Scientific & Technical Review Panel

Final Report for OSAC 2022-S-0024 Best Practice Recommendations for Evaluative Forensic DNA Testimony

Organization of Scientific Area Committees (OSAC) for Forensic Science

STRP Final Report
OSAC 2022-S-0024
Best Practice Recommendations for Evaluative Forensic DNA Testimony

Organization of Scientific Area Committees (OSAC) for Forensics Science
July 11, 2023

Disclaimer:

This report was produced by an independent Scientific and Technical Review Panel (STRP). The views expressed in the report do not necessarily reflect the views or policies of the U.S. Government. Visit the OSAC website for more information on OSAC's STRP process.

Scientific & Technical Review Panel Members

- Hillel Bavli, SMU
- Frederick R. Bieber, Harvard Medical School
- Sara Walker, Intermountain Forensics
- Jonathan Koehler, Northwestern University
- Greggory LaBerge, Denver Police Department
- Susan Pope, Principal Forensic Services
- Adele Quigley-McBride, Duke University
- Tiffany Roy, ForensicAid
- Joel Sutton, USACIL
- Sandy Zabell, Northwestern University

Report Summary:

The Scientific and Technical Review Panel (STRP) for "Best Practice Recommendations for Evaluative Forensic DNA Testimony" is an independent panel appointed by the National Institute of Standards and Technology (NIST). A STRP is established with a range of experts to consider how well a standard meets the needs of the forensic science, law enforcement, and legal communities, and to recommend improvements to the standards under review. The STRP appreciates the efforts of Tim Kalafut, Human Forensic Biology Subcommittee member, while serving as the subcommittee liaison to this STRP during the review process.

The STRP began its review process with a kickoff meeting on February 11, 2022, and concluded with this STRP final report. The panel reviewed the draft standard and prepared comments for the Human Forensic Biology Subcommittee.

Note Regarding Supplemental Comments – This STRP Report includes "Comments submitted during the voting process" that represent the view(s) of an individual(s). The comments were proposed after the STRP discussion was completed, and therefore, were not subjected to discussion amongst the STR panelists. All the comments submitted for the respective sections are included to ensure the views of the STR panelists are accurately represented.

Report Components:

The STRP reviewed this draft standard against OSAC's *STRP Instructions for Review* which include the following content areas: scientific and technical merit, human factors, quality assurance, scope and purpose, terminology, method description and reporting results. The details below contain a brief description of each reviewed content area and the STRP's assessment of how that content was addressed in the draft OSAC Proposed Standard.

1. **Scientific and Technical Merit:** OSAC-approved standards must have strong scientific foundations so that the methods practitioners employ are scientifically valid, and the resulting claims are trustworthy. In addition, standards for methods or interpretation of results must include the expression and communication of the uncertainties in measurements or other results.

    1.1 **Consensus View** - Upon review, the STRP believes that the proposed Best Practice Recommendation (BPR) serves to address an important topic and is based on a logical framework well-characterized in the global forensic community. This framework, known as the "hierarchy of propositions", is intended to assist the scientist with when and how to properly evaluate the DNA results during expert testimony while considering the key question being asked and its implications in a

judicial setting. For example, the proposed BPR rightly points out that moving up the hierarchy (e.g., sub-source to activity level questions) requires more case information and expert knowledge. The proposed BPR also provides practical applications of the method in "Supporting Information" and through the use of real-world case examples. Uncertainty with this method is explicitly expressed through the use of conditional probabilities as described in the document.

Although the BPR attempts to address all levels of the hierarchy of propositions encountered in forensic DNA analysis, the STRP recognizes the differing probabilistic and statistical issues that arise at each level of the hierarchy; from sub-sub-source to sub-source, to source, and to activity and crime levels. Accordingly, the STRP recommends that activity and crime level propositions should be handled separately in a stand-alone document. This would best provide guidance through both a framework and discussion of reporting requirements and testimony guidelines. Research is lacking for many conclusions and opinions that could arise from these types of analyses. Therefore, it would be a service to the forensic DNA community to handle such issues separately in detail to avoid confusion in the current document between the levels of hierarchy.

