SWGDAM GUIDELINES FOR REPORTING LIKELIHOOD RATIOS

The Scientific Working Group on DNA Analysis Methods (SWGDAM) Working Group for reporting of likelihood ratios (LRs) was reconvened for the purposes of reviewing and updating the previously published recommendations. This group was again composed of experts in the application of statistical principles to forensic evidence and forensic practitioners with expertise in the interpretation of mixed DNA specimens and probabilistic genotyping (PG).

The current document provides updates and additional information with regards to the original recommendations. Some of this additional information came from the Forensic Technology Center of Excellence webinar, provided in 2018 when the original recommendations were published.

The purpose of these guidelines is to promote consistency among laboratories when reporting the results of direct comparisons of evidentiary and reference profiles. These guidelines apply to LRs derived from probabilistic and binary interpretation approaches, as well as kinship analyses. These recommendations are not intended to be applied to the results of familial and other database searching.

This document was accepted by the membership of SWGDAM, received approval of the Executive Board of SWGDAM on Month DD, YYYY, and is not intended to be applied retroactively. This document supersedes the previously published recommendations.

1. REPORTING OF QUANTITATIVE AND QUALITATIVE STATEMENTS TO CONVEY LIKELIHOOD RATIOS

1.1: The numerical value for an LR shall be reported as a quantitative estimate of statistical weight, whether it supports the numerator (referred to as H1 in this document; often thought of as the prosecutor's proposition) or denominator (referred to as H2 in this document; often thought of as the defense proposition) proposition, with the exception of results deemed exclusionary as discussed in Recommendation 2.1.

LRs >1 provide greater support for the H1 proposition than for the H2 proposition. LRs <1 may be reported as the reciprocal of the LR to indicate the degree of support for H2 relative to H1. In this manner, an LR of 0.01 (1/100), for example, would reflect that the DNA evidence is 100 times more likely if it originated from an unknown, unrelated individual (H2) than if it originated from the person of interest (H1).

1.1.1 LRs exist in distributions, and no calculated LR value can be assumed to be the true LR for a particular comparison. Several ways of reporting LRs are valid, although the options available to the laboratory will be limited by the capabilities of the software being used. To ensure transparency the laboratory must disclose the reporting option used (e.g., in a report appendix) and the value(s) (e.g., lowest) being reported. All calculated values must be retained in the case record.

1.1.1.1 Reporting point estimate LR(s) for one or multiple populations: report calculated LRs for all population groups, or if simplifying and reporting a single value, the laboratory should generally choose the single lowest value from all populations.

1.1.1.2 Reporting one-sided interval(s) of LR distributions (e.g., 95 or 99% lower HPD[1]) for one or multiple populations: report calculated one-sided intervals for all population groups, or if simplifying and reporting a single value, the laboratory should generally choose the single lowest value from all populations.

1.1.1.3 Reporting two-sided interval(s) of LR distributions (e.g., 95 or 99% interval) for one or multiple populations: report the upper and lower values of the chosen (e.g., 99%) interval for all population groups, or if simplifying and reporting a single interval, report the upper and lower values from a single population's interval, generally the one with the lowest lower bound.

1.1.1.4. Stratified or unified LRs may be reported, but the underlying assumptions (i.e., population data, or average number of children) for those calculations must be included in the case record.

1.1.1.5. Reporting LRs from multiple analyses of the same data using different seeds under the same parameters: report calculated LRs for all population groups for all analyses, or if simplifying and reporting a single value, the laboratory should generally choose the single lowest value from all analyses.

Note that SWGDAM does not recommend reporting a single LR value closest to 1 among population groups when it is not the lowest. For example, if LRs among population groups are 4.0, 1.0, and 0.10 (i.e., 1/10), in cases where the lab only reports one LR, the LR of 0.10 should be reported.

Reporting a single LR value closest to one:

- Ignores potentially exculpatory LR values;
- May understate the exclusionary support for non-contributors when using population groups disparate from the sources of DNA in the evidence samples [Rohlfs RV, Fullerton SM, Weir BS (2012) Familial Identification: Population Structure and Relationship Distinguishability. PLoS Genet 8(2)]; and
- Could be mistaken as an upper bound of the LR for values below 1.

1.1.2 There is no scientific necessity to cap an LR value (i.e., set an upper bound on reported LR values).

[1] HPD = Highest posterior density

1.1.2.1 If a laboratory elects to cap the value(s) of reported LRs, it is recommended a cap not be less than one trillion (10^{12}).

1.1.2.2 If an LR cap is employed, the calculated LR values must be maintained in the case record.

1.1.2.3 Laboratories employing a cap should take care that the cap value not be misinterpreted as an identity threshold (e.g., source attribution), or a threshold above which any association is definitive.

