Human Forensic Biology

Organization of Scientific Area Committees (OSAC) for Forensic Science

Human Forensic Biology Subcommittee Response to NIST DNA Mixture Interpretation Foundation Review (NISTIR 8351 DRAFT)

Submitted August 23, 2021

OSAC has a mission to strengthen forensic science through the development and promotion of the use of high quality technically sound standards and best practices for forensic laboratories. With regard to this mission, there are validation, reporting, testimony, and research needs related issues in the NISTIR 8351 draft report on DNA Mixture Interpretation that we feel should be addressed as identified by the OSAC Human Forensic Biology Subcommittee. We have outlined the main topics of concern below.

Chapter 2

1. Exhaustive propositions

Principle 16 (lines 2350-2354) states "Assessing the strength of evidence in favor a [sic] proposition (hypothesis) H1 requires at least one other proposition (hypothesis) H2. These propositions H1 and H2 are required to be mutually exclusive and exhaustive. Strength of evidence assessments depend on the framework of circumstances within which they are evaluated." Lines 3572-3575, appear to cite Gittelson et al. 2018 out of context: "...as it has been noted: "The truth lies in the propositions: either the prosecution proposition is true or the [defense] proposition is true" (Gittelson et al. 2018). The implicit assumption in this statement is that the propositions are exhaustive."

The assertion that propositions need to be exhaustive is incorrect. These propositions do not need to be exhaustive. They are more useful when they represent each of the competing views of the parties, i.e. "exhaustive within the context of the case", but this is not the same as "being exhaustive". This is explained in the literature, as well as in numerous guidance documents.

Following currently available guidance documents and literature, the OSAC-developed standard pending at ASB (ASB Standard 041) Assigning Propositions for Likelihood Ratios and a best practice document currently in development in the OSAC Human Forensic Biology Subcommittee regarding evaluative Forensic DNA testimony explain that the propositions do not need to be exhaustive.

Chapter 4

1. Bracketing and factor space related to validation

The attempt in this report to propose validation studies through the concepts of "factor space (coverage)" and "bracketing" are confusing and potentially misleading.

Section 4.1.4. seems to suggest validation experiments that test the entire factor space, which is potentially misleading. This idea is revoked later in Chapter 4, e.g., in lines 3465-3466: "It is unrealistic to obtain and examine the volume of samples needed in order to provide complete coverage of the potential factor space with DNA mixture interpretation." Instead, the authors propose "bracketing", which is described as "consider(ing) results from samples that are both more complex or less complex than the casework sample of interest as a pragmatic way of understanding case-specific reliability of an interpretation system." If the authors are suggesting "case-type profiles (…) that represent (in terms of number of contributors, mixture ratios, and total DNA template quantities) the range of scenarios that would likely be encountered in casework" that "include compromised DNA samples (e.g., low template, degraded, and inhibited samples)", this is already covered in SWGDAM documents as well as ANSI/ASB Standard 020, Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol, First Edition, 2018 and in ANSI/ASB Standard 018, Standard for Validation of Probabilistic Genotyping Systems, First Edition, 2020 (both published and on the OSAC Registry). If this is not the intent, the actual intent and reasoning as well as specific gaps that are not covered by the aforementioned documents should be specifically addressed.

2. ROC plots (i.e. ROC curves)

ROC plots are mentioned as an acceptable way to assess the reliability of PGS. In this document, the authors write that ROC plots "have been used in evaluation of PGS systems previously (e.g., Bleka et al. 2016b, You & Balding 2019)" (lines 3722-3723). This statement is confusing, as both of the cited studies use ROC plots only for the comparison of models on a specific dataset, and not for the evaluation of a PGS system per se. We were unable to find any study where a PGS system was validated through the presentation of ROC plots. ROC plots can be useful (e.g., for the comparison of different models all run on the same dataset), but the literature is full of alternative methods, such as violin plots, LOWESS plots, Hd-True testing, and more. However, with such stress on using ROC plots to test the reliability of PGS, it would seem counterintuitive to rely on the output of the PGS system being tested to generate the ROC plots.

ROC plots present many limitations. Currently, there are no SWGDAM standards, no OSAC-developed standards, or as far as we are aware, any other documents that specifically call for the use of ROC plots in PGS validation studies. There are a variety of valid statistical approaches and tools available for laboratories to use in order to understand the limits of PGS systems. A foundational review should discuss other reliable and valuable options found in the literature.

