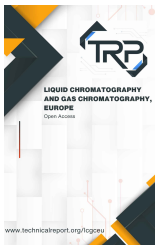

Journal of Liquid Chromatography and Gas Chromatography, Europe
https://technicalreport.org/lcgceu/

LC GC Eu., Vol. 39, No.1, 16-22, 2026

Technical Reports Press

RESEARCH ARTICLE

Uncertainty-Aware Likelihood Ratios for Dyed-Hair SERS Evidence: Propagating Pathway, Colorant, Sampling, and Density-Estimation Error

Fatima Zainab[1*]

[1]*Deaprtment of Chemistry, University of Education, Township Campus Lahore, Pakistan*
[*]*Corresponding author*

Corresponding author:
fatimazaib@uoel.edu.pk

Abstract
Exogenous hair dyes provide valuable information about the origin of a fiber since the applied colorants and products of dyeing process are able to preserve the relevant spectral fingerprint. SERS coupled with probabilistic methods allows the characterization of hair dye pathway and colorant composition, as well as estimating the degree of similarity between two dyed hair samples in terms of dye mixtures used with minimal hair sample consumption. DyeSPY-LINK converts similarity measures between hair dyed with the same dyes into likelihood ratios that represent evidential statements concerning common origin of dyes. A single LR, however, can convey misleadingly precise information when dye pathway classification, colorant prediction, and replicative sampling are considered deterministic operations. This paper develops an uncertainty-aware likelihood ratio framework that accounts for multiple error sources throughout the evidence evaluation process. Hair dye pathway and colorants become latent variables; spectra are resampled within strand-within-dye design; match and non-match score distributions are re-estimated in the Monte Carlo procedure; and LR distribution rather than a single LR value is provided as a result of the analysis. Validation in the new context is conducted at dye source level to avoid spectral leakage; measures of discrimination power, calibration, misleading evidence rate, uncertainty interval coverage, uncertainty partition, and decision category stability are proposed as quality indicators of the developed framework. Output LR is reported together with its calibrated log-scale uncertainty interval and posterior probability that evidential category remains unchanged after propagating uncertainty.

Keywords: forensic chemistry; dyed hair; surface-enhanced Raman spectroscopy; score-based likelihood ratio; uncertainty propagation; kernel density estimation; trace evidence; calibration.

Received: 25 July 2025
Accepted: 22 February 2026
Available online: 30 March 2026

1. Introduction

Interpretation of forensic evidence has moved away from categorical source declarations towards a probabilistic framework based on explicitly formulated propositions. The likelihood-ratio (LR) method lies at the heart of this development because it separates scientific assessment of observations from judicial assessment of prior probability and ultimate conclusion [1–4]. Trace comparison evidence is particularly appropriate for such an interpretation because hair, fibers, glass, paint, and chemical residues might be useful in assessing contact or common origin, but usually cannot be used to individualize without further assumptions regarding the population [5, 6]. Thus, in order to interpret a comparison forensically, it is necessary to specify both the propositions under consideration and the extent to which observations favor one of those propositions over the other.

Comparison of dyed hair is ideal for this LR framework. Whereas raw hair can only produce biologically informative signals, dyed hair will typically have signals related to the exogenous dye, intermediate, and reaction products introduced during the dying process. Different formulations have different dye mechanisms, precursors, couplers, amounts of direct-dye colorants, and target shades [7–9]. SERS is particularly useful in the study of dyed hair due to its sensitivity to residual dyes and capacity for generating detailed spectra with minimal material requirement. The development of the DyeSPY platform

Cite as: F. Zainab (2026). Uncertainty-Aware Likelihood Ratios for Dyed-Hair SERS Evidence: Propagating Pathway, Colorant, Sampling, and Density-Estimation Error. LC GC Eu., 39(1) (2026) 16-22.

made it possible to classify dyes and their constituent colorants using SERS-based spectra, alongside perceptual predictions of color [10]. DyeSPY-LINK then interpreted dyed-hair comparisons in a probability framework by translating cosine-similarity scores into SLRs using kernel density estimation (KDE) [11]. The SLR of a comparison with score δ is thus

$$\mathrm{SLR}(\delta) = \frac{f_{\mathrm{KM}}(\delta)}{f_{\mathrm{KNM}}(\delta)}, \tag{1}$$

where f_{KM} and f_{KNM} represent the KDE of the score density obtained from known-match (KM) and known-non-match (KNM) comparisons. DyeSPY-LINK represents a solid theoretical framework for interpreting dyed hair in evidential terms. Among the most strongly discriminating results, nonoxidative dyed hairs produced better results than oxidative samples, in part because oxidative formulations tend to be more spectrally and chemically diverse [11]. While the SLR is an interpretation of evidence comparing two hairs, the LR in the case of dyed hair specifically relates to the question of shared dye-mixture origins. Thus, LR calculations depend upon the proposition framework used in the analysis.

