INNOCENCE PROJECT

Executive Director
Christina Swarns, Esq.

Co-Founders & Special Counsel
Barry C. Scheck, Esq.
Peter J. Neufeld, Esq.

INNOCENCE PROJECT PUBLIC COMMENT ON
NISTIR 8351-DRAFT
DNA MIXTURE INTERPRETATION: A NIST Scientific Foundation Review
August 23, 2021

The Innocence Project is pleased to respond to the National Institute of Standards and Technology (NIST) call for public comments regarding the NISTIR 8351-DRAFT report, *DNA Mixture Interpretation: A NIST Scientific Foundation Review*. For nearly 30 years, the Innocence Project has worked to exonerate the innocent and to prevent wrongful convictions through systemic reform. The vast majority of our exonerations were achieved by the power and strength of forensic DNA evidence. Based on our decades of experience and success, we respectfully submit that as DNA analysis and interpretation becomes more complex, it must be applied with transparency and proper safeguards in order to ensure that forensic DNA technology serves its full potential to exonerate.

This scientific foundation review ("the report") is the first of the series produced by NIST and it may prove to be among the landmark publications in forensic science scholarship. In carrying out the work of this review, we commend the authors for operating with transparency, actively disseminating information regarding its process at conferences across the country, and now holding a public comment period to receive feedback. The feedback we respectfully offer addresses: (1) parts of the report we believe are critical to scientific rigor in forensic science and should therefore be retained, (2) concepts in the report for which stronger language or clearer directives are needed, and (3) parts of the report that we believe may be misinterpreted, manipulated, or create problems for the justice process without more context. Our comments below are listed using a chart that emulates the public comment process for standards development work and is organized by these three categories.

Critical Report Components to Retain

The following comments reference language in the report that are important to retain to ensure that the practice of forensic DNA testing is based on policies and protocols that promote a sound, quality, and just enterprise. With respect to concepts and language repeated or used multiple times throughout the report, we may reference a selection of excerpts but intend for our comments to apply globally. When new language or edits are suggested to resolve comments regarding excerpts

of the report, ~~strikethroughs~~ are used to indicate text that should be deleted and **[bracketed and bold text]** indicate text that should be added.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
512-514	All scientific methods have limits. One must understand those limits to use a method appropriately. This is especially important in forensic science as critical decisions impacting life and liberty are often based on the results of forensic analysis.	This critical opening line in the Executive Summary sets the tone for the rest of the report.	This language should be retained in the final report.
699 and 1039 704-709	Reliability is not a yes or no question, but a matter of degree. Reliability centers on trustworthiness established through empirical assessments of available data to evaluate the degree of reliability of a system or its components. We use the term "factor space" to describe the factors that influence complexity, measurement, and interpretation reliability – these factors include the number of contributors, the degree of allele sharing, the ratios of mixture components, and the amount and quality of the DNA tested.	These lines denote the complexity of evaluating reliability of forensic science methods and are deeply connected to the reasons why forensic laboratories must practice transparency in every aspect of their work.	This language should be retained in the final report.
714-717	In addition, reliability cannot be established without validation tests using known samples of similar complexity. The results of such tests provide data that are considered accurate and reliable; only with such valid results can comparisons be made as to the reliability of unknown casework samples.	Validation studies are not typically provided in the discovery process. These lines demonstrate that one cannot properly interpret the findings of a test without first establishing the range of the forensic laboratory's valid testing capacity. It is important for courts to recognize this concept as attorneys seek validation study data. If validation studies were published online it would conserve criminal process resources and ensure that attorneys are able to access the data.	This language should be retained in the final report and a recommendation should be made for forensic laboratories to publish their validation studies online.
677-682 and 2123-2124	KEY TAKEAWAY #2.6: Likelihood ratios are not measurements. There is no single, correct likelihood ratio (LR). Different individuals and/or PGS systems often assign different LR values when presented with the same	Likelihood ratios (LRs) are difficult to communicate and are not well understood by the public. Great care needs to be taken when presenting LRs, especially when LR values are low and when these values are uninformative for	This language should be retained in the final report and a recommendation should be made that forensic laboratories set a threshold below which a

