Department of Justice
Forensic Science Discipline Review of Testimony:
Draft Methodology

To comment, please access this document through www.regulations.gov, OLP Docket No. 158. Comment is open through August 1, 2016. For more information on the Forensic Science Discipline Review of testimony, please contact the Office of Legal Policy at 202-514-4601 or FSDR.OLP@usdoj.gov.

Jonathan Wroblewski
Principal Deputy Assistant Attorney General
Office of Legal Policy

George Tillery
Office Director, Office of Science & Technology
National Institute of Justice

Shimica Gaskins
Acting Deputy Assistant Attorney General
Office of Legal Policy

Joel Hunt
Senior Computer Scientist
National Institute of Justice

Kira Antell
Senior Counsel
Office of Legal Policy

Linda Truitt
Senior Social Science Analyst
National Institute of Justice

Kevin Scott
Director, Policy Analysis Unit
Office of Legal Policy

Matthew Durose
Statistician
Bureau of Justice Statistics

Project Abstract

The goal of the Forensic Science Discipline Review (FSDR) is to advance the use of forensic science in the courtroom by understanding its use in recent cases and to facilitate any necessary steps to ensure that expert forensic testimony is consistent with scientific principles and just outcomes. To accomplish this goal, the Department of Justice (Department) is planning a Department-level review of forensic testimony by Department personnel, beginning with an examination of FBI testimony. The Department is undertaking this review because it is good management to conduct macro-level program reviews and not because of known or suspected problems with particular forensic science disciplines.

The following document is a draft methodology for the FSDR. Elements in the draft methodology are subject to revision, and comment is invited. The Department proposes to review and evaluate trial testimony provided by FBI forensic examiners in several forensic disciplines in state and federal cases for a five-year period (2008–2012). All cases in which an FBI examiner testified in these forensic disciplines—and for which a transcript can be obtained—are proposed to be reviewed. It is anticipated that the FSDR will review all cases meeting these criteria in at least some of the following disciplines: fiber, firearms, footwear, glass, questioned documents, latent prints, paints, polymers, tire tread, and toolmarks. The order in which disciplines will be reviewed has not been determined, in part because the development of the FSDR testimonial standards, against which the testimony of Department personnel will be compared, is ongoing. The FSDR's initial level will identify whether an examiner provided at least one statement describing a relationship between at least two items (Statement of Relationship).[1] If at least one Statement of Relationship exists, testimony will be compared

[1] For purposes of the FSDR, a Statement of Relationship is considered to be any statement regarding the relationship of an unknown item to a known item (such as bullets to a firearm) or to another unknown item (such as two sets of

against criteria established by the FSDR testimonial standards, which will vary for each discipline and will be based on language from the Uniform Language for Testimony and Reports, being developed by the Department in a public process.[2]

I. Purpose, Goals, and Learning from Past Research

The Department of Justice (Department) provides forensic science services through its forensic laboratories. The Department relies on forensic science through its prosecution of criminal and civil cases and is a supporter of forensic science through its research sponsored by the National Institute of Justice and its grant programs supporting state and local forensic science practitioners. The Department engages with the forensic science community through its relationship with the National Commission on Forensic Science (NCFS), a federal advisory committee, and the Organization of Scientific Area Committees (OSAC). The NCFS was established by the Department, in partnership with the National Institute of Standards and Technology (NIST), to enhance the practice of forensic science. The NCFS includes federal, state, and local forensic science service providers; research scientists and academics; law enforcement officials; prosecutors, defense attorneys, and judges; and other stakeholders from across the country. OSAC is part of an initiative by NIST and the Department to strengthen forensic science in the United States. OSAC is a collaborative body of more than 500 forensic science practitioners and other experts who represent local, state, and federal agencies; academia; and industry.

---

fingerprints) or of an unknown substance to a particular substance (such as whether a stain on a shirt contains semen). This includes statements of exclusion, inconclusive statements, and statements of association, all three of which may exist in a single testimony.

[2] See *Justice Department Issues Draft Guidance Regarding Expert Testimony and Lab Reports in Forensic Science* (June 3, 2016), https://www.justice.gov/opa/pr/justice-department-issues-draft-guidance-regarding-expert-testimony-and-lab-reports-forensic.

As a forensic science provider, the Department has numerous forensic laboratories across three agencies and employs thousands of dedicated forensic science personnel. Department forensic laboratories complete tens of thousands of forensic examinations a year, making the Department, in the aggregate, one of the largest providers of forensic services in the world. At the same time, Department forensic laboratories are located in separate agencies, with separate accrediting and quality management systems, and they support different missions. Nevertheless, because the Department is a single agency, it is appropriate for it to develop Department-wide policies for forensic science practice reflecting Department-wide goals.

The Department's commitment to advancing forensic science led it to conceive of the Forensic Science Discipline Review (FSDR) and subsequently to the proposed methodology to conduct a retrospective analysis of forensic testimony to focus on the use of forensic science in the courtroom. The goal of the FSDR is to advance the use of forensic science in the courtroom by understanding its use in recent cases and to facilitate any necessary steps to ensure that expert forensic testimony is consistent with scientific principles and just outcomes. To accomplish this goal, the Department is planning a Department-level review of forensic testimony by Department examiners. The Department proposes to accomplish this by assessing whether Department forensic examiners have testified consistent with applicable scientific language as established by the FSDR testimonial standards. The Department believes it is critical to have means to review internal procedures and practices to ensure the organization is performing at the highest level possible. Consistent with this best practice, Deputy Attorney General (DAG) Sally Yates

announced on February 24, 2016, that the Department would conduct a quality assurance-like review of some forensic science disciplines practiced at the Department.[3]

The FSDR is a Department-level review of forensic testimony provided by Department examiners in recent cases with the aim of improving the quality of such testimony, beginning with an examination of FBI testimony. The review is another step toward institutionalizing continuous quality improvement in the practice of forensic science that is supplementary to numerous other accreditation requirements and is in harmony with the existing quality assurance programs at Department forensic laboratories. The Department is undertaking this review because it is good management to conduct macro-level program reviews and not because of known or suspected problems with particular forensic science disciplines. Although testimony provided by Department examiners is just a very small part of the forensic science services provided by the Department, it may be the most visible. The Department is conducting this review with the expectation that it will help build greater public confidence in the forensic work done at the Department and the testimony of its experts, and that it will put the entire field of forensic science on a stronger footing.[4]

Accreditation is one critical way the Department engages in oversight of the testimonial evidence provided by examiners.[5] Accreditation requires a laboratory to periodically review its examiners' testimony according to requirements that the laboratory developed, but there is no

[3] *See Deputy Attorney General Sally Q. Yates Delivers Remarks During the 68th Annual Scientific Meeting Hosted by the American Academy of Forensic Science* (Feb. 24, 2016), https://www.justice.gov/opa/speech/deputy-attorney-general-sally-q-yates-delivers-remarks-during-68th-annual-scientific.

[4] Id.

