Public comments received

on Draft NISTIR 8225
NIST Scientific Foundation Reviews
Published February 12, 2019

DRAFT NISTIR 8225: NIST Scientific Foundation Reviews was published on September 24, 2018. That document is available at:
https://www.nist.gov/topics/forensic-science/draft-nistir-8225-nist-scientific-foundation-reviews

The public comment period closed on November 19, 2018. This document lists all comments received. These comments will be included in the final version of the document, which will be published at a later date.

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

There are two places in the DRAFT NISTIR 8225 that strike me as having the wrong tone for a NIST document that lays out a scientific foundation:

1. The verb "opines" on line 460 seems inappropriate in describing a scholarly analysis. The two passages quoted in that paragraph are findings based on thorough investigation and analysis, not simply opinions. I suggest changing "opines" to "finds" or "found".
2. I would delete the list of six criticisms of the PCAST report from lines 887-895. Some of them sound antagonistic to PCAST and even to general best practices of science. The inclusion of this list does not add useful information about scientific foundations, and it seems to cast unsupported doubt on the validity of the PCAST report. If the list is retained, the draft document should be augmented with discussion of these criticisms and the arguments for and against each. But I think it's better to simply delete this list and reference 29.

These opinions are my own and not of the NAS.

Scott Weidman, Ph.D.
Deputy Executive Director, Division on Engineering and Physical Sciences
National Academies of Sciences, Engineering, and Medicine

Subject: NISTIR 8225 comment

Date: Thursday, September 27, 2018 at 9:52:47 AM Eastern Daylight Time

From: Bruce J Heidebrecht -State Police-

To: ScientificFoundationReviews

Line 250 needs an end quotation mark, and maybe a period.

 3 We recognize that methods used in various forensic disciplines may have differing levels of supporting background information available. As noted in the NRC 2009 report (p. 39): "…the term 'forensic science' is used with regard to a broad array of activities, with the recognition that some of these activities might not have a well-developed research base, are not informed by scientific knowledge, or are not developed within the culture

Bruce J. Heidebrecht
DNA Technical Leader
Maryland State Police
Forensic Sciences Division
████████ ████████████

Subject: Re: NISTIR 8225 comment

Date: Monday, October 1, 2018 at 8:29:25 AM Eastern Daylight Time

From: Bruce J Heidebrecht -State Police-

To: ScientificFoundationReviews

Line 675 needs bold font for "3.7.3.", as consistent with font for remainder of numbering system.

Bruce J. Heidebrecht
DNA Technical Leader
Maryland State Police
Forensic Sciences Division

███████████████████

On Thu, Sep 27, 2018 at 9:52 AM Bruce J Heidebrecht -State Police- <bruce.heidebrecht@maryland.gov> wrote:
Line 250 needs an end quotation mark, and maybe a period.

3 We recognize that methods used in various forensic disciplines may have differing levels of supporting background information available. As noted in the NRC 2009 report (p. 39): "…the term 'forensic science' is used with regard to a broad array of activities, with the recognition that some of these activities might not have a well-developed research base, are not informed by scientific knowledge, or are not developed within the culture

Bruce J. Heidebrecht
DNA Technical Leader
Maryland State Police
Forensic Sciences Division

███████████████████

Subject: DRAFT NISTIR 8225: NIST Scientific Foundation Reviews

Date: Monday, October 1, 2018 at 10:49:41 AM Eastern Daylight Time

From: Dror, Itiel

To: Butler, John M. (Fed), Taylor, Melissa (Fed)

CC: Dror, Itiel, ███████████

Hi John & Melissa,

I read with great pleasure that DRAFT NISTIR 8225: NIST Scientific Foundation Reviews.
It is very much a welcome & important effort.

I wanted to send to you directly two thoughts/comments about the document and review
effort. When looking at the things you will be reviewing and considering, and the 'Terminology
and Concepts', I see two things that are missing:

1. Cognitive bias: How much is the domain susceptible to influences from irrelevant contextual
information.
This is an important issue in any scientific endeavour and is in the NAS & PCAST reports,
but not really sufficiently in your document and review effort. The issue of bias is
particularly important in forensic science due to the crime related content of the domain, the
work conducted often within law enforcement, the nature of the adversarial legal system,
etc.

2. When thinking and considering the issue of repeatability, reproducibility, etc., you need to
tease apart the 'observations' from the 'interpretations'. You lump together these two
elements, but any differences & variability you find (within or between measures --what you
call repeatability and reproducibility) may be due to differences in either what is
perceived/observed as the data, VS. the interpretation of the data (e.g., in
fingerprinting, repeatability and reproducibility issues may stem from different
interpretations of what the minutia data mean, VS. due to what minutia data are
observed).

Happy to discuss further, or/and send you relevant supporting materials.

Itiel

Subject: NIST foundational review comments

Date: Thursday, October 4, 2018 at 3:01:05 AM Eastern Daylight Time

From: Norah Rudin

To: ScientificFoundationReviews

Thank you for the opportunity to review the draft document NIST foundational review.

I have two main comments:

1. Section 1.2 How will we evaluate the data.
It seems to me that an important aspect is missing: Is the underlying data provided as supplementary material so that others can replicate the experiments? [This same comment applies to section 3.7.1 Appropriate Scientific Literature]

2. Section 5 Terminology and Concepts, line 1061 Validation.
Of course it must be understood that "validation" is wholly a forensic concept, with no meaning in basic science. The definition provided in the document is the one that has historically led, and continues to lead, to weaknesses in forensic work. A better definition, that would be more useful and rigorous, is the following:
"Validation is performed to establish the capabilities and limitations of both the method, and of the samples, analyzed using the method. It must use samples for which ground truth is known to determine when the method or system will fail (support an incorrect inference). As such, the samples used to perform the validation must be examples of marginal/extreme samples that exceed the complexity expected to be encountered in casework. The information resulting from such validation must be used to inform interpretation guidelines. Casework samples exceeding the complexity of those samples used for validation must not be interpreted without performing additional validation work."

