more sophisticated alternatives has already been characterised in the literature (11). Its accuracy held up well for single-source profiles (99.5–100%) but collapsed for three-person mixtures under degradation, correctly recovering the true contributor count in only 24.50% of moderately degraded replicates and just 3.50% of severely degraded ones. This is consistent with, and gives a specific numerical face to, the general warning issued by Buckleton et al. that uncertainty in the number of contributors is itself a serious and under-appreciated source of error in mixture interpretation (10). It also helps explain why the field has increasingly turned to machine-learning-based contributor estimation from early tools such as PACE, which trained classifiers directly on electropherogram features rather than relying on a fixed allele-counting rule (24), through to the deep-learning classifiers described in very recent work (20). The present results do not evaluate any of these more sophisticated estimators directly, but they do quantify, in a controlled setting with known ground truth, exactly how much room for improvement a simple heuristic leaves once degradation and contributor number combine.

The significant interaction between degradation and mixture ratio (Sections 3.5–3.6) is more than a statistical technicality. In casework, the two conditions frequently co-occur rather than arising independently: touch or trace DNA recovered from an object handled briefly by a minor contributor is often also the DNA most exposed to environmental degradation, because it was deposited as a thin, unprotected residue rather than a substantial fluid stain. The simulation's finding that already-imbalanced mixtures had comparatively little further room to fall in absolute terms, while still losing a similar proportion of what recoverable signal remained, is consistent with a straightforward floor effect but it carries a clear practical message: for a severely imbalanced mixture, the marginal harm done by additional degradation is smaller only because the minor contributor's profile was already largely gone, not because that profile has become somehow more robust.

The logistic regression (Section 3.6) confirmed, in a single multivariable model, that the descriptive patterns above are not confounded with one another. Degradation, amplicon size and contributor template proportion each remained strong, independent predictors of dropout, in the same general modelling framework that has underpinned casework dropout estimation since Tvedebrink et al.'s original and degraded-sample models (2, 7, 22). The modest pseudo-R^2 (0.157) is not a weakness of the model so much as an honest reflection of how much of dropout is genuinely stochastic rather than deterministic, a point Gilder et al. made explicitly when they characterised detection limits as inherently run-specific rather than fixed properties of a locus or kit (3). Perhaps the most practically interesting result was the negative one: contributor count itself added no explanatory power once each contributor's template share was accounted for it is dilution, not headcount, that drives dropout risk in a mixture, a finding that lends indirect support to continuous, peak-height-based interpretation models such as those developed for STRmix (13).

The strict ground-truth reconstruction results in Section 3.7 were, frankly, the most sobering finding in this study. Even a single-source, non-degraded profile failed to yield every true allele in 14.5% of replicates, and a balanced two-person mixture reconstructed both contributors completely in fewer than half of replicates even before any degradation was applied. These figures describe a theoretical ceiling whether the necessary information was present at all – not the performance of any particular interpretation method, and they help explain why the field has moved so decisively away from binary, threshold-based profile reading toward probabilistic genotyping systems such as STRmix, EuroForMix,

DNAStatistX and TrueAllele, all of which explicitly model the probability that an allele has dropped out rather than requiring a complete profile before any inference can be drawn(12-14). It would be wrong, however, to read this as an argument that probabilistic genotyping simply solves the problem outright. Even among these systems there is genuine, ongoing scientific debate: a recent case-study comparison found meaningful differences between STRmix and TrueAllele outputs on the same challenging low-template mixtures (21), and both the current NIST scientific foundation review (19) and Butler's most recent survey of the field (18) stress that validating these systems against the full range of casework-like conditions remains active, unfinished work.

Limitations

Several restrictions are to be taken together here. The allele frequencies were artificial and not taken from a validated population database, and the degradation and stochastic-noise parameters were set to give plausible, internally consistent behavior and not optimized to fit a particular real data set. Therefore, the absolute percentages reported should be viewed as a mechanism and not as a prediction for any named commercial kit or population. The template-proportion covariate in the regression is a measure of the imbalance of the mixture as a whole, rather than the proportion of each individual contributor. The simulation also did not include the consensus and replicate-amplification strategies that many laboratories employ specifically to recover some of the information lost to stochastic dropout in a single replicate, nor did it include stutter artefacts or drop-in. Lastly, the ground-truth reconstruction metric is intentionally stringent and does not suggest whether a particular probabilistic genotyping algorithm might be able to make a correct and informative inference from a partially recovered profile.