1.2 **Minority View** - This alternative view concurs with the implicit suggestion in the Consensus View noted above that activity and crime level propositions should not be part of the "Best Practice Recommendations for Evaluative Forensic DNA Testimony." But this alternative view takes the position that the focus in the document on activity level propositions is misplaced. Instead, the focus should be on providing weight of evidence testimony for source propositions alone. A source proposition (e.g., "Whose DNA is on the knife?") lends itself to scientific weight of evidence testimony in the form of empirically validated conditional probabilities. This is not the case for an activity level proposition (e.g., "How did the defendant's DNA get on the knife?"). The problem is not with the logic of the hierarchy of propositions framework. Rather, the problem is with the absence of statistical databases that speak to the creation of the conditional probabilities of interest when dealing with activity level propositions. A DNA examiner could be in a scientifically tenable position to estimate the probability of seeing an evidentiary DNA profile that matches a defendant's own DNA profile given that the defendant is not the source of the evidentiary DNA. However, the examiner cannot be in a similar position to estimate the probability of seeing matching DNA profiles given that the match was the result of transfer from the hand of a different person who held the knife. In the former case (source level), databases can be constructed, consulted, and subject to examination at trial. In the latter case (activity level), there would be no such databases, and therefore the basis for an examiner's conditional probabilities related to the activity level propositions would not rest on firm footing and would likely be idiosyncratic.

1.3 **Comment Received During Voting** - The Consensus View requires a minor, but important, modification to be acceptable: Add a sentence that explicitly rejects

speculative testimony regarding probabilities for activity-level propositions in the absence of empirical support or a substantial logical basis.

1.4 **Comment Received During Voting** - DNA analysts should never address the offense/crime level. This is the province of the trier. This document appropriately addresses sub-source and sub-sub source level evaluations. It should be noted that evaluations on these levels are absolutely required by accreditation to be included in written, technically reviewed and administratively reviewed reports. This document should break out evaluation of findings at the source level and activity level as separate documents because we are currently not operating at those levels of the hierarchy in the U.S. and more guidance will be required.

The title of this document would need to change in order for the good parts to stay and the bad parts to go. As a suggestion, "Evaluations of findings given sub-sub source and sub-source level propositions."

2. **Human Factors:** All forensic science methods rely on human performance in acquiring, examining, reporting, and testifying to the results. In the examination phase, some standards rely heavily on human judgment, whereas others rely more on properly maintained and calibrated instruments and statistical analysis of data.

2.1. **Consensus View** - Although the subcommittee has worked hard to address issues that stem from human factors concerns, there are still several serious problems that remain in the standard. The two main human factors concerns are noted here.

First, the document does little to incorporate the existing empirical research on how to communicate forensic evidence, such as results of DNA comparisons, to judges and jurors effectively. Communicating complex information so that it can be understood by the factfinder is one of the primary goals of expert testimony, so it is a serious oversight for the subcommittee to ignore an entire body of work addressing exactly that issue. For instance, the only empirical article from human factors researchers that is provided in this standard is more than three decades old (Thompson and Schumann, 1987). There are many studies examining how people understand likelihood ratios or verbal equivalents in the context of forensic testimony, but none of these are cited in the document or the normative references (e.g., Martire et al., 2014: https://doi.org/10.1016/j.forsciiint.2014.04.005). The recommendations contained in many of these articles, at times, run counter to the recommendations in the standard. For example, the standard advocates for the use of verbal equivalents in testimony, which has little empirical support from statistics and human factors research. Words can mean different things to different people and research suggests that using equivalents can mislead jurors rather than clarify the value of the results.

Second, the document allows for a single analyst to perform complex analyses on the stand without time to prepare, consider all factors, and conduct research. Results relevant to activity-level propositions require specific, higher-level experience and

training before an analyst can perform those analyses and interpret those results appropriately. In addition, part of that experience and training involves knowing when it is inappropriate to even attempt an analysis at the activity-level. Given the multitude of factors that can influence transfer, persistence, and other activity-relevant circumstances, these analyses are far too complex to be done on the stand and, in many cases, too complicated and uncertain to be done at all. From the perspective of a human factor's researcher, the strong effect of such statements on the fact finder and chance of error and bias is too high for it to be acceptable for DNA analysts to engage with activity-level hypotheticals on the stand with no preparation, and no technical or quality review, verification, or other oversight.