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
868-871 and 4646-4647	evidence because they base their judgment on different kits, protocols, models, assumptions, or computational algorithms. Empirical data for assessing the fitness for purpose of an analyst's LR are therefore warranted. Takeaway 5.4: DNA statistical results such as a sub-source likelihood ratio does not provide information about how or when DNA was transferred, or whether it is relevant to a case. Therefore, using the likelihood ratio as a standalone number without context can be misleading.	decision making. Forensic laboratories should heed these recommendations when developing guidelines and standards for LR testimony.	result should be reported as uninformative.
1141-1144	First, forensic genetics is an evolving field, and this study can only provide a snapshot of the state of the science at a particular moment in time. Therefore, the literature and empirical evidence we discuss in this review will be incomplete as soon as it is published, as is the case with evidence reviews in other evolving fields such as medicine and public health.	The authors model the scientific principle that knowledge accumulates, and that science is ever evolving and changes over time. This concept contrasts with legal traditions and its inclusion here is an important assertion of values that are central to science.	This language should be retained in the final report.
1495-1501	This overall process can be divided into two parts (Figure 2.1): (1) measurement that involves a series of steps to generate a DNA profile and (2) interpretation of the DNA profile to help fact finders understand the value of the evidence. The measurement steps result in an electropherogram (EPG), which is a representation of the DNA profile observed from the test sample at specific DNA locations. Interpretation of the EPG concludes with a written report describing a strength-of-evidence statistic for Q-to-K comparison with the POI(s), and in some cases, court testimony.	The separation of the DNA testing process into measurement and interpretation phases is important framing for communicating the critical importance of validation studies.	This language should be retained in the final report.
1510-1511	Measurements reflect the physical properties of the sample while	These lines communicate that interpretation is subjective. The	This language should be retained in the final report.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
658-662 and 1528-1529	interpretation depends on the DNA analyst assigning values that are not inherent to the sample. KEY TAKEAWAY #2.2: Generating a DNA profile involves measuring the inherent physical properties of the sample. Interpreting a DNA profile involves assigning values that are not inherent to the sample. To do this, the DNA analyst uses their judgment, training, tools (including computer software), and experience, and considers factors such as case context.	measurement must first be done correctly, but the interpretation is based on many more factors. Analysts should be encouraged to be honest about the subjective nature of interpretation. During testimony, analysts frequently present their interpretation of the data as an objective process to avoid the accusation that they misinterpreted the evidence.	
775-777 and 3594-3595	KEY TAKEAWAY #4.8: We encourage a separate scientific foundation review on the topic of likelihood ratios in forensic science and how LRs are calculated, understood, and communicated.	Given the complexities of explaining the likelihood ratio and recommendations by Lund and Iyer (2017) to limit its use to personal decision making, a deeper evaluation of the LR would be beneficial to the field of forensic science.	We encourage NIST to conduct a separate scientific foundation review on the topic of LRs and their impact on the presumption of innocence.
2943-2946	On the other hand, when serving as an expert witness in a court setting, a forensic scientist is the provider of information while a trier of fact (judge or jury) and lawyers asking questions in the admissibility hearing or trial are users of the provided testimony. In this case, the judge, jury, and lawyers determine whether sufficient information has been provided.	The discussion of providers and users provides an important foundation for the discussion of transparency and data sharing. Importantly, the report also includes criminal justice stakeholders among the users of the data. Stakeholders missing from this list include defendants and persons of interest. People accused of crime should have access to all the data relevant to any evaluation of the evidence against them. Although a forensic scientist is responsible for answering the questions that are asked of them, when they are asked a question that is misleading or will lead a factfinder to an incorrect conclusion about the facts in a case, a forensic scientist should respond beyond the direct question asked to ensure the results are communicated accurately and clearly.	Defendants and persons of interest should be included among the defined “users” in section 4.1.5.
2991-2993	When publishing developmental	These passages demonstrate that as	This language should be