[5] Accreditation is the means through which a forensic service provider establishes that its practices and procedures (such as testing, instrument calibration, reagent storage) are consistent with standards and requirements. ASCLD/LAB is a large forensic laboratory accreditor; it defines its services as "programs of accreditation in which a forensic service provider delivering covered services may participate to demonstrate that its technical operations and overall management system meet ISO/IEC 17025:2005 requirements and applicable ASCLD/LAB-International supplemental requirements." https://ascld-lab.qualtrax.com/ShowDocument.aspx?ID=1588, accessed June 8, 2016.

systemized Department-level process for periodic review to determine whether such testimony complied with applicable scientific and legal standards. The current Department-level feedback mechanisms—through appeal, written judicial decision, and anecdotal reports—are narrow and limited to the legal perspective, and they do not permit laboratory directors or Department leadership to oversee this process at a macro-level. Although the FSDR is beginning with a retrospective analysis of recent FBI cases in some disciplines, the FSDR and the final methodology could be subsequently implemented Department-wide as an additional layer of review and evaluation of forensic science testimony in the courtroom. The frequency and scoping of such a review will depend on the results of the FSDR's initial phases. If successful, the methodology and testimony review process may prove beneficial and provide guidance for adoption outside the Department by other federal, state, and local forensic science laboratories.

The information obtained in the FSDR of testimony will be used to: (1) promote uniform testimony by Department forensic science examiners; (2) identify whether there are systemic problems within certain forensic science disciplines requiring consideration of policy or practice changes; (3) assist state and local forensic science providers in developing appropriate review of their testimony practices; and (4) inform stakeholders on the application of forensic science in the courtroom.

A. *Objectives in FSDR Methodology Development: Transparency and Independence*

The Department believes it is best positioned to develop an initial draft methodology within a process that allows for robust public comment and appropriate refinement of the methodology. The principles underlying the FSDR methodology development process are transparency and independence.

The FSDR methodology development is predicated on transparency of process. All elements of the draft methodology (save budget and resources) will be fully available for review and comment by the public. In addition, by inviting the public—including experts in forensic science, statisticians, academics, and practitioners—to weigh in on elements of the FSDR in the drafting process, the Department can expand input to develop an optimal program and limit potential bias. The unprecedented step of posting the draft methodology online and inviting public comment to participate in the development is beyond that offered through a formal peer review process and unique in its transparency.

Independence is also a critical element of the FSDR methodology development process. To ensure independence in the methodology development phase, those engaging in the design have examined categories of threats to independence, including: self-interest, self-review, bias, familiarity, undue influence, management participation, and structural bias.[6] The development of the FSDR methodology is appropriately independent because the individuals developing the FSDR: (1) are accountable to the DAG; (2) will report results to the DAG, to the reviewed laboratory, and to the public; (3) are located organizationally outside the Department forensic laboratories; (4) have access to laboratory documents, materials, and staff; and (5) are career Department employees and are thus removed from political pressures.[7] In addition, none of the people engaged in developing the methodology are current forensic science practitioners, and none have ever been employed by Department forensic laboratories as forensic practitioners. Nevertheless, to ensure the forensic science community perspective was appropriately represented, former full-time forensic practitioners were engaged to provide insight in the practice of forensic science. Moreover, and perhaps most important to ensure independence

[6] U.S. Government Accountability Office, Government Auditing Standards, §3.14 (rev. 2011).

[7] GAS §3.31 (internal auditor independence).

from internal Department pressures, the methodology will be submitted to the NCFS for its consultation and review prior to adoption by the DAG.

B. *Literature Review*

1. Previous Reviews of Forensic Testimony

Several forensic laboratories over the past 20 years have found their work scrutinized due to concerns over quality, and many local and state governments (and the federal government) have reviewed these laboratories' work, occasionally focusing on the work of specific laboratory employees. Despite this scrutiny, there has been little comprehensive and empirical analysis that would permit generalized conclusions to be safely drawn regarding forensic testimony. As a result, as late as 2007, scholars would conclude that "the evidence concerning over-claiming . . . is generally anecdotal or indirect" (Cole 2007, 822). For example, the National Research Council (NRC) report on compositional bullet lead analyses concluded that "in many cases, the experts apparently have not, in their testimony, recognized the limitations of such evidence" (2004, 90), despite the fact that it the NRC's report did not systematically sample or code report language or trial transcripts when examiners testified about the composition of bullet lead.

Those studies that have sought to systematically evaluate expert reporting or testimony have necessarily relied on convenience sampling.[8] Cole collected 34 transcripts of latent fingerprint testimony through a variety of methods.[9] He found that statements of association tended to fall into three categories:

[8] Convenience sampling is a form of nonprobability sampling. Any form of probability sampling requires knowing the population from which the sample is drawn, and defining the population of forensic reports and testimony across all laboratories would be extremely difficult to accomplish.

[9] Some suggest that Cole's findings are subjective because his work lacked independent review. Without remarking on those concerns, this document notes that it references Cole's work primarily for a discussion of methodology.

- Process statements, in which "the witness characterized the evidence not in terms of probative value, but as the outcome of process" (Cole 2007, 831), were found in 18 of the 34 transcripts;
- Source attribution statements, in which "flat assertions that the defendant 'made' the print, that the print 'is' the print of the defendant or that the defendant 'could be the only source' of the print" (Cole 2007, 832), were found in 16 of the 34 transcripts; and
- Identity statements, in which "the expert witness states that the latent print and the reference print were 'identical' or 'one and the same'" (Cole 2007, 834), were found in 4 of the 34 transcripts.

Eight of the 34 transcripts included bolstering, and four included quantification of "the witness's certainty as '100%'" (Cole 2007, 836).

In 2009, Garrett and Neufeld published a study evaluating the forensic testimony that occurred in the trials of 137 individuals who had been exonerated by DNA evidence. They found that 82 of the 137 cases (60 percent) had invalid testimony (Garrett and Neufeld 2009, 14), in a significant range across forensic disciplines.[10] Garrett and Neufeld argued that invalid testimony fell into two basic categories: "(1) the misuse of empirical population data; and (2) conclusions regarding the probative value of evidence in the absence of empirical data" (2009, 16).[11]

[10] There were 205 testimonies given in the 137 cases. Of the 205, 92 (45 percent) were invalid by Garrett and Neufeld's coding. Garrett (2011) updated these data to include 153 cases with 245 testimonies and found that "in 61% of the trials where a forensic analyst testified for the prosecution, the analyst gave invalid testimony" (p. 90).

[11] Garrett and Neufeld further break down the two classes into six types of invalid testimony. For Garrett and Neufeld, and for Cole, the categories are not mutually exclusive. Garrett and Neufeld also attempted to evaluate the extent to which invalid testimony was correlated with wrongful conviction by assessing the incidence of testimony in "matched" nonexonerations; they found, in a sample of 30 cases where the defendant was not exonerated by DNA evidence, that 63 percent of the cases involved invalid forensic science testimony (2009, 29).