One way to reduce these human factors concerns would be to have a training standard designed to arm analysts with the necessary knowledge and experience to perform these analyses correctly (and also know when they are inappropriate), as well as test their ability to do these analyses. The required training and expertise to perform these types of analyses are separate and distinct from the expertise that is required to perform evaluations of findings on the lower levels of the hierarchy, a fact that U.S. DNA analysts may not intuitively know.[1] There would also need to be a standard guiding analysts' evaluation and reporting procedure for DNA findings given proposed activities. Such a standard would specify the supporting notes and data to be provided alongside the evaluation, as well as the procedure for technical and administrative review of the work prior to testifying to activity-level results.

2.2. **Minority View** - Many laboratories in the U.S. have taken steps to move toward evaluation of DNA profiles given sub-source level propositions. Labs that have implemented STRmix and the likelihood ratio should already be familiar with the well-established framework of the Hierarchy of Propositions, first described by Cook et. al in 1998.[2] This STRP feels this document has significant merit in clarifying for the community the framework for performing evaluations of DNA findings given proposed activities, including clarifications on how to avoid common pitfalls such as the prosecutor's fallacy. There were some discussions regarding language and definitions for clarity from a human factors perspective, but the overall STRP was supportive of the document as it applies to sub-sub-source, sub-source and source level evaluations of DNA results given propositions on these levels.

2.3. **Comment Received During Voting** - The draft BPR addresses a number of important human-factors concerns. However, there are serious concerns that have not been adequately addressed. For example, the draft standard does not account for a body of recent literature regarding how people understand likelihood ratios or

[1] Roland A.H. van Oorschot, Bianca Szkuta, Kaye N. Ballantyne, Mariya Goray, Need for dedicated training, competency assessment, authorizations and ongoing proficiency testing for those addressing DNA transfer issues, Forensic Science International: Genetics Supplement Series, Volume 6, 2017,Pages e32-e34, ISSN 1875-1768, https://doi.org/10.1016/j.fsigss.2017.09.013.

[2] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, Science & Justice, Volume 38, Issue 4, 1998, Pages 231-239, ISSN 1355-0306, https://doi.org/10.1016/S1355-0306(98)72117-3.

verbal expressions of such measures in the context of forensic testimony. This literature arguably runs counter to certain recommendations in the standard. In addition, while the draft standard focuses heavily on how an expert should phrase a statement to avoid transposing a conditional, it arguably does not adequately focus on phrasing testimony such that jurors, judges, attorneys, and other legal actors will not misunderstand the testimony so as to transpose a conditional. Finally, the BPR seems to support the use of notional probabilities and other testimony regarding activity-level propositions that are not necessarily based on firmly grounded data. The draft BPR seems to substantially underestimate the danger of jurors attaching undue weight to and otherwise misusing this evidence.

3. **Quality Assurance:** Quality assurance covers a broad range of topics. For example, a method must include quality assurance procedures to ensure that sufficiently similar results will be obtained when the methodology is properly followed by different users in different facilities.

    3.1. **Consensus View** - A process for the provision of an evaluative expert opinion and testimony on DNA activities must include quality assurance that:

    - The process follows the recognized framework for casework assessment and interpretation and has been validated by a suitable accreditation body for the laboratory and experts using it.
    - The robustness of the expert's opinion sits on a base of verified data, applicable published, or publicly available validation studies, as well as demonstrable recorded experience.
    - The initial results and verification of those results are both considered equally important when reviewing the evidence.

    This STRP finds that there are still several elements that need to be addressed from a quality assurance perspective. Specifically, without guidance for training and assessment, as well as feedback on use of this framework, there are no provisions in this document for assuring the quality of its use.

    There are also potential quality assurance issues associated with analysts engaging with hypotheticals on the stand and performing analyses (in particular, activity-level analyses) during testimony. This would be equivalent to reporting/ testifying to analyses that have not been through quality review or technical review. In addition, activity-level analyses require a level of specialized training not typically offered or undertaken by analysts in the U.S. (or even internationally), so there remains questions about whether analysts are trained and accredited to perform these analyses even in the lab, let alone while on the stand. These analyses often also lack the necessary published data and recorded experience to support the findings.