The calculation of $\mathrm{SLR}(\delta)$ is a simplification of what goes on when analyzing dyed-hair data. In practice, a hair sample will have to be preprocessed to remove extraneous signals, have its dye pathway (oxidative or nonoxidative) classified, have colorant composition predicted, and be used to calculate KM and KNM score KDEs. If any of these steps introduce uncertainty into the analysis, there is a chance of misinterpretation. When only a single value of $\mathrm{SLR}(\delta)$ is calculated, a comparison score conditioned on an uncertain factor may suggest a much stronger interpretation of evidence than the available information warrants. This is an issue in the broader LR context as well, when score-based methods are used in forensic analysis in a machine learning pipeline [12–15].

Here, we retain the proposition structure of the dye analysis provided by DyeSPY-LINK and develop an LR distribution model by introducing uncertainty in each of the factors described above. As a result, rather than a single SLR value, we obtain an LR ensemble representing uncertainty in upstream labeling and processing steps:

$$\left\{\mathrm{UALR}^{(1)}, \mathrm{UALR}^{(2)}, \ldots, \mathrm{UALR}^{(B)}\right\}, \tag{2}$$

where each ensemble element corresponds to an independently sampled realization of pathway classification, colorant labels, strand sampling, spectrum sampling, and reference KDEs. The resulting output is an overall LR value along with an uncertainty interval and a stability measure reflecting changes in the corresponding evidential category.

This work offers four contributions. First, we develop a distributional LR framework for dyed hair comparisons, accounting for uncertainty at multiple stages in the analysis. Second, we describe a validation procedure preventing bias towards the known outcome from leaking into an LR analysis. Third, we propose using stability as a reporting diagnostic indicating the consistency of evidential categories generated from uncertain upstream decisions. Finally, we establish the source of LR variability, making uncertainty quantification actionable at the lab level.

2. Background

2.1. Likelihood-ratio interpretation in trace evidence

Probability-based interpretation of forensic evidence depends upon the formulation of two hypotheses (propositions) describing the events that would explain a piece of evidence. In dyed-hair analysis, the competing propositions could be written as

$$H_s : \text{same dye-mixture origin}, \tag{3}$$

$$H_d : \text{different dye-mixture origins}. \tag{4}$$

The likelihood ratio is

$$\mathrm{LR} = \frac{\mathbb{P}(E \mid H_s, I)}{\mathbb{P}(E \mid H_d, I)}, \tag{5}$$

where E denotes the observed evidence and I is the collection of relevant information, e.g., analytical methodology and modeling assumptions. An LR value above 1 is supportive of H_s versus H_d, while LR ¡ 1 suggests that H_d is more likely than H_s. It should be noted that the likelihood ratio is not a posterior probability but rather a measure of evidence, as it does not represent a probability statement about any of the propositions considered [1, 16].

One challenge posed by high-dimensional spectroscopic data, e.g., in dye analysis, is density estimation. This is why score-based likelihood ratios have been proposed as a means of overcoming the dimensionality issue [12, 17]. The score-based LR assumes that a high-dimensional comparison can be reduced to a comparison score with corresponding KDE estimates from KM and KNM populations. This reduction should be performed assuming clearly specified score and its conditioning information. If conditioning information is itself uncertain, it should be reflected in the LR calculation.

2.2. Dyed-hair SERS and the DyeSPY-LINK analytical context

The DyeSPY data structure includes sets of SERS spectra from commercial hair dyes, separated into two pathway types: nonoxidative and oxidative [10, 11]. DyeSPY data structure facilitates uncertainty propagation because it allows for uncertainty to be sampled separately for the spectrum, strand, and dye-source levels. This hierarchical data organization makes it possible to distinguish between real agreement and artificial signal due to sampling or parameter choice. In particular, the difference between oxidative and nonoxidative dyes stems from the chemical processes occurring during dying, in addition to differing colorant structures. Therefore, the uncertainty of pathway assignment is important for an analysis.

DyeSPY-LINK applies preprocessing, cosine-similarity scoring, and KDE to produce an SLR for comparisons. It makes use of the hierarchical structure of the data, where each dye-source has multiple strands, which in turn are represented by multiple spectra. This hierarchical structure makes it possible to propagate upstream uncertainty from labeling to score comparison. However, a calibrated point SLR can still be prone to producing an overly precise evidential assessment. First, the uncertainty around SLR estimates may go unnoticed when an LR is based solely on a single pathway label, colorant set, or KDE fit. Second, verbal LR categories used in forensic science can amplify slight variations around key boundaries such as 10 or 100 [18]. Third, if a procedure requires a clear choice between oxidative and nonoxidative pathways, LR calculation may underrepresent the uncertainty.

