

Department Response to the PCAST Report

From: "Hunt, Ted (ODAG)" (b)(6)
To: "Donoghue, Richard (ODAG)" (b)(6)
Date: Wed, 21 Oct 2020 14:16:01 -0400
Attachment Final DOJ Draft Statement on the PCAST Report ODAG Review pdf (538 26 kB)

Rich,

Attached is a draft Department response to certain key aspects of the 2016 PCAST (President's Council of Advisors on Science and Technology) Report released during the last Administration. The PCAST Report that has been used by defense attorneys over the last 3 years (especially over the last year) to argue that traditional pattern comparison methods (firearms ID, latent print ID, etc.) should not be admitted in court.

These efforts have met with some success over the past year. Admissibility rulings on firearms/toolmarks comparison testimony in five federal district court cases during 2019-2020 have prohibited "identification" testimony ("this gun fired that bullet"), and have limited those opinions to something akin to "this gun can't be excluded from having fired that bullet," or "it's more likely than not that this gun fired that bullet."

(b) (5)

Incidentally, these are some of the same issues currently before the Federal Evidence Rules Advisory Committee, and the Committee Reporter, Dan Capra, has drawn heavily on the PCAST Report to urge that Rule 702 be amended due to forensic "overstatements."

I authored this paper, but it has been reviewed by the Department's Forensic Science Working Group, and some minor post-review edits have been incorporated into the attached draft. The FSWG consists of attorneys and laboratory directors/system directors from FBI, OTD, ATF, and DEA, as well as DOJ officials from OJP, CRM, EOUSA, BOP, and USMS. So if you'd rather not slog through this whole thing (and I wouldn't blame you for that), just be advised that the paper is limited to a technical response to points raised by PCAST, and our position on those points has FSWG support.

I'd like to get this out fairly quickly, as this is an ongoing issue affecting the admissibility of evidence in pending prosecutions. After ODAG approval, I'll talk to OPA about the best path for general dissemination.

My apologies for the long email!

Ted

nuts and bolts of tomorrow's meeting

From: "Shapiro, Elizabeth (CIV)" (b)(6)
To: "Goldsmith, Andrew (ODAG)" (b)(6); "Hunt, Ted (ODAG)" (b)(6); "Donoghue, Richard (ODAG)" (b)(6)
Date: Thu, 12 Nov 2020 15:44:27 -0500
Attachments: Letter to Evidence Committee re FRE 702_11.6.20 FINAL.pdf (237.47 kB); U.S. v. Harris.pdf (219.23 kB); U.S. v. Hunt.pdf (230.68 kB); Beretta Barrel Fired Bullet Open Set Validation Study (JFS 2020).pdf (871.08 kB)

Rich, Andrew and Ted:

I wanted to touch base on how tomorrow will likely unfold. We start at 10:00 am, and can expect a break around 11:15, and a lunch break around 12:30. We are unlikely to go past 3.

The meeting will begin with introductions (we have a new chair), approval of the minutes, and any other housekeeping.

Rich, I will introduce you (or the Chair may introduce you) so that members know who you are. Feel free to say hi, but don't feel like you need to say anything more; no one will expect it. Feel free to stay for as much or as little of the meeting as you'd like or have time for. Everyone understands that staying the entire meeting is a big time commitment, and in the virtual setting, with cameras off (as they will be most of the meeting), no one will notice anyway.

I expect the order of substantive business will be (I) 702; (II) 106; (III) 615; (IV) Possible future agenda items (Tab 6 in the agenda book).

After soliciting comments and input from the Criminal, Civil, and Appellate working groups, as well as from each litigating division at DOJ, we are pretty unified in our positions:

I. 702

The Reporter will outline the arguments and options in his memo. (b)(5) per CIV

The civil components of the Department and Civil Working group (b)(5) per CIV

I intend to offer a line or two when called on, but otherwise I will turn it over to Ted to address/rebut specific points of contention or answer questions from the group. (b)(5) per CIV

II. 106

(b)(5) per CIV

We will be pushed on oral statements, as the Reporter feels strongly that they are admissible now, but under a common law theory, and that the current system punishes the unwary. (b)(5) per CIV

I will be interested to hear Judge Schiltz's views; he is a known scholar on the Rules of Evidence.

III.

(b)(5) per CIV

IV. New Agenda Items

Tab 6 lists over a dozen rules where the Reporter has found circuit splits. (b)(5) per CIV

Looking forward to seeing everyone (virtually) tomorrow. Feel free to send questions that come up in advance, or even during the meeting. I'll have email up and open.

Thanks,
Betsy



November 6, 2020

Hon. Patrick Schiltz
United States District Judge
United States Courthouse
300 South Fourth Street, Room 14E
Minneapolis, MN 55415

Re: Possible Amendment to Federal Rule of Evidence 702

Dear Judge Schiltz:

We write respectfully, in advance of our upcoming November 13 meeting, to supplement the agenda materials with some additional reference materials and thoughts. Since the virtual nature of our meeting may make free-flowing discussion more difficult, we hope that having our views in advance will help further the conversation.

Uniform Language for Testimony and Reports

As the Committee will recall, the Department has proposed that the Committee table any amendment to Rule 702 in order to gauge the effectiveness of Department's initiatives with respect to Uniform Language for Testimony and Reports ("ULTRs"). The Department's Forensic Science webpage currently contains 16 ULTRs, many updated this past summer to further address important qualifications and limitations of expert testimony in various forensic disciplines.

In the forensic geology discipline, for example, an examiner may testify to a (1) Fracture fit; (2) Inclusion (i.e., included); (3) Exclusion (i.e., excluded); or (4) Inconclusive. When explaining his or her conclusion, "[a]n examiner shall not assert that two or more geologically-derived materials were once part of the same object unless the materials physically fit together." In addition, when offering a conclusion, an examiner shall not assert that a fracture fit is based on the "uniqueness" of an item of evidence; use the term "individualize" or "individualization;" or claim that the geologically-derived materials originated from the same object "to the exclusion of all other objects." Nor may an examiner assert absolute or 100% certainty or claim that forensic geology examinations are infallible or have a zero-error rate. Moreover, the ULTRs make clear that an examiner's source identification opinion is not based on a statistically derived or verified measurement or comparison to all other potential sources of a questioned sample. *See* <https://www.justice.gov/olp/page/file/1284776/download>.

Beginning in 2018, and continuing to the present, there are ample examples of federal, state, and D.C. courts that have limited or excluded testimony regarding the source of a spent bullet or shell casing. These cases, some of which are included in the case law digest, include:

United States v. Jovon Medley, No. PWG 17-242 (S.D. Md. April 24, 2018)
Williams v. United States, 210 A.3d 734 (D.C. Ct. App. June 27, 2019)
United States v. Tibbs, 2019 D.C. Super. LEXIS 9 (D.C. Sup. Ct. September 5, 2019)
United States v. Davis, 2019 U.S. Dist. LEXIS 155037 (W.D. Va. September 11, 2019)
United States v. Shipp, 2019 U.S. Dist. LEXIS 205397 (E.D.N.Y. November 26, 2019)
United States v. Adams, 2020 U.S. Dist. LEXIS 45125 (D. Oregon March 16, 2020)
People v. A.M., 2020 N.Y. Misc. LEXIS 2961 (Sup Ct. Bronx June 30, 2020)

In each of these cases—whether or not one agrees with the analysis and ultimate decision—the court used the existing rules of evidence to preclude the examiner from offering identification testimony. In contrast, the meeting memo (“Memo”) discusses *U.S. v. Simmons*, 2018 U.S. Dist. LEXIS 18606 (E.D. Va), decided January 12, 2018, as an example of a case that failed to heed the Department’s directives. *Simmons*, however, predated the publication of the ULTR documents. In addition, *Simmons* was a case in which the government—not the witness—offered alternative formulations of the expert’s conclusion for the court’s consideration during pretrial proceedings.

Although the Memo correctly notes that the ULTRs are not binding on state laboratories or state courts, neither are the Federal Rules of Evidence. Nevertheless, the ULTRs may well have an important impact on the states. The Organization of Scientific Area Committees¹ (“OSAC”), whose primary mission is to develop uniform national standards across forensic disciplines, and whose membership includes experts from federal, state, county, and local government, academia, and the private sector, has drawn from language provided in the ULTRs to draft national forensic standards. By allowing this industry-wide standards-building process to continue and develop, the guidance articulated in ULTRs may take hold faster and more effectively than any federal rule change. Indeed, in two recently published opinions, one from the D.C. District Court and another from the Western District of Oklahoma, the court utilized the Department’s ULTRs to properly limit the scope of firearms-toolmarks testimony.²

The Conceptual and Practical Differences Between “Match” and “Source Identification”

The conceptual formulation of a “match” and a “source identification” opinion is not the same. The traditional “match” paradigm in the forensic pattern comparison disciplines employed an essentially deductive reasoning process in which a sufficient combination of corresponding features was considered to be “unique” in the natural world. It followed that if a questioned sample exhibited a sufficient combination of features that corresponded to those observed in the known item, then the questioned sample (pattern) was considered “unique.” As such, an examiner “individualized” the questioned sample “to the exclusion of all other” such items (e.g. fingerprints, shell casings).

¹ <https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science>

² *U.S. v. Harris*, 2020 U.S. Dist. LEXIS 205810 (D.D.C.) (Nov. 4, 2020); *see also U.S. v. Hunt*, 2020 U.S. Dist. LEXIS 95471 (W.D. Okla.) (June 1, 2020).

In contrast to the “match” paradigm, a “source identification”³ conclusion is the result of an inductive reasoning process that makes no universal claims of deductive certainty. During an examination, a known item and a questioned sample are examined for a sufficient combination of corresponding features.⁴ If an examiner determines that there is sufficient correspondence such that she (based on her knowledge, training, experience, and skill) would not expect to find the same combination of features repeated in another source, and there is insufficient disagreement to conclude that the combination of features came from a different source, then the correspondence provides extremely strong support for the proposition that the questioned sample came from the known item. Similarly, it provides extremely weak or no support for the proposition that the questioned sample came from a different source. The examiner then inductively infers (from the observed data) that the questioned sample originated from the known item.⁵ The resulting classification as a “source identification,” “source exclusion,” “inconclusive,” is ultimately an examiner’s skill and experience-based opinion.