    3.2. **Minority View** - Quality assurance topics are properly covered in this draft standard. There are references for international standards and guidelines giving advice on suitable validation processes including risk assessment, setting user

requirements, specifications and acceptance criteria for a subjective evaluation process.

3.3. **Comment Received During Voting** - The Consensus view must incorporate a recommendation regarding communication of relevant error rate information and estimates to triers of fact.

4. **Scope and Purpose:** Standards should have a short statement of their scope and purpose. They should list the topics that they address and the related topics that they do not address. Requirements, recommendations, or statements of what is permitted or prohibited do not belong in this section.

4.1. **Consensus View** - Upon review, the STRP believes that the "Scope" of the document is clearly detailed as written but does not include a specific "Purpose" statement. The "Foreword" does provide an excellent summary of background information to include the underlying scientific principles behind the framework needed for evaluative forensic DNA testimony. The purpose is addressed at length in this same section beginning with the second sentence (line 2) and further summarized with the "goal" (lines 72-74). Perhaps a single purpose statement could be added under the "Scope" section using this same language from the "Foreword". Annex A (supporting documentation) is an important section in this proposed BPR as it often provides the needed detail and basis for many of the recommendations listed. It should be mentioned in the foreword with its purpose to better direct the reader for the appropriate context with each recommendation. Currently, the foreword only mentions the recommendations (lines 76-82).

4.2. **Comment Received During Voting -** *Recommended Modification to the Consensus View* – Add the following: "Best practice is to include the evaluation of DNA evidence given activity level propositions in a laboratory report. This will allow for any required technical reviews of the findings that would be reported on the witness stand. However, currently very few laboratories are including this in their reports, yet questions regarding activity level propositions show up routinely in trials. The goal of this document is to address immediate needs for best practices in testimony when the expert witness is not given the opportunity to provide their results in a written report, and how to communicate both the evaluation and the limitations of that evaluation."

If the expert has not been given the opportunity to provide their results in a written report, they should not be performing these types of high-level analyses, especially on the witness stand. Evaluations of findings given proposed activities require formal consideration far beyond what would be possible during a testimony.

5. **Terminology:** Standards should define terms that have specialized meanings. Only rarely should they give a highly restricted or specialized meaning to a term in common use among the general public.

5.1. **Consensus View** - The STRP finds that the draft standard requires substantial revision with respect to its terminology. First, the draft fails to define the term "notional probability." The intended meaning and substantive application of notional probabilities are fundamental to the scientific and technical merit of the draft and should be defined and explained carefully.

Second, while the draft correctly emphasizes the substantial risk of experts, attorneys, and other legal actors inadvertently transposing a conditional or of presenting testimony in a way that jurors are likely to transpose a conditional—that is, to confuse the probability of A given B as the probability of B given A—it makes the mistake of equating a "transposed conditional" with the "prosecutor's fallacy." These are distinct concepts—with the prosecutor's fallacy being a particular type of transposed conditional—and should be treated as such in the draft.

Third, the draft frequently includes unnecessarily extreme language—for instance, "It is always desirable to put things in terms of probability"; "It is logically meaningless to suggest that any evidence has value in itself as support for any particular proposition in isolation"; It is "not useful," or "mathematically meaningless," to say that something is "possible." Reasonable scientists could, and likely would, disagree with some of these statements. For purposes of accuracy and buy-in, the draft should employ more moderate language—and adopt more flexible positions (e.g., lines 932-35: "*the expert should be cautious* of assigning a probability of 1 or 0").

Fourth, and relatedly, the draft should adopt a more flexible position regarding an expert's response to questions about whether something is "possible." Although the STRP consensus view agrees that one's acknowledgment that something is "possible" generally has little probative value, depending on the circumstances, it may be appropriate for the expert to answer the question and consider providing further explanation. In addition, the "note" that accompanies the definition of the term "possible" (lines 249-53) should be removed.

Fifth, the concept of a "verbal scale" for communicating a likelihood ratio should be defined and developed more fully. The intended meaning and substantive application of such scales are fundamental to the scientific and technical merit of the draft, and the utility of any qualitative terms employed to describe various ranges of likelihood ratios should have empirical support or a substantial logical basis.