3. Independent assessment and data availability

Key Takeaway #4.3 states, "Currently, there is not enough publicly available data to enable an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software (PGS) systems." Upwards of 60 published papers and studies are cited in this report, encompassing 1000's of samples and many millions of probabilistic

comparisons. It would be helpful to provide information as to how much data would be enough. Additionally, along with all the cited publications, SWGDAM guidelines for the Validation of Probabilistic Genotyping Systems as well as ANSI/ASB Standard 018, Standard for Validation of Probabilistic Genotyping Systems, First Edition, 2020 have been published. If followed, these promote reliable DNA mixture interpretation practices and would also be assessed during accreditation.

The second part of this Key Takeaway #4.3 states "….we encourage forensic laboratories to make their underlying PGS validation data publicly available….". Traditionally, journals have not been receptive to the publication of multiple internal validation studies. This NIST report suggests addressing this issue by stating laboratories need to be willing to let anyone access their data, for example, from a website. Web hosting and curation could represent other resource limitations (budgetary and personnel) that many, particularly smaller laboratories, cannot accommodate. There are also ethical concerns related to the publication of DNA profiles. Most validation samples are collected from volunteers giving informed consent, possibly following Institutional Review Board review, approval and documentation. There is no expectation that such volunteers would agree en masse to have their DNA profiles publicly available on the internet. Government collected samples must remain private according to the Privacy Act of 1974 (5, U.S.C. 522a). Placing validation samples on the internet disregards ethical considerations of genetic privacy.

4. Validation performance results in the case file and report

Key Takeaway #4.7 states, "… To enable users of results to assess the degree of reliability in the case of interest, it would be helpful to include these validation performance results in the case file and report."

We acknowledge that no current standard, guideline or best practice document from any organization or accrediting body suggests this but it is unclear if this means to include a validation summary or all of the validation data. In practical terms, providing gigabytes of data along with the report, is not practical. The legal mechanism of discovery is put in place for exactly this and there are numerous ethical issues as stated above.

The authors should instead consider suggesting the development of a best practice document addressing how to provide a reasonable summary of validation studies and limitations that could be provided in the case record or published with the report (i.e. such as an appendix to a lab report). Not only is this more realistic, but it would also provide a framework that could be followed by all forensic DNA laboratories.

5. Reliability

Key Takeaway #4.4 states, "Additional PGS validation studies have been published since the 2016 PCAST Report. However, publicly available information continues to lack sufficient details needed to independently assess reliability of specific LR values produced in PGS systems for complex DNA mixture interpretation. Even when a comparable reliability can be assessed (results for a two-person mixed sample are generally expected to be more reliable than those for a four-person mixed sample, for example), there is no threshold or criteria established to determine what is an acceptable level of reliability."

Key Takeaway #4.4 focuses on the difference in reliability of simple mixtures and more complex mixtures. However, the report does not mention that the uncertainty of such samples is accounted for

in the magnitude of the LR. For example, in simple two-person mixtures, LRs are commonly of magnitudes outside the scope of the understanding of a lay person. For difficult 4-person mixtures with degraded DNA and drop-out issues, the LR can be in the hundreds or tens. The difference is clear in the magnitude of the LR.

Reliability is a two-way street. DNA has exonerated hundreds of falsely incarcerated individuals and perhaps millions of falsely accused persons. It is unclear if this report means to infer that those exonerations should not have occurred if anyone cannot independently assess the reliability of mixtures both more and less complex for the factor space of difficult mixtures. Numerous organizations have put forth guidelines and standards dealing with validation and mixture interpretation for forensic DNA crime laboratories. These documents cover what is required and laboratories following these are assessed as such and are required to meet an acceptable level of reliability for accreditation. A foundational review raising questions about the reliability of mixture interpretation validations and PGS should reference specific gaps and expected outcomes not covered by current guidelines and standards It is ineffective to state that there is insufficient data to support reliability without defining the criteria necessary to make that determination.

Chapter 5

The majority of Chapter 5 relates to elements beyond mixture interpretation (such as the transfer of DNA and the hierarchy of propositions) and it is suggested that these items be removed from this document and considered for an additional document. It may be better to create a separate foundational review for source and activity level propositions as well as the transfer of DNA.

Although the hierarchy of propositions, Case Assessment and Interpretation (CAI) approach and Bayesian Networks (BN) are mentioned in this report, portions of Chapter 5 are in direct conflict with a document that is currently being drafted by the OSAC Human Forensic Biology Subcommittee regarding Best Practice Recommendations for Evaluative Forensic DNA Testimony which is in line with current published literature on this topic.

1. Hierarchy of propositions

Chapter 5 should draw a boundary between an evaluation for activity level propositions and an evaluation for offense level propositions.

This report appears to blend relevance, transfer, persistence and background DNA all together, making it unclear to know which discussions refer to activity level evaluations and what discussions refer to offense level evaluations. A clear, unambiguous distinction is crucial, because forensic scientists do not evaluate DNA results with regard to offense level propositions as the evaluation at the offense level is for the trier of fact alone.