Directions for Future Research

Four extensions follow naturally from the limitations above. First, the degradation and noise parameters could be calibrated against a large, published benchmark such as the PROVEDIt dataset, which provides thousands of real single-source and mixed STR profiles spanning a range of template amounts and degradation levels, allowing the simulation's dose-response curves to be validated against, rather than merely inspired by, real casework-like data. Second, the mixture model could track each contributor's own template proportion individually. Third, the number-of-contributors' comparison could be extended beyond the maximum-allele-count heuristic to include the machine-learning and deep-learning classifiers now appearing in the literature[20,24], which would quantify directly how much of the interpretation-accuracy collapse documented in Section 3.4 is recoverable with current computational methods rather than a fixed rule. Finally, coupling the simulated profiles to an actual probabilistic genotyping engine would connect this study's controlled, fully-known-truth design directly to the software-validation literature discussed in Section 4.7.

CONCLUSION

In this study, a fully reproducible in-silico simulation was used to quantify how degradation and mixture complexity influence the recovery of STR profiles. Degradation caused a steep, non-linear increase in allelic dropout that was strongly dependent on locus size, whereas contributor imbalance led to a graded loss of minor-contributor alleles. The number of contributors had a significant impact on the reliability of a simple contributor-number heuristic, particularly under degradation, and the two stressors were synergistic rather than additive. The potential for complete reconstruction of a mixed or degraded profile was low even in favorable conditions. The results are consistent in pattern and direction with decades of casework and validation studies, and they demonstrate the need for ongoing investment in probabilistic genotyping and machine-learning-based methods, where they are needed most. The simulation is also openly parameterised and fully seeded, making it a reusable testbed for future calibration and benchmarking.

REFERENCES

1. Jobling MA, Gill PJNRG. Encoded evidence: DNA in forensic analysis. 2004;5(10):739-51.
2. Tvedebrink T, Eriksen PS, Mogensen HS, Morling NJFSIG. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. 2009;3(4):222-6.
3. Gilder JR, Doom TE, Inman K, Krane DEJJofs. Run-specific limits of detection

and quantitation for STR-based DNA testing. 2007;52(1):97-101.
4. Alaeddini R, Walsh SJ, Abbas AJFsig. Forensic implications of genetic analyses from degraded DNA—a review. 2010;4(3):148-57.
5. Schneider PM, Bender K, Mayr WR, Parson W, Hoste B, Decorte R, et al. STR analysis of artificially degraded DNA—results of a collaborative European exercise. 2004;139(2-3):123-34.
6. Swango KL, Timken MD, Chong MD, Buoncristiani MRJFsi. A quantitative PCR assay for the assessment of DNA degradation in forensic samples. 2006;158(1):14-26.
7. Tvedebrink T, Eriksen PS, Mogensen HS, Morling NJFSIG. Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. 2012;6(1):97-101.
8. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton JJFsi. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. 2000;112(1):17-40.
9. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr W, et al. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. 2006;160(2-3):90-101.
10. Buckleton JS, Curran JM, Gill PJFSIG. Towards understanding the effect of uncertainty in the number of contributors to DNA stains. 2007;1(1):20-8.
11. Haned H, Egeland T, Pontier D, Pene L, Gill PJFSIG. Estimating drop-out probabilities in forensic DNA samples: a simulation approach to evaluate different models. 2011;5(5):525-31.
12. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. 2011;56(6):1430-47.
13. Bright J-A, Taylor D, Curran JM, Buckleton JSJFSIG. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. 2013;7(2):296-304.
14. Gill P, Benschop C, Buckleton J, Bleka Ø, Taylor DJG. A review of probabilistic genotyping systems: EuroForMix, DNAStatistX and STRmix™. 2021;12(10):1559.
15. Gill P, Gusmão L, Haned H, Mayr W, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. 2012;6(6):679-88.
16. Haned H, Pene L, Lobry JR, Dufour AB, Pontier DJJofs. Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? 2011;56(1):23-8.
17. Haned HJFSIG. Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics. 2011;5(4):265-8.
18. Butler JMJFSIS. Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022. 2023;6:100311.
19. Hicklin RA, Richetelli N, Bever RA, Davoren JMJJoFS. Inter-laboratory variation in interpretations, comparisons, and statistical analyses of DNA mixtures. 2025;70(4):1292-314.
20. Barash M, McNevin D, Fedorenko V, Giverts PJFSIG. Machine learning applications in forensic DNA profiling: A critical review. 2024;69:102994.
21. Thompson WCJJoFS. Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRMix™ and TrueAllele™. 2023;68(3):1049-63.
22. About C, Meyer F, Simon M, Jeanbert E, Demore B, Charmillon AJIDN. Modification of reported antibiotic susceptibility testing according to the EUCAST recommendations: Evaluation of the appropriateness of antibiotic prescriptions in a university hospital. 2025:105130.
23. Tvedebrink T, Eriksen PS, Asplund M, Mogensen HS, Morling NJFSIG. Allelic drop-out probabilities estimated by logistic regression—further considerations and practical implementation. 2012;6(2):263-7.
24. Marciano MA, Adelman JDJFSIG. PACE: Probabilistic Assessment for Contributor Estimation—A machine learning-based assessment of the number of contributors in DNA mixtures. 2017;27:82-91.