The objective of this paper is to build an LR model that can be applied directly to the empirical framework of dyed hair provided by DyeSPY/DyeSPY-LINK. In other words, the LR method developed here preserves the same proposition structure and analysis procedure provided by DyeSPY. Specifically, this LR method seeks to address the following research questions:

RQ1. What kind of uncertainty propagation model can be constructed for score-based LRs of dyed-hair evidence comparisons?

RQ2. What evidence interpretations are stable after uncertainty propagation?

RQ3. What factors contribute most significantly to variation in the $\log_{10}$ LR values?

RQ4. What is the optimal LR output format?

It is expected that uncertainty propagation will make interpretation less overconfident in borderline cases, while maintaining the strong evidential support in more clear-cut cases.

3. Materials and methods

3.1. Data source and empirical context

The model specification is for the publicly available DyeSPY/DyeSPY-LINK data structure with dyed-hair SERS spectra. The empirical design includes 17 nonoxidative dyes and 43 oxidative dyes. Each dye is represented by three strands of hair, with five spectra obtained per strand, leading to nested spectra-strands-dye sources structure [11]. The same-dye-source spectral comparisons correspond to the KM type of comparison, while different-dye-source spectral comparisons are defined as the KNM type of comparison.

As for the data preparation stage, the protocol includes applying the preprocessing routine specific to the DyeSPY-LINK study. This consists of spectral range trimming to the 450–1650 cm^{-1} band, baseline correction, smoothing, and normalization by area. The analysis is entirely computational and employs the experimental design provided. The hypothesis to test here concerns whether the same or different dye-mixes are used rather than human source identification.

Taking into account the hierarchical structure of the reported dataset, each dye source provides 15 spectra, which yields the following figures for the nonoxidative dyes subset: 255 spectra, 1,785 KM comparisons, and 30,600 KNM comparisons. As for the oxidative dyes subset, the figures amount to 645 spectra, 4,515 KM comparisons, and 203,175 KNM comparisons. Thus, it becomes inevitable to conduct source-level partitioning, as otherwise, replicated dye sources will end up in both partitions.

Table 1: Dye-source hierarchy and potential within-pathway comparison counts before validation splitting.

Pathway	Dyes	Spectra	KM pairs	KNM pairs
Nonoxidative	17	255	1,785	30,600
Oxidative	43	645	4,515	203,175
Total	60	900	6,300	233,775

3.2. Point score-based likelihood ratio

For two preprocessed spectra $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, cosine similarity is

$$\delta_{ij} = \frac{\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\|\boldsymbol{x}_i\|_2 \|\boldsymbol{x}_j\|_2}. \tag{6}$$

For dye pathway $m \in \{\mathrm{NOX}, \mathrm{OX}\}$, the KM and KNM training scores define pathway-specific densities $\hat{f}_{\mathrm{KM},m}(\delta)$ and $\hat{f}_{\mathrm{KNM},m}(\delta)$. The point SLR is

$$\widehat{\mathrm{SLR}}_m(\delta) = \frac{\hat{f}_{\mathrm{KM},m}(\delta) + \varepsilon}{\hat{f}_{\mathrm{KNM},m}(\delta) + \varepsilon}, \tag{7}$$

where $\varepsilon > 0$ is a small tail regularization constant. The constant prevents numerical instability when KDE estimates approach zero in the tails; its influence is evaluated during sensitivity analysis rather than treated as a fixed assumption.

3.3. Sources of uncertainty

The uncertainty-aware model propagates four error sources that arise in a multi-stage dyed-hair SERS procedure.

3.3.1. Dye-pathway uncertainty

Let $M \in \{\mathrm{NOX}, \mathrm{OX}\}$ denote the dye pathway. Rather than fixing M to a single class, the model uses calibrated posterior weights:

$$w_m = \mathbb{P}(M = m \mid \boldsymbol{x}_i, \boldsymbol{x}_j, I), \qquad \sum_m w_m = 1. \tag{8}$$

The weights may be obtained from calibrated pathway-classifier probabilities, from cross-validated confusion matrices, or from Bayesian model averaging. When only hard labels are available, w_m is estimated from pathway-specific sensitivity and specificity, with conservative inflation for ambiguous classifications. In Monte Carlo iteration b, a pathway is sampled as

$$M^{(b)} \sim \mathrm{Cat}(w_{\mathrm{NOX}}, w_{\mathrm{OX}}). \tag{9}$$

3.3.2. Colorant-label uncertainty

Let $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iK})$ and $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jK})$ denote latent binary vectors for the presence of K possible colorants. Predicted colorant labels are treated as uncertain:

$$\begin{aligned} \boldsymbol{z}_i^{(b)} &\sim \mathbb{P}(\boldsymbol{z}_i \mid \boldsymbol{x}_i, M^{(b)}, I), \\ \boldsymbol{z}_j^{(b)} &\sim \mathbb{P}(\boldsymbol{z}_j \mid \boldsymbol{x}_j, M^{(b)}, I). \end{aligned} \tag{10}$$