1.2: A qualitative (verbal) statement that conveys the degree of support indicated by the LR may be reported in addition to the numerical value for the LR. The qualitative statement, if provided, should be reported in accordance with the verbal scale provided herein.

LRs are not probabilities, nor are they frequencies, and they may be difficult to conceptualize for lay people. To aid the court or other laypersons in understanding evidential strength, Ian Evett (1987) suggested a scale of verbal qualifiers to convey the degree of support for a given proposition, providing context to the magnitude of the LR. The scale categorized LR values as limited, moderate, strong and very strong in support of one proposition relative to an alternative proposition. The use of a verbal scale is supported across various disciplines of forensic science and has been adopted by the Association of Forensic Science Providers (AFSP, 2009) and the European Network of Forensic Science Institutes (ENFSI, 2015).

There are many published and unpublished verbal scales in use that SWGDAM considered in making these guidelines. Verbal scales are conventions that arise through a consensus process; a single verbal scale promotes the use of the same language for the same numerical values within and across jurisdictions. When used in reports and testimony by forensic analysts within and among different laboratories, the use of the same verbal scale promotes a consistent representation of evidential weight.

Table 1. Scale of verbal qualifiers for reporting LRs

LR for H1 Support and 1/LR for H2 Support	Verbal Qualifier
1	Uninformative
2 – <100	Limited Support
100 – <10,000	Moderate Support
10,000 – <1,000,000	Strong Support
≥1,000,000	Very Strong Support

LR results may be reported using the following quantitative and qualitative statements demonstrating application of the SWGDAM verbal scale, as exemplified for a two-person mixture:

The DNA typing results for Item 1 are 23 billion times more likely if they originated from SMITH and an unknown, unrelated individual than if they originated from two unknown,

unrelated individuals. This analysis provides very strong support for the proposition that SMITH is a contributor to the DNA obtained from Item 1 rather than the alternate proposition.

1.2.1 If a verbal qualifier is reported, the laboratory report should include the entire scale for purposes of providing context to any numerical value and may include an explanation of the scale, such as follows:

Equal support for both propositions results in an LR of 1, which is qualified as Uninformative. As LRs increase in magnitude, the scale reflects stronger degrees of support. LRs occur on a continuum; the categories recommended here have been chosen in part based on the observation that adventitious support for a proposition (e.g., LR >1 for an individual whose DNA is not present in the sample; or LR <1 for an individual whose DNA is present in the sample) is most commonly observed within the Limited Support category and generally not expected within the Very Strong Support category.

1.2.2 Additional context (e.g., text or images) should be provided whenever results fall into the "limited support" range.

A phenomenon termed the "weak evidence effect" has been described in the literature (Martire et al., 2013). It has been observed that the recipient of verbal scale information may interpret "weak evidence" for one proposition to mean "strong evidence" for the alternate proposition. In the Martire study, this appeared to be directional, where weakly inculpatory evidence was seen as strongly exculpatory. For this and other reasons, SWGDAM's verbal qualifier scale (Table 1) replaces the term "weak" with "limited." This change alone may not correct the misconception. Such text could alert readers to the issue, and emphasize the true meaning by tying the support statement back to the LR (see also Guidance Note 4 of ENFSI 2016). For example:

The use of the phrase 'limited support' for one proposition does not indicate or imply that there is more support for the alternate proposition. The first proposition still explains the evidence [LR] times better than the alternate proposition.

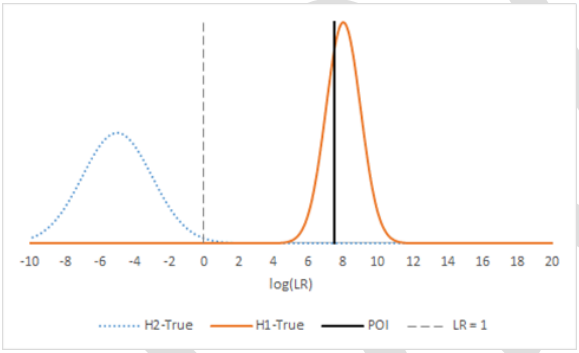
1.2.3 The verbal qualifier should not be communicated without a numerical value for the LR.