Thanks,

Norah Rudin

~~~~~~~~~~~~~~~~
Dr. Norah Rudin, Ph.D.
Forensic DNA Consultant
██████████████

www.forensicdna.com

███████████
████████████████████████
~~~~~~~~~~~~~~~~
Download or support Lab Retriever at scieg.org
~~~~~~~~~~~~~~~~
The Science in Forensic Science keeps the Right Guys on the Right Side of the Bars

Subject: Comments on NISTIR 8225

Date: Thursday, November 8, 2018 at 1:07:08 PM Eastern Standard Time

From: James L. Wayman

To: ScientificFoundationReviews

I certainly applaud this effort. My chief concern is that the elephant in the room has been ignored, although probably for good reason. Much of the forensic community seeks to do "source identification", seeking to answer the question "Are these two samples from the same source?" Those sciences are strongly divided between those that want to express conclusions as (a) the probability of the hypothesis knowing the evidence and those that express (b) the probability of evidence under various hypotheses.   Both approaches have very serious limitations. Assessing the "scientific foundations" of (a) and (b) would seem to require very different methods, as (a) can only be the examiner's subjective probability, hiding the past experimental learnings , while (b) can only be aleatory outcomes of physical experiments, hiding the subjective input of the experiment designer. Only those practicing naïve Bayesianism believe that the two approaches can be reconciled.

Ultimately the trier of fact wants to know (a): the probability of the hypothesis. If stating that subjective probability (maybe using a 5 point verbal scale) is the job of the forensic scientist, we can develop quantitative measures of the performance of the forensic examiners under a variety of conditions in a "black box test". The responses of the examiners are the data points. These tests are favored by the NIST fingerprint and face identification communities, for example.

But some argue that the job of the forensic scientist is limited to (b) developing the probability of evidence when tested under various hypotheses. In this case, we can develop quantitative measures of the evidence variation under differing test conditions, none exactly reproducing the unknown conditions of any real case and all reflecting the subjective nature of the measurement protocols. The metrics deriving from repeated experiments are the data points. Such physical data collection exercises are favored by the NIST DNA and speaker recognition communities.

So it seems to me that "black box testing" is by necessity an exercise in social science, while physical testing is an exercise in physical data collection with the social or subjective elements buried in the experimental methods.   The document "NIST Scientific Foundation Reviews" paper does not address the elephant. I suggest that the NIST Foundations paper discuss how scientific foundations of the two approaches differ and hence why current foundations of face identification and DNA analysis must be assessed differently.

James L. Wayman

Subject: Public Comment

Date: Tuesday, November 13, 2018 at 5:46:12 PM Eastern Standard Time

From: James Johns (MCSO)

To: ScientificFoundationReviews

Please accept my comments that follow on the <u>Draft NISTR 8225: NIST Scientific Foundation Reviews</u>:

I would like to respectfully suggest that the terminology used in the <u>Draft NISTR 8225: NIST Scientific Foundation Reviews</u> referencing friction ridge examination should use the broad term "friction ridges", rather than the narrower term "latent fingerprints", which appears multiple places in the draft. It is my opinion that the term "latent fingerprints" refers only to friction ridge impressions that are from crime scene investigations and are not readily visible and not intentionally recorded. The term "latent fingerprints" is narrow in scope and excludes the entirety of friction ridge examination, including tenprint and other intentionally captured friction ridge events. Further, use of the term "friction ridges" allows the entirety of the friction ridge examination discipline to be included in and make use of the literature included in the draft.

Respectfully,
James Johns
Fingerprint Analyst
Maricopa County Sheriff's Office/Records and AFIS Division/AFIS Unit
███████████████████████████
Phoenix, AZ 85003
████████████████
███████████████████

Subject: how to make validations retrievable

Date: Sunday, November 18, 2018 at 2:56:44 PM Eastern Standard Time

From: Mechthild Prinz

To: ScientificFoundationReviews

This may be beyond the scope off this foundational review, but I am sending this email to find out if there is anything that NIST could do to make validation and internal data retrievable.

I was happy to see that information gathering will not be limited to peer reviewed literature but include other data sources such as internal validations and training material e.g. workshops. This approach will come very close to covering the available body of knowledge.

But then most internal validations and other inhouse data sets will not meet the requirements listed under "retrievable" and "respected" and will have to be dismissed in the data evaluation phase. Could NIST create an online repository for the information? The foundational review could include a process where data sets and documents are externally reviewed and approved under the reliability criteria before they are allowed on the repository. This would increase the amount of foundational science for forensic disciplines lacking enough peer-reviewed research. If this controlled approach is not feasible, NIST could help facilitate online availability by providing a public online platform organized by disciplines for what in the document is called "open peer review".

Mechthild Prinz, PhD
Associate Professor Science Dept
John Jay College of Criminal Justice

AMERICAN SOCIETY OF
CRIME LABORATORY DIRECTORS, INC.

65 Glen Road, Suite 123, Garner, NC 27529

ASCLD BOARD OF DIRECTORS

Matthew Gamette, President
Idaho State Police

Brooke Arnone, President-Elect
Arizona Department of Public Safety

Ray Wickenheiser, Past President
New York State Police Crime Laboratory System

Linda Jackson, Secretary
Virginia Department of Forensic Science

Rita C. Dyas, Treasurer
Chandler Police Department

Kris Deters
Minnesota Bureau of Criminal Apprehension Forensic Science Service

Erin Forry
Boston Police Department

Arlene Hall
Illinois State Police

Bruce Houlihan
Orange County Crime Laboratory

Jennifer McNair
Utah Bureau of Forensic Services

Scott O'Neill
New York City Police Department

Timothy Scanlan
Jefferson Parish Sheriff's Office

Christian Westring
Member at Large

ASCLD STAFF

Jean Stover
Executive Director

Ramona Robertson
Administrative Assistant

November 16, 2018

Re: ASCLD Comments to NIST Draft NISTIR 8225, Scientific Foundation Reviews

The American Society of Crime Lab Directors (ASCLD) is a nonprofit professional society of over 600 crime laboratory directors and forensic science managers across the United States and worldwide, dedicated to providing excellence in forensic science through leadership and innovation. The purpose of the organization is to foster professional interests, assist the development of laboratory management principles and techniques; acquire, preserve and disseminate forensic based information; maintain and improve communications among crime laboratory directors; and to promote, encourage and maintain the highest standards of practice in the field.