The exoneration of individuals convicted in cases in which FBI examiners had testified regarding microscopic hair comparison led the FBI and, eventually, the states of New York and Texas, to initiate reviews of testimony, and in some cases testimony and reports, provided by laboratory examiners.[12] The FBI, in conjunction with the Innocence Project (IP) and the National Association of Criminal Defense Lawyers (NACDL), undertook a review of cases in which FBI examiners had conducted microscopic hair comparisons. Specifically, the FBI's Microscopic Hair Comparison Analysis review (FBI MHCA review), which is ongoing, encompasses reports and testimony in which an FBI hair examiner made a positive, probative hair association; the defendant was convicted; DNA analysis was not conducted on the evidentiary hair; and the hair analysis occurred prior to December 31, 1999.[13]

The FBI MHCA review classifies statements as erroneous if they meet predefined categories that exceed the limits of science, and developed three categories of "errors": statements that stated or implied that evidentiary hair could be associated with a specific individual to the exclusion of all others; statements that assigned a statistical weight or probability that the evidentiary hair originated from a particular individual; and statements that compared the number of times an examiner or the laboratory were unable to differentiate known hair samples from different people to the total number of cases worked in an attempt to bolster an association between the evidentiary and known hairs. The FBI MHCA review involves sequential review by FBI laboratory personnel, FBI Office of General Counsel personnel (for instances in which laboratory personnel identified no errors), and the IP and NACDL. In 2015, the FBI, IP, and NACDL released review results of the first 500 cases and concluded that, "in the

[12] Several other states have begun considering similar reviews, including: Arizona, Illinois, Massachusetts, Minnesota, Oklahoma, and Pennsylvania.

[13] See https://www.fbi.gov/about-us/lab/scientific-analysis/fbi-doj-microscopic-hair-comparison-analysis-review. Accessed May 10, 2016.

268 cases where examiners provided testimony used to inculpate a defendant at trial, erroneous statements were made in 257 (96 percent) of the cases."[14] One of the distinguishing characteristics of the FBI MHCA review is that it treats any statement that reviewers feel may fall within one of the three defined error types as an "error." This is done by considering statements on their own in a "line-by-line" approach, and the review gives no weight to attempts to qualify, correct, mitigate, or offer context for those statements.

Following the FBI announcement regarding the review of microscopic hair testimony, the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) contacted its accredited forensic laboratories, informing them of the FBI MHCA review, and recommending that "each laboratory, in consultation with the appropriate legal authorities, consider whether there may be past cases, specifically involving convictions, in which it would be appropriate to evaluate the potential impact of the reported conclusions and/or related testimony on the conviction."[15] The Texas Forensic Science Commission (TXFSC) appears to have considered and adopted this recommendation. The New York Commission on Forensic Science (NYCFS) considered this recommendation and conducted a preliminary review, referenced herein.[16]

The TXFSC's hair microscopy review looks at cases in which a hair examiner made a positive, probative association. The review team, consisting of four subject matter experts and three attorneys, flags cases for notification in which either the transcript or the laboratory report

[14] https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review. Accessed May 10, 2016.

[15] http://www.ascld-lab.org/notification-from-the-ascldlab-board-of-directors-to-interested-parties-concerning-potential-issues-with-hair-comparison-testimony/. Accessed May 10, 2016.

[16] At its June 17, 2016, meeting, the NYCFS did not reach a formal conclusion on the cases it had reviewed during the preliminary review and decided not to conduct a more fulsome review. The processes described in this document refer only to the preliminary review. For more information or to watch recordings of that meeting, please see http://www.criminaljustice.ny.gov/pio/openmeetings.htm.

contain a statement of identification, assign a probability or statistical weight, or contain any other "potentially misleading statements or inferences" (Texas Forensic Science Commission 2015, 17). The TXFSC review also samples cases (rather than looking at the population) from the 20 different state, city, and county laboratories that have performed hair microscopy in Texas over time, and it will request additional cases where it observes patterns of concern. The TXFSC team reviews cases (after they have been screened for possible probative associations) collaboratively (rather than sequentially) and looks at the entire testimony rather than identifying single statements of association. As of this writing, the TXFSC review was ongoing and had determined that roughly 40 percent of the testimonies it had reviewed had exceeded the standards of science as understood today.

The NYCFS review began in 2013. Laboratories under the NYCFS purview were asked to provide the first five rape or homicide cases in the 1980s and the 1990s that had inculpatory reports or testimony. The NYCFS then reviewed the transcripts of those cases where charges were filed, a plea was not taken, and the defendant was found guilty. The NYCFS review team identified 11 cases (of the 75) that met those criteria and were eventually presented to the NYCFS, but, as of this writing, the NYCFS has made no determination of the rate of error in those 11 cases.

2. Juror Perception of Forensic Evidence

Understanding how jurors perceive the testimony offered by forensic examiners may serve to inform decisions about how best to evaluate that testimony. The limited research on the topic crosses forensic science pattern disciplines but may nonetheless help indicate the best way

to structure a review of testimony.[17] McQuiston-Surrett and Saks (2008) surveyed undergraduates and found that "match" was perceived as more certain than "consistent with," "reasonable scientific certainty," or "probable" (1162), a finding they confirmed in an experimental setting using potential jurors as subjects. In a second experiment, McQuiston-Surrett and Saks found that informing the jurors of the limitations of forensic testimony (in this experiment, microscopic hair) either in cross-examination or in juror instructions "had little measurable or meaningful impact on their judgments about the likelihood that the defendant was the source of the crime-scene hair or their perceived understanding of the expert's testimony" (McQuiston-Surrett and Saks 2008, 1169; McQuiston-Surrett and Saks 2009).

The work of McQuiston-Surrett and Saks suggests that individual statements of association may be the appropriate unit of analysis, as jurors' assessment of the statements made was not undermined by later cross-examination or judges' instructions. However, Koehler (2011), using footwear evidence, found that jurors are sensitive to the raising of concerns on cross-examination and attempts by examiners to acknowledge the limitations of match statistics led jurors to give less weight to that evidence (relative to the examiner merely offering the match statistics without discussing potential sources of error).[18]

Garrett and Mitchell, using fingerprint identification as their pattern subject, differentiated experimental subjects into "simple positive match," "bolstered positive match," "qualified or inconclusive match," and "exclusion" treatments (2013, 489); they found that bolstered positive match statements were no more likely than simple positive match statements

[17] There is a robust literature on how forensic examiners convey statistical probabilities to jurors (*see e.g.*, Thompson and Newman 2015; Neumann, Ranadive, and Kaye 2015), but much of that literature focuses on DNA rather than pattern or trace evidence.