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	validation results with a new STR typing kit, 2992 the goal of mixture studies is typically *to demonstrate detection of minor alleles rather than accuracy with interpreting and/or deconvoluting mixture profiles*.	currently produced, mixture studies are overly simplified and insufficient for assessing the reliability of a testing system. More attention needs to be paid to validation studies for complex samples.	retained in the final report.
3009-3011	With these developmental validation studies, rarely is more than a single two-person mixture examined with the mixture ratio being the primary variable explored.		
3412-3414	Information on DNA quantities examined, mixture ratios studied, and degree of allele sharing in these five-person mixture samples was not explicitly stated in the referenced public sources.		
3343-50	After comparing results from 15 contributing laboratories, all laboratories could only identify every minor allele in the prepared mixtures between mixture ratios of 2:1 and 1:2. They could detect ~50% minor alleles at a 9:1 ratio and ~17% at a 19:1 ratio (Krenke et al. 2002). Instrument and assay sensitivity have improved in the past two decades so it is expected that lower-level minor contributors are detectable now across multiple laboratories. This aspect has not been specifically explored in published STR typing kit developmental validation studies or DNA mixture interpretation interlaboratory studies.		
3192-3195	The factor space for DNA mixture interpretation is vast and increases significantly with more contributors (Lynch & Cotton 2018). It is therefore practically impossible to demonstrate reliability across the full extent of any factor space.	These passages are akin to setting off alarm bells on complex DNA mixture interpretations. The LRs derived from complex mixtures have not been demonstrated to be reliable because the factor space is so vast that it cannot be captured. Without established criteria for assessing reliability and when bracketing approaches and sanity checks do not solve the problem, the	These passages and Key Takeaways #4.3 and #4.4 are important to retain in the final report.
3201-3207	Based on an examination of publicly available information reviewed during		

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	the time frame of this study, there is not enough information for the authors of this report to independently assess the degree of reliability of DNA mixture interpretation at any one point in the factor space. This is particularly true without an established and accepted criteria for reliability with complex mixtures involving contributors containing low quantities of DNA template (e.g., Benschop et al. 2015a) or where there is a high degree of allele overlap among contributors (e.g., Bright et al. 2018, Lin et al. 2020).	user must rely on "an element of trust in the DNA interpretation system used." This is a considerable ask in an adversarial system in which forensic laboratories produce DNA testing for law enforcement customers, and in an opaque system that does not freely share foundational data.	
3209-3212	A bracketing approach (discussed in Section 4.4.5) may provide a pragmatic way to infer reliability for DNA mixtures in a region of the factor space, *but will still require an element of trust in the DNA interpretation system used* since the entire factor space may not be covered with previously collected validation data		
3221-3222	However, such "sanity checks" with observed trends in LR values do not demonstrate the reliability of a specific LR number.		
868-871 and 4646-4647	KEY TAKEAWAY #5.4: DNA statistical results such as a sub-source likelihood ratio do not provide information about how or when DNA was transferred, or whether it is relevant to a case. Therefore, using the likelihood ratio as a standalone number without context can be misleading.	The vulnerability and the potential for miscommunication when using the LR as a standalone number exemplifies why transparency in criminal cases is so important. The context of a case cannot be understood without all the available information. Analysts should also keep in mind that tunnel vision or other confirmatory biases can inadvertently impact reporting and testimony. Transparency is a form of insurance to guard fulsome testimony.	This language should be retained in the final report.