Importantly, at the conclusion of this process, an examiner makes no claim that the observed combination of corresponding features in the questioned sample (class and individual

³ “Identification is the decision process of establishing with sufficient confidence (not absolute certainty), that some identity-related information describes a specific entity in a given context, at a certain time.” Casey Eoghan & David-Oliver, *Do Identities Matter?* 13 *Policing: A Journal of Policy & Practice* 21, 21 (March 2019).

⁴ “The question for the scientist is not ‘are this mark and print identical’ but, ‘given the detail that has been revealed and the comparison that has been made, what inference might be drawn in relation to the propositions that I have set out to consider.’” Christophe Champod & Ian Evett, *A Probabilistic Approach to Fingerprint Evidence*, *Journal of Forensic Identification*, 101-22, 103 (2001).

⁵ See David Kaye, *Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies*, 75 *Brooklyn L. Rev.* 1163, 1176 (2010) (“In appropriate cases . . . it is ethical and scientifically sound for an expert witness to offer an opinion as to the source of the trace evidence. Of course, it would be more precise to present the random-match probability instead of the qualitative statement, but scientists speak of many propositions that are merely highly likely as if they have been proved. They are practicing rather than evading science when they round off in this fashion.”).

Most inferential reasoning in forensic contexts is inductive. It relies on evidential propositions in the form of empirical generalisations . . . and it gives rise to inferential conclusions that are ampliative, probabilistic and inherently defeasible. This is, roughly, what legal tests referring to “logic and common sense” presuppose to be the lay fact-finder’s characteristic mode of reasoning. Defeasible, ampliative induction typifies the eternal human epistemic predicament, of reasoning under uncertainty to conclusions that are never entirely free from rational doubt.

Paul Roberts & Colin Aitken, *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice*, 3. *The Logic of Forensic Proof—Inferential Reasoning in Criminal Evidence and Forensic Science, Guidance for Judges, Lawyers, Forensic Scientists, and Expert Witnesses*, Royal Statistical Society 43 (2014) <https://www.maths.ed.ac.uk/~cgga/Guide-3-WEB.pdf>.

Events or parameters of interest, in a wide range of academic fields (such as history, theology, law, forensic science), are usually not the result of repetitive or replicable processes. These events are singular, unique, or one of a kind. It is not possible to repeat the events under identical conditions and tabulate the number of occasions on which some past event actually occurred. The use of subjective probabilities allows us to consider probability for events in situations such as these.

Colin Aitken & Franco Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists* (Wiley 2nd Ed. 2004).

characteristics) is “unique”⁶ in the natural world, or that the examiner can universally “individualize”⁷ the item or person from which the questioned sample originated. Moreover, given the limitations of inductive reasoning, an examiner cannot logically “exclude all other” potential sources of the questioned sample with certainty.⁸ Accordingly, ULTR documents that authorize a “source identification”⁹ conclusion also prohibit an examiner from asserting that a questioned sample originated from a known source “to the exclusion of all other sources.” They also disallow claims of absolute or 100% certainty, infallibility, or a zero-error rate.¹⁰

From a legal perspective, a “source identification” conclusion is properly characterized as technical or specialized knowledge under Rule 702,¹¹ as it is based on an examiner’s training, skill, and experience—not statistical methods or measurements. As such, the PCAST Report erred when it claimed that all forensic pattern comparison disciplines are “metrology” (measurement science).¹² Although many of these disciplines are grounded in scientific principles, source identification conclusions provided by forensic examiners are “skill and experience-based”

⁶ “Every entity is unique; no two entities can be ‘Identical’ to each other because an entity may only be identical to itself. Thus, to say ‘this mark and this print are identical to each other’ invokes a profound misconception: they might be indistinguishable but they cannot be identical.” Champod, *supra* note 4, at 103.

⁷ “[I]ndividualization—the conclusion that ‘this trace came from this individual or this object’—is not the same as, and need not depend on, the belief in universal uniqueness. Consequently, there are circumstances in which an analyst reasonably can testify to having determined the source of an object, whether or not uniqueness is demonstrable.” Kaye, *supra* note 5, at 1166. The Department uses the term “identification” rather than “individualization.”

⁸ “We cannot consider the entire population of suspects - the best we can do is to take a *sample*... We use our observations on the sample, whether formal or in formal, to draw inferences about the *population*. No matter how large our sample, it is not possible for us to say that we have eliminated every person in the population with certainty. . . . This is the classic scientific problem of *induction* that has been considered in the greatest depth by philosophers.” Champod, *supra* note 4, at 104-105.

⁹ See also Kaye, *supra* note 5, at 1185 (“Radical skepticism of all possible assertions of uniqueness is not justified. Absolute certainty (in the sense of zero probability of a future contradicting observation) is unattainable in any science. But this fact does not make otherwise well-founded opinions unscientific or inadmissible. Furthermore, whether or not global uniqueness is demonstrable, there are circumstances in which an analyst can testify to scientific knowledge of the likely source of an object or impression.”).

¹⁰ <https://www.justice.gov/olp/uniform-language-testimony-and-reports>.

¹¹ See, e.g. *U.S. v. Herrera*, 704 F.3d 480 (7th Cir. 2013) (“[E]xpert evidence is not limited to ‘scientific’ evidence, however such evidence might be defined. . . . It includes any evidence created or validated by expert methods and presented by an expert witness that is shown to be reliable.” (Latent print decision); *Restivo v. Hessemann*, 846 F.3d 547, 576 (2d Cir. 2017) (“Rule 702 ‘makes no relevant distinction between ‘scientific’ knowledge and ‘technical’ or ‘other specialized’ knowledge,’ and ‘makes clear that any such knowledge might become the subject of expert testimony.’ *Kumho Tire Co.*, 526 U.S. at 147”); see also *U.S. v. Harris*, 2020 U.S. Dist. LEXIS 205810 (D.C. November 4, 2020) (characterizing firearms-toolmarks testimony as technical/specialized knowledge); *Accord U.S. v. Hunt*, 2020 U.S. Dist. LEXIS 95471 (W.D. Okla.); *U.S. v. Johnson*, 2019 U.S. Dist. LEXIS 39590 (S.D.N.Y. 2019); *U.S. v. Otero*, 849 F. Supp. 2d 425 (D.N.J. 2012); *U.S. v. Mouzone*, 696 F. Supp. 2d 536 (D. Md. 2009); *U.S. v. Monteiro*, 407 F. Supp. 2d 351 (D. Mass. 2006).

¹² *President’s Council of Advisors on Sci. & Tech., Executive Office of the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods* 23, 44, 143 (2016) (original emphasis) at 23, 44 n.93, 143.

opinions, similar to those offered by an electrical engineer, and discussed in the meeting Memo (pp. 132-33). It is also important to note that the PCAST Report chose to use the term “proposed identification” as the appropriate way for a forensic pattern examiner to articulate his or her conclusion. By adding the word “proposed,” PCAST meant to convey the possibility that the opinion might be incorrect¹³ As such, a “proposed identification” is essentially equivalent to a “source identification” conclusion. Both formulations recognize that an examiner’s opinion is potentially fallible.

Cross-Examination as a Solution to Perceived “Overstatement”

The meeting Memo suggests that empirical studies have shown that cross-examination is an ineffective means by which to challenge the credibility of expert witnesses—citing a 2008 study by McQuiston-Surrett & Saks. That study, however, is inconsistent with more recent research, including a 2013 paper authored by Professor Brandon Garrett. That study found that

[p]articipants exposed to the examiner who testified on direct that his method was reliable and then acknowledged on cross a possible misidentification rated the general reliability of fingerprint identifications the lowest. Thus, our results suggest that an examiner who claims infallibility on direct will be viewed skeptically after a cross that elicits error-risk concessions, but an examiner who on direct describes her method in reasonable terms, including acknowledging some risk of error, may be able to limit the negative impact of an effective cross-examination or contrary fingerprint evidence presented by the defense.¹⁴

In another study published in 2015, Joseph Eastwood and Jiana Caldwell found that educating jurors about the limitations of forensic procedures by presenting opposing expert witnesses can be effective in raising legitimate doubts about the forensic conclusions.¹⁵

A 2019 study—authored by PCAST contributor William Thompson—reported that participants found an expert less credible and were less likely to convict when the expert admitted that his interpretation rested on subjective judgment and when he admitted to having been exposed to potentially biasing task-irrelevant contextual information.¹⁶ Thompson found that,

¹³ *Id.* at 46. (“We suggest the term “*proposed* identification” to appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong. We will use this term throughout this report.”) (original emphasis).

¹⁴ Brandon Garrett & Gregory Mitchell, *How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment*, J. of Empirical Legal Stud., 484, 505-06 (2013); see also Brandon Garrett & Gregory Mitchell, *How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment*, Journal of Empirical Legal Studies, 484, 507 (“[W]hen the fingerprint examiner admitted that his method is not foolproof and that his conclusion in this case could be in error, that disclosure had a significant negative impact on the evidence.”).

¹⁵ Joseph Eastwood & Jiana Caldwell, *Educating Jurors About Forensic Evidence: Using an Expert Witness and Judicial Instructions to Mitigate the Impact of Invalid Forensic Science Testimony*, 60 J. Forensic Sci. 1523, 1528.

¹⁶ William Thompson & Nicholas Scurich, *How Cross-Examination on Subjectivity and Bias Affect Jurors’ Evaluations of Forensic Evidence*, 64 J. Forensic Sci. 1379-88 (2019).

[o]verall, the results indicate that jurors were skeptical of the expert's claim that he had ignored the task-irrelevant information, and this skepticism increased when the expert also admitted that his interpretation of the findings required subjective judgment in the absence of objective standards for interpretation.¹⁷

* * *

From a legal perspective, the finding suggests that lawyers can successfully challenge the credibility of a non-blind forensic expert in two ways: either by revealing the subjectivity of the expert's methods or by revealing the expert's exposure to task irrelevant information.¹⁸

Accordingly, recent research supports the position that conceding the fallibility of forensic findings on direct examination, during cross-examination, or through contrary evidence by an opposing expert, *does* affect the persuasiveness of a forensic examiner's opinion. Moreover, cross-examination is enhanced by the timely production of information underlying the expert's opinion. This was the reason that the Criminal Rules Committee—with the Department's support—has worked on a proposed amendment to Rule 16. The proposed timeliness requirement in Rule 16 is also being supplemented with additional DOJ training to ensure that prosecutors understand and adhere to their disclosure obligations.