[18] Specifically, Koehler finds that cross-examination focusing on sources of error in the forensic testimony undermines the strength of the evidence, but "jurors who read the detailed cross-examination of the shoeprint expert were no less likely to believe the defendant's shoes made the shoeprint or that the defendant was guilty of the burglary than were jurors who did not read any cross-examination testimony" (Koehler 2011, 46).

to produce jurors who thought the defendant left his prints at the crime scene or that the defendant was the robber (495).[19] Like Koehler, though, Garrett and Mitchell found that "when the examiner admitted that fingerprint examiners sometimes make mistakes and that the identification could be wrong, [experiment] participants reduced their judgments about the likelihood the defendant committed the crime, reduced their estimates of the probability that the defendant left his prints at the crime scene, and had less confidence in their guilt judgments" (2013, 505). Finally, they found that examiner acknowledgment on cross-examination that misidentification was possible reduced the probative value of the fingerprint identification (2013, 505).

II. FSDR Proposed Methodology

As stated above, the goal of the FSDR is to advance the use of forensic science in the courtroom by understanding its use in recent cases and to facilitate any necessary steps to ensure that expert forensic testimony is consistent with scientific principles and just outcomes. To accomplish this goal, the Department is planning a Department-level review of forensic testimony by beginning with FBI examiner testimony in specific disciplines; but in its development stage, the FSDR requires a well delineated focal point around which the methodology can be developed. The research question is listed below, and elements of the methodology to address the question are described in this section. To reiterate, elements in the draft methodology are subject to revision, and comment is invited.

Research Question: How closely do FBI examiner Statements of Relationship from 2008–2012 in select disciplines conform to current FSDR testimonial standards?

[19] Jurors gave less weight to testimony that offered a qualified or inconclusive match.

A. *Process for Development*

At the National Commission on Forensic Science (NCFS) meeting on March 21, 2016, the Office of Legal Policy (OLP) presented a framework for the FSDR, outlining critical steps in development of the FSDR methodology. Soon after the framework presentation, OLP began conducting intensive outreach with stakeholders and experts from across the forensic science community. OLP posted the FSDR framework in the Federal Register on April 13, 2016, seeking public comment, and accepted comments through May 9, 2016.[20]

OLP has led a team of Department research scientists to develop the draft methodology described herein. OLP presented the draft methodology at the NCFS meeting on June 21, 2016. A notice will be posted in the Federal Register[21] and this document will be posted on www.regulations.gov, OLP Docket No. 158, to further solicit public comment.

B. *Proposed FSDR of Testimony Draft Methodology*

Previous reviews focused on testimony of forensic hair analysis in light of an acknowledgement of specific issues in that discipline. By contrast, the FSDR is being undertaken to advance the use of forensic science in the courtroom by understanding its use in recent cases and to facilitate any necessary steps to ensure that expert forensic testimony is consistent with scientific principles and just outcomes—and not because of specific concerns with other disciplines. As such, where previous reviews have examined only statements that include positive, probative associations made by examiners, the FSDR will encompass all Statements of Relationship—statements of exclusion, inconclusive statements, and statements of association. This is consistent with the intent of the review to assess the language examiners used in their expert testimony.

[20] *See* 81 Fed. Reg. 20675 (Apr. 8, 2016). Comment is available to be viewed online. *See* https://www.regulations.gov/#!docketDetail;D=DOJ-OLP-2016-0003.

[21] *See* 81 Fed. Reg. _____ (forthcoming June 2016).

The proposed FSDR methodology outlines the Department proposal for how cases for review are selected, at what level of detail testimony will be examined, the standard by which those statements and testimony are to be evaluated, who will participate in the review process, how parties to the cases reviewed are to be notified, and what considerations may affect any expansion of the proposed review.

1. Case Selection

a. Case Selection—Previous Reviews

The Texas and New York reviews both relied on samples of cases, and in both states, that sample was a nonprobability sample.[22] The TXFSC initially requested the first several cases of a particular decade from each laboratory. Although some Texas laboratories followed those instructions exactly, others provided all hair microscopy cases they could identify in their records for which a positive association was made. The NYCFS requested the first five cases of a particular decade from each laboratory with which the state commission works. Although drawing such samples is sensible, given that the population of cases that meet the criteria is unknown, drawing inferences can be problematic, because some cases had no probability of being selected for review. By asking each laboratory to provide the same number of transcripts, a greater percentage of cases are reviewed from laboratories conducting fewer hair examinations, whereas laboratories conducting more hair examinations have a smaller percentage of their testimonies and reports reviewed.

[22] A probability sample would require that every individual in a population have a "known, non-zero chance of being selected." See http://www.people-press.org/methodology/sampling/why-probability-sampling/. Because cases reviewed, for example, after the fifth case in each laboratory in each decade in New York labs have a zero chance of being selected, the sample drawn is a nonprobability sample, and inferences about the population of cases cannot be made.

Although drawing a random sample might be an option if the population is known and the number of cases in the population is manageable, reviewing the entire population of cases eliminates any need for—and any uncertainty created by—drawing statistical inferences.

b. Case Selection—FSDR Proposal

The FSDR draft methodology proposes to review transcript testimony in all closed cases (acquittals, convictions, and hung juries) in which an FBI examiner testified in state or federal court between 2008 and 2012 for which a transcript can be obtained.[23] It is anticipated that the FSDR will review all cases meeting this criteria in at least some of the following disciplines: fiber, firearms, footwear, glass, handwriting, latent prints, paints, polymers, tire tread, and toolmarks. The first level of FSDR evaluation will establish whether the examiner provided at least one Statement of Relationship. The time period of 2008 to 2012 was selected because it is anticipated to be long enough to provide a sufficient number of transcripts, is recent enough that transcripts may still exist, but is sufficiently historical such that cases are anticipated to be closed.

The review is being initiated with the FBI—how and whether to extend the FSDR to testimony provided by other Department forensic examiners will be considered after the FSDR has completed an initial implementation. The FBI testimony was selected for an initial implementation for several reasons. First, FBI examiners practice a range of disciplines and would provide the most varied population. Second, the FBI has a large forensic laboratory

[23] The FSDR draft methodology anticipates a several-step process to ensure the review of as many transcripts as possible: (1) FBI will provide case information (defendant name, date of testimony, location) for those cases for which they have a record that an examiner testified; (2) FBI will provide transcripts for all of the cases for which they have transcripts (a subset of the cases in which testimony was provided); (3) FSDR methodology implementation staff will attempt to locate transcripts for any case where no transcript was provided (this may include contacting courts, prosecutors, defense counsel); if a transcript cannot be located but the court reporter's notes exist, the FSDR staff will typically seek transcription. In addition, when the first discipline(s) to be reviewed is announced, the FSDR will ask members of the public to submit case information for cases where an FBI examiner testified that will be compared against FSDR records to verify that the population of cases is complete.

system in the Department, with approximately 1,100 employees. The Drug Enforcement Administration has a very large laboratory presence but conducts forensic analysis in a limited number of disciplines. The Bureau of Alcohol, Tobacco, Firearms and Explosives has fewer forensic laboratory staff, and although it does conduct several kinds of analyses, it focuses in particular disciplines. Additionally, the FBI conducts analyses in the largest variety of federal, state, and local cases, including violent offenses in Indian Country.