INNOCENCE PROJECT

Concepts that Deserve Stronger Language or Clearer Directives

The following comments reference concepts in the report that deserve to be elevated or stated more directly and forcefully to ensure the weight of these passages is fully communicated to criminal justice stakeholders. With respect to concepts and language repeated or used multiple times throughout the report, we may reference a selection of excerpts but intend for our comments to apply globally. When new language or edits are suggested to resolve comments regarding excerpts of the report, ~~strikethroughs~~ are used to indicate text that should be deleted and **[bracketed and bold text]** indicate text that should be added.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
1075-1083 4578-4579	Forensic laboratories have been using procedures to avoid contamination since the advent of DNA methods. However, because the likelihood of detecting contaminating DNA has increased with highly sensitive DNA methods, contamination avoidance in forensic laboratories is more important than ever. Furthermore, contamination avoidance procedures should be used during all stages of an investigation, including at the crime scene. Elimination databases that include DNA profiles of laboratory staff and police who go to crime scenes can help identify contamination and should be maintained. Therefore, relevance should be carefully assessed and considered by both the DNA analyst and users of the DNA results, especially when an evidence item contains very small amounts of DNA. KEY TAKEAWAY #5.3: Highly sensitive methods increase the likelihood of detecting contaminating DNA that might affect an investigation. Contamination avoidance procedures should be robust both at the crime scene and in the laboratory.	This recommendation should be strengthened into a takeaway. As an ethical matter, if the government can compel DNA samples from defendants for the purposes of crime investigations, then its own agents should also provide their DNA profiles for elimination databases. Elimination profiles may simplify the interpretation of a complex mixture, are important when minor donor(s) are present in small amounts, and can prevent wrongful convictions in the event a DNA mixture is inadvertently misinterpreted.	Strengthen the importance of elimination databases by integrating this concept into Key Takeaway #5.3: KEY TAKEAWAY #5.3: Highly sensitive methods increase the likelihood of detecting contaminating DNA that might affect an investigation. Contamination avoidance procedures should be robust both at the crime scene and in the laboratory. **[These procedures include the maintenance of elimination databases that include both analysts and police and improved protocols for evidence collection.]**
902-905	An overall assessment of 1) how a new technology works, 2) what its limitations are, and 3) how it might	In order to fully evaluate whether the implementation of a new technology is worthwhile, the overall assessment	These lines should be revised to state the following:

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	specifically address the problem to be solved (e.g., DNA mixture interpretation) is important and a key component of evaluating whether implementation will be worthwhile.	must also include a fourth item —-the social impact of the technology, especially with regard to justice and equity. We should be asking ourselves what it would mean to implement a technology in an American policing system beset by structural racial biases and absence of oversight and whether there are sufficient mitigations or interventions that would ensure just and equitable implementation. Even if NIST does not believe that this fourth item is within its purview, it should reference social impact as an essential component.	An overall assessment of 1) how a new technology works, 2) what its limitations are, ~~and~~ 3) how it might specifically address the problem to be solved (e.g., DNA mixture interpretation)**[, and 4) whether this new technology can be justly and equitably implemented in an American policing system beset by structural racial biases and absence of oversight]** is important and a key component of evaluating whether implementation will be worthwhile.
1731-1742 664-666 and 1743-1744	Section 2.3.3. Mixture Complexity Increases as Number of Contributors Increase KEY TAKEAWAY #2.4: DNA mixtures vary in complexity, and the more complex the sample, the greater the uncertainty surrounding interpretation. Factors that contribute to complexity include the number of contributors, the quantity of DNA from each contributor, contributor mixture ratios, sample quality, and the degree of allele sharing.	These sections of the report recognize that the complexity of a mixture is impacted by various factors, however, it does not indicate when a sample should be deemed uninterpretable. This report does not take a position on the number of contributors to a sample that would make it uninterpretable. Previously, PCAST set a limit of interpretability at a 3-person mixture where the minor contributor was 20% of the mixture. Probabilistic genotyping software programs make varied claims with regard to the number of contributors they can deconvolute, with one vendor claiming to be able to deconvolute up to seven contributors. It would be helpful for this report to provide users with guidance on how to evaluate the reliability of those claims.	Include information on the type of data that users need to evaluate the claims of probabilistic genotyping software programs, how to use that data to do so, and the current limits of deconvolution software.
2891-2895	With higher-order DNA mixtures, the potential factor space becomes vast (e.g., consider one aspect of the factor space with possible genotyping combinations as described in Lynch & Cotton 2018).	Without clear requirements for the scope and breadth of factor spaces that need to be tested, the reliability of a forensic laboratory's testing cannot be assessed and their capabilities cannot be compared with other forensic	Key Takeaway #4.4 should be edited to state the following: KEY TAKEAWAY #4.4: Additional PGS validation