The Department recognizes that a forensic examiner's past performance on relevant, skill-based testing is an important measure for evaluating her performance in a given case. As such, FBI proficiency test results are routinely provided to defense counsel upon request. The FBI Laboratory will soon begin disclosing proficiency test results without a specific defense request as part of their general discovery and disclosure procedures. In addition, Department laboratory quality assurance manuals, standard operating procedures, testing methodologies, and other laboratory policies are currently available online to defense attorneys and the general public.¹⁹ Moreover, the Department's ULTRs, which set forth the qualifications and limitations for sixteen forensic disciplines, are available to defense counsel in each case and are available on-line.²⁰

In a recent study, Professor Garrett examined the impact of proficiency test results and laboratory error rates on jury-eligible adults. His study found that,

[w]hen jurors receive information about flaws or weaknesses in a forensic method or receive general information about a field's error rates, the juror cannot be sure how that information applies to the particular analyst in the case at hand. But when jurors receive information about the testifying expert's own performance on a proficiency test that simulates the task involved in the case at hand, the relevance of this information is easy to comprehend and hard to ignore.²¹

¹⁷ *Id.* at 1386.

¹⁸ *Id.*

¹⁹ <https://www.justice.gov/olp/forensic-science#posting>.

²⁰ See <https://www.justice.gov/olp/uniform-language-testimony-and-reports>. “This document is intended to describe and explain terminology that may be provided by Department examiners. *It shall be attached to, or incorporated by reference in, laboratory reports or included in the case file.*” (Emphasis added).

²¹ Gregory Mitchell & Brandon Garrett, *The Impact of Proficiency Testing Information and Error Aversions on the Weight Given to Fingerprint Evidence*, 37 *Behav. Sci. Law*, 1, 14 (2019).

Regarding the impact of proficiency test information in particular, the Garrett study found that,

[t]he fingerprint examiner's level of performance on a proficiency test (high, medium, low, or very low), but not the type of error committed on the test (false positive identifications, false negative identifications, or a mix of both types of error), affected the weight that jury-eligible adults gave to an examiner's opinion that latent fingerprints recovered from a crime scene matched the defendant's fingerprints, which in turn affected judgments about the defendant's guilt.²²

Collectively, these recent studies undermine the position that cross-examination is an ineffective means of challenging the credibility of a forensic examiner. Instead, the findings clearly support the position that conceding the potential fallibility of forensic results on direct examination or during cross-examination, or challenging forensic evidence by use of an opposing expert, impacts the credibility of a forensic examiner's opinion.

Strength of Evidentiary Support versus Opinion Testimony

The meeting Memo appears to favor “strength of evidence” testimony over an expert’s opinion about the source of a questioned item. Memo at 110. Recent research, however, has shown that jurors do not correctly discern differences between subtle gradations of evidentiary strength, such as those endorsed by the American Statistical Association and described in the Memo.

Specifically, Eleanor Arscott found that study participants performed poorly when attempting to distinguish between strength of evidence expressions at the strong end of the scale (“strong,” “very strong,” and “extremely strong”).²³ As a result, she concluded that it was possible “to question the effectiveness of the scale of verbal expressions in communicating the intended evidential strength at the higher end of the scale.”²⁴ Arscott also noted the same can be argued for distinctions between “weak” and “moderate” strength, and between “moderate” and “moderately strong” evidence.²⁵ She concluded that “[t]hese results suggest we may not be able to assume that decision makers will be able to discern between these expressions.”²⁶

Separate research by Kristy Martire²⁷ on verbally described gradations in evidentiary strength revealed what she described as “the weak evidence effect.” That is, study participants presented with evidence that weakly supported guilt tended to invert that finding and wrongly think that “weak” evidence in support of the prosecution’s case actually meant that the evidence favored the accused.²⁸ Participants presented with weakly exculpatory evidence, however, were

²² *Id.* at 1.

²³ Eleanor Arscott et al., *Understanding Forensic Expert Evaluative Evidence: A Study of the Perception of Verbal Expressions of the Strength of Evidence*, 57 *Sci. and Just.* 222, 224, n.13 (2017).

²⁴ *Id.* at 224.

²⁵ *Id.*

²⁶ *Id.* at 227.

²⁷ Kristy Martire et al., *The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect*, 37 *Law and Hum.* 197, 205-06 (2013).

²⁸ *Id.* at 205-06.

not affected in the same way.²⁹ These studies demonstrate that testimony based on gradations of evidentiary support may actually confuse rather than clarify the intended meaning of an examiner's conclusion. This is surely not the intended result of a proposed rule change to FRE 702.

Assumptions Underlying the Proposed Rule Change and Note

1. Studies on the Baseline Valuation of Forensic Evidence by Potential Jurors: The So-Called "CSI Effect"

The draft Committee Note that accompanies the proposed amendments to FRE 702 suggests that jurors may overvalue scientific evidence and either unquestionably accept it or fail to understand expert testimony. *See, e.g.*, Memo at p. 143. ("Just as jurors are unable to evaluate meaningfully the reliability of scientific and other methods underlying expert opinion, jurors lack a basis for assessing critically the conclusions of an expert that go beyond what the expert's methodology may reliably support.").

Recent research, however, contradicts the notion that jurors overvalue forensic evidence. To the contrary, the findings show that jurors approach forensic evidence with a critical eye and tend to underweight its probative value. For example, a 2020 study by Jacob Kaplan and colleagues³⁰ reached the following conclusion:

We find that individuals in the United States hold a pessimistic view of the forensic science investigation process, believing that an error can occur about half of the time at each stage of the process. We find that respondents believe that forensics are far from perfect, with accuracy rates ranging from a low of 55% for voice analysis to a high of 83% for DNA analysis, with most techniques being considered between 65% and 75% accurate.³¹

The results differed from the researchers' expectations:

While we expected respondents to have a high level of confidence in the forensic science investigation process and for the accuracy of each forensic science technique (Hypothesis 1), our results suggest that members of the US public hold significant doubts about the accuracy of forensic techniques and believe that each technique contains high levels of human judgement. The technique perceived to be most accurate was DNA evidence at 83% accuracy, while voice analysis at 55% and footwear analysis at 57% were perceived to be least reliable. Most forensic techniques were considered to be in the range of 65–75% accurate. Our results align with prior work indicating that DNA is often perceived to be among the most accurate forensic techniques, though our study yields lower perceptions of accuracy for DNA than reported elsewhere. Additionally, respondents indicated that they

²⁹ *Id.* at 205.

³⁰ Jacob Kaplan et al., *Public Beliefs About the Accuracy and Importance of Forensic Evidence in the United States*, 60 *Sci. & Just.* 263-72 (2020).

³¹ *Id.* at 263.

believed there was a substantial risk of error at each stage of the forensic science process, and that each stage involves a large amount of human judgement.³²

In short, the authors found that,

US respondents believe that there is a high degree of human judgement involved and high risk of an error occurring at each stage of the forensic science process. When considering forensic science techniques specifically, those in the US hold a skeptical view of the vast majority of techniques, viewing some of them as little more accurate than a coin flip, and no technique more than 84% accurate.³³

Kaplan's results corroborate the findings of a similar study from Australia. In that work, Gianni Ribeiro and colleagues³⁴ found, contrary to their expectations, that study participants believed that the forensic process involved considerable human judgment and was relatively prone to error. Specifically, the researchers found:

[P]articipants had wide-ranging beliefs about the accuracy of various forensic techniques, ranging from 65.18% (document analysis) up to 89.95% (DNA). For some forensic techniques, estimates were lower than that found in experimental proficiency studies, suggesting that our participants are more skeptical of certain forensic evidence than they need to be.³⁵

Ribeiro concluded that, “[i]n this study, we have demonstrated that participants do not just blindly believe that all forensic techniques are highly accurate, which has previously been assumed in the CSI effect literature. Instead, our participants believe that the forensic science process is error prone and involves a considerable amount of human judgment at each and every stage.”³⁶

As surprising as these findings may be, they are not anomalous. Indeed, they are consistent with other research finding that study participants consistently undervalue the significance of forensic evidence. For example, Dale Nance, in a study that involved people called for jury service in Illinois, concluded that, “[l]ooking at the forest rather than the trees, the dominant problem the empirical research reveals is that jurors as a group tend to undervalue the scientific evidence.”³⁷

In a separate large-scale empirical study—again using members of an Illinois jury pool—Nance confirmed the findings of his earlier research that jurors tend to minimize forensic

³² *Id.* at 270.

³³ *Id.* at 271.

³⁴ Gianni Ribeiro et al., *Beliefs About Error Rates and Human Judgment in Forensic Science*, 297 *Forensic Sci. Int'l.* 138-47 (2019).

³⁵ *Id.* at 138.

³⁶ *Id.* at 146.

³⁷ Dale Nance & Scott Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability*, 42 *Jurimetrics J.* 403 (2002).

evidence.³⁸ Specifically, he found that “for the most part jurors’ innate skepticism and need to be convinced create a dominating undervaluation of the evidence.”³⁹

In a later study, Jason Schklar⁴⁰ found that “[a]lthough no published study has reported jurors’ naive expectancies of how likely it was that a DNA match report could have resulted from either random chance or a laboratory error, some evidence indicates that people think human errors in the DNA lab are *more likely* than proficiency test results have revealed.”⁴¹ In addition, Schklar concluded, “[t]he results of this study also suggest that jurors may not infer that DNA test results are error-free when they do not receive an LE [error rate] estimate.”⁴²

Most recently, William Thompson and Edward Newman⁴³ found that study participants undervalued forensic footwear evidence.⁴⁴ Their findings “indicate that perceptions of forensic science evidence are shaped by prior beliefs and expectations as well as expert testimony and consequently that the best way to characterize and explain forensic evidence may vary across forensic disciplines.”⁴⁵ The authors concluded, “The complexity of our findings suggests that the problem of how “best” to present forensic evidence to lay audiences may not have a single, simple solution.”⁴⁶

2. Error Rates

Professor Brandon Garrett, in a letter to the Committee, claimed that “[n]o conclusion can be reached about a method without qualification or discussion of error rates, because there is no type of expertise that does not have some error rate.” Memo, p. 121. The draft Committee Note reflects this view. See Memo, p. 143 (“Accurate testimony will ordinarily include a fair assessment of the rate of error of the methodology employed, based where appropriate on empirical studies of how often the method produces correct results, as well as other relevant limitations inherent in the methodology.”). But it is scientifically incorrect to assume that a single error rate can be attributed to a particular method or generally applied to all forensic examiners who practice that method.⁴⁷

³⁸ Dale Nance & Scott Morris, *Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability*, 34 J. Legal Stud. 395 (2005).