The decision that the FSDR methodology would review all cases in selected disciplines over a particular time frame was made due to the relative difficulty of drawing a representative sample. As a preliminary matter, the FBI was asked to identify the number of instances in which FBI examiners provided testimony in several disciplines over a period of years. As **Table 1** indicates, the FBI reported that examiners testified relatively infrequently.[24]

Table 1. Instances in Which FBI Examiners Provided Testimony for Certain Disciplines, FY 2008–2012[25]

Latent prints	Firearms & Toolmarks	General Documents	Shoeprint & Tire Tread	Paints & Polymers
132	45	46	25	17

According to the FBI, its examiners individually are likely to testify less often than state and local forensic examiners due to the expanded FBI mission in counterterrorism and intelligence as these types of cases have fewer trials. Since 2001, the FBI's violent crime casework (much of which is provided by state and local jurisdictions)[26] has decreased substantially, and this has directly impacted the number of testimonies provided across most disciplines. Between 2008 and 2012, the number of testimonies varied across disciplines, ranging from a high of 132 in the

[24] The FBI notes that frequency of testimony provided has not been a critical statistic to track historically, but the FBI instituted a "Testimony Tracker" database in 2013 to make the collection of testimony data more consistent.

[25] The number of testimonies provided from FY 2008 to 2015 in various disciplines was obtained from individual unit/discipline databases maintained by the FBI. The FBI has no method for ensuring complete accuracy of this data through other methods.

[26] There are also requirements in some states to use state or local laboratories where possible in violent offenses.

latent print discipline to a low of 17 in the paints and polymers discipline.[27] For all of these disciplines, the sample size necessary to make valid inferences about the population quickly approaches the size of the population, so reviewing the entire population of cases in each discipline offers much more confidence for only a little more work.[28] Relatively low numbers of testimonies were compounded by the fact that not all testimony had at least one Statement of Relationship. For example, a cursory review of firearms transcripts found that a significant percentage of testimonies related to shooting incident reconstruction or whether a particular firearm meets charged statutory or sentencing elements and did not contain Statements of Relationship.

2. Review of Examiner Statements

a. Review of Examiner Statement—Previous Reviews

The FBI MHCA review treats individual positive, probative statements of association as the unit of analysis without considering the context in which the statement was offered. This approach likely requires the least judgment on the part of the individual reviewer, but it also can produce misleading conclusions about the accuracy of examiner testimony in context. The TXFSC approach has been to read the entire testimony of cases where positive, probative associations were made and determine, based on context, whether the statements made exceeded the limits of science.

The FBI MHCA approach leads to a relatively high "error" rate; the initial report indicated that "in the 268 cases where examiners provided testimony used to inculpate a

[27] The FBI was unable to disaggregate testimonies given in some disciplines (for example, serology testimony from nDNA testimony), and rather than reporting artificially inflated numbers of testimonies here, it was decided to omit them from this table. When the FSDR methodology is finalized and implemented, the Department will report on the total numbers of cases in which testimony was provided and the number of transcripts reviewed.

[28] For example, if a sample of 10 (of the 17) paints and polymers testimonies was drawn, the sampling error (with 95% confidence and assuming a sample proportion of .5) would be 20.5%. It stands to reason that "sampling" the additional seven testimonies would permit more robust conclusions.

defendant at trial, erroneous statements were made in 257 (96 percent) of the cases." [29] Further, the FBI MHCA approach does not credit the examiner for attempts to qualify or correct a statement. The Texas approach, on the other hand, lowers the number of testimonies viewed as erroneous but requires reasonably robust assumptions about how jurors perceive the testimony they are hearing.[30] This approach may also include task irrelevant information that may be the source of cognitive bias in the review process and may require a standard that is difficult to apply consistently across a large number of cases and forensic science disciplines.[31]

b. Review of Examiner Statements—FSDR Proposal

The FSDR of testimony proposes to use a middle ground approach to review examiner statements. Previous reviews have either focused on a line-by-line review that ignores any relevance of context of comments or a completely contextual review that fails to address individual statements. For the purpose of this review, a testimony thread is defined as the complete, even if not continuous, discussion of a piece of evidence (or collection of pieces of evidence that are discussed as a group) that has at least one Statement of Relationship.

Although there is no consensus in the literature, the studies point in the direction of jurors weighing not only what the examiner says about the relationship but also how the statement is made, what caveats or enhancements that testimony includes, and general statements about the possibility of error (McQuiston-Surrett and Saks 2008, 2009; Garrett and Mitchell 2013; Koehler 2011). Further, jurors appear cognizant of what happens during both direct and cross-

[29] https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review. Accessed May 10, 2016. What remains unclear from that finding is the number of positive probative statements of association that were made in the 268 cases and how many of those statements were erroneous under the FBI MHCA review standard.

[30] Academic reviews of trial testimony (Cole 2007; Garrett and Neufeld 2009) ostensibly take the entire testimony as their unit of analysis but do so in a way that any single statement can characterize the entire testimony.

[31] The NYCFS has elected not to proceed with a further review at this time.

examination, so evaluation of an examiner's statements should take into account the entire testimony, including any cross-examination.

At the same time, any given examiner may be testifying about a series of relationships; more than one fingerprint may have been left at the crime scene, for example, and an examiner may testify about them serially or collectively (or in some combination of the two). To properly gauge the quality of examiners' testimony, context— the statements made by the examiner about the strengths or weaknesses of a particular pattern discipline, whether made on direct or cross— influences how jurors react. But the experimental conditions the literature relies on all envision only one piece of forensic testimony, so applying this understanding to the real world of courtroom testimony requires some adjustments.

To meet what the literature appears to suggest is the optimal approach, it may be useful to view a forensic examiner's testimony as a series of statements, some of which are about relationships between items that aggregate into threads. In any given testimony, the examiner may be asked to compare an individual questioned object (Q) to a known object (K), but the examiner may be asked to do so for a series of Qs (and Ks) or may be asked if a collection of Qs can be matched to a given K. If one treats the examiner's testimony as a series of threads, then it may be possible to separate out the related Statements of Relationship an examiner makes and evaluate each of those threads separately.