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
2915-2917	Therefore, it is unlikely that laboratories have explored every possible region of this factor space and may not be comfortable commenting on the degree of reliability with especially complex samples. It is recognized that each laboratory has to demonstrate their own degree of reliability and that we must be careful not to pool data from different sources that may come with different assumptions and caveats.	laboratories. Key Takeaway #4.4 should be modified to call for the establishment of comparable factor space criteria that should be assessed across forensic laboratories.	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. **[Validation standards that articulate specific factor space testing are needed to establish comparable assessments of forensic laboratories' reliability.]** 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. **[These thresholds need to be defined.]**
2460-2493 2495-2584 2800-2804	3.1.2. Available Internal Laboratory Data 3.1.3. Available Proficiency Test Data and its subsections For DNA mixture interpretation, this means that samples with known genotypes, known number of contributors, known mixture ratios, known degrees of degradation, etc., have been tested using the process of measurement and interpretation, and results from such tests are available to provide the basis for stakeholders to assess the degree of reliability of the process	An insufficient number of forensic laboratories publish internal validation studies and proficiency test data. The availability of internal validation summaries as provided by Table 3.2 are insufficient as the lack of meta data makes independent evaluation impossible. Proficiency tests are woefully inadequate for fully assessing the performance of forensic analysts because they are simple and do not reflect the complexities of casework. Unless proficiency tests and their data are shared, users would not be able to assess the performance capabilities of a forensic laboratory.	Include an additional "Key Takeaway" in Chapter 3 that states: **[Transparency of laboratory data is a best practice and a professional obligation to the criminal legal system. Forensic laboratories have a professional responsibility to publish or share empirical data from their validation studies and proficiency tests in an accessible format and provide them upon request.]**

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
2909-2912	To assess reliability of any system, the factors that impact that system's performance need to be studied and evaluated. In attempting to address the question of reliability, we need to first understand what portions of the factor space have been explored and what were the experimental outcomes.	Given the overwhelming volume of content provided in the report regarding the need for transparency and how essential it is for users to interpret the reliability of a forensic laboratory's results, this section should include a "Key Takeaway" that states that transparency is not just a best practice, but a professional obligation. If data cannot be interpreted reliably without validation information and if performance of analysts can be demonstrated by proficiency tests, then there is no credible scientific or justice reason for hiding this data.	
2929-2931	A provider of information delivers this information and accompanying data in an accessible format to be used for assessment by the user. The provider also explains the relevance and significance of the information and data.		
3080-3082	If proficiency tests are representative of commonly seen casework in a forensic laboratory, then these results can also help assess what PCAST termed "validity as applied" (PCAST 2016).		
3127-3129	These CTS DNA mixture PTs involve single-source or two-person mixtures created from large quantities of DNA (hundreds to thousands of cells). In other words, the mixtures in the Forensic Biology, DNA Semen, and DNA Mixture PT exams (Table 4.6) are not complex.		
3190-3191	Demonstrating reliability requires that the provider provide empirical data that is accessible to users of the information for independent assessments of reliability.		
3061-3063	During our discussions on the topic of available data to assess PGS systems for DNA mixture interpretation performance, the DNA Resource Group (see Table 1.2) underscored that additional PGS data exists in forensic laboratories as part of their internal validation studies.		

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
3227-3230	However, to independently assess the degree of reliability of PGS models, metadata associated with specific sample results and the corresponding specific log(LR) value datapoints are needed. Data of this nature are not generally shared in publications or validation summaries.		
3367-3370	For example, many internal validation studies described in Table 4.5 do not clearly state the number of samples tested, making it difficult to assess the extent of the studies. The lack of availability of underlying data prevents independent assessments of reliability.		
3419-3424	Although more validation studies (see Tables 4.3 and 4.5) have been performed since the 2016 PCAST Report was released almost five years ago, in their present form, publicly available internal validation summaries often do not provide sufficient information to assess factor space coverage. Further, these summaries typically do not provide data points (e.g., LR values) and associated information (see Box 4.1) necessary to assess the degree of reliability and performance under potential case scenarios.		
3446-3451	Potential reasons why forensic laboratories choose not to make their internal validation data publicly available include: (1) the information from a study itself may not be publishable[23] due to lack of novelty (e.g., Buckleton 2009), (2) genotype data may include information from donors who did not consent to public sharing of their DNA profiles (e.g., Manabe et al. 2017), and (3) sharing foundational data is not required by current accreditation or guidance documents.		