If the colorant classifier returns calibrated probabilities, independent or dependence-adjusted Bernoulli sampling is used. If only binary predictions are available, colorant-specific sensitivity, specificity, and false-discovery estimates from cross-validation define the sampling probabilities. The colorant-overlap statistic in iteration b is

$$\eta^{(b)} = \phi\left(\boldsymbol{z}_i^{(b)}, \boldsymbol{z}_j^{(b)}\right), \tag{11}$$

where $\phi(\cdot)$ may be the shared-colorant count, the Jaccard index, or another pre-specified overlap function. This statistic allows the score density to incorporate chemical-label agreement without treating the labels as error-free.

3.3.3. Spectral and strand-level uncertainty

Let $\boldsymbol{x}_{d,s,r}$ denote spectrum r from strand s of dye source d. The nested sampling structure can be represented as

$$\boldsymbol{x}_{d,s,r} = \boldsymbol{\mu}_d + \boldsymbol{a}_{d,s} + \boldsymbol{e}_{d,s,r}, \tag{12}$$

where $\boldsymbol{\mu}_d$ is the dye-source spectral profile, $\boldsymbol{a}_{d,s}$ is a strand-level deviation, and $\boldsymbol{e}_{d,s,r}$ is spectrum-level noise. Bootstrap resampling is performed first at strand level and then at spectrum level within selected strands. This preserves within-source dependence and avoids treating replicate spectra from the same strand as independent dye sources.

3.3.4. KDE estimation uncertainty

KDE uncertainty is propagated by resampling the KM and KNM score sets and refitting the densities:

$$\hat{f}^{(b)}_{\mathrm{KM},m}(\delta), \quad \hat{f}^{(b)}_{\mathrm{KNM},m}(\delta), \qquad b = 1, \ldots, B. \tag{13}$$

Bandwidth uncertainty is evaluated through Scott's rule, Silverman's rule, and cross-validated bandwidth selection. Conclusions are considered robust only when the direction and verbal category of the evidential statement do not depend on a single bandwidth rule.

3.4. Conditional score-density estimation

To connect colorant uncertainty with the score-based LR, each Monte Carlo draw estimates a score density conditional on the sampled colorant-overlap statistic. For hypothesis class $H \in \{\mathrm{KM}, \mathrm{KNM}\}$ and pathway m, the conditional KDE is

$$\widehat{f}^{(b)}_{H,m}(\delta \mid \eta) = \frac{\sum_{r \in H_m^{(b)}} K_{h^{(b)}}(\delta - \delta_r) K_{\lambda^{(b)}}(\eta - \eta_r)}{\sum_{r \in H_m^{(b)}} K_{\lambda^{(b)}}(\eta - \eta_r)}. \tag{14}$$

Here, $K_{h^{(b)}}$ is the score kernel, $K_{\lambda^{(b)}}$ is the colorant-overlap kernel, and $H_m^{(b)}$ is the resampled reference set for pathway m. If colorant labels are unavailable, $K_{\lambda^{(b)}}(\eta - \eta_r)$ is set to one and the estimator reduces to ordinary score KDE.

3.5. Uncertainty-aware likelihood ratio

For Monte Carlo iteration b, the model samples a dye pathway, colorant labels, spectra, strands, and KDE densities. The uncertainty-aware LR draw is

$$\mathrm{UALR}^{(b)} = \frac{\widehat{f}^{(b)}_{\mathrm{KM},M^{(b)}}\left(\delta^{(b)}_{ij} \mid \eta^{(b)}\right)+\varepsilon}{\widehat{f}^{(b)}_{\mathrm{KNM},M^{(b)}}\left(\delta^{(b)}_{ij} \mid \eta^{(b)}\right)+\varepsilon}. \tag{15}$$

The reported central estimate is the median

$$\widetilde{\mathrm{UALR}} = \mathrm{median}\left\{\mathrm{UALR}^{(1)},\ldots,\mathrm{UALR}^{(B)}\right\}, \tag{16}$$

with a central uncertainty interval

$$\left[Q_{0.025}\left(\mathrm{UALR}^{(1:B)}\right), Q_{0.975}\left(\mathrm{UALR}^{(1:B)}\right)\right]. \tag{17}$$

Because LR distributions are commonly skewed, summaries are computed and visualized on the $\log_{10}$ scale and then translated back to the LR scale for reporting.