5.2. **Comment Received During Voting –** One commenter agrees with this view and further; DNA has always been the gold standard because we have avoided the use of subjective, personal and notional terms. We have always applied frequencies derived from scientific study to our sub-source and sub-sub source analysis. Firearms and toolmarks and fingerprints have worked for the last decade to move away from notional language for reporting the weight of a comparison. Why would

we allow this backslide in DNA at the highest level of analysis that we can possibly evaluate evidence on?

Notional probabilities could not be applied to sub-sub source and sub source level analyses because that would violate accreditation standards and legal case law. Numerical probabilities will also be required to accompany source and activity level evaluations. One of the STRP members raised concerns over the verbal scale because it has no statistical support for it's use, and I would make the same argument against notional probabilities. The term should be removed from this document and from DNA lexicon, in my opinion.

6. **Method Description:** There is no rule as to the necessary level of detail in the description of the method. Some parts of the method may be performed in alternative ways without affecting the quality and consistency of the results. Standards should focus on standardizing steps that must be performed consistently across organizations to ensure equivalent results. Alternatively, standards can define specific performance criteria that are required to be demonstrated and met rather than specifying the exact way a task must be done. For example, it may be enough to specify the lower limit for detecting a substance without specifying the equipment or method for achieving this limit of detection.

    6.1. **Consensus View** – Upon review, the STRP believes that the proposed BPR addresses a well-described, scientifically sound, and comprehensive method. However, in regard to reporting results with this method there are concerns with the noted differences with the models mentioned for assigning probabilities to the LRs that will be reported for DNA results given source level and activity level propositions. For source level propositions, Recommendation 4.3.1.1.1 does not give a description of the model to be used. Recommendation 4.3.2.1 later describes a process for how to use probabilities or an LR to communicate source level findings, but it is still unclear on the model itself for how these probabilities are to be assigned. A better description of the model (e.g., Bayesian networks) should be included here along with any published references.

    For activity level propositions, the STRP has similar concerns with Recommendation 4.4.2.1 and the subsequent "Note" as written. This proposed BPR states that one may use a model (not defined) or may use notional or personal probabilities. This potentially conflicts with the ISFG recommendations (2020) which state that whenever possible, relevant published data should be used. In the absence of published data, calibrated experience, case-tailored experiments, and peer consultation can be used but this is not intended to be "personal" to the scientist. Instead, ISFG emphasizes that it is the responsibility of the scientist to represent the view of the community of scientists such that similar views would be expected by other informed individuals as far as it is possible. If notional probabilities are to be used in the LR that will be reported, they must be accompanied by data sources that can be independently reviewed which support these probabilities.

The proposed BPR also notes in 4.4.2.2 that the LR should be communicated numerically *OR* by using a qualitative verbal statement. If a verbal statement is to be used, it must come with the assignment of a numerical LR as mentioned with sub-sub-source, sub-source, and source level LRs. Again, the ISFG recommendations (2020) state that a verbal scale is optional but cannot be used by itself for evaluating DNA results considering activity level propositions.

The methodology described in this document of rendering conclusions on the activity level while testifying is not in line with the guiding principles of interpretation: balance, logic, **robustness** and **transparency**. Laboratories in the U.S. commonly report at the sub-sub-source or sub-source level. At the sub-sub source, sub-source and source levels of the hierarchy, we recognize that we would perform all analysis in advance of trial, write reports, submit our work for technical and administrative review and issue a written statement describing our opinions, in line with common accreditation standards. So, the recommendation to perform analysis at an activity level is concerning. The provisions in the activity level section including evaluations of findings given proposed activities for the first time in court and ad-hoc assignment of notional probabilities that do not involve frequencies or calculations are not **robust**. This practice does not allow for an expert to research literature, locate relevant data or perform experiments. Another qualified analyst does not review this analysis to ensure it is scientifically sound, which is not **transparent**. Relying on unspecified training and experience, as well as unspecified resources, is not recognized by the scientific community or the normative references listed in this document.[3] The suggested diversion from published, recognized best practices for communicating evaluations of findings given proposed activities to address a perceived knowledge gap requires revision and further discussion before it is suitable to publish as a standard.