³⁹ *Id.* at 436.

⁴⁰ Jason Schklar & Shari Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 Law & Hum. Behav. 159 (1999).

⁴¹ *Id.* at 165 (emphasis added).

⁴² *Id.* at 178.

⁴³ See William Thompson & Eryn Newman, *Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents*, 39 Law & Hum. Behav. 332 (2015).

⁴⁴ Consistent with these results, other research has also found that study participants underutilize forensic evidence. See William Thompson & Edward Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defense Attorney’s Fallacy*, 11 Law & Hum. Behav. 167 (1987); David Faigman & A.J. Baglioni, *Bayes’ Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 Law & Hum. Behav. 1 (1988); Jane Goodman, *Jurors’ Comprehension and Assessment of Probabilistic Evidence*, 16 Am. J. Trial Advoc. 361 (1992).

⁴⁵ *Id.* at 332.

⁴⁶ *Id.* at 348.

⁴⁷ See, e.g., William Thompson et al., American Academy for the Advancement of Science *Forensic Science Assessments: A Quality and Gap Analysis* (2017) (“[I]t is unreasonable to think that the “error rate” of latent fingerprint examination can meaningfully be reduced to a single number or even a single set of numbers. At best, it might be

First, many experts, including skill and experience-based experts, will be unable to testify to a specific error rate. Consider the brain surgeon testifying in a medical malpractice suit. Based on the surgeon's experience performing and observing a procedure thousands of times, she opines that the failure to correctly clamp a particular artery led to the plaintiff's excess bleeding and subsequent paralyzing stroke. The surgeon's opinion, and her confidence in that opinion, may be tested on cross-examination and through rebuttal experts. But there is no error rate that accompanies the methodology used to reach that opinion. Similarly, the structural engineer who studies the collapse of a bridge and testifies that, in his opinion, the bridge had a specific design flaw need not provide an error rate in order to offer his skill and experience-based opinion.

Second, even error rate advocates concede that it is exceedingly difficult to accurately establish scientifically valid and generally applicable figures. PCAST contributor and Boston College Symposium participant Itiel Dror addressed this point in a recent paper in which he discussed the complexities and practical difficulties of establishing a valid error rate.⁴⁸ These include knowing ground truth facts, establishing appropriate databases, determining what counts as an error, deciding on an acceptable metric, and problems with the external or ecological validity⁴⁹ of generalizing a given rate to different situations and circumstances.⁵⁰ Dror observed that, "[p]roviding 'an error rate' for a forensic domain may be misleading because it is a function of numerous parameters and depends on a variety of factors."⁵¹ He then posed the following rhetorical question:

The need to properly establish error rates in forensic science is clear. But, given the time and effort it requires, as well as the inherent limitations of the very notion of error rates, is it worth it? And, how does it compare (or complement) other measures of performance (e.g., effective proficiency testing, quality assurance checks such as dip sampling and blind verification, accreditation, and ongoing training and development).⁵²

Given these limitations, perhaps the best one can do is to examine the compendium of relevant studies and view them as a composite measure of the potential range of error rates across a discipline⁵³—but one that is not necessarily applicable to any particular case or examiner (due

possible to describe, in broad terms, the rates of false identifications and false exclusions likely to arise for comparisons of a given level of difficulty.”).

⁴⁸ Itiel Dror, *The Error in Error Rate: Why Error Rates Are So Needed, Yet So Elusive* 65 *J. Forensic Sci.*, 1034 (2020).

⁴⁹ Ecological validity refers to “a kind of external validity referring to the generalizability of findings from one group to another group.” W. Paul Vogt, *Dictionary of Statistics and Methodology* 78 (Sage Publications 1993).

⁵⁰ Dror, *supra* note 48, at 1034.

⁵¹ *Id.* at 1037.

⁵² *Id.* at 1038.

⁵³ *Daubert* discussed the known or *potential* rate of error. See also The American Association for the Advancement of Science (AAAS) recently published a study on latent fingerprints (William Thompson et al., *Forensic Science Assessments: A Quality and Gap Analysis* (2017)) that discussed the concept of “convergent validity,” an approach that draws conclusions about method validity from the body of relevant literature *as a whole*, recognizing that various study designs have different strengths and weaknesses. It also recognized that some studies can reinforce others and collectively support conclusions not otherwise warranted. Thompson, at 44. See also NAT'L RESEARCH COUNCIL, NAT'L ACADS., *THE EVALUATION OF FORENSIC DNA EVIDENCE* 85, 87 (1996) (“The question to be

to the scientific limitations imposed by external/ecological validity). For example, the composite false positive error rate derived from extant firearms-toolmarks studies is at or below 1%—a rate consistent with that detected by the largest latent fingerprint study to date.⁵⁴

The Department provided the Committee with the results of an ongoing firearms-toolmarks experiment by Mark Keisler and Stacey Hartman, *Isolated Pairs Research Study*, 50 AFTE Journal 56 (Winter 2018). The false positive error rate for that study is currently zero. This finding is consistent with the low false positive error rates recorded by numerous research studies in the firearms-toolmarks discipline of various experimental design.

A new firearms-toolmarks open-set black box study conducted by Jamie A. Smith was recently accepted for publication in the peer-reviewed *Journal of Forensic Sciences*.⁵⁵ The study was undertaken in response to the PCAST Report’s criticism of closed set experimental designs used in some past firearms-toolmarks studies. Smith’s study involved 72 qualified firearms examiners who compared bullets fired from 30 consecutively manufactured barrels (which makes comparisons much more difficult than those typically encountered during casework). The study’s false positive error rate was calculated to be 0.08% with only 1 false association recorded in 1,250 comparisons.⁵⁶

Finally, consider that the PCAST Report said the following about forensic error rates: “To be considered reliable, the FPR [false positive rate] should certainly be less than 5 percent and it may be appropriate that it be considerably lower, depending on the intended application.”⁵⁷ The extant studies (including black box and other designs) for firearms-toolmarks and latent fingerprints consistently record false positive error rates at or less than 1%—well below PCAST’s recommended 5% upper threshold.

A table of firearms-toolmarks studies that have measured false positive error rates for examiner-participants who conducted forensic comparisons of spent bullets and/or shell casings is appended to this letter as Attachment A.

decided is not the general error rate for a laboratory or laboratories over time but rather whether the laboratory doing DNA testing in this particular case made a critical error.”) and (“The risk of error is properly considered case by case, taking into account the record of the laboratory performing the tests, the extent of redundancy, and the overall quality of the results”).

⁵⁴ For latent prints, in the largest-scale study to date, involving 169 examiners and 17,121 total decisions, the false positive error rate was 0.1%. Bradford Ulery et al., *Accuracy of Forensic Latent Fingerprint Decisions*, 108 Proceedings of the National Academy of Sciences 7733-38 (2011).

⁵⁵ Jamie A. Smith, Beretta Barrel Fired Bullet Validation Study, *Journal of Forensic Sciences* (accepted for publication October 2, 2020).

⁵⁶ It is important to note that experimental study error rates do not translate to laboratory error rates, as comparisons performed during studies do not have the benefit of verification performed by a second examiner or a laboratory’s quality assurance measures. In this regard, see BALDWIN ET AL., A STUDY OF FALSE POSITIVE AND FALSE NEGATIVE ERROR RATES IN CARTRIDGE CASE COMPARISON 18 (2014), <https://www.ncjrs.gov/pdffiles1/nij/249874.pdf>: (“This finding [a 1.0% false positive error rate] does not mean that 1% of the time each examiner will make a false-positive error. Nor does it mean that 1% of the time laboratories or agencies would report false positives, since this study did not include standard or existing quality assurance procedures, such as peer review or blind reanalysis.”).

⁵⁷ PCAST Report, *supra* note 12, at 152.

The Transactional Cost of a Rule Change to FRE 702

During the October 2017 roundtable that the Committee hosted in Boston, there seemed to be a consensus among participants that Rules 702 and 104(a) already provide the correct standard by which courts should assess the admissibility of expert testimony. The discussion was more focused on whether there was value in tweaking the rules to emphasize that courts should follow the existing rules, and in so doing, use the rule change to more broadly discuss the topic in a committee note. On the issue of admissibility versus weight, Judge James O. Browning—a participant in the Boston roundtable—subsequently wrote the following in a published opinion:

Rule 702’s most prominent hurdle is the sufficiency of basis. Yet the judiciary’s uncomfortableness with analyzing an opinion’s basis can be seen in the conflict in the cases. The current conflict is whether the questions of sufficiency of basis, and of application of principles and methods, are matters of weight or admissibility.

*** There should not be a conflict. Rule 702 states that these are questions of admissibility. Yet many courts treat them as questions of weight. *** The Court is concerned that the federal courts will overact to the wayward opinions that have created a split whether sufficiency of basis and application of methods is for the court or goes to the evidence’s weight. The Court is concerned that the federal courts are going in the direction of new rules. *** The development of new rules burdens the federal judiciary and the bar -- all of which are overworked -- with mandatory changes each year, often constituting little more than stylistic changes. Everyone has to get new rule books every year. The burden of new rules often does not justify the meager benefits of the changes.

Walker v. Spina, et al, Civil Action No. 17-0991 JB\SCY (D.N.M. Jan. 9, 2019) (Doc. 111), p. 32, n. 11 (internal citations omitted).

Judge Browning’s observation is especially apt here, where proposed textual changes are not strictly necessary, but open the door to sweeping commentary in the note. Here, the proposed note is already obsolete, and would only become further outdated by the time an amendment takes effect. Forensic science is a quickly evolving discipline where new studies constantly add to a growing body of knowledge. *See, e.g., Harris, supra* at *2 (“recent advancements in the field in the four years since the PCAST Report address many of Mr. Harris’s concerns). Studies conducted in the last few years already undermine the lead premise of the proposed note, *i.e.*, that jurors overvalue forensic testimony. Given the swift pace of forensic and social science research, the slow pace of rulemaking, and the permanence of Committee notes, we propose restraint. Other methods exist to educate courts on the correct application of Rule 702. The language of the Federal Rules already provide courts the tools necessary to regulate expert testimony, and many courts are actively doing so.