3.6. Decision-category stability

Forensic communication often uses verbal categories associated with LR ranges. Let $\mathscr{C}(\cdot)$ denote the category assigned to an LR:

$$\mathscr{C}(L) = \begin{cases} \text{strong support for } H_d, & L \le 0.01, \\ \text{support for } H_d, & 0.01 < L \le 0.1, \\ \text{inconclusive}, & 0.1 < L < 10, \\ \text{support for } H_s, & 10 \le L < 100, \\ \text{strong support for } H_s, & L \ge 100. \end{cases} \tag{18}$$

The decision-category stability index is

$$S_{\mathscr{C}} = \max_c \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left\{\mathscr{C}(\mathrm{UALR}^{(b)}) = c\right\}. \tag{19}$$

A stable conclusion has $S_{\mathscr{C}}$ close to one. An unstable conclusion places substantial probability mass in more than one evidential category. Strong verbal support is reported only when $S_{\mathscr{C}} \ge 0.80$ and the central interval does not span both sides of the inconclusive region.

3.7. Computational procedure

Figure 1 summarizes the analysis sequence. The point-SLR density ratio is retained, but classification, colorant-label, spectral, and density-estimation error are propagated before the final evidential statement.

3.8. Validation strategy

Validation is performed with dye-source grouping so that spectra from the same dye source cannot appear in both training and test partitions. For each split, complete dye sources are assigned to the training or test set. KM and KNM reference scores are generated only within the training set, KDE models are fitted on training scores, and test comparisons are evaluated without refitting on test sources. This design evaluates generalization to unseen dye sources rather than memorization of replicate spectra.

The primary discrimination metric is the area under the receiver-operating curve (AUC) computed from held-out KM and KNM comparisons. Calibration is measured using C_{LLR} and $C^{\mathrm{min}}_{\mathrm{LLR}}$, which distinguish the quality of the raw LR output from the best calibration attainable by monotonic transformation [15, 19]. For N_{KM} KM and N_{KNM} KNM

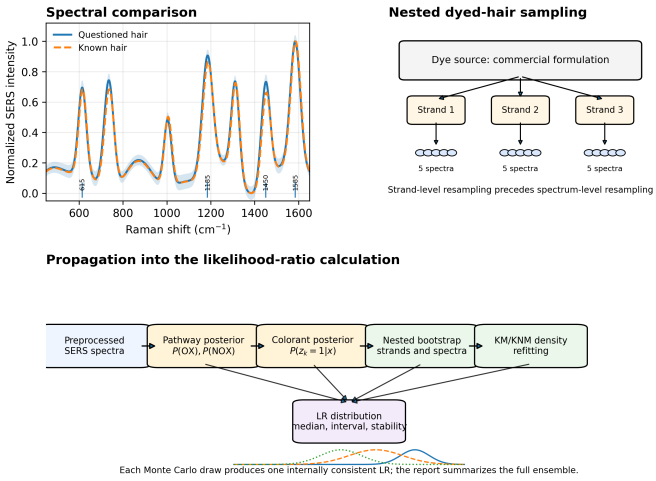

Figure 1: Uncertainty-aware SERS likelihood-ratio procedure for dyed-hair evidence. Representative SERS spectra, the strand-within-dye sampling hierarchy, pathway and colorant posterior probabilities, nested bootstrap resampling, KM/KNM density refitting, and LR-distribution reporting are shown as a single evidential calculation.

test scores with reported likelihood ratios L_i, the log-likelihood-ratio cost is

$$C_{\mathrm{LLR}} = \frac{1}{2}\left[\frac{1}{N_{\mathrm{KM}}}\sum_{i\in\mathrm{KM}}\log_2\left(1+\frac{1}{L_i}\right) + \frac{1}{N_{\mathrm{KNM}}}\sum_{j\in\mathrm{KNM}}\log_2\left(1+L_j\right)\right]. \tag{20}$$

Misleading-evidence rates are evaluated at threshold T as

$$m_{\mathrm{KM}}(T) = \mathbb{P}(\mathrm{LR} \le 1/T \mid \mathrm{KM}), \tag{21}$$

$$m_{\mathrm{KNM}}(T) = \mathbb{P}(\mathrm{LR} \ge T \mid \mathrm{KNM}). \tag{22}$$

The uncertainty-aware calculation additionally reports

$$p_{\mathrm{cat}} = \mathbb{P}\left(\mathscr{C}(\mathrm{UALR}^{(b)}) \ne \mathscr{C}(\widetilde{\mathrm{UALR}})\right), \tag{23}$$

$$p_{\mathrm{inc}} = \mathbb{P}\left(0.1 < \mathrm{UALR}^{(b)} < 10\right), \tag{24}$$

which quantify how frequently the evidential statement changes under propagated uncertainty or falls within the inconclusive interval.