6.2. **Minority View** – This draft section touches on important points but requires simplification. For example, the description can be simplified to say as follows: The draft BPR applies a scientifically sound and well-grounded method. However, it is concerning in a number of respects. First, the draft BPR is vague in its descriptions of models for assigning probabilities—e.g., for source-level propositions in section 4.3.1.1.1 (*see also* section 4.3.2.1) and activity-level propositions in section 4.4.2.1. Second, the draft BPR seems to support the use of notional probabilities without providing a well-supported empirical basis for such testimony. The STRP views this as problematic and arguably contrary to accepted principles and practices (*see, e.g.*, 2020 ISFG recommendations). Similarly, the draft BPR supports the use of qualitative verbal statements as an alternative to numerical expressions of likelihood ratios. However, the BPR does not provide or require a well-supported

[3] DNA commission of the International society for forensic genetics_ Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II_ Evaluation of biological traces considering activity level propositions (isfg.org): "*Whenever possible, relevant published data should be used. The source of the knowledge should be disclosed and the limits of the data discussed. In the absence of published data, calibrated experience, case tailored experiments, peer consultation can be used. In any case, the use of the data needs to be justified and the expert should be transparent on the limitations of the data. Viewpoints, based solely on personal casework experience should be avoided*."

empirical basis for such standalone expressions. The STRP views this as problematic and arguably contrary to accepted principles and practices (*see id.*).

7. **Reporting Results:** Methods must not only be well described, scientifically sound, and comprehensive but also lead to reported results that are within the scope of the standard, appropriately caveated, and not overreaching.

    7.1. **Consensus View** – The STRP believes that this standard only provides substantive guidance for reporting via verbal communication of results in testimony. Although the written reports are mentioned throughout as a supplement to verbal communication/testimony, there are no requirements or guidelines in this document to help analysts and laboratories with formal report writing, required components of written reports, and the requirement of peer review. The standard also does not refer the reader to another standard or document where such guidance can be found.

    For the reasons already described in the Methodology section of this STRP report, any opinions not reported in a formal written report and disclosed to both legal parties in advance of trial should *not* be considered robust, transparent, or legally admissible. The STRP believes that the framework described in this standard could be misused if there is no accompanying requirements or guidance for the written reports on the results that form the basis for such testimony.

    In addition, there needs to be an adequate description of who should and should not be using the framework outlined in the standard, and clear definitions about when this framework should and should not be used. Without these elements, there is a serious risk of personnel testifying to results that they are not trained to complete, or testifying to results that have not undergone formal review or formal reporting procedures. In addition to potentially leading to miscarriages of justice, this would do little to change the recognized training gap that exists presently—the gap this standard seeks to address. In fact, it may actually exacerbate the issues that arise from the training gap.

    If the subcommittee intends any guidance on formal report writing and associated elements to be left out of this proposed BPR, then this should be stated in the "Scope" for clarity, especially given there are references to formal reports in the document itself. The same is true for any training and authorization required for a DNA scientist to use this method—this is also noticeably absent and necessary for appropriate application of the guidance in this standard. For example, adding language to the Scope indicating that "this standard was not created to provide guidance on formal written reports that should be used to support the testimony described in this framework, or the training and assessment that should occur before an analyst employs this framework when giving testimony. These necessary elements are outside the Scope of this document and should be addressed by the laboratory prior to encouraging analysts to apply this framework to their oral communication of their results."

7.2. **Minority View** - This section touches on important points but requires simplification and modification. For example, the description can be simplified and narrowed (although with one additional point) to say as follows: The draft BPR provides guidance for reporting results via testimony. There is a significant potential for misuse in the absence of guidance for written reports that form the basis of such testimony. The draft BPR does not refer the reader to other standards or sources for such guidance, and it does not adequately address written components of an expert's analysis. The BPR should clarify its intended scope in this regard and should discuss the importance of a written report with respect to an expert's testimony. In addition, as suggested above, the STRP finds problematic the draft BPR's support of an expert's use of notional probabilities and standalone qualitative verbal statements to describe likelihood ratios. Testimony in this regard in the absence of a firm empirical basis is problematic and arguably contrary to accepted principles and practices (*see id.*). Finally, the draft BPR provides inadequate clarity regarding when and by whom the framework described in the draft BPR should be used.

Comment Received During Voting - The BPR needs to be explicit about how weight of evidence should be described to maximize the chance that triers of fact understand the value of the evidence.