Respectfully,

/s/ Elizabeth J. Shapiro
Elizabeth J. Shapiro, Deputy Director
Ted R. Hunt, Senior Advisor on Forensic Science
U.S. Department of Justice

Appendix A

Significant Firearms-Toolmarks False Positive Error Rate Studies

Lead Author	Source	Year	Number of Participants	False Positive Rate (%)	Comparison Type Cases/Bullets
*Brundage	AFTE Journal	1998	30 (Plus 37 Informal Participants)	0	Bullets
Bunch	AFTE Journal	2003	8	0	Cartridge Cases
DeFrance	AFTE Journal	2003	9	0	Bullets
Smith	AFTE Journal	2004	8	0	Both
*Hamby	AFTE Journal	2009	507 (Includes *Brundage (1998) Participants)	0	Bullets
Lyons	AFTE Journal	2009	22	1.2 ^a	Cartridge Cases
Mayland	AFTE Journal	2010	64	1.7 ^b	Cartridge Cases
Cazes	AFTE Journal	2013	68 (or 69)	0	Cartridge Cases
Fadul	AFTE Journal	2013	Phase 1: 217 Phase 2: 114	Phase 1: .064 ^c Phase 2: 0.18 ^c	Cartridge Cases
Fadul	NIJ (NCJRS)	2013	183	0.40 ^d	Bullets
Stroman	AFTE Journal	2014	25	0	Cartridge Cases
Baldwin	NIJ (NCJRS)	2014	218	1.0	Cartridge Cases
Kerkhoff	Science & Justice	2015	11	0	Both
Smith	JFS	2016	31	0.14 Cases 0 Bullets	Cartridge Cases Bullets
Duez	JFS	2018	46 Examiners 10 trainees	0 ^e	Cartridge Cases
Keisler	AFTE Journal	2018	126	0	Cartridge Cases
*Hamby	JFS	2019	619 (Includes *Brundage (1998) and Hamby (2009) Participants)	0.053% ^f	Bullets
Smith	Journal of Forensic Sciences (Accepted)	2020	72	0.08%	Bullets

*Brundage study was continued by Hamby who added additional participants and reported the combined data in fall 2009 and 2019.

^a The error rate reported by the author appears to be (1-True Positive Rate). There were three false positive identifications made but the number of true negative comparisons is not reported. 259

correct positive identifications were made. The False Discovery Rate (FDR) for the study is $3/(3+259) = 1.1\%$.

^b The false positive error rate is not reported by the authors. There were three false positive identifications and 178 correct positive identifications made. The False Discovery Rate (FDR) for the study is $3/(3+178) = 1.7\%$ and is reported in the table above.

^c The error rates reported by the authors are roughly equivalent to the False Discovery Rates (FDR) for each of the study phases (FDR = .062% and 0.18% respectively).

^d Eleven false positives occurred. The false positive error rate is not reported by the authors. The error rate quoted is equivalent to the False Discovery Rate = $11/(11+2734) = 0.40\%$.

^e Two false positives were made by one trainee. None were made by the qualified examiners. The false positive rate does not include the trainee errors. If trainee data is included with that submitted by examiners, the False Positive Rate is $(2/112) = 1.8\%$.

^f The empirically observed false positive rate is 0%. Using Bayesian estimation methods, the authors' most conservative (worst case) estimate of the average examiner false positive error rate for the study is .053% with a 95% credible interval of $(1.1 \times 10^{-5}\%, 0.16\%)$.

List of References

1. Brundage, D. (Summer 1998). The Identification of Consecutively Rifled Gun Barrels, *AFTE Journal*, 30(3), 438-44 (Bullets).
2. Bunch, S.G., & Murphy, D.P. (Spring 2003). A Comprehensive Validity Study for the Forensic Examination of Cartridge Cases, *AFTE Journal*, 35(2), 201-03 (Cartridge Cases).
3. DeFrance, C.S. & Van Arsdale, M.D. (Winter 2003). Validation Study of Electrochemical Rifling, *AFTE Journal*, 35(1), 35-37 (Bullets).
4. Smith, E.D. (Fall 2004). Cartridge Case and Bullet Comparison Validation Study with Firearms Submitted in Casework, *AFTE Journal*, 36(4), 130-35 (Bullets and Cartridge Cases).
5. Hamby, J.E., Brundage, D.J., & Thorpe, J.W. (Spring 2009). The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries, *AFTE Journal*, 41(2), 99-110 (Bullets).
6. Lyons, D.J. (Summer 2009). The Identification of Consecutively Manufactured Extractors, *AFTE Journal*, 41(3), 246-56 (Cartridge Cases).
7. Mayland, B. & Tucker, C. (Spring 2012). Validation of Obturation Marks in Consecutively Reamed Chambers, *AFTE Journal*, 44(2), 167-69 (Cartridge Cases).
8. Cazes, M. & Goudeau, J. (Spring 2013). Validation Study Results from Hi-Point Consecutively Manufactured Slides, *AFTE Journal*, 45(2), 175-77 (Cartridge Cases).

9. Fadul Jr., T.G., Hernandez, G.A., Wilson, E., Stoiloff, S., & Gulati, S. (Fall 2013). An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides, *AFTE Journal*, 45(4), 376-93 (Cartridge Cases).
10. Fadul Jr., T.G., Hernandez, G.A., Wilson, E., Stoiloff, S., & Gulati, S. (December 2013). An Empirical Study to Improve the Foundation of Firearm and Tool Mark Identification Utilizing Consecutively Manufactured Glock EBIS Barrels with the Same EBIS Pattern. <https://www.ncjrs.gov/pdffiles1/nij/grants/244232.pdf> (Bullets).
11. Stroman, A. (Spring 2014), Empirically Determined Frequency of Error in Cartridge Case Examinations Using a Declared Double Blind Format, *AFTE Journal*, 46(2), 157-75 (Cartridge Cases).
12. Baldwin, D.P., Bajic, S.J., Morris, M., & Zamzow, D. (April 7, 2014). A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a611807.pdf> (Cartridge Cases).
13. Kerkhoff, W. et al. (2015). Design and Results of an Exploratory Double Blind Testing Program in Firearms Examination, *Science & Justice*, 55, 514-19 (Bullets and Cartridge Cases).
14. Smith, T.P., Smith, A.G., & Snipes, J.B. (July 2016). A Validation Study of Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework, *Journal of Forensic Sciences*, 61(4), 939-45 (Cartridge Cases).
15. Duez, P. et al. (July 2018). Development and Validation of a Virtual Examination Tool for Firearm Forensics, *Journal of Forensic Sciences*, Vol. 63(4), 1069-1084 (Cartridge Cases).
16. Keisler, M. et al. (Winter 2018). Isolated Pairs Research Study, *AFTE Journal*, 50(1), 56-58 (Cartridge Cases).
17. Hamby, J. et al. (March 2019). A Worldwide Study of Bullets Fired From 10 Consecutively Rifled 9MM Ruger Pistol Barrels—Analysis of Examiner Error Rates, *Journal of Forensic Sciences*, 64(2), 551-57 (Bullets).
18. Smith, J. (2020). Beretta Barrel Fired Bullet Validation Study, *Journal of Forensic Sciences* (accepted for publication October 2, 2020) (Bullets).

**UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA**

UNITED STATES OF AMERICA, :
 :
 v. : Criminal Action No.: 19-358 (RC)
 :
 DEMONTRA HARRIS, : Re Document No.: 22
 :
 Defendant. :

MEMORANDUM OPINION

**DENYING DEFENDANT’S MOTION IN LIMINE TO EXCLUDE EXPERT TESTIMONY AS TO
FIREARM EXAMINATION TESTING**

I. INTRODUCTION

Defendant Demontra Harris is charged with unlawful possession of a firearm as a person previously convicted of a felony, assault with a dangerous weapon, and possession of a firearm during a crime of violence. Superseding Indictment at 1–2, ECF No. 39. On July 24, 2019, the D.C. Metropolitan Police Department (“MPD”) responded to a report of gunshots and recovered four 9mm shell casings from the incident scene, which were then entered into the National Integrated Ballistic Information Network (“NIBIN”). A witness later provided MPD with a video filmed that night that allegedly shows Mr. Harris holding and then discharging a firearm in the location where the shell casings were later discovered. No firearm was recovered at the time. Roughly six weeks later on September 8, 2019, during a response to a call for service for a person with a weapon, MPD recovered a Glock 17 Gen4 9x19 pistol (“Glock 17”). This recovered firearm was test-fired and the resulting casings were entered into the NIBIN, where a match was identified with the casings recovered on the night of July 24, 2019. The Government then submitted the relevant evidence to an independent firearms examiner for forensic examination. Chris Monturo, a tool mark examiner who operates the Ohio-based forensic

services firm Precision Forensic Testing, examined the evidence and concluded in a report that he believed the four recovered casings from the July 24, 2019 incident scene were fired by the recovered Glock 17. *See* March 14, 2020 Report of Chris Monturo (“Monturo Report”), ECF No. 22-2. The Government intends to call Mr. Monturo to testify regarding these findings at the upcoming trial in this matter.

This opinion addresses Mr. Harris’s *motion in limine* to Exclude Expert Testimony as to Firearm Examination Testing (“Def.’s Mot.”), ECF No. 22, pursuant to *Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579 (1993), Federal Rule of Evidence 702, and Federal Rule of Evidence 403. Def.’s Mot. at 1–2. The motion has been fully briefed, with both parties also filing supplemental motions. *See generally* Def.’s Mot.; Govt.’s Opp’n to Def.’s Mot. to Excl. Firearm and Toolmark Testimony (“Govt. Opp’n”), ECF No. 28; Def.’s Supp. Mot. to Excl. Expert Testimony as to Firearm Exam. Testing (“Def.’s Supp. Mot.”), ECF No. 32; Govt.’s Opp’n to Def.’s Supp. to Excl. Firearm and Toolmark Testimony (“Govt. Supp. Opp’n”), ECF No. 33. In addition, the Court conducted a Daubert hearing on October 15, 2020 to consider this issue, taking the testimony of Todd Weller, an expert in the field. A jury trial in this matter is currently scheduled to begin on November 12, 2020.