The validation procedure tests the point-SLR algorithm against the uncertainty-aware LR on the same splits of dye source. The results will be considered separately for oxidative and nonoxidative dyes due to differences in chemical complexity and discrimination between the two dye pathways. Restricted propagation experiments are also conducted by enabling one uncertainty source at a time. A conclusion is experimentally validated only if discrimination is acceptable, calibration does not degrade significantly in comparison with point SLR calculation, misleading rates do not grow and evidential strength declines predominantly in low stability comparisons.

4. Results and discussion

4.1. Behavior relative to point-SLR evaluation

An uncertainty-aware LR may not always increase discrimination with respect to the corresponding point-SLR evaluation. It serves to distinguish evidence which is supported by a robust uncertainty structure from those that hinge on the uncertain upstream decisions. Highly concordant KM dye pairs should yield the same score, the same dye pathway assignment and the same colorant overlap statistic during

Monte Carlo draws. Their $\log_{10}(\mathrm{UALR})$ distribution is expected to be tight and $S_{\mathscr{C}}$ large. KNM chemically dissimilar dye pairs should produce either distributions below one or those shifting towards inconclusive region instead of yielding misleading evidential support for H_s.

The most interesting behavior takes place in boundary-sensitive comparisons. In this case, small shift in dye pathway, label sampling or KDE tails results in change of the category to which the point-SLR estimate belongs. This feature of the uncertainty-aware LR allows identification of boundary sensitive evidential support by extending uncertainty interval and lowering category stability. This effect is desired since it prevents misreporting based on numeric thresholding.

4.2. Use of uncertainty intervals in evidential reporting

Point SLR equal to 150 with 95% interval of [95,240] has an entirely different meaning from point-SLR of 150 with interval of [4,310]. While the former result stays within the range of strong evidential support through all MC iterations, the latter includes significant portions of weaker or inconclusive categories. Ignoring the full uncertainty interval means treating these two dye comparisons equally despite having distinct uncertainty structures.

Figure 2 illustrates four types of $\log_{10}(\mathrm{UALR})$ distributions of LR values and the associated behavior. The tightest distribution belongs to stable same-dye pairs, staying far away from zero and above the LR = 100 level. Boundary sensitive evidence covers the area around the LR = 10 and LR = 100 boundaries. Inconclusive evidential support is clustered near LR = 1. Different-dye comparison evidences are tightly grouped below one.

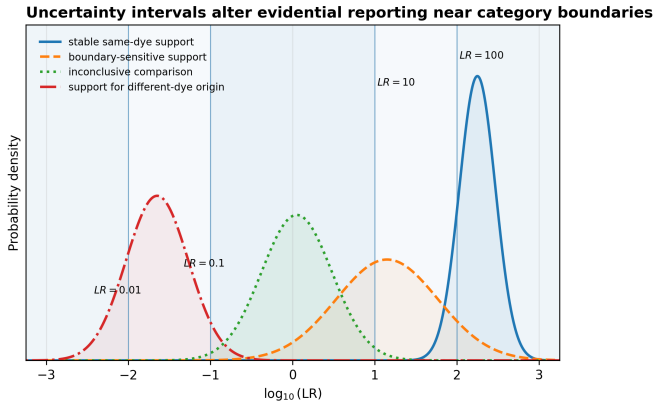

Figure 2: Representative uncertainty-aware LR distributions on the $\log_{10}$ scale. Vertical regions correspond to common verbal evidential ranges. The LR interval and category-stability index distinguish stable support from boundary-sensitive and inconclusive comparisons even when point estimates are similar.

4.3. Component-wise uncertainty attribution

The Monte Carlo structure permits uncertainty to be decomposed by source. Four restricted analyses are used:

(i) fix dye-pathway weights and propagate only colorant-label uncertainty;
(ii) fix colorant labels and propagate only dye-pathway uncertainty;
(iii) fix pathway and labels but bootstrap spectra and strands;
(iv) fix spectra and labels but bootstrap KDE densities.

For uncertainty source u, the restricted variance contribution is summarized as

$$R_u = \frac{\mathrm{Var}\{\log_{10}(\mathrm{UALR}_u^{(1:B)})\}}{\sum_v \mathrm{Var}\{\log_{10}(\mathrm{UALR}_v^{(1:B)})\}}, \tag{25}$$

where v varies over all four uncertainty factors. High pathway factor means either poor oxidative and nonoxidative separation or poor probability estimation. High colorant label factor means uncertain colorant assignment. High sampling factor means insufficient replicates or strand variation. High KDE factor means lack of supporting population data or instable behavior on tail distribution. The above allocation helps to relate the uncertainty bound to improvements at lab procedures.