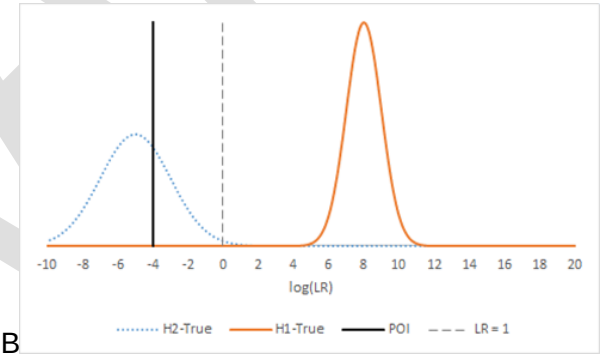
1.3: Qualifiers other than a verbal scale may be used to provide context for LRs in addition to the numerical value of the LR. The following are examples:

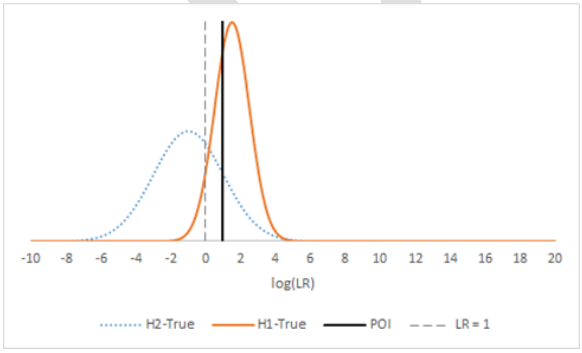
- Turing's rule ($P(LR > x|H2 \text{ true}) \leq 1/x$) states that the expected rate of non-contributor profiles that would be expected to provide LRs the same magnitude (or greater) as that of a tested individual is roughly equivalent to the reciprocal of the LR. For example, if an LR for a POI was calculated to be 1000, it would be expected that approximately one in a thousand non-contributors would have an LR of the same magnitude or greater (i.e., ≥1000).

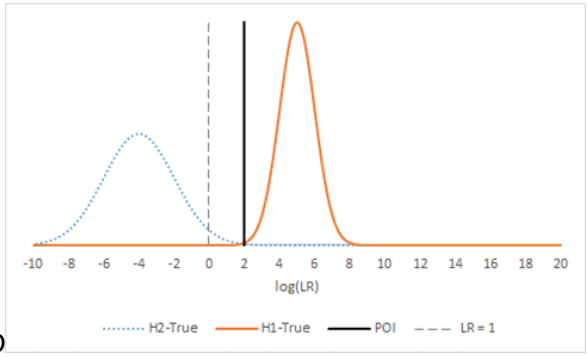
- H2-True testing (i.e., non-contributor testing) provides an interpretation-specific distribution of LRs for non-contributors. Similar to Turing's rule, this can be used to determine the proportion of non-contributor LRs that would be the same magnitude, or greater, as that of a tested individual. H2-True testing of a particular evidence profile interpretation involves using non-contributor profiles as the person of interest (POI_{NC}) in the calculation of LRs. POI_{NC} profiles are typically created *in silico* in proportion to allele frequencies in a relevant database. The distribution of LRs obtained from a large number of H2-True tests can provide context to the LR of the POI in relation to LRs of people known not to be contributors to the evidence.
- H1-True testing (i.e., true-contributor testing) provides an interpretation-specific distribution of LRs for possible contributor profiles as the person of interest (POI_{PC}) in the calculation of LRs. POI_{PC} profiles are typically created *in silico* using genotypes that have been determined to potentially contribute to a sample through use of a probabilistic genotyping system. The distribution of expected LRs obtained from a large number of H1-True tests can provide context to the LR of the POI in relation to LRs from profiles that could fit as contributors to the evidence.

These distributions do not replace the LRs reported for the POI(s). Interpretation-specific H1-True and H2-True testing can, however, provide context on whether the LR of the POI falls within the typical range for possible contributors (Figure A), non-contributors (Figure B), both (Figure C), or neither (Figure D).

A 

B 

C 

D 

Turing's rule and H2-True or H1-True tests speak to the expectations of the scientist about the data producing an LR of a certain value. They also relate directly to the propositions used in the

calculated LR. For example, if the H2 proposition of the original interpretation included a single unrelated, unknown individual, the statement applying Turing's rule would apply to the rate of unrelated non-contributors expected to produce an LR of the same, or greater, magnitude as the POI. Alternatively, if the H2 proposition of the original interpretation included a single untested sibling of the POI, the statement applying Turing's rule would apply to the rate of non-contributing siblings expected to produce an LR of the same, or greater, magnitude as the POI.

2. REPORTING AN EXCLUSION BASED ON LIKELIHOOD RATIOS THAT SUPPORT THE ALTERNATE PROPOSITION

2.1: As a matter of policy, a laboratory may establish an LR value below which an individual may be reported as excluded as a possible contributor rather than reporting an LR value that supports exclusion.