On September 24th, NIST published *Draft NISTIR 8225, Scientific Foundation Reviews*. This publication describes NIST's approach to conducting scientific foundation reviews in forensic science, which seek to document and evaluate the body of scientific data underpinning forensic methods. This letter includes ASCLD's input to NIST in preparing its final document.

ASCLD is supportive of the need to demonstrate the scientific foundation of forensic disciplines and of NIST leading the initiative in true partnership with the forensic science community. As a neutral and objective scientific body, NIST is well positioned to conduct studies regarding the level of documentation existing to provide the current state, as well as highlight those areas needing additional research and documentation. Critical to this endeavor, is the extensive involvement, inclusion, input and direction from the relevant forensic disciplines and expertise to ensure all practical aspects and knowledge is included in this evaluation. Studying the foundation of heart surgery while including extensive involvement of heart surgeons would be expected as a given. This is not to understate the importance of having a neutral and objective scientific based organization conducting the study, but to illustrate the pivotal nature of specific experience and expertise needed to ensure the study is valid.

In fact, the inclusion of forensic expertise is noted in line 705-706 on Page 12 (point 2) specifically addresses the input from "a variety of outside experts". Lines 758-760 follows up with the need for forensic expertise involvement. ASCLD feels the need for forensic expertise is absolutely fundamental to ensure the resulting review is well informed, valid and accepted within the forensic community. Failure to involve a large component of forensic expertise would undermine what we believe could be a very positive step forward for the direction of forensic science disciplines.

Of note, line 802 begins a recommendation that those disciplines that do not pass muster should not have standards placed on the OSAC registry. ASCLD believes this point should be reconsidered. For disciplines that are in the process of developing their foundation, the creation of standards represents a key component of their development. Therefore, holding back disciplines from raising their science, processes, and procedures is counter-intuitive, and should be re-thought.

Some previous attempts at evaluating and guiding forensic science have suffered from severe shortcomings, the most critical of which were the lack of intimate involvement of relevant forensic experts and the inclusion of their input as the final document was being developed through its completion. Input must not only be sought, but taken into account, through a thorough vetting process of the comments received and the provision of multiple levels of draft documents for review. Inclusion of content from a well-seasoned and broad-based group of forensic experts not only provides input and guidance, it provides relevance, validity and acceptance with the community tasked with its execution and stewardship. Draft documents must be provided at multiple stages during the process to ensure that input opportunities are real and included. Properly done, the proposed foundational studies must truly be done in partnership with the forensic community.

There are approximately 400 plus crime laboratories across the US, with federal, state and local labs conducting a wide range of analyses under varying conditions and procedures. This broad base of representation will ensure relevance to the laboratories that are tasked with making improvements arising from the studies proposed. Therefore, ASCLD supports the NIST foundational review concept, with the caveat that extensive involvement of forensic expertise from our community is absolutely critical. As the organization which represents the leadership of the forensic community, ASCLD welcomes the opportunity to assist in leading this valuable initiative to ensure that it is valid and appropriately conducted.

Comments on: NISTIR 8225 DRAFT - NIST Scientific Foundation Reviews

by: Geoffrey Stewart Morrison

http://geoff-morrison.net/

geoff-morrison@forensic-evalauation.net

The draft is well written and the criteria for evaluating data reasonable.

I have two potentially substantive comment:

The draft does not appear to include a criterion as to whether the interpretation of observations (interpretation of evidence) is logically correct. I recommend that this be added at 1.2(2).

In may be alluded to under "validation in 5, but in 1.2(2) it is not made explicit that data used for testing should reflect casework conditions. I recommend that this be made explicit at 1.2(2). If not, test results will not reflect expected performance under casework conditions and thus will not be suitable for determining whether the system tested is fit for purpose.

Some minor points:

At 1.2(2) "Are statistically significant sample sizes used?" In context, I interpret this to mean something like "Are the data used for testing of sufficient size that the test results (whatever they may be) will be reasonably representative estimated of performance." I recommend that it be made explicit that the "sample" in question is a set of test data. Even if one were to take a frequentist null-hypothesis testing approach (see the American Statistical Association's statement on p-values, Wasserstein & Lazar, 2016), would it be correct to talk of a "statistically significant sample size"?

The section in the grey box "What is Science?" appears to tackle head on a much debated philosophical question. To avoid this debate, I would recommend rephrasing this and recasting the section as "What are the norms of modern scientific practice?"

I recognize that it is in parenthesis and thus perhaps should not be taken as an explicit invocation of frequentist null-hypothesis testing, and that it may also be a quote, but "(Falsifiable) hypotheses" seems out of place in a document related to forensic science, in which a likelihood-ratio approach would be more appropriate. I would recommend "Testable hypotheses".

If logically correct evaluation of evidence is required and meaningful empirical validation is required, then validation metrics should be consistent with the logically correct framework for evaluation of

evidence. False positive and false negative error rates (listed in 5) are not appropriate. A number of metrics and graphical representations have been proposed (see Morrison, 2011; Meuwly et al., 2017). The most widely accepted appear to be the log likelihood ratio cost (C_{llr}) and Tippett plots. I would recommend consideration of whether validation results have been appropriately quantified and represented.

Potential sources of bibliographies not included in the draft:

ENFSI Guidlines http://enfsi.eu/documents/forensic-guidelines/

OSAC documents, including some that may be drafts by subcommittees (for example the Speaker Recognition Subcommittee decently approved a document "Key literature related to human-supervised automatic approaches to forensic speaker recognition").

References:

Meuwly D., Ramos D., Haraksim R. (2017). A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Science International*, 276, 142–153. http://dx.doi.org/10.1016/j.forsciint.2016.03.048