4.4. LR decision category stability as a safeguard in reporting results

The forensic report should be careful in interpreting LR values whenever the LR probability distribution overlaps two categories verbally. Table 2 shows some guidelines for reporting LR while maintaining the interpretation of LR but avoiding the instability of distribution with regard to the high point estimate.

Table 2: Reporting rules for uncertainty-aware dyed-hair LR interpretation.

Condition	Reported interpretation	Reason
$S_{\mathscr{C}} \geq 0.80$ and the interval remains within one support region	Report median LR, interval, and stable verbal category	The evidential category is robust to propagated uncertainty
$S_{\mathscr{C}} < 0.80$ but most mass favors one proposition	Report limited support with an instability statement	The direction of support is present but category strength is uncertain
Interval crosses the inconclusive region	Report limited or inconclusive support	The evidence is sensitive to threshold placement
Large point LR with wide interval	Do not report the point LR alone	The point value gives false precision
Ambiguous dye-pathway posterior	Report model-averaged and pathway-specific LRs	The evidential value depends on pathway assignment

Figure 3 translates the numerical output into reporting decisions. Comparisons with narrow intervals and high $S_{\mathscr{C}}$ support direct LR reporting. Comparisons with wide intervals, low stability, or substantial mass in the inconclusive region require a cautious statement. The same display also identifies the uncertainty source most responsible for instability, which supports targeted method improvement.

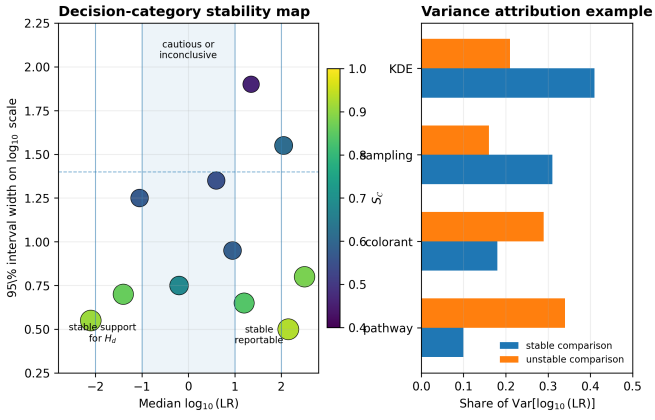

Figure 3: Validation and reporting stability map for uncertainty-aware dyed-hair LR interpretation. The left panel relates median $\log_{10}(\mathrm{LR})$, uncertainty-interval width, and category stability. The right panel illustrates component-wise variance attribution for stable and unstable comparisons.

4.5. Robustness and sensitivity analyses

Robustness validation verifies that modeling choices drive the evidential conclusion, not the spectral evidence itself. The point SLR and the uncertainty-aware approach produce summaries on equal splits of the dye source data. KDE bandwidths span rule-based and cross-validated selections. The density regularizing constant ε varies by many orders of magnitude. Classifier calibration is tested on dye sources held out from training, and performance summaries are stratified according to oxidative and nonoxidative pathways. Reports of case-level analyses are created for the stable, unstable, and inconclusive comparisons.
This collection of tests addresses reproducibility, overfitting, classifier calibration, and practical interpretability. It provides a pragmatic basis for determining whether an uncertainty-aware interval would lead to the conclusion of different strength of evidence or confirm the stability of the point-SLR outcome.

5. Practical forensic reporting

The uncertainty-aware model does not alter the definition of the LR. It alters the amount of uncertainty associated with that LR. A clear report might read:

> The SERS spectra of the questioned and known dyed hairs were evaluated using a likelihood-ratio model conditioned on the dyed-hair reference database and on the propositions under evaluation. The median likelihood ratio was $\widetilde{L}$, with a 95% uncertainty interval from $L_{0.025}$ to $L_{0.975}$. In the set of uncertainty evaluations, p_c of the outcomes fell into decision category c. This result concerns shared dye mixture origins and should not be interpreted as the identification of a particular individual.

This statement avoids the common incorrect phrasing of hairs being "L times more likely to come from the same person," which misrepresents the propositions and the conditional nature of the LR.

6. Contributions and implications

The primary theoretical contribution made by this study is the development of a likelihood-ratio distribution based on the uncertainties in a multi-stage forensic spectroscopy analysis pipeline:

$$p(\mathrm{LR} \mid \boldsymbol{x}_i, \boldsymbol{x}_j, I). \tag{26}$$

Uncertainty in classification, colorant predictions, spectral comparison scores, and density estimation are all included in this LR distribution. This approach generalizes score-based LR interpretation models beyond a simple score-density ratio but maintains the LR proposition framework.
The practical contribution consists of a reporting framework that enables additional information to be provided to the analyst: an interval of LRs, a category-stability index, and breakdown of LR uncertainty. These metrics allow forensic labs to judge how stable their evidential conclusions are and whether they need to be reported at full strength, at limited strength, or as inconclusive. It allows them to diagnose problem areas in their methods and allocate further method development resources accordingly.