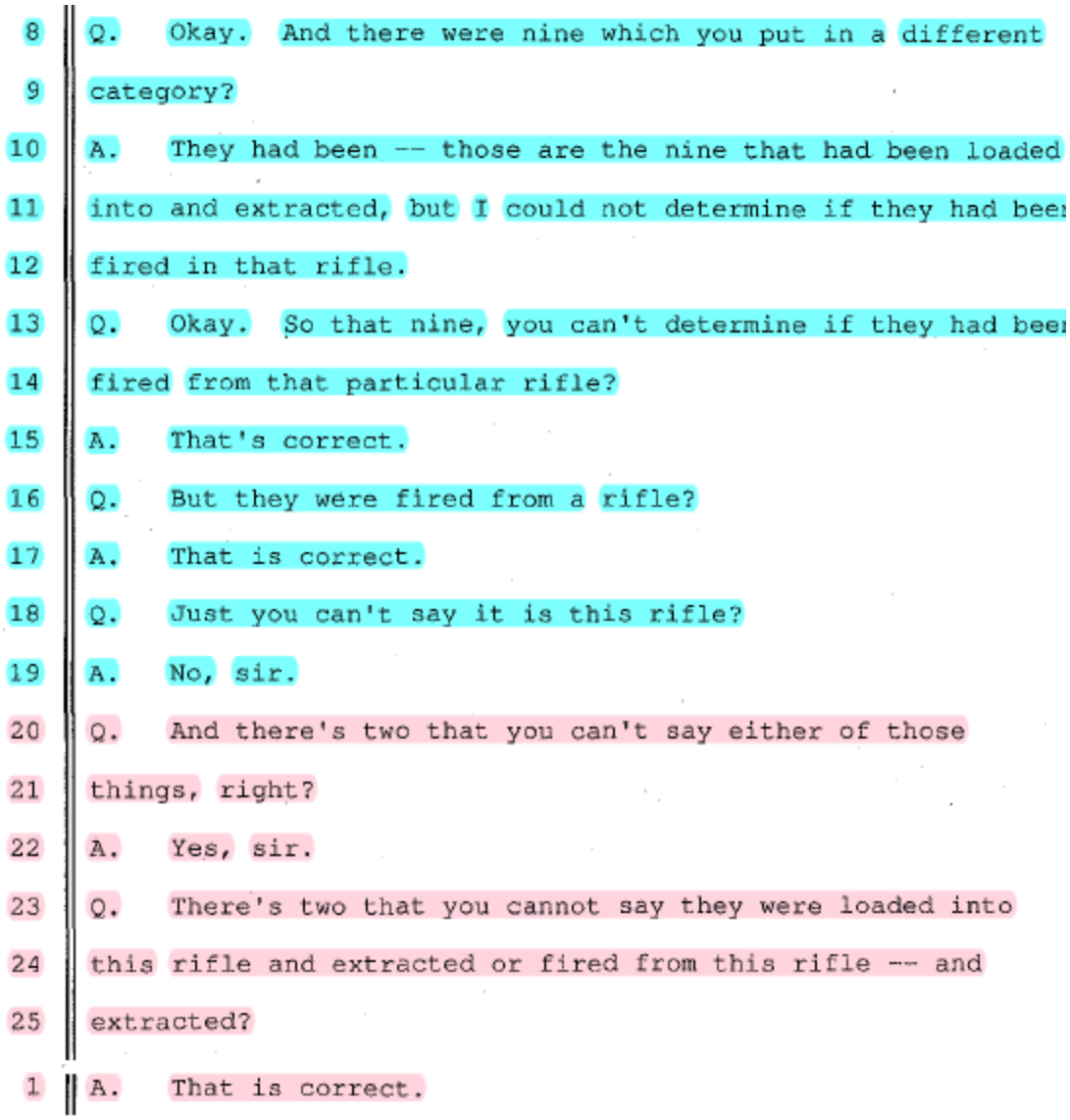
To be sure, such an approach comes with its own challenges. First, although identifying individual threads may sound analytically appealing, doing so in practice may be complex. Nonetheless, adequate training and familiarity with the patterns of testimony should allow it to be done consistently. Second, the threading permits the same thread to be assigned to multiple

raters and can be used to measure inter-rater agreement.[32] This approach avoids judging the appropriateness of the entire testimony at once and avoids the task-irrelevant information for purposes of this review. This approach also avoids evaluating each Statement of Relationship without the context that the literature indicates is relevant to how jurors process testimony of forensic examiners.

To identify a thread, evaluators would first read the entire testimony to (a) identify if any Statements of Relationship are made and then (b) identify threads. It may be the case, for example, that the examiner is asked about Q_1 and Q_2 as they relate to K_1 in direct testimony, implying one thread. But if the cross-examination asks separately about the relationship between Q_1 and K_1 and Q_2 and K_1, then there would be two threads, with overlap in the threads on direct examination. Figure 1 illustrates an example of how this would work in practice.

[32] If the same rater gets two threads from the same testimony, the places of overlap can also be used to measure intra-rater agreement to ensure consistency within a rater.

Q. Okay. So as far as your examination goes -- well, strike that. Let me ask you then about the casings. You were provided with 32 casings for examination, right?
A. Yes, sir.
Q. And again, one of the things that you wanted to determine was whether the casings were associated with this particular rifle, right?
A. Yes, sir.
Q. And the examination there entails an analysis of the back of the casing, right, where the firing pin hits the casing?
A. That's part of the analysis, yes, sir.
Q. It's different from an analysis of the bullet fragments, right?
A. Yes, sir.
Q. And in that case, you were able to determine that 22 of them -- I'm sorry, 21 of them were in fact bore characteristics that they were fired from this particular rifle?
A. That's correct. 21 of the cartridge cases were fired in the submitted rifle.



Q. Okay. And there were nine which you put in a different category?
A. They had been -- those are the nine that had been loaded into and extracted, but I could not determine if they had been fired in that rifle.
Q. Okay. So that nine, you can't determine if they had been fired from that particular rifle?
A. That's correct.
Q. But they were fired from a rifle?
A. That is correct.
Q. Just you can't say it is this rifle?
A. No, sir.
Q. And there's two that you can't say either of those things, right?
A. Yes, sir.
Q. There's two that you cannot say they were loaded into this rifle and extracted or fired from this rifle -- and extracted?
A. That is correct.

Figure 1. Example of Testimony Broken Down by Threads

In the Figure 1 example, there are three threads: one discussing the 21 casings that the examiner testifies are identified with the submitted firearm (text highlighted in yellow); one discussing the nine casings that had been loaded and fired from a firearm but the examiner could not determine if they were fired from the submitted firearm (text highlighted in blue); and two casings where the examiner could not determine if they had been fired at all (text highlighted in

pink). There is also text in green that offers context for all three threads so would be included in all three of the threads as reviewed.

3. Standards

a. Standards—Previous Reviews

The FBI MHCA review seeks to identify which statements by microscopic hair examiners exceeded the limits of science; the TXFSC review applies the same standard; the New York review has not determined a review standard. As straightforward as such a standard may appear, it translates into applying the current state of accepted practice to reporting and testimony offered decades ago. Using such a standard raises the challenge of determining whether statements that exceed the limits of science are the responsibility of the examiner, who spoke (or wrote) beyond what could be claimed given the state of the discipline at the time the testimony was given, or whether the examiner was correctly applying what was understood at the time to be the consensus of the discipline.[33]

b. Standards—FSDR Proposal

The FSDR testimonial standard for each discipline will vary and is expected to be based on language from the Uniform Language for Testimony and Reports (Uniform Language) but to differ somewhat to account for the fact that the standards will be applied retrospectively.