Morrison G.S. (2011). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, 51, 91–98. http://dx.doi.org/10.1016/j.scijus.2011.03.002

Wasserstein R.L., Lazar N.A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. http://dx.doi.org/10.1080/00031305.2016.1154108

INNOCENCE PROJECT PUBLIC COMMENT ON
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
DRAFT NISTIR 8225: NIST SCIENTIFIC FOUNDATION REVIEWS
November 16, 2018

The Innocence Project submits this statement in response to the opportunity to provide public comments to the National Institute on Standards and Technology (NIST) on the *DRAFT NISTIR 8225: NIST Scientific Foundation Reviews*.[1] This submission begins with comments regarding the clarification of the purpose and scope of the reviews and then addresses the questions posted: (1)"Are the data sources identified in this document appropriate?" (2) "Are the criteria identified in this document for evaluating the data appropriate?" and (3) "Should additional elements be included in the scientific foundation review?"

Founded in 1992, the Innocence Project harnesses the power of DNA technology to free innocent people wrongfully convicted of crimes. The misapplication of forensic science is the second most common contributing factor to wrongful convictions, found in nearly half (45%) of the 362 DNA exoneration cases documented to date. In order to address and prevent the misapplication of forensic science, the Innocence Project's forensic science priorities focus on improving the empirical basis of and supporting initiatives that enhance the scientific foundations of forensic methods used in criminal proceedings. We are deeply committed to efforts that improve the accuracy, reliability and accountability of science used by the justice system. The seeds of the Scientific Foundation Reviews were planted by the National Commission on Forensic Science (NCFS).[2] On September 12, 2016, the NCFS recommended that NIST "should establish an in-house entity with the capacity to conduct independent scientific evaluations of the technical merit of test methods and practices used in forensic science disciplines." The implementation of this recommendation through the Scientific Foundation Reviews is a positive development. We believe that NIST is the right scientific agency for this important task and we applaud it for its continued pursuit of this commitment.

In anticipation of future Scientific Foundation Reviews, the feedback provided here is offered with the understanding that this document will serve as the underlying guideline for all future Scientific Foundation Reviews, which can potentially span across a wide range of forensic science

---

[1] Press, R. (2018, September 24). DRAFT NISTIR 8225: NIST Scientific Foundation Reviews. Retrieved October 31, 2018, from https://www.nist.gov/topics/forensic-science/draft-nistir-8225-nist-scientific-foundation-reviews

[2] National Commission on Forensic Science. (2016, September 12). Recommendation to the Attorney General Technical Merit Evaluation of Forensic Science Methods and Practices. U.S. Department of Justice. https://doi.org/10.17226/12589

disciplines. Rather than issuing a broadly stated document, we believe that the *Draft NISTIR 8225* would benefit from including more detailed guidance.

Clarification of the scope and purpose of the reviews.

- The document would be strengthened by a clearer description of the purpose of a Scientific Foundation Review. The purpose is described in the abstract as "to document and consolidate information supporting the methods used in forensic analysis and identify knowledge gaps where they exist." Section 1 frames the scope in these ways: "What empirical data exist to support the methods that forensic science practitioners use to analyze evidence?" and "…we will evaluate whether the selected features are characterized and measurable; to what extent the discriminating power of those features is known; and whether the factors that affect the transferability and persistence of those features are understood." Section 1.1, however, is framed as an inquiry into a method's reliability (presumably including measures of repeatability, reproducibility and accuracy), although this was not explicitly stated. Clarifying the purpose, and the questions to be addressed in a foundational review, is essential: the questions that are asked will determine what literature (studies) and data are to be sought; this needs to be specified with sufficient detail and clarity to allow the methods used to be understood and reproduced.

  The scope or parameters of the Scientific Foundation Review should also be described. Will a review assess currently available (state of the art) methods, or will it also undertake an assessment of methods that have previously been used, or that may currently be used but that are not the most technically advanced method available? For example, will the DNA mixture review chronicle the evolution of DNA interpretation, starting with manual deconvolution of two or more-person DNA mixtures, and continuing to the use of the Combined Probability of Inclusion/Exclusion and to the use of Probabilistic Genotyping Software? Will the review discuss the algorithmic differences in Probabilistic Genotyping Software that can lead to different conclusions for the same sample? Consequently, will the review provide guidance on how to handle circumstances when different software programs return contradicting results?

- If the goal of the document is to describe the methods to be used in a foundational review, it needs greater detail. We suggest that NIST revise and post a general description of the purpose and methods of the Scientific Foundation Reviews, and

that NIST also disseminates a more specific protocol in advance of commencing a discipline-specific review that includes topic-specific information on:

- Scope
- Specific questions to be addressed
- Criteria used to evaluate studies (published and unpublished)
- Any plans to conduct research as a part of the review. [This will depend on the question(s) you are asking and the data you find]

Our comments in the following section are bulleted below the relevant questions, as delineated in the request for comments:

1. **Are the data sources identified in this document appropriate? Should additional sources be used?**

   - Based on section 1.1 (What Data Sources Will We Use?) it appears that NIST will not be conducting its own intramural studies in its Scientific Foundation Reviews. In the second recommendation of the NCFS Recommendation for NIST to conduct Technical Merit Studies, the NCFS recommended that NIST include its intramural research programs to provide data for recommendations. Without the use of empirical data targeted to assessing a forensic discipline's scientific foundations, the NIST reviews may not be that much different than the many studies it references in Section 3 of the *Draft NISTIR 8225*. Some of these data may already be available for DNA mixtures, but where gaps in data exist, NIST should use its research capacity to fill those gaps.