7. Conclusion

This uncertainty-aware LR framework addresses a key weakness of point SLRs in SERS evidence for dyed hair: Uncertainty due to dye pathway classification, colorant prediction, spectral sampling, and KDE can have significant effects on the resulting evidential conclusion. The incorporation of this uncertainty into an LR distribution enables a more reliable forensic reporting process. The primary outputs of the methodology include the median LR value, an uncertainty interval, a decision category-stability index, and component-wise decomposition of the LR variance. The robustness testing design is safeguarded against spectral leakage by the use of complete dye source splits rather than partial ones. Reporting procedures avoid the problem of threshold-dependent conclusions. This framework is especially useful for borderline comparisons, in which a point LR estimate could overstate evidential strength; but it retains the strong-evidence category when spectral agreement, colorant overlap, and density estimates are stable. Future validation will include a wide range of dye library samples, aging, contamination, and inter-laboratory spectral acquisition systems.

References

[1] Aitken, C., & Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*. John Wiley & Sons.

[2] Association of Forensic Science Providers. (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, *49*(3), 161–164.

[3] European Network of Forensic Science Institutes. (2015). *ENFSI guideline for evaluative reporting in forensic science*. ENFSI.

[4] Morrison, G. S. (2018). The impact in forensic voice comparison of lack of calibration and validation. *Science & Justice*, *58*(5), 338–340.

[5] Christensen, A. M., Crowder, C. M., Ousley, S. D., & Houck, M. M. (2014). Error and its meaning in forensic science. *Journal of Forensic Sciences*, *59*(1), 123–126.

[6] Sauzier, G., van Bronswijk, W., & Lewis, S. W. (2021). Chemometrics in forensic science: Approaches and applications. *Analyst*, *146*(8), 2415–2448.

[7] Da Franca, S. A., Dario, M. F., Esteves, V. B., Baby, A. R., & Velasco, M. V. R. (2015). Types of hair dye and their mechanisms of action. *Cosmetics*, *2*(2), 110–126.

[8] Holman, A., & Kurouski, D. (2023). The effects of sun exposure on colorant identification of permanently and semi-permanently dyed hair. *Scientific Reports*, *13*, 2168.

[9] Holman, A. P., & Kurouski, D. (2024). Surface-enhanced Raman spectroscopy in forensic analysis. *Reviews in Analytical Chemistry*, *43*(1), 20230079.

[10] Holman, A. P., Maalouf, A., & Kurouski, D. (2025). DyeSPY: Establishing the first forensic SERS reference for hair dye colorant evidence. *Analytical Chemistry*.

[11] Holman, A. P., & Kurouski, D. (2026). DyeSPY-LINK: The first likelihood-based inference of near-source kinship for dyed hair evidence comparisons. *Forensic Chemistry*, *49*, 100746.

[12] Garton, N., Ommen, D., Niemi, J., & Carriquiry, A. (2020). Score-based likelihood ratios to evaluate forensic pattern evidence. *arXiv preprint*, arXiv:2002.09470.

[13] Johnson, M. Q., & Ommen, D. M. (2022). Handwriting identification using random forests and score-based likelihood ratios. *Statistical Analysis and Data Mining*, *15*(3), 357–375.

[14] Veneri, F., & Ommen, D. M. (2023). Ensemble learning for score likelihood ratios under the common source problem. *Statistical Analysis and Data Mining*, *16*(6), 528–546.

[15] van Lierop, S., Ramos, D., Sjerps, M., & Ypma, R. (2024). An overview of log likelihood ratio cost in forensic science–where is it used and what values can we expect? *Forensic Science International: Synergy*, *8*, 100466.

[16] Caliebe, A., Walsh, S., Liu, F., Kayser, M., & Krawczak, M. (2017). Likelihood ratio and posterior odds in forensic genetics: Two sides of the same coin. *Forensic Science International: Genetics*, *28*, 203–210.

[17] Martyna, A., Zadora, G., Neocleous, T., Michalska, A., & Dean, N. (2016). Hybrid approach combining chemometrics and likelihood ratio framework for reporting the evidential value of spectra. *Analytica Chimica Acta*, *931*, 34–46.

[18] Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*, 61–68.

[19] Brummer, N., & du Preez, J. (2006). Application-independent evaluation of speaker detection. *Computer Speech & Language*, *20*(2–3), 230–275.