Mr. Harris argues that the field of firearm and toolmark identification lacks a reliable scientific basis and is not premised on sufficient facts or data, is not the product of reliable principles and methods, and was not applied properly by Mr. Monturo to the facts of the case. Def.’s Mot. at 1–2. The Court disagrees, and will admit Mr. Monturo’s testimony to the extent it falls within the Department of Justice’s Uniform Language for Testimony of Reports for the Forensic Firearms/Toolmarks Discipline – Pattern Matching Examination (“DOJ ULTR”). While Mr. Harris raises important issues as to the reliability of firearm and toolmark

identification, memorialized most notably by the 2016 President’s Council of Advisors on Science and Technology Report (“PCAST Report”), these issues are for cross-examination, not exclusion, as recent advancements in the field in the four years since the PCAST Report address many of Mr. Harris’s concerns. Mr. Harris also remains free to have his own expert examine the firearm and ballistics evidence and contradict the Government’s case.

II. ANALYSIS

A. Legal Standard

“Motions *in limine* are designed to narrow the evidentiary issues at trial.” *Williams v. Johnson*, 747 F. Supp. 2d 10, 14 (D.D.C. 2010). “While neither the Federal Rules of Civil Procedure nor the Federal Rules of Evidence expressly provide for motions *in limine*, the Court may allow such motions ‘pursuant to the district court’s inherent authority to manage the course of trials.’” *Barnes v. District of Columbia*, 924 F. Supp. 2d 74, 78 (D.D.C. 2013) (quoting *Luce v. United States*, 469 U.S. 38, 41 n.4 (1984)).

Federal Rule of Evidence 702 provides that qualified expert testimony is admissible if “(a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.” Fed. R. Evid. 702. “In general, Rule 702 has been interpreted to favor admissibility.” *Khairkhwa v. Obama*, 793 F. Supp. 2d 1, 10 (D.D.C. 2011) (citing *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 587 (1993); Fed. R. Evid. 702 advisory committee’s note to 2000 amendment (“A review of the caselaw after *Daubert* shows that the rejection of expert testimony is the exception rather than the rule.”). Indeed, the Supreme Court has clarified that it is not exclusion, but rather “vigorous

cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof” that “are the traditional and appropriate means of attacking shaky but admissible evidence.” *Daubert*, 509 U.S. at 596.

When considering the admissibility of expert evidence under Federal Rule of Evidence 702, district courts are required to “assume a ‘gatekeeping role,’ ensuring that the methodology underlying an expert’s testimony is valid and the expert’s conclusions are based on ‘good grounds.’” *Chesapeake Climate Action Network v. Export-Import Bank of the U.S.*, 78 F. Supp. 3d 208, 219 (D.D.C. 2015) (quoting *Daubert*, 509 U.S. at 590–97). This gatekeeping analysis is “flexible,” and “the law grants a district court the same broad latitude when it decides how to determine reliability as it enjoys in respect to its ultimate reliability determination.” *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141–42 (1999) (emphasis omitted). While district courts may apply a variety of different factors to assess reliability, in *Daubert* the Supreme Court provided a non-exhaustive list of five factors to guide the determination, including: (1) whether the technique has been or can be tested; (2) whether the technique has a known or potential rate of error; (3) if the technique has been subject to peer review and publishing; (4) the existence of controls that govern the technique’s operation; and (5) whether the technique has been generally accepted within the relevant scientific community. *See Daubert*, 509 U.S. at 593–94. In contrast, expert testimony “that rests solely on ‘subjective belief or unsupported speculation’ is not reliable.” *Groobert v. President & Directors of Georgetown Coll.*, 219 F. Supp. 2d 1, 6 (D.D.C. 2002) (citing *Daubert*, 509 U.S. at 590).

“The burden is on the proponent of [expert] testimony to show by a preponderance of the evidence that . . . the testimony is reliable.” *Sykes v. Napolitano*, 634 F. Supp. 2d 1, 6 (D.D.C. 2009) (citing *Meister v. Med. Eng’g Corp.*, 267 F.3d 1123, 1127 n.9 (D.C. Cir. 2001)). Even if

the proposed expert testimony is reliable, the Court may nonetheless exclude it “if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.” Fed. R. Evid. 403; *see Bazarian Int’l Fin. Assocs., LLC v. Desarrollos Aerohotelco, C.A.*, 315 F. Supp. 3d 101, 128 (D.D.C. 2018) (analyzing expert testimony under Rule 403).

B. Firearm and Toolmark Identification

1. Firearm and Toolmark Identification Science

Mr. Harris’s motion challenges the reliability of the Government’s proposed use of firearm toolmark identification as a discipline for expert testimony. Firearm identification began as a forensic discipline in the 1920s, *see* James E. Hamby, *The History of Firearm and Toolmark Identification*, 31 Ass’n of Firearm and Tool Mark Examiners J. 266, 266–284 (1999), and “for decades” has been routinely admitted as appropriate expert testimony in district courts. *United States v. Taylor*, 663 F. Supp. 2d 1170, 1175 (D.N.M. 2009); *see also United States v. Brown*, 973 F.3d 667, 704 (7th Cir. 2020) (noting firearm and toolmark identification has been “almost uniformly accepted by federal courts”) (citations omitted).

Firearm and toolmark identification “is used to determine whether a bullet or casing was fired from a particular firearm.” *Brown*, 973 F.3d at 704. A firearm and toolmark examiner will make this determination “by looking through a microscope to see markings that are imprinted on the bullet or casing by the firearm during the firing process,” which will include marks left on the bullet by the firing pin as well as scratches that occur when the bullet travels down the barrel. *Id.*

A firearm examiner is trained to observe and classify these marks into three types of characteristics during a firearm toolmark examination, which include:

(1) Class characteristics: i.e., the weight or caliber of the bullet, the number of lands and grooves, the twist of the lands and grooves, and the width of the lands and grooves, that appear on all bullet casings fired from the same type of weapon and are predetermined by the gun manufacturer;

(2) Individual characteristics: unique, microscopic, random imperfections in the barrel or firing mechanism created by the manufacturing process and/or damage to the gun post-manufacture, such as striated and/or impressed marks, unique to a single gun; and

(3) Subclass characteristics: characteristics that exist, for example, within a particular batch of firearms due to imperfections in the manufacturing tool that persist during the manufacture of multiple firearm components mass-produced at the same time.

Ricks v. Pauch, No. 17-12784, 2020 WL 1491750, at *8–9 (E.D. Mich. Mar. 23, 2020).

A qualified examiner can conclude that casings were fired by the particular firearm by “comparatively examining bullets and determining whether ‘sufficient agreement’ of toolmarks exist,” which occurs when the class and individual characteristics match. *Id.* at *9; *see also Brown*, 973 F.3d at 704. The methodology of determining when sufficient agreement is present is detailed by the Association of Firearm Toolmark Examiners (“AFTE method”), and is “the field’s established standard.” *United States v. Ashburn*, 88 F. Supp. 3d 239, 246 (E.D.N.Y. 2015). Under the governing AFTE theory, no two firearms will bear the same microscopically identical toolmarks due to differences in individual characteristics. *United States v. Otero*, 849 F. Supp. 2d 425, 427 (D.N.J. 2012).

In recent years three scientific reports have examined the underlying scientific validity of firearm and toolmark identification. They include the 2008 Ballistic Imaging Report, Def.’s Supp. Mot. Ex. 1, ECF No. 32-1, the 2009 National Academy of Science Report, Def.’s Supp. Mot. Ex. 2, ECF No. 32-2, and the 2016 President’s Council of Advisors on Science and Technology Report (“PCAST Report”), Def.’s Supp. Mot. Ex. 3, ECF No. 32-3. Mr. Harris argues that these reports “reject the claim that firearms identification is a valid and reliable

science.” Def.’s Supp. Mot. at 2–3. The Court is generally convinced by the Government’s arguments and ample citations to case law that the 2008 Ballistic Imaging Report and the 2009 National Academy of Science Report are both “outdated by over a decade” due to intervening scientific studies and as a result have been repeatedly rejected by courts as a proper basis to exclude firearm and toolmark identification testimony. Govt. Supp. Opp’n at 2–4 (collecting cases holding firearms identification evidence admissible after considering these reports). The PCAST Report provides better support for Mr. Harris’s arguments, given its more recent origin and use in recent opinions that have interrogated the danger of subjectivity in this discipline. *See, e.g., United States v. Tibbs*, No. 2016-CF1-19431, 2019 WL 4359486 (D.C. Super. Ct. Sept. 5, 2019).

The PCAST Report ultimately concluded that firearm and toolmark identification fell “short of the criteria for foundational validity,” after raising a number of critiques of the science. PCAST Report at 11. Chief among them was that the report concluded that “foundational validity can only be established through multiple independent black-box studies”¹ and at the time the report was published in 2016, there had only been one black-box study conducted on the discipline to date. Def.’s Supp. Mot. at 4 (citing PCAST Report at 106, 111). In response, the Government has put forth sworn affidavits from researchers that speak to post-PCAST Report scientific studies that they argue contradicts the PCAST Report’s conclusions. The Government’s Daubert hearing expert, Todd Weller, devoted much of his testimony to

¹ The PCAST report defined a black-box study as “an empirical study that assesses a subjective method by having examiners analyze samples and render opinions about the origin or similarity of samples.” PCAST Report at 48. Mr. Weller added at the Evidentiary Hearing that a black-box study is one in which there are “question samples [given to examiners] that have a matching known, and question samples that do not have a matching known, and also that each of those comparisons is independent from each other.” October 15, 2020 Evidentiary Hearing Tr. (“Evid. Hr’g Tr.”) 49:6-12.

discussing the scientific advances that have occurred since the PCAST Report was published in 2016, all of which he posited affirms the discipline's validity. *See generally* Evid. Hr'g Tr.

2. Mr. Monturo's Report Methodology

Mr. Harris's *motion in limine* specifically challenges the proposed testimony of the Government's firearm and ballistics expert Chris Monturo, who examined the firearms evidence at issue in this case. In creating his report for the Government, Mr. Monturo first test fired the Glock 17 and found it to be operable. Monturo Report at 2. He then used the Glock 17 to create test-fired cartridge cases. *Id.* Mr. Monturo then microscopically compared his test-fired cartridge cases to the cartridge cases recovered from the crime scene on July 26, 2019, and found the two sets of cartridges "to have corresponding individual characteristics." *Id.* These results were then verified that same day by Calissa Chapin, another qualified firearm and ballistics expert from Mr. Monturo's lab. March 14, 2020 Report of Chris Monturo Notes ("Monturo Report Notes") at 3, ECF No. 22-3. As a result, Mr. Monturo is expected to testify that "[b]ased upon these corresponding individual characteristics. . . namely aperture sheer marks,"² "along with Mr. Monturo's training and experience, [he] is of the opinion that the Glock firearm fired" the cartridge casings recovered from the July 26, 2019 crime scene. Govt. Opp'n at 11–12.