   - The *Draft NISTIR 8225* states in Section 1.1 that NIST may rely on data from laboratory validation studies, proficiency tests, and other data sources that may not be available through published, peer-reviewed literature. We understand this wider approach to identifying relevant data can be useful in the study of forensic science as there are many factors, unrelated to the quality of the data, that influence if, when, and where a study is published. However, the use of unpublished data, as well as published data, should be limited to those collected using a methodologically sound design, for which the study methodology and results are or can be made publicly available to allow external review. As described in bullet #2 below, an evaluation by NIST scientists with expertise in methodology and statistical analyses using relevant criteria (see bullet #2, below) is necessary for both unpublished and published studies. These steps would justify the use of unpublished data and provide a process that would raise public confidence regarding its use.

One example of unpublished data is the study conducted by Adam Freeman and Iain Pretty examining reliability of bite mark evaluations by forensic odontologists. This study was presented at the 2015 annual meeting of the American Academy of Forensic Sciences as well as at other public meetings, and the methods and results are available through the Texas Forensic Science Commission reports.[3] The data from this study are clearly relevant to the Scientific Foundation Review of bitemark analysis, and so should be evaluated for this purpose.

2. **Are the criteria identified in this document for evaluating the data appropriate?**
   - The discussion of data sources in Section 1.1 and data evaluation in Section 1.2 places too much reliance on publication and number of citations as an indicator of the quality or soundness of a source of data. Additionally, reliance on "published" or "peer-reviewed" as an indicator of quality isn't adequate, given the vagaries of what passes for peer-review. "Retrievable" doesn't provide any assurance of the soundness of the design, implementation, analysis, and presentation of results. We also can find little support for the use of citation frequency as a measure of value. Publications used in the Scientific Foundations Review should be evaluated by NIST researchers with expertise in methodology and statistical analyses to meet the standard set by the National Commission on Forensic Science's View on Scientific Literature in Support of Forensic Science and Practice.[4] This independent assessment should address methodological strengths and limitations, appropriateness of the analytic strategy, and completeness of reporting.

3. **Should additional elements be included in the scientific foundation review?"**
   - A discussion of historical methods would be important to include in the resulting report. Such a discussion would put a method currently used in context with its position within the progression of technology in the field. The report should include a summary evaluation of these methods in addition to the method that is the focus of the review (e.g. the manual deconvolution of two or more-person DNA mixtures, the Combined Probability of Inclusion/Exclusion, Probabilistic Genotyping).

---

[3] Freeman A, Pretty I. *Construct Validity Bitemark Assessments Using the ABFO Bitemark Decision Tree*. Study details and results at: http://www.fsc.texas.gov/sites/default/files/FinalBiteMarkReport.pdf

[4] National Commission on Forensic Science (2015, January 30). Views of the Commission Scientific Literature in Support of Forensic Science and Practice. Department of Justice. https://www.justice.gov/archives/ncfs/file/786591/download

- In the course of evaluating literature or data for each Scientific Foundation Review, the researchers will identify (or suggestions will be offered for) data sources that do not address the specific question(s) under review or did not meet specific evaluation criteria. To avoid some of the critiques of past forensic science reviews, it would be useful to provide an inventory of the studies that were identified, noting which question(s) each study addressed. This will provide a transparent record of the studies used for each question, and importantly, a record of studies that were not used because they did not address a relevant question considered by the review. It does not have to be an overly burdensome process, as there are software options available to facilitate this type of task.

- The review should also include details and documentation of the evaluation process used and a summary of the results of the evaluation of the soundness of the methodology of each of the studies or data sources used in the review. The software used for systematic reviews in clinical medicine can be a good resource for this kind of record keeping.

- The criteria used for reaching the conclusions in the Scientific Foundation Review should also be clearly described.

High quality and accurate forensic science will help to identify, remediate, and prevent wrongful convictions, as well as help identify the true perpetrators of crime - providing true justice to victims of crime, the wrongfully accused, the wrongfully convicted, and society at large. NIST is well-positioned to take on this important task as both a substantively appropriate scientific research agency with forensic expertise and as an entity that has convened and is trusted by the diverse stakeholders of the criminal justice system. We greatly appreciate NIST's commitment to transparency and stakeholder engagement by providing an opportunity for public comment on an effort of such importance. The Innocence Project stands ready to help if we may provide additional context or assistance regarding this submission.

Comment on NISTIR 8225 DRAFT NIST Scientific Foundation Reviews

https://doi.org/10.6028/NIST.IR.8225-draft

Line 159: These reviews seek to answer the question: "What empirical data exist to support the methods that forensic science practitioners use to analyze evidence?"

The application of probability theory to draw inferences from DNA data is not new. Consider the match probability (in the US RMP) for a single source stain. This is assigned by the use of a population genetic model (usually either recommendation 4.1 or 4.2 of NRC II [1]), a coancestry coefficient (theta), and allele probability assignments. The RMP is therefore the result of a fusion of modelling and some assignments that are partially empirically based. The allele probabilities are drawn from population samples and the relevance of the population sample to the case in question is determined using judgement, knowledge of population differences, and the case circumstances.

The value for theta may be informed by studies of a diverse set of populations.

The total RMP can only be tested empirically for fewer loci than used in the current megaplexes.

We use this simple example to show that inference in forensic science uses probability and genetic theory, judgement, and empirical data, not simply empirical data. The overall RMP is not, or barely[1], testable empirically.

If we turn to the use of probabilistic genotyping[2] the situation is the same except that there are more, and more complex, models. Black box testing is valuable and is routinely undertaken in both the developmental and internal validation of STRmix™ [7-9], however restricting any assessment of validity to this type of testing would be incorrect. Validity of the assigned *LR* arises from belief that the models are at least good approximations of reality, and that the models are applied correctly. Recall that the models themselves are nuisance parameters. It is not necessary to show that the model is approximately correct or that it is the only model that could exist. These models are nuisance parameters en route to the *LR*.