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	fn23: The willingness of journals to publish validation studies is a separate issue from the willingness of laboratories to make data available on their website for anyone to download or at least sharing full data sets with credible parties in a timely manner when requested.		
3315-3325	As described earlier in Section 4.1**.6, a determination of whether the amount and type of data available is satisfactory or sufficient to the user of the information is something that must be decided by the user of the information (e.g., the DNA analyst), not the provider (e.g., the software developer).** It is not helpful for the provider to describe a method as "validated" without providing context around the method's use and access to data to support claims of validity and reliability. Instead, it might be more appropriate to state "the following developmental validation studies have been conducted and here is the complete collection of results obtained, which can be examined by users to make reliability judgments." Internal validation studies provide an opportunity for the user (e.g., DNA analyst) to understand performance of a method in their forensic laboratory environment rather than trusting the provider's (e.g., the software developer) claim that everything works fine.	This very strong statement in the report is not integrated into a "Takeaway" or key principle. The general policy of forensic laboratories is not to provide public access to this information and Table 3.2 provides evidence of this fact. Among the approximately 400 publicly funded forensic laboratories in the United States (DuRose, 2016), the NIST report was only able to identify eight forensic laboratories or laboratory systems that post internal validation study information online. To ensure fairness, transparency, and equity among jurisdictions, forensic laboratory policies should default to full disclosure. From the strong language in this report, we would assert that data sharing extends to proprietary software as well. Reference: DuRose, M. R., Burch, A. M., Walsh, K., & Tiry, E. (2016). Publicly Funded Forensic Crime Laboratories: Resources and Services, 2014 (p. 12). U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. https://www.bjs.gov/content/pub/pdf/pffclrs14.pdf	Revise Key Takeaway #4.2 to integrate the concept that data availability is based on user demand rather than provider prerogative: KEY TAKEAWAY #4.2: To enable effective use of any information, responsibilities exist with both providers and users of that information. While a provider explains the relevance and significance of the information and data, only the user can assess the degree of reliability, validity, and whether that information is fit-for-purpose. **[A determination of whether the amount and type of data available is satisfactory or sufficient to the user of the information is something that must be decided by the user of the information, not the provider. This provision applies to validation data as well as proprietary software.]**
732-755	KEY TAKEAWAY #4.1: The degree of reliability of a component or a system can be assessed using empirical data (when available) obtained through	Chapter 4 and Takeaways #4.1-4.4 discuss the critical need for and absence of empirical data. Users (defense attorneys) cannot evaluate the	Integrate language in this report to acknowledge that transparency of data is a professional obligation of all