C. The Subject Matter of Mr. Monturo's Testimony Meets Rule 702's Standards

Mr. Harris argues that the Government's proposed expert must be excluded under Rule 702 and *Daubert* because the underlying firearm and toolmark identification discipline "is based

² As defined in the AFTE Glossary, 6th Edition, a firing pin aperture shear is "[s]triated marks caused by the rough edges of the firing pin aperture scraping the primer metal during unlocking of the breech." Govt. Supp. Opp'n, Ex. 15, ECF No. 33-15. It is these individual characteristics Mr. Monturo used to classify the cartridge cases at issue.

not upon science but rather ‘subjectivity.’”³ Def.’s Supp. Mot. at 2. To address Mr. Harris’s concerns about the admission of Mr. Monturo’s expert testimony, the Court will undertake a factor-by-factor analysis of the discipline’s reliability, using *Daubert* as a guide. Complicating this process is the fact that Mr. Harris did not specifically address the *Daubert* criteria in his briefing on this topic, so the Court will instead rely on the implications raised by the PCAST Report and other scientific reports he has brought to the Court’s attention.

1. Whether the methodology has been tested

As previously noted, the first *Daubert* factor asks whether the technique in question has been or can be tested. *See Daubert*, 509 U.S. at 593–94. This “testability” inquiry, as articulated in the Advisory Committee Notes to Rule 702, concerns “whether the expert’s theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot be reasonably assessed for reliability.” Fed. R. Evid. 702 advisory committee’s note to 2000 amendment. Mr. Harris argues that firearm and toolmark identification is “unavoidably subjective,” and also cites to the 2008 Ballistics Imaging Report which expressed concerns about “the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks.” Def.’s Supp. Mot. at 2–3. In response, the Government has put forth evidence to show “[f]irearms and toolmark identification has been thoroughly tested with

³ Based on remarks such as these and his citation to *United States v. Glynn*, Mr. Harris appears to be peripherally raising the point that firearm and toolmark identification cannot “fairly be called ‘science,’” *United States v. Glynn*, 578 F. Supp. 2d 567, 570 (S.D.N.Y. 2008), a preliminary inquiry some courts have investigated before proceeding to the *Daubert* analysis. The Court does not believe such an inquiry is required here, given that, as other courts have also found, firearm and toolmaking identification is “clearly is technical or specialized, and therefore within the scope of Rule 702.” *United States v. Hunt*, No. CR-19-073-R, 2020 WL 2842844, at *3 n.2 (W.D. Okla. June 1, 2020) (citing *United States v. Willock*, 696 F. Supp. 2d 536, 571 (D. Md. 2010), *aff’d sub nom. United States v. Mouzone*, 687 F.3d 207 (4th Cir. 2012)).

ground-truth experiments designed to mimic casework.” Govt. Opp’n at 1. The Court agrees with the Government that this factor supports admissibility.

A number of courts have examined this factor in depth to conclude that firearm toolmark identification can be tested and reproduced. *See, e.g., Otero*, 849 F. Supp. 2d at 432 (“The literature shows that the many studies demonstrating the uniqueness and reproducibility of firearms toolmarks have been conducted.”); *Taylor*, 663 F. Supp. 2d at 1175–76 (noting studies “demonstrating that the methods underlying firearms identification can, at least to some degree, be tested and reproduced.”); *United States v. Diaz*, No. CR 05-00167, 2007 WL 485967, at *6 (N.D. Cal. Feb. 12, 2007) (holding that “the theory of firearms identification, though based on examiners’ subjective assessment of individual characteristics, has been and can be tested.”). Indeed, even Judge Edelman in the *Tibbs* opinion relied on by Mr. Harris concluded that “virtually every court that has evaluated the admissibility of firearms and toolmark identification has found the AFTE method to be testable and that the method has been repeatedly tested.” *Tibbs*, 2019 WL 439486 at *7 (collecting cases).

The fact that there are subjective elements to the firearm and toolmark identification methodology is not enough to show that the theory is not “testable.” Indeed, studies have shown that “the AFTE theory is testable on the basis of achieving consistent and accurate results.” *Otero*, 849 F. Supp. 2d at 433; *see also* July 7, 2017 Decl. of Todd Weller (“Weller I”) at 2–6, ECF No. 28-5 (describing various studies that support the reproducibility of the AFTE identification theory). This conclusion has only been further strengthened in recent years due to advances in three-dimensional imaging technology, which has allowed the field to interrogate the process and sources of “subjectivity” behind firearm and toolmark examiners’ conclusions. For example, Mr. Weller testified regarding a study which used 3D image technology to assess the

process used by trained firearm examiners when identifying casings to a particular firearm. *See* Sept. 19, 2019 Decl. of Todd Weller (“Weller II”) at 15–16 (citing Pierre Duez et al., *Development and Validation of a Virtual Examination Tool for Firearm Forensics*, 63 J. Forensic Sci, 1069–84 (2018), (“Heat Map Study”)), ECF No. 28-6. The Heat Map Study indicated that firearm examiners from fifteen different laboratories, all conducting an independent assessment, were “mostly using the same amount and same location of microscopic marks when concluding identification.” Weller II at 16. Critically, the trained examiners also correctly reported 100% of known matches while reporting no false positives or false negatives. *Id.*

It is also important to note that the testability criticism leveled at the firearm and toolmark field in the PCAST Report—that at the time of publishing “there [was] only a single appropriately designed study to measure validity and estimate reliability”—appears to now be out of date. PCAST Report at 112. As previously discussed, the PCAST Report only considered studies that were a “black-box” or “open-set” design, disregarding hundreds of validation studies in the process. *See* Evid. Hr’g Tr. 48:9-17 (noting that PCAST only evaluated nine of the hundreds of studies that were submitted for review). Setting aside for the moment the utility of this “black-box” requirement— which goes beyond what is required by Rule 702— the Government has provided to the Court three recent scientific studies that meet the PCAST’s black-box model requirements and demonstrate the reliability of the firearm and toolmark identification method. These include one of the tests administered during the Heat Map Study detailed above, *see* Weller II at 16 n. 84, along with another recent black box study testing the identification of fired casings, which resulted in a .433% false positive error rate from three errors among 693 total comparisons. *See* Lilien et al., *Results of the 3D Virtual Comparison*

Microscopy Error Rate (VCMER) Study for Firearm Forensics, J. of Forensic Sci. Oct. 1, 2020 (“Lilien Study”) at 1, ECF No. 41. A third post-PCAST Report study also followed the PCAST recommended black-box model and found that of 1512 possible identifications tested, firearms examiners correctly identified 1508 casings to the firearm from which the casing was fired. Keisler et. al., *Isolated Pairs Research Study*, Ass’n of Firearm and Tool Mark Examiners J. 56, 58 (2018) (“Keisler Study”), ECF No. 33-9; *see also* Evid. Hr’g Tr. 65:3-11. This evidence indicates that even under the PCAST’s stringent black-box only criteria, firearm and toolmark identification can be tested and reasonably assessed for reliability.

A final factor demonstrating the strength of the testability prong is that firearm and toolmark examiners are required, as Mr. Monturo has done here, to document their results and findings through written reports and photo documentation, and have these results validated by another qualified examiner. These elements “ensure sufficient testability and reproducibility to ensure that the results of the technique are reliable.” *Diaz*, 2007 WL 485967 at *5 (citing *United States v. Monteiro*, 407 F.Supp.2d 351, 369 (D. Mass. 2006)).⁴ For all of these reasons, the Court concludes that the testability factor supports admissibility of Mr. Monturo’s testimony.

2. The known or potential error rate

The second *Daubert* factor inquires as to whether the technique has a known or potential rate of error. *See Daubert*, 509 U.S. at 594. The PCAST Report concluded that non-black box

⁴ Mr. Harris’s only explicit acknowledgement of this *Daubert* factor is an assertion in a parenthetical that the court in *United States v. Green* found that “ballistic evidence fails to meet *Daubert* criteria regarding . . . testability.” Def.’s Mot. at 7 (citing *United States v. Green*, 405 F. Supp. 2d 104, 120–22 (D. Mass. 2005)). But the facts at issue in *Green* were quite different than the instant case. *Green*’s holding that the methods at issue could not be tested rested on an absence of notes and photographs from the initial examination that “made it difficult, if not impossible” for another expert to verify the examination. *Green*, 405 F. Supp. 2d at 120. In contrast, Mr. Monturo documented his work in addition to having it verified that same day by another certified firearms analyst. Accordingly, reproducibility is not at issue here.

studies had “inconclusive and false-positives rate that are dramatically lower (by more than 100-fold)” compared to partly black-box or fully black-box designed studies. PCAST Report at 109. The Government counters that “collectively, th[e] body of scientific data demonstrate[s] a low rate of error” for firearm and toolmark identification, and provides several recently published studies to refute the PCAST Report’s finding of differences in rate of error tied to study design. Govt. Opp’n at 2; Govt. Supp. Opp’n at 13–14.

First, as the Government argues and this Court agrees, the critical inquiry under this factor is the rate of error in which an examiner makes a false positive identification, as this is the type of error that could lead to a conviction premised on faulty evidence. *See Otero*, 849 F. Supp. 2d at 434 (noting, “the critical validation analysis has to be the extent to which false positives occur”).⁵ Mr. Weller testified that “over the past couple of decades in research” he had seen a rate of false positives in research studies ranging from 0-1.6 percent. Evid. Hr’g. Tr. 84:19–22. To support this assertion, the Government provided the false positive error rates for nineteen firearm and toolmark validation studies conducted between 1998 and 2019, of which eleven studies had a false positive error rate of zero percent, and the highest false positive error rate calculated was 1.6%. Govt. Opp’n at 27–29. Other federal courts have also recognized that validation studies as a whole show a low rate of error for firearm and toolmark identification. *See, e.g., United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1119 (D. Nev. 2019) (“[T]he studies cited by [the firearms examiner] in his testimony and by other federal courts examining the issue universally report a low error rate for the AFTE method.”); *Taylor*, 663 F. Supp. 2d at 1177 (“[T]his number [less than 1%] suggests that the error rate is quite low”).

⁵ Perhaps the false negative rate could be important in a case where a defendant asserts his co-defendant (or a third party) was the culprit and examination of that person’s firearm tested negative. But that situation does not apply here.