Line 162: The central activity of forensic science is to make associations between pieces of evidence or between evidence and known items in order to shed light on past events and action.

Firstly, we object to the one sided nature of this sentence. At the very least we need to recognise the ability of forensic science to exclude.

In response to this statement, we confine ourselves to trace evidence. A useful distinction has been drawn between investigative and evaluative considerations. For a discussion refer to Section 2.10 of [10]. We discuss only evaluative considerations.

At the evaluative stage years of scholarship (we largely follow Evett) have established that the weight of evidence is expressed as an LR. This directs us to consider the probability of the findings

---

[1] There are some outstanding papers in this area pioneered by Weir [2-6]
[2] We confine ourselves to comments about the software STRmix™

given each of two exclusive and practically exhaustive propositions. Evidence has evidential weight if, and only if, these two probabilities differ.

Hence, rather than "to make associations" we would suggest that the central activity of forensic scientist is to characterise traces, form reasonable propositions about their possible origin, and assign probabilities to these findings given these propositions.

Line 168:...to what extent the discriminating power of those features is known…

This appears to be formulated for categorical features characterized with absolute certainty. In more generality if the measurements on the samples are $\bar{x}$ and $\bar{y}$ then we need to evaluate

$$\frac{p(\bar{x}, \bar{y} \mid H_p)}{p(\bar{x}, \bar{y} \mid H_a)} \text{ where}$$

H_p: The two samples are from the same source
H_a: The two samples are from different sources

This implies a need to study the distribution of features within and between sources.

Line 210: Are statistically significant sample sizes used?

First, if statistical significance is to be used at all then it is the inference that is significant not the sample size. We suggest an abandonment of statistical significance altogether. Rather it is the strength and reliability of the inference(s) that can be drawn that is of interest.

General: Bibliography

We would suggest abandonment of the bibliography. It will age too quickly and the "top ten" [11] is simply too subjective, and filled with non-primary and unrefereed sources.

John Buckleton and Jo-Anne Bright

References

[1] National Research Council II. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press; 1996.
[2] Weir BS. Matching and partially-matching DNA profiles. Journal of Forensic Sciences. 2004;49:1009-14.
[3] Lauc G, Dzijan S, Marjanovic D, Walsh S, Curran J, Buckleton J. Empirical support for the reliability of DNA interpretation in Croatia. Forensic Science International: Genetics. 2008;3:50-3.
[4] Curran J, Walsh SJ, Buckleton JS. Empirical support for the reliability of DNA evidence interpretation in Australia and New Zealand. Australian Journal of Forensic Sciences. 2008;40:99-108.
[5] Curran JM, Walsh SJ, Buckleton JS. Empirical testing of estimated DNA frequencies. Forensic Science International: Genetics. 2007;1:267-72.
[6] Tvedebrink T, Eriksen PS, Curran JM, Mogensen HS, Morling N. Analysis of matches and partial-matches in a Danish STR data set. Forensic Science International: Genetics. 2012;6:387-92.
[7] Bright J-A, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix; A multi laboratory response to PCAST. Forensic Science International: Genetics. 2018;34:11-24.

[8] Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. Internal validation of STRmix; for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2017;29:126-44.

[9] Taylor D, Buckleton J, Evett I. Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics. 2015;16:165-71.

[10] Jackson G, Aitken C, Roberts P. Case assessment and interpretation of expert evidence. Guidance for judges, lawyers, forensic scientists and expert witnesses. Practitioner guide 4. 2015.

[11] Butler JM, Iyer H, Press R, Taylor M, Vallone PM, Willis S. DNA mixture interpretation principles: insights from the NIST scientific foundation review. 29th International Symposium on Human Identification. Phoenix, AZ2018.

National Association of Criminal Defense Lawyers
Response to NIST Request for Comment on
NISTIR 8225: NIST Scientific Foundation Reviews

November 19, 2018

The National Association of Criminal Defense Lawyers (NACDL) thanks the National Institute of Standards and Technology (NIST) for the opportunity to comment on the *DRAFT NISTIR 8225: NIST Scientific Foundation Reviews.* NACDL commends NIST for seeking to fulfill the National Commission on Forensic Science's (NCFS) call for a technical merit evaluation of forensic methods as a necessary first step to the Organization for Scientific Area Committees (OSAC) developing standards.[1] The process of evaluation and the results of the evaluation have the potential to promote an empirical approach to forensic science and to better inform stakeholders quantitatively of the limits of the methods under review. But NACDL has concerns about the plan to use unpublished data, about the evaluation of published studies, and about the absence of an intramural research program as part of the evaluation.

NACDL is the largest organization in the United States advancing the mission of the criminal defense bar to ensure justice and due process for persons accused of crime or wrongdoing. A professional bar association founded in 1958, NACDL's approximately 9,000 direct members in 28 countries—and 90 state, provincial, and local affiliate organizations totaling up to 40,000 attorneys—include private criminal defense lawyers, public defenders, military defense counsel, law professors, and judges committed to preserving fairness and promoting a rational and humane criminal justice system. NACDL has a strong interest in ensuring the accuracy and reliability of all evidence that may be introduced to support a criminal prosecution.

NACDL has played a vital role in several significant historic reviews of flawed forensic science evidence. In 2007, NACDL partnered with the Innocence Project and the FBI to review comparative bullet lead analysis cases, following the FBI's admission that its agents potentially gave flawed or misleading testimony in thousands of such cases. NACDL worked with the Department of Justice (DOJ) Office of Enforcement Operations to correct the serious injustice caused by the failure to notify thousands of defendants whose cases were affected by the findings of wrongdoing in the 1996 Office of the Inspector General Report and FBI Task Force investigation. In addition, NACDL and the Innocence Project partnered with the FBI and the DOJ in their review of criminal cases in which the FBI conducted microscopic hair analysis in

---

[1] *See* Views on Technical Merit Evaluation of Forensic Science Methods and Practices *(Adopted at NCFS Meeting #10 - June 21, 2016)* and Recommendation on Technical Merit Evaluation of Forensic Science Methods and Practice *(Adopted at NCFS Meeting #11 - September 12, 2016)*

order to identify cases in which FBI hair examiners made scientifically invalid statements in testimony or lab reports. While the review is ongoing, the results thus far have conclusively documented the extraordinary frequency of exaggerated testimony.