2.1.1 It is recommended that this value be at most 1/100. This ensures that any reported "exclusion" falls outside the limited support range of the verbal scale.

2.1.2 While the LR need not be reported for an exclusion, the upper bound below which exclusions are made should be specified in the report. For example, it could be specified as part of the verbal scale, or the report may include a statement such as, "LRs less than 0.01 are reported as exclusions."

2.1.3 All calculated values must be maintained in the case record.

3. REPORTING LIKELIHOOD RATIO VALUES THAT ARE CLOSE TO ONE

3.1: An "inconclusive zone" or other similarly named range (e.g., "uninformative zone" other than LRs of approximately 1) should not be used.

As LRs approach 1, the support for a given proposition decreases, and per the Turing expectations the probability of adventitious support for an incorrect proposition increases. However, with the exception of results deemed exclusionary as discussed in 2.1, LRs appropriately express the strength of the evidence and should be reported no matter how low or high the numerical value.

In general, LRs close to 1 indicate that the data is less informative, although not inconclusive, and may be due to lower template amounts for contributors, potential allelic drop-out, when few obligatory alleles are present, and/or allele masking. This is a known phenomenon and the LRs obtained generally speak to the quality of the data. LR values should not be looked at to determine whether a POI is "included" or whether a particular conclusion is correct. Instead, the data is providing the trier of fact logically relevant (e.g., Federal Rules of Evidence 401), albeit limited, information for the evaluation of the inclusionary/exclusionary hypotheses.

3.1.1 LRs should not be deemed inconclusive to mitigate a potential risk of adventitious support for either proposition.

In general, analysts should be comfortable explaining the meaning of LRs close to 1 and reasons there may be false support for either proposition, rather than rely on an "inconclusive zone" to buffer expectations. As an example, overestimating the number of contributors may provide false support for the inclusionary proposition for true non-contributors, while underestimating the number of contributors may provide false support for the exclusionary proposition for a true contributor. Note that this false support for a proposition may go beyond the limited support range.

Numerical values in the Limited Support for H1 range are comparable to Random Match Probabilities (RMPs) or Combined Probabilities of Inclusion (CPIs) that have been reported irrespective of magnitude (e.g., 1 in 5 or 1 in 100) despite the possibility that a true non-contributor might be included as a possible contributor to the evidence.

3.1.2 Analyses that provide Limited Support for H2, should be reported as support for H2 rather than as inconclusive. These LRs are potentially exculpatory and should be reported for transparency.

3.1.3 Calculations performed using different populations or multiple analyses of the same data (i.e., input file) with different seeds that result in LRs supporting opposing hypotheses (e.g., 10, which supports H1 and 0.1, which supports H2) should not be deemed inconclusive. Reporting these results should be done in accordance with section 1.1.

3.1.4 Specificity studies should not be used to establish an inconclusive zone.

Non-contributor testing has often been misunderstood as a reason to determine LRs of various magnitudes "inconclusive" because non-contributors providing LRs of the same magnitude were thought to be indicative of uncertainty of a POI's "inclusion" in the sample. Inconclusive zones implemented for the purposes of limiting or mitigating the chance of false "inclusions" are attempting to put binary answers on an infinite scale of LR magnitudes.

Non-contributor studies are ill-suited to designating "inconclusive zones". Non-contributor testing generally confirms the expectation that LRs supporting the inclusionary proposition are more common when there is less information in the data. If the laboratory were to use the highest LR value observed from a non-contributor to define an "inconclusive zone", the range of a given "inconclusive zone" will be dependent upon the number of profiles in the non-contributor tests. Those with sample sizes of hundreds of profiles may have inconclusive zones in the 100s to 1000s, while labs using several thousands of non-contributor profiles may generate inconclusive zones orders of magnitude wider (see Table 2).

Table 2. Example of Maximum LR values based on database size.

H2 True DB Size	5:1 Mixture Max LR
100	0.002
1,000	0.28

10,000	85
100,000	3.40E+04
1,000,000	2.70E+05

In addition, regardless of the range or the method of development (e.g., a percentile of non-contributor LRs) of the inconclusive zone, the presence of this zone perpetuates a myth that LR values outside of this zone are more conclusive with respect to a POI's "inclusion" in a sample. This unintended consequence of using an "inconclusive zone" undermines the reason to use one in the first place, namely, to prevent conveying a certainty that isn't present in the LR value.

In contrast, non-contributor testing conducted during validation may help inform a laboratory how well their probabilistic genotyping system and the model used within it is performing relative to expectations (i.e., Turing's rule). This testing may also provide information on the magnitude of the LR values expected given the quality of data present in a sample.