The Uniform Language, once approved for each discipline, is intended to establish a prospective standard for Department personnel in all forensic disciplines practiced at the Department. The Proposed Uniform Language is based on the FBI's Approved Scientific Standards for Testimony and Reports (ASSTRs) and is currently in the process of development through a collaborative process across Department components. The Department is seeking

[33] For example, there were different interpretations of whether "consistent with" was a potentially misleading statement or inference.

public comment on Proposed Uniform Language for seven disciplines.[34] Public comment closes for those disciplines on July 8, 2016. An internal Department group will review comments and provide Proposed Uniform Language to the DAG. If the DAG adopts that language, it will become Approved Uniform Language for Testimony and Reports.

It is not anticipated that the FSDR testimonial standard will be identical to the Uniform Language for any discipline. Commenters to the FSDR framework[35] specifically note the challenge associated with utilizing a current standard for a retrospective analysis. Some commenters suggest that the FSDR should use a two-question approach and identify both whether any testimony was not in conformity with internal laboratory policies in effect at the time of the testimony and whether and to what extent testimony is not in conformity with a current standard.[36]

Using a historical standard is problematic. Policies for particular aspects of testimony in most forensic disciplines did not exist at all time periods that are to be reviewed by the FSDR. Moreover, where policies did exist, they were revised with varied frequency. The challenges associated with identifying which standard was in effect at a moment in time militate against adherence to a shifting preexisting standard across many forensic science disciplines over a five-year time span. As such, the FSDR methodology contemplates use of a single standard that incorporates some flexibility to address the retrospective application.

It is expected that the FSDR testimonial standard for at least one discipline will be sufficiently developed to be presented to the NCFS at its September 2016 meeting.

---

[34] *See* 81 Fed. Reg. 37,642 (June 10, 2016), online at www.justice.gov/forensics

[35] 81 Fed. Reg. 20,675 (Apr. 8, 2016).

[36] *See* Innocence Project, *Public Comment on the Proposed Forensic Science Discipline Review Framework*, at 3 (May 9, 2016), https://www.regulations.gov/#!documentDetail;D=DOJ-OLP-2016-0003-0012; E.G. Morris, National Association of Criminal Defense Lawyers, *Letter from NACDL to J. Wroblewski*, at 3 (May 9, 2016), https://www.regulations.gov/#!documentDetail;D=DOJ-OLP-2016-0003-0008.

4. Sequencing and Personnel

a. Sequencing and Personnel—Previous Reviews

The FBI MHCA review occurs sequentially–laboratory personnel first review the case, next the FBI's Office of General Counsel reviews the case if laboratory personnel do not identify any errors, and finally the IP and NACDL independently review the case and provide the FBI with a consolidated evaluation. In Texas, the prosecutors' representatives, laboratory personnel, criminal defense attorneys, and IP staff collectively review testimony to make a determination as to whether there are errors. Laboratory examiners and defense lawyers play a role in the FBI MHCA Review, the Texas review, and the New York review. In the Texas and New York reviews, prosecutors are also represented.

All three reviews rely on subject matter experts (forensic examiners and attorneys) as reviewers and then seek to address potential bias either via collective decision making (TXFSC) or by conducting the review sequentially (FBI MHCA). The group decision-making approach may promote understanding across various stakeholders and prevent potential stray outliers and rogue evaluations, but it eliminates the possibility that inter-rater reliability can be measured.

b. Sequencing and Personnel—FSDR Proposal

The Department proposes to adopt a validated protocol for reviewing transcripts and training raters to use that protocol to identify and categorize Statements of Relationship in each transcript (Garrett and Neufeld 2009, 16). The FSDR threading methodology can be accomplished in discrete actions (reviewing transcripts for Statements of Relationship, identifying threads, separating the transcript into threads, evaluating statements and threads). Each of those actions can be assigned to multiple coders to permit evaluation of inter-rater fidelity to protocol and of inter-rater reliability.

The FSDR threading methodology is designed to limit human bias, and trained raters are intended to be used as the thread reviewers.[37] As noted above, laboratory examiners are used in FBI MHCA, Texas, and New York processes, and defense lawyers play a role in all three processes. In the Texas and New York processes, prosecutors have also been represented in the review, accounting for possible bias by seeking consensus among the represented parties. Although collective review may promote comity, it may be the case that group dynamics play a role in determining the outcome, as is common with collegial decision making (Forsyth 2013).

5. Notification Protocol and Reporting of Results

a. Notification and Reporting—Previous Reviews

The FBI MHCA review notifies all prosecutors, defense counsel if known, and any defendant in cases where a report or testimony is reviewed, including whether the report testimony is deemed to have exceeded the limits of science as defined in the FBI MHCA standard. The TXFSC notifies the prosecutor, defendant, court, and laboratory, including providing a final report if the review finds that the testimony contained a statement of identification, assigned probability or statistical weight, or contained any other potentially misleading statements or inferences. New York has not, as of this writing, agreed on a notification protocol.

The FBI MHCA policy of broadly notifying defendants where the review identified an "error" also resulted in many notices in forensic scientists' files, even where the forensic scientist was explicitly complying with the report or testimony standard then in effect. The use of the term "error" when applying a retrospective standard may have exacerbated this problem.

[37] Because the threading process can be separated from the review of Statements of Relationship, and the review of Statements of Relationship can be separated from evaluating threads and testimony, the draft methodology offers tremendous flexibility in terms of staffing and measuring inter-rater reliability.

b. Notification and Reporting—FSDR Proposal

The FSDR draft methodology contemplates notification and reporting in two senses. First, the Department intends to make the outcome of any implementation of the FSDR of testimony available to the public. The Department hopes its review will serve as a model for developing routine management-level review of forensic testimony beyond that provided by accreditation. Although it is unclear whether the FSDR will result in a robust data set, it is hoped that at least some findings and information regarding variables of interest will be available.

Second, although the methodology will begin from a neutral standpoint (with no expectation of inaccuracies), there may be some cases where testimony not in conformity with the FSDR testimonial standard is identified that triggers a legal or ethical obligation to notify court officials (prosecutors, defense counsel) and individual defendants who were found guilty.[38] That may involve notification where a certain threshold of nonconforming statements in multiple threads are identified, but it may be the case that some threads or even individual statements of association are sufficiently problematic that, regardless of the content of the balance of the testimony, the defendant should receive notification.

The Department is working to develop an appropriate notification system that prioritizes appropriate notification of individuals but is cognizant of applying standards retrospectively that could have unintended professional consequences. The Department anticipates an individual notification system to draw on the FBI MHCA review notification system and anticipates it to include representation from organizations representing prosecutors and defense counsel. Independent of the Department's own publication and publicity efforts, when the FSDR methodology had been finalized and is closer to implementation, the Department plans to seek

[38] As previously stated, acquittals, convictions, and hung juries are reviewed, but individual notification will be provided only where there was a conviction.

assistance in notifying legal practitioners as to which disciplines are being reviewed and any anticipated timeframe.