The statement made at line 810: “Relevance should be assessed. If not, the evidence can be misleading.” is an example of contradicting the individual levels of the hierarchy of propositions. This is outside the scope of the forensic scientist/expert witness. The judge is the gatekeeper for the admission of evidence at court, and the jury is the only one that determines relevance. We suggest re-writing this chapter specifically to remove discussion of “relevance” and replace it with clear discussion focusing on the activity level of the hierarchy.

2. Investigative vs. Evaluative mode

We recommend adding discussion related to the differences between the investigative role of the forensic expert and the evaluative role. All examples and concerns seem to address the investigative mode but this is not clear in this document.

Page 7, lines 792, 803, 806 use the terms “readily”, “might have been”, and “might pick up”. These are all phrases that may be somewhat useful when a forensic expert is engaging with a case investigator, but we feel are inappropriate for the evaluative role of the expert at court. There is no discussion of this in the report.

In Key Takeaway #5.5, the transfer rate is not addressed in the statement “transfer easily between objects”. Without discussion of the evaluative role of the forensics expert, there is a very real danger that investigative terms such as “easy transfer” will be used in court. As transfer depends on numerous things (case-specific and otherwise), we feel this is an overstatement and is a research gap (i.e. under what specific case information and propositions can one base the assessment of “easy transfer”? What is the value of an appropriate alternate proposition?) rather than a given occurrence. Additionally, Key Takeaway #5.5, is an example of a one-sided DNA approach. Although this Chapter discusses the CAI Approach, (P 140, line 4911) there should be clear examples of a balanced approach where the evidence is considered under competing propositions. See further discussion below.

3. The danger of the transposed conditional in evaluative mode

There are sections of Chapter 5 that run the risk of being incorrectly incorporated into an evaluative role at court instead of investigative. An example is found in the Executive Summary on Page 7 of the full report, Lines 802-804: “First, that DNA might have been deposited before or after the crime was committed and therefore may not be relevant to the crime.”

There is no mention of investigative vs. evaluative mode related to this statement. If such a statement is made at court, this is an example of the transposed conditional; it is an evaluation of the following propositions: P1 = “before the crime” and P2 “after the crime”. There is no evaluation of the DNA evidence “if” or “given” those propositions. Evaluations of the propositions like this are for the trier of fact.

4. Formulation of activity level propositions

Activity level propositions stated throughout Chapter 5 appear to be in conflict with the ISFG Guidelines published in 2020 (Gill et al., (2020), '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.', Forensic Sci Int Genet 44, 102186) and with current OSAC documents in production (which are based on guidelines from ISFG, ENFSI, and others). On page 6 of the ISFG 2020 Guidelines, Recommendation 6 states: “Results or factors that scientists take into account in their evaluation should not be interwoven into the propositions. The scientist should avoid the use of the term ‘transfer’ in propositions. Instead, there should be a focus on the alleged activities.” The following statements in this report are examples that appear to violate this recommendation:

> "For example, in the case of a stabbing, the prosecution hypothesis might be that the DNA was transferred to the handle of a knife during the activity of stabbing, while the defense hypothesis might be that the DNA was deposited due to contamination or secondary transfer." (page 136, lines 4727-4730)
>
> "An activity proposition might be, for instance, that DNA collected during a sexual assault examination was deposited during sexual activity, or that DNA found on the handle of a knife was deposited during the act of stabbing a victim." (page 135, lines 4691-4693)

The correct formulation of activity level propositions for the former case would be "The POI stabbed the victim with this knife" for the prosecution's proposition, and "The POI did not touch this knife" for the defense proposition. The correct formulation of activity level propositions for the latter statement would be "that sexual activity occurred between the POI and the complainant, or that the POI stabbed the victim with this knife".

5. Significant knowledge gaps

Key Takeaway #5.6 states, "There is a growing body of knowledge about DNA transfer and persistence, but significant knowledge gaps remain." Unfortunately, it fails to give any specific details about where the significant knowledge gaps lie. One of the tasks of OSAC consists of identifying specific research needs. Without specific details, it is left to the reader to guess exactly what they are.

Conclusion:

In general terms, this NIST report reads more like a Standard or Best Practice document rather than a Foundational Review. We feel that OSAC would be able to incorporate the gaps established from a foundational review into best practice documents, standards and research needs to better serve the Forensic DNA community if the report was more specific about the validation requirements, research needs and knowledge gaps that exist. Finally, we believe a separate foundational review should be done for source and activity level propositions as well as the transfer of DNA in lieu of inclusion in this document.