APPENDIX: EXAMPLE CONCLUSION STATEMENTS

An example of statements that could be used to report and further contextualize an LR result is presented below:

A. The profile is assumed to be a mixture of DNA from two individuals.
B. Inclusionary Hypothesis (HI): The DNA originated from Joe Smith and one unrelated, unknown individual.
C. Exclusionary Hypothesis (HE): The DNA originated from two unrelated, unknown individuals.
D. The DNA profile is 1.2 trillion times more likely if it originated from Joe Smith and one unrelated, unknown individual than if it originated from two unrelated, unknown individuals.
E. Based on this calculation, there is very strong support for the inclusion of Joe Smith as a possible contributor to the DNA profile obtained from the evidence.
F. The probability of an unrelated individual in the population, who has not contributed DNA to this sample, yielding this level of support or greater, is less than 1 in 1.2 trillion.

LR for H1 Support and 1/LR for H2 Support	Verbal Qualifier
1	Uninformative
2 – <100	Limited Support
100 – <10,000	Moderate Support
10,000 – <1,000,000	Strong Support
≥1,000,000	Very Strong Support

In the example above, statements E and F may be used to provide additional context to the value of the likelihood ratio presented in statement D. E and F may be presented in a report or offered to explain the likelihood ratio during testimony.

REFERENCES:

Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. Science and Justice 49:161–164.

Bright, J., Curran, J. M., Hopwood, A. J., Puch-Solis, R. and Buckleton, J. S. (2013b) 'Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts I (corrigendum)', Science & Justice, vol. 53, p 371

Buckleton, J. S., Pugh, S. N., Bright, J., Duncan A. Taylor, D. A., Curran, J. M., Kruijver, M., Gill, P., Budowle, B., Cheng, K. Are low *LR*s reliable? Forensic Science International: Genetics Volume 49, November 2020, 102350 https://doi.org/10.1016/j sigen.2020.102350

Cook, R., Evett, I.W., Jackson, G., Jones, P.J., Lambert, J.A. (1998) A model for case assessment and interpretation, Science and Justice 38:151–156.

ESR, DBLR V1.1 User's Manual,, issued December 1, 2020, page 50

European Network of Forensic Science Institutes (2016) ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE). Approved version 3.0. Accessed January 2, 2018 at: http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

Evett, I.W., Jackson, G., Lambert, J.A., McCrossan, S. (2000) The impact of the principles of evidence interpretation on the structure and content of statements, Science and Justice 40:233–239.

Evett, I.W. (1987) Bayesian inference and forensic science: problems and perspectives. Journal of the Royal Statistical Society, Series D 36:99–105.

Evett, I.W. (1998) Towards a uniform framework for reporting opinions in forensic science casework. Science and Justice 38:198–202.

Federal Bureau of Investigation (2011) Quality Assurance Standards for Forensic DNA Testing Laboratories. Effective 9/1/2011. Accessed January 2, 2018 at: http://media.wix.com/ugd/4344b0_4a22824ce56f43d4b1a4d2486409f95d.pdf

Forensic Science Regulator. Allele frequency databases and reporting for DNA profiling. The regulator's DNA specialist group has produced allele frequency databases and reporting guidance for DNA profiling.

Gill, P., Hicks, T., Butler, J. M. , Connolly, E., Gusmão, L., Kokshoorn, B., Morling, N., van Oorschot, R., Parson, W., Prinz, M., Schneider, P.M., Sijen, T., Taylor, D., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions, Part I: evaluation of DNA profiling comparisons given (sub-) source propositions, FSI: Genet. 36 (2018) 189-202

Hopwood, A. J., Puch-Solis, R., Tucker, V. C., Curran, J. M., Skerrett, J., Pope, S. and Tully, G. (2012) 'Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts', Science & Justice, vol. 52, pp 185–190

Marquis, R., Biedermann, A., Cadola, L., Champod, C., Gueissaz, L., Massonnet, G., Mazzella, W.D., Taroni, F., Hicks, T. (2016) Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings, Science and Justice 56:364–370.

Martire, K., Kemp, R., Watkins, I., Sayle, M., Newell, B. (2013) The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect, Law and Human Behavior 37(3):197–207

Myers, S., Searching CODIS with binary conversions of STRmix interpretations, FSI: Genet. 55 (2021) 102569

Perlin, M., Efficient construction of match strength distributions for uncertain multi-locus genotypes, Heliyon 4 (2018) e00824

Scientific Working Group on DNA Analysis Methods (2015) SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Accessed January 2, 2018 at: https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf

Triggs, C., Harbison, S.A., Buckleton, J. (2000) The calculation of DNA match probabilities in mixed race populations, Science & Justice 40:33-38.