C. *Additional FSDR Draft Methodology Elements for Comment*

1. Evaluating Statements of Relationship

Within a given thread, Statements of Relationship (identified by page and line number of the transcript) can be measured on:

- Where in the trial the Statement of Relationship occurred (direct, cross, re-direct, re-cross);
- Who spoke the words that provided the relationship that was affirmed by the examiner (prosecutor, defense attorney, court, examiner);[39]
- The nature of the relationship (identification, inconclusive, exclusion [Garrett and Mitchell 2013]), including the actual transcript language (e.g., "consistent with," "identified as");
- Whether the Statement of Relationship was bolstered (Cole 2007; Garrett and Mitchell 2013) by some statement of probability not supported by any scientific foundation;
    - Whether that probability statement was ostensibly objective ("there is a 1 in 10,000 chance") or personal ("based on my examination of thousands of similar cases"), including reporting the actual language in the transcript; and
- Whether there is a qualification of an earlier Statement of Relationship, and whether that qualification was adequate.

[39] The FSDR methodology considered this an important variable. Although statements spoken by other individuals will need to be affirmed by an examiner to become a Statement of Relationship, it may be that statements that do not conform to the FSDR standard are more likely to have been generated by nonexaminers. The Department anticipates that how questions are phrased will affect examiners' answers, especially in cases where attorneys use words and phrases not of the witness's choosing.

2. Evaluating Threads and Testimonies

The FSDR threading methodology will have a protocol to evaluate entire testimonies based on the evaluation of Statements of Relationship. How testimonies will be evaluated and such evaluation reported has not yet been determined; this is a subject on which the Department is explicitly seeking input.

Threads will vary significantly in the number of Statements of Relationship and the presence or absence of qualifications as to the limits of testimony. Further, the FSDR testimonial standards discussed above may acknowledge that, whereas a given Statement of Relationship is permissible in one forensic science discipline, it may not be acceptable in another discipline. For example, the statement "Q_1 is consistent with [treads, handwriting, glass fragments, shell casings] made by K_1" may be supported by science in some disciplines but not supported in others.

Because the number of testimonies is so low, assigning a numerical value during evaluation (e.g., a "grade" or a "score") may not accurately convey the accuracy (and the range of accuracy) of testimony, whereas a thorough review and reporting may offer the opportunity for more robust analysis. One option consistent with such an evaluation would be to review and code each Statement of Relationship against the FSDR testimonial standard, and once the data collection is complete, conduct exploratory analysis to determine what information can be gleaned from testimonies as a whole. The Department specifically seeks comment on how to meaningfully evaluate the data collected and whether the levels of analysis contemplated (Statements of Relationship, threads, and testimony) are analytically useful.[40]

[40] For example, a simple average, or even a weighted average (weighting each thread score by the number of relationship in that thread) may not produce a useful metric because the threads overlap and a given Statement of Relationship may appear in more than one thread. Simply averaging the thread scores would give undue influence to statements of association that appear in more than one thread. One possible solution to this issue is to score

3. Possible Expansion of FSDR

The FSDR framework identified that a certain threshold of testimonial inaccuracies would trigger a "secondary" review of additional cases. This was predicated on an expectation that the FSDR methodology would rely on sampling rather than a review of the population of cases. Because, for reasons explained above, the FSDR draft methodology proposes review of all cases meeting certain criteria, the utility of a secondary review, regardless of outcome, may be less clear. Nevertheless, there could be certain findings that require a response that could include an expansion of cases to be reviewed.

Such situations include:

- If, after implementing the FSDR methodology in several disciplines and completing those reviews, it is determined that testimony in a particular discipline has less conformity with the FSDR testimonial standard than testimony in other disciplines, it may be an indication that a more fulsome review of that discipline is warranted to determine what changes may be made prospectively to ensure conformity with standards.
- If testimony provided earlier in the FSDR timeframe has more instances of nonconformity than testimony provided more recently, it may be an indication that the Department has addressed the nonconformities but may need to notify individuals who could have been affected prior to the FSDR timeframe.
- If a particular examiner has more instances of nonconformity with the FSDR testimonial standard than other examiners in a discipline, it may indicate the need for further review of that examiner's body of work.

---

testimony on unique Statement of Relationship (which would include efforts to qualify, correct, or provide context) rather than on threads.

D. *Notice and Comment*

As noted above, and consistent with the principles of transparency and independence, this document will be posted in the Federal Register—and on www.regulations.gov (OLP Docket No. 158)—to solicit public comment. Elements in the draft methodology are subject to revision, and comment is invited. OLP plans to review submitted comments (which will also be publicly available to any interested party) and refine the methodology. A revised draft methodology will be presented at the NCFS meeting in September 2016 along with at least one proposed FSDR standard. If the revised draft methodology is assessed to be largely complete at that time, the methodology will be submitted to the DAG for adoption and implementation.

The Department encourages broad distribution of this document to interested parties. Although individuals and organizations are welcome to contact the OLP directly, comments should be provided through www.regulations.gov so they can be publicly accessed.

Works Cited

Cole, Simon. "Where the Rubber Meets the Road: Thinking About Expert Evidence as Expert Testimony." *Villanova Law Review* 52, no. 4 (2007): 803-842.

Forsyth, Donelson R. *Group Dynamics.* Belmont, CA: Cengage Learning, 2013.

Garrett, Brandon L. *Convicting the Innocent: Where Criminal Prosecutions Go Wrong.* Cambridge, MA: Harvard University Press, 2011.

Garrett, Brandon L., and Peter J. Neufeld. "Invalid Forensic Science Testimony and Wrongful Convictions." *Virginia Law Review* 95, no. 1 (2009): 1-97.

Garrett, Brandon, and Gregory Mitchell. "How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment." *Journal of Empirical Legal Studies* 10, no. 3 (2013): 484-511.

Koehler, Jonathan J. "If the Shoe Fits They Might Acquit: The Value of Forensic Science Testimony." *Journal of Empirical Legal Studies* 8, no. s1 (2011): 21-48.

McQuiston-Surrett, Dawn, and Michael J. Saks. "Communicating Opinion Evidence in Forensic Identification Sciences: Accuracy and Impact." *Hastings Law Journal* 59 (2008): 1159-1190.

McQuiston-Surrett, Dawn, and Michael J. Saks. "The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear." *Journal of Law and Human Behavior* 33, no. 5 (2009): 436-453.

National Research Council Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison. *Forensic Analysis: Weighing Bullet Lead Evidence.* Washington, DC: National Academies Press, 2004.

Neumann, Cedric, Anjali Ranadive, and David H. Kaye. "Communicating the Results of Forensic Science Examinations." Final Technical Report for NIST Award 70NANB12H014, 2015.

Texas Forensic Science Commission. *2015 Annual Report.* Austin, TX, 2015.

Thompson, William C., and Eryn J. Newman. "Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents." *Law and Human Behavior* 39, no. 4 (2015): 332-349.