Commensurate with the views and recommendations of the NCFS, NACDL supports NIST's role as the entity best suited to conduct a technical merit evaluation of existing and future forensic methods, and that this evaluation should precede and form the foundation for standards development. If such an evaluation is undertaken with the guidance set forth by the NCFS, the standards articulated in the draft proposal, including employing the definition of science and scientists set forth in the draft proposal, the result could provide three substantive benefits to forensic science.

First, the interdisciplinary process would continue the cross discipline and scientist/practitioner dialogue emblematic of the NCFS that led to the production of more than 40 collaborative work products. Second, the undertaking would produce an empirical evaluation of each method under a variety of "case-work-like" circumstances, providing a foundation from which the OSAC can develop practice standards. Third, the review would provide an empirically based assessment for the public of the reliability, accuracy, and limitations of the methods under inquiry.

The proposal as drafted, however, includes a review of unpublished data as part of the technical review process and does not reference any effort by NIST to undertake its own intramural research program. NACDL understands the interest in assessing unpublished data. Given the dearth of published data, it is currently impossible to make an empirical evaluation of "case-work-like" applications of many forensic disciplines. If NIST uses unpublished data, NIST must work closely with its statisticians, CSAFE, academics, and experts to assess the source and suitability of the data for a "case-work-like" evaluation and make any unpublished data relied upon available for outside review.

An example of the merits of NACDL's concern about unpublished data is NIST's proposal to examine proficiency testing data. As NIST is aware, the rigor of proficiency testing has been called into question, and not surprisingly so, since the development of proficiency tests is a for-profit business with obvious market incentives not to make the tests too hard. As the President of Collaborative Testing Services told the NCFS during its seventh meeting (August 10, 2015), he has been "under commercial pressure to make proficiency tests easier."

Given reasonable concerns about unpublished data and the dearth of published data, NIST should not ignore the NCFS recommendation No. 2 that encourages NIST to undertake "its own intramural research program" in which it can better control study design and the development of

data.[2] In addition, an intramural research program would prevent NIST from essentially replicating the work of past entities identified in Section 3 of the draft proposal—which have undertaken literature reviews combined with meetings with and presentations from practitioners— and instead add to the empirical understanding of the performance of the disciplines under evaluation.

NACDL also has concerns about existing published studies. Given the critique by several independent scientific committees, NIST must review the published literature with strict adherence to the scientific principles it enumerates in this proposal and cognizant that much of what has been published has been published in a limited set of journals dominated by practitioners with few—if any—individuals with expertise in research design and statistical modeling.

To use the Association of Firearm and Toolmark Examiners (AFTE) and the AFTE Journal as an example, AFTE's membership is limited to "individual[s] who derive[ ] a substantial portion of [their] livelihood from the examination identification, and evaluation of firearms and related materials and/or toolmarks."[3] Its Editorial Board is made up of AFTE members and the reviewers are also predominantly, if not exclusively, AFTE members.[4] Such a journal does not meet the criteria set out by the NCFS and NIST. A membership-only review process is neither independent nor external. And a body comprised of individuals limited to those whose livelihoods depend on the acceptance of the discipline within the criminal justice system is not conflict-free. Thus, if NIST relies on research published in such journals, NIST must assemble its own body of independent external reviewers to assess the studies.

As participants in the criminal justice system, NACDL is all too aware of the limitations of many judges, prosecutors, and defense lawyers in understanding science, statistics, research design, and the scientific method. And we daily experience the allegiance to law enforcement by many (but not all) forensic practitioners. Thus, NACDL fully supports the NCFS's call that "all forensic science methodologies should be evaluated by an independent scientific body to characterize their capabilities and limitations in order to accurately and reliably answer a specific and clearly defined forensic question." NACDL is hopeful that NIST will undertake this challenge with a rigorous assessment that advances measurement science, standards, and technology in each discipline it evaluates.

---

[2] See Recommendation on Technical Merit Evaluation of Forensic Science Methods and Practices *(Adopted at NCFS Meeting #11 - September 12, 2016)*

[3] See AFTE membership requirements found at https://afte.org/membership/membership-requirements.

[4] See AFTE Board of Directors and Editorial Committee, *Comments on NCFS Views Document: "Scientific Literature in Support of Forensic Science and Practice"*, AFTEJ Vol. 47, No 2 (Spring 2015)

Response to NIST Scientific Foundation Reviews

Christian Cole, Alexander Gray, Lucina Hackman, Hervé Ménard and Niamh Nic Daéid
Leverhulme Research Centre for Forensic Science, University of Dundee, Dundee, UK

Summary

The NIST draft document NISTIR 8225 "NIST Scientific Foundation Reviews" published 24th September 2018 develops the framework by which the agency will review specific disciplines of Forensic Science.

The framework outline is broken down into three sections which cover the main aspects of how the Foundation Reviews will be assessed and reported.

1. What Data Source Will We Use?
2. How Will We Evaluate the Data?
3. What Information Will We Report?

The bulk of the draft document covers the history and mechanics of why NIST is performing these Foundation Reviews in detail. The final section discusses similar processes occurring in countries outside of the United States; specifically Australia and the United Kingdom.

Overall, this a very welcome document which should have a strongly positive influence in how Forensic Science is evaluated in future allowing the respective areas to show confidence in their expertise.

Data Sources

It is good that NIST will not be limiting their scope to finding data sources in only peer-reviewed journals as not all high quality research data is published. For example, negative data and validation experiments are particularly difficult to publish in traditional journals due to lack of "novelty" or "interest". Coupled to this is the question of the openness of the data: it is clear that research work published within the Forensic Science domain open access publications appear to be largely absent.