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	validation studies, interlaboratory studies, and proficiency tests. KEY TAKEAWAY #4.2: To enable effective use of any information, responsibilities exist with both providers and users of that information. While a provider explains the relevance and significance of the information and data, only the user can assess the degree of reliability, validity, and whether that information is fit-for-purpose. KEY TAKEAWAY #4.3: 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. To allow for external and independent assessments of reliability going forward, we encourage forensic laboratories to make their underlying PGS validation data publicly available and to regularly participate in interlaboratory studies. KEY TAKEAWAY #4.4: 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 of a system without data. For this reason, validation studies must be made public. This report stops short of calling for forensic laboratories to publish this data. While NIST may avoid mandates or the use of "shall" or "must" language, this report takes great pains to describe the scientific cost of not making data public. NIST should establish transparency as a professional obligation to crystallize this responsibility for providers and users.	public forensic laboratory scientists and customers of public forensic laboratories. These stakeholders have an ethical obligation to eliminate barriers to data transparency.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	reliability.		
769-773 and 3487-34888	KEY TAKEAWAY #4.7: The degree of reliability of a PGS system when interpreting a DNA mixture can be judged based on validation studies using known samples that are similar in complexity to the sample in the case. 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.	Takeaway 4.7 represents critical information needed to assess the conclusions reached by the forensic laboratory. Validation data is not typically provided in the case file or the report and should become a regular practice as a matter of scientific principles, transparency, and justice. In *United States v. Gissantaner*, the trial court correctly identified that the Michigan State Police was conducting analyses outside the limits of what its DNA laboratory had appropriately validated. In *People v. Collins-Peaks*, the court excluded high sensitivity DNA testing and the NYC Office of Chief Medical Examiner's (OCME) Forensic Statistical Tool after a Frye hearing where the validation data for each method was assessed. However, it was not until the hearing that validation data was shared by the OCME. This language needs to be strengthened in proportion to the importance of validation studies on the assessment of the data. Validation studies are not "helpful," for interpretation, they are "essential" for evaluating the reliability of the results.	This takeaway should be edited to the following text: KEY TAKEAWAY #4.7: The degree of reliability of a PGS system when interpreting a DNA mixture can be judged based on validation studies using known samples that are similar in complexity to the sample in the case. **[Validation studies are essential to]** ~~To enable~~ users of results to assess the degree of reliability in the case of interest**[.]** ~~, it would be helpful to~~ **[Forensic laboratories that seek to fully and scientifically communicate their results will]** include these validation performance results in the case file and report.
3047-3048	*Table 4.3* includes a list of published validation data from peer-reviewed literature	The 2016 President's Council of Advisors on Science and Technology's report on forensic science raised the concern that published validation studies for probabilistic genotyping software were primarily authored by software developers and expressed the need for groups independent of these developers to publish such studies. Most of the publications referenced in Table 4.3 are published by software developers. Does NIST share PCAST's concern? <u>Reference:</u> President's Council of Advisors on	Include a discussion about the disproportionate availability of validation data from software developers compared to independent researchers or laboratories and how this may impact the body of knowledge in DNA mixture scholarship.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
		Science and Technology. (2016). Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. President's Council of Advisors on Science and Technology. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf	
913-915 and 5344-5345	KEY TAKEAWAY #6.2: Implementation requires a thorough understanding of the benefits and limitations of the new technology as well as the practical investment of time and effort put forth for its adoption by the laboratory.	Implementation of a forensic technology does not take place in siloes, but rather in a criminal legal system beset by adversarial tensions, structural racism, and deep disparities in how communities of color are policed and criminalized. The social implications of a technology and how it will be implemented within the criminal process needs to be understood as well. Simply ensuring validity, reliability, and application within limits does not mean that a technology should be unleashed freely.	Revised Key Takeaway #6.2 to reflect the need to understand the social impact of technologies: KEY TAKEAWAY #6.2: Implementation requires a thorough understanding of the benefits**[,]** ~~and~~ limitations**[, and social implications]** of the new technology as well as the practical investment of time and effort put forth for its adoption by the laboratory **[and the American policing system in which it will be implemented]**.

Report Components that may be Misapplied in the Justice Process

The following comments reference specific passages in the report that are vulnerable to misinterpretation and can potentially be manipulated to serve a purpose beyond what we presume to be the authors' intent. When new language or edits are suggested as resolutions to comments regarding excerpts of the report, ~~strikethroughs~~ will be used to indicate text that should be deleted and **[bracketed and bold text]** will indicate text that should be added.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
531-532	The findings described in this report are meant solely to inform future work in the field.	This report identifies critical best practices that forensic science service providers may not be following. These lines undermine the recommendations of	This sentence should be removed and replaced with the following text:

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
		the report, the ethical and legal obligations of forensic science service providers, and excuses forensic laboratories from taking responsibility and being held accountable for their past work. Not only do these lines appear to permit unacceptable work product that may exist, they dismiss the obligation of forensic laboratories to take action to remedy the problems as mandated by quality management principles. This line also offers attorneys and courts an argument against correcting mistakes or errors that impact the criminal legal system. As referenced in lines (513-514), this language may influence "critical decisions impacting life and liberty."	**[The findings in this report offer forensic laboratories an opportunity to reflect upon their current policies and protocols. Forensic laboratories should be reminded of their ethical and professional obligations when their policies and protocols fall short of best practices. Please refer to item #15 and #16 in the National Commission on Forensic Science's National Code of Ethics and Professional Responsibility for the Forensic Sciences.]**
1911-1914 1953-1956	That is, different people may assign different values to the same evidence. Concerns have been raised that the LR framework applies only to personal decision making and cannot automatically be used for the transfer of information from one expert to a separate decision maker (Lund & Iyer 2017). A common problem known as "transposing the conditional" (Evett 1995) or committing the "prosecutor's fallacy" (Thompson & Schumann 1987) can lead to a misunderstanding of the meaning of an LR result. In these situations, a user confuses "the probability of the evidence given the propositions" with "the probability of the propositions given the evidence."	These important passages help us understand that Bayesian frameworks and the use of LRs can have varying degrees of subjectivity or provide incomplete analyses of a phenomena. The difficulties of communicating the Bayesian framework and transposing the conditional is also described using the example of a cow and its four legs (Lines 1961-1964). Notably, however, Bess Stiffelman has written about how the Bayesian framework can violate presumptions of innocence in the forensic setting. Not only are results only given in degrees of inclusion, thereby eliminating the ability to exclude a person of interest, but one must assume that the prior probability of a defendant contributing to a DNA sample is greater than zero. Therefore, any presentation of results assumes a level of guilt and this is especially true when only LRs are presented. It is not clear how one can mitigate this problem given the legal and	Include a discussion of how the Bayesian framework impacts the presumption of innocence that include the following resources: Stiffelman, B. (2109). No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials. https://doi.org/10.15779/Z384Q7QQ6X Presser, J. R., & Robertson, K. (2021). AI Case Study: Probabilistic Genotyping DNA Tools in Canadian Criminal Courts. Law Commission of Ontario.

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
		mathematical complications of presenting a prior probability, but this complex issue should be raised in the report.	
2087-2089 2206-2215	Another possible source of variation in LR results comprises the estimated degree of co-ancestry in observed alleles, which involves using a subpopulation correction factor typically symbolized by the Greek letter theta (Balding & Nichols 1994, NRC 1996). Principle 6 [Genetics]: DNA profiles from close relatives are more similar than DNA from unrelated people DNA profiles from close relatives are expected to be more similar than DNA profiles from unrelated individuals (Li et al. 1993). There are a limited number of alleles at each locus, and even individuals who are not closely related will share alleles and genotypes. The frequency of occurrence of specific alleles and genotypes varies. *This principle is a reminder that while statistical models typically assume individuals are unrelated, if case context suggests closely related individuals may have contributed to the sample in question, then performing calculations assuming individuals are related may be helpful to decision makers.*	The problem of relatedness and the potential impact of this problem deserves more discussion. Forensic laboratories should account for relatedness in their testing protocols. Principle 6 should also explicitly discuss the real risk of falsely including a non-contributor relative in a DNA mixture.	Expand Principle 6 by adding the following text to (Line 2215): **[Not accounting for relatedness can increase the risk of falsely including a non-contributor relative in the DNA mixture.]**
3264-3270	We, the authors of this NIST report, emphasize that publicly available data from validation studies, whether or not this information has been published in a peer-reviewed journal, enable a user (e.g., the DNA analyst when the provider is the PGS developer or the court when the analyst is providing their results) to scrutinize the underlying data and supporting details for what is currently possible in research	The report goes into great detail regarding the inadequate scope and breadth of DNA mixture validation studies. This passage may be misused by stakeholders to erase the concerns raised by Key Takeaways #4.3 and #4.4. Taken out of context, the language can appear to convey that the state of DNA mixture interpretation is sound. This statement should be revised to more carefully reflect the intent of the authors' position and reduce its potential for manipulation	Revise this passage to the following: We, the authors of this NIST report, emphasize that publicly available data from validation studies, whether or not this information has been published in a peer-reviewed journal, **[can be used by]** ~~enable~~ a user

Lines	NISTIR 8351-DRAFT Language	Comment	Resolution
	settings (what PCAST terms "scientific or foundational validity") and what is actually happening in casework settings (what PCAST calls "validity as applied").	when quoted out of context.	(e.g., the DNA analyst when the provider is the PGS developer or the court when the analyst is providing their results) to scrutinize the underlying data and supporting details for what is currently possible in research settings (what PCAST terms "scientific or foundational validity") and **[ascertain whether]** what is actually happening in casework settings (what PCAST calls "validity as applied") **[reflects the claims of the validation data].**