Not all data are created equally, however, and often the source can be used to assess the quality of the information it contains. A suggestion to assist the understanding of the differences in the potential standard of the data contained in each source type would be to establish a method to rank the types of sources of data guiding reviewers in future discipline Foundation Reviews. For example, top ranked sources will have been peer reviewed, have a formal mechanism for data submission with associated metadata, are able to be individually cited, and have a commitment to long-term storage. Whereas sources which have none of the above score less. As part of the review all the data sources and their rankings should be shown (see Table 1 as an example).

Table 1. Example scoring matrix for source of data. 'X' characters indicate whether a source satisfies the criterion fully. 'x' indicates that the criterion is satisfied in part.

Source	peer review	formal data sub	metadata	citation	storage
Journal SI	X	x	x	X	X
Data Dryad		X	X	X	X
Company			x		

This list of criteria is not exhaustive and could additionally include licensing restrictions, cost, appropriate use of data formats, etc.

No mention is made of data repositories as sources of data within the draft document and we would suggest that this is reconsidered. Institutional and third-party repositories are a rapidly growing source of high quality scientific data for use by others. The National Science Foundation has an open data policy(1) and the National Academies of Science, Engineering and Medicine has published a report promoting data sharing and openness "by design"(2). In the UK the funding bodies via UK Research and Innovation requires universities to make research data available for 10 years after publication(3) and EU research organisations have a similar commitment via their Open Science initiatives(4).

Data durability should be a key requirement, for if a dataset is defined as foundational there ought to be

a commitment to make it retrievable for the foreseeable future. Data that are at risk of disappearing due to funding limitations or policy decisions cannot be the highest category of quality as noted in point 1.2 of the draft document.

Respectability

The NIST draft mentions that the data and information should be respected (paragraph 1.2). Similarly to the above attribution of quality scores to data, a set of criteria would be useful for assessing their respectability. Citations of journal articles are often used to determine their impact/importance relative to other publications. The same can be done for datasets if they are appropriately maintained such as is the case for data repositories. These data repositories are a relatively new resources (such as Data Dryad(5) or Zenodo(6)) and largely unknown in Forensic Science circles, meaning that the majority of data are not available via this mechanism. A benefit of open data repositories is that they can become community resources for the continuous updating of datasets with the most up-to-date information. Just like the Genbank or Protein Data Bank repositories are commonly used in the biological sciences. Looking forward it should be acknowledged and encouraged that foundational data is a citable and reviewable resource in their own right.

Altmetrics, or article level metrics, fill the gap where traditional citations are missing: the time lag between publication and first citations, and resources not published via journal articles. Altmetrics measure the impact of citable data resources via mentions on all types of online media from news articles to social media posts and blogs. They are now commonly reported by journal sites e.g. *Forensic Science International: Genetics* uses the PlumX metric, and can be used for any resource with a citable Digital Object Identifier (DOI). In using DOIs data, media, articles, sites, etc can all be compared for respectability using the same criteria and reported in a similar fashion to the data sources. For example, the "Open by Design" publication mentioned previously(2) was published in July 2018 and has only been cited three times according to Google Scholar. It is too early to say from citations alone how impactful it will be. Whereas the NAS report(7) was published February 2009 and has been cited 753 times. The Altmetric score for the two publications is 346 and 650, for the "Open by Design" and NAS publications respectively, putting both in the top 5% of all research tracked by Altmetric.

A strong support for respectable data is whether the data are reproducible. A core requirement in science, in order to be reproducible, a dataset needs to have clear and transparent methodology and all resources (e.g. materials and software) required to recreate it should be readily available. Again this information should be made available to allow this information to be assessed by those using the data be it practitioners, the judiciary or scientists.

Core resources

Using the above set of criteria together with other information the NIST draft highlights in section 1.3 "What Information Will We Report?" a core set of resources could be identified for practitioners, investigators and the judiciary essentially as a 'go to' reference most appropriate for the needs of the different users. For example, practitioners would be most interested in appropriate data to apply to their cases from the most high quality sources, whereas a judge may be more interested in a pertinent review of the underlying science to assess its admissibility.

Historical Overview

It is our view that there is too much time spent on this, often repetitive, aspect covering well-known events such as the NAS(7) and PCAST(8) reports that do not require more attention. Having said that, on p. 17 lines 887-888 public criticisms of the PCAST are reported as per Koehler(9) without any of the clear rebuttals made in the reference. The neutral stance in this report does not accurately convey the lack of merit the criticisms have. It is correct to highlight there were criticisms to the PCAST report but the authors should also reflect how meritorious they are.

References

1. NSF Data Sharing Policy [Internet]. 2017 [cited 2018 Nov 2]. Available from:
   https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_11.jsp

2. National Academies of Sciences E. Open Science by Design: Realizing a Vision for 21st Century Research [Internet]. 2018 [cited 2018 Nov 2]. Available from: https://doi.org/10.17226/25116

3. Rylance R, Wingham D, Wright N, Bruce R, Hammonds W, Arrowsmith J, et al. Concordat on Open Research Data [Internet]. 2016 [cited 2018 Nov 2]. Available from: https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/

4. EOSC Declaration [Internet]. 2017 [cited 2018 Nov 2]. Available from: https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

5. https://datadryad.org/

6. https://zenodo.org/

7. Council NR. Strengthening Forensic Science in the United States: A Path Forward [Internet]. 2009 [cited 2018 Nov 2]. Available from: https://doi.org/10.17226/12589

8. President's Council of Advisors on Science and Technology. Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods [Internet]. Executive Office of The President's Council of Advisors on Science and Technology, Washington DC; 2016. Available from here.

9. Koehler JJ. How Trial Judges Should Think About Forensic Science Evidence [Internet]. Rochester, NY: Social Science Research Network; 2018 Feb [cited 2018 Sep 28]. Report No.: ID 3118745. Available from: https://papers.ssrn.com/abstract=3118745