As was the case under the testability prong of the *Daubert* analysis, here too recent studies have resolved some of the concerns raised by the PCAST Report. Mr. Weller described for the Court how three black box studies that post-date the PCAST Report all have extremely low rates of error. Govt. Supp. Opp'n at 14, Evid. Hr'g Tr. 65:2-77:8. The Heat Map and Keisler studies both had an overall error rate of zero percent, and the Lilien study produced a false positive rate of only 0.433%. Govt. Supp. Opp'n at 14. Because the evidence shows that error rates for false identifications made by trained examiners is low—even under the PCAST's black-box study requirements—this factor also weighs in favor of admitting Mr. Monturo's expert testimony.

3. Whether the methodology has been subject to peer review and publication

The third *Daubert* factor concerns if the methodology has been subject to peer review and published in scientific journals, a component the Supreme Court emphasized as critical to “good science” since “it increases the likelihood that substantive flaws in methodology will be detected.” *See Daubert*, 509 U.S. at 593–94. The Government contends that scientific data concerning firearms and toolmark identification “have been published in a multitude of scientific peer-reviewed journals,” Govt. Opp'n at 1, and Mr. Weller presented evidence to this effect at the evidentiary hearing, describing the variety of scientists from different disciplines who have published on the topic in several different peer-reviewed journals. *See Weller I* at 9–10. The Court agrees with the Government that this factor weighs in favor of admissibility.

Much of the literature in this discipline has been published in the AFTE Journal, a peer-reviewed journal that “publishes articles, studies and reports concerning firearm and toolmark evidence.” *United States v. McCluskey*, No. CR 10-2734 JCH, 2013 WL 12335325, at *6 (D.N.M. Feb. 7, 2013). The AFTE Journal uses a formal process for article submissions,

including “specific instructions for writing and submitting manuscripts, assignment of manuscripts to other experts within the scientific community for a technical review, returning of manuscripts to other experts within the scientific community for clarification or re-write, and a final review by the Editorial Committee.” *Id.* (quoting Richard Grzybowski, et al., *Firearm/Toolmark Identification: Passing the Reliability Test Under Federal and State Evidentiary Standards*, 35 AFTE J. 209, 220 (2003)).

Other courts have examined the scientific credibility of the AFTE Journal. Notably, the court in *Tibbs* concluded that the AFTE Journal’s lack of a double-blind peer review process along with the fact that it is published by the group of practicing firearms and toolmark examiners could create an “issue in terms of quality of peer review.” *Tibbs*, 2019 WL 4359486, at *10. In response, the Government asserts, citing to testimony from Dr. Bruce Budowle, “the most published forensic DNA scientist in the world,” that there is far from consensus in the scientific community that double-blind peer review is the only meaningful kind of peer review. Govt. Supp. Opp’n at 23; *see also* Affidavit of Bruce Budowle at 2, ECF No. 33–17. To this point, Mr. Weller described the various advantages and disadvantages of each type of peer review. *Weller II* at 22–24. Compellingly, the Government also refuted the allegation by Judge Edelman in *Tibbs* that the AFTE Journal does not provide “meaningful” review, by bringing to the Court’s attention a study that was initially published in the AFTE Journal, and then was subsequently published in the *Journal of Forensic Science* with no further alterations. Govt. Supp. Opp’n at 27. Because the *Journal of Forensic Science* employs a double-blind peer review process, this indicates that at least in this instance, the open peer review process of the AFTE Journal led to the same outcome as a double-blind peer review. *Id.* In addition, numerous courts have concluded that publication in the AFTE Journal satisfies this prong of the *Daubert*

admissibility analysis. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1119; *United States v. Johnson*, No. 16 Cr. 281, 2019 WL 1130258, at *16 (S.D.N.Y. Mar. 11, 2019); *Ashburn*, 88 F. Supp. 3d at 245–46; *Otero*, 849 F. Supp. 2d at 433; *Taylor*, 663 F. Supp. 2d at 1176; *Monteiro*, 407 F. Supp. 2d at 366–67. The Court queries whether excluding certain journals from consideration based on the type of peer review the journal employs goes beyond a court’s appropriate gatekeeping function under *Daubert*.

And even if the Court were to discount the numerous peer-reviewed studies published in the AFTE Journal, Mr. Weller’s affidavit also cites to forty-seven other scientific studies in the field of firearm and toolmark identification that have been published in eleven other peer-reviewed scientific journals. *Weller II* at Ex. A. This alone would fulfill the required publication and peer review requirement.

Because the toolmark identification methodology used by Mr. Monturo has been subject to peer review and publication, the Court finds this *Daubert* factor to also weigh in favor of admission.

4. The existence and maintenance of standards to control the methodology’s operation

The fourth *Daubert* factor inquires as to whether there are proper standards and controls to govern the operation of the technique in question. *See Daubert*, 509 U.S. at 594. Mr. Harris argues that there are insufficient objective standards in place, citing to the PCAST Report to claim that the AFTE’s “sufficient agreement” analysis that is used by examiners to reach their conclusions is subjective and impermissibly based on the “personal judgment” of each examiner. *Def.’s Supp. Mot.* at 4 (citing PCAST Report at 47, 60, 104, 113). In opposition, the Government argues that “the firearms community has implemented standards,” citing to a

number of industry guidebooks and regulations. Govt. Opp'n at 2. While a close call, the Court finds that the lack of objective standards ultimately means this factor cannot be met.⁶

The Government identifies a number of what they refer to as “standards for professional guidance” for the firearm and toolmark profession, Govt. Opp'n at 32–33, but the primary standard that governs the discipline is the AFTE Theory of Identification, which describes the methodology examiners should undertake when “pattern matching” between firearms and cartridges. *See, e.g.*, Govt. Opp'n at 8 (explaining that Theory of Identification was created “to explain the basis of opinion of common origin in toolmark comparisons”). According to the AFTE Theory of Identification, examiners can conclude that a firearm and cartridges have a common origin when a comparison of toolmarks shows there is “sufficient agreement” between “the unique surface contours of two toolmarks.” The Association of Firearm and Tool Mark Examiners, *AFTE Theory of Identification as It Relates to Toolmarks*, <https://afte.org/about-us/what-is-afte/afte-theory-of-identification> (last visited November 4, 2020). This theory of identification dictates that “sufficient agreement” between two toolmarks exists only when “the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.” *Id.* The Court finds this standard to be generally vague, and indeed, the AFTE Theory acknowledges that “the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience.” *Id.* As other courts have found, under this method “matching two tool marks essentially comes down to the examiner's subjective judgment based on his training, experience, and knowledge of firearms.”

⁶ This *Daubert* factor is, as the Government concedes, “the only *Daubert* factor that some courts have found lacking” in firearm toolmark identification. Govt. Opp'n at 33. This makes it all the more puzzling that the Government fails entirely to address this factor in its reply.

Romero-Lobato, 379 F. Supp. 3d at 1121; *Glynn*, 578 F. Supp. 2d at 572 (“[T]he standard defining when an examiner should declare a match – namely ‘sufficient agreement’ – is inherently vague.”).

Accordingly, it is evident and hardly disputed that the “AFTE theory lacks objective standards.” *Ricks*, 2020 WL 1491750, at *10. The entire process of reaching a conclusion regarding the “sufficient agreement in individual characteristics” is one that relies wholly on the examiner’s judgment, without any underlying numerical standards or guideposts to direct an examiner’s conclusion. *See* Evid. Hr’g Tr. 37:16–38:25 (noting the absence at this time of objective standards to guide an examiner’s findings). And as Mr. Weller testified, even in contrast to other subjective disciplines such as fingerprint analysis, firearm toolmark identification does not provide objective standards even as a quality control measure, such as a baseline to trigger further verification. *See* Evid. Hr’g Tr. 112:18-113:17 (explaining that while fingerprint testing does not have an agreed-upon standard for the number of matching points required for an identification, it does use matching points as a quality control measure that triggers further verification if below a certain threshold). While Mr. Monturo’s additional use of “basic scientific standards” through taking contemporaneous notes, documenting his comparison with photographs, and the use of a second reviewer for verification surely assist in maintaining reliable results, without more the Court cannot conclude this *Daubert* factor is met.

It should be noted, however, that even if this factor cannot be met, a partially subjective methodology is not inherently unreliable, or an immediate bar to admissibility. Rule 702 “does not impose a requirement that the expert must reach a conclusion via an objective set of criteria or that he be able to quantify his opinion with a statistical probability. *Romero-Lobato*, 379 F. Supp. 3d at 1120. And indeed, “all technical fields which require the testimony of expert

witnesses engender some degree of subjectivity requiring the expert to employ his or her individual judgment, which is based on specialized training, education, and relevant work experience.” *Johnson*, 2019 WL 1130258 at *18 (citations omitted); *see also* Evid. Hr’g Tr. at 30:14–31:6 (Mr. Weller testified that “all science involves some level of interpretation,” and went on to describe subjective components to both drug testing and DNA interpretation). Accordingly, this factor weighs against the admission of Mr. Monturo’s testimony, but does not disqualify it.

5. Whether the methodology has achieved general acceptance in the relevant community

Finally, the fifth and last *Daubert* factor asks whether the technique has been generally accepted within the relevant scientific community, reasoning that “a known technique which has been able to attract only minimal support within the community, may properly be viewed with skepticism.” *See Daubert*, 509 U.S. at 594. The Court finds that the Government has put forth more than sufficient evidence to show that the AFTE theory as used by Mr. Monturo enjoys widespread scientific acceptance. *See* Govt. Opp’n at 2; Govt. Supp. Opp’n at 28.

Mr. Weller testified that firearm and toolmark identification is practiced by accredited laboratories in the United States and throughout the world, including England (Scotland Yard), New Zealand, Canada, South Africa, Australia, Germany, Sweden, Greece, Turkey, China, Mexico, Singapore, Malaysia, Belgium, Netherlands, and Denmark. *See* Weller II at 30. In the United States alone, there are 233 accredited firearm and toolmark laboratories, that often operate within a larger forensic laboratory providing chemistry, DNA, and fingerprint identification, and scientists from a variety of disciplines author studies within the area of firearms and toolmark identification. *Id.*

The criticism contained in the PCAST Report does not undermine this factor, as “techniques do not need to have universal acceptance before they are allowed to be presented before a court.” *Romero-Lobato*, 379 F. Supp. 3d at 1122. Even courts that have been critical of the validity of the discipline have conceded that it does enjoy general acceptance as a reliable methodology in the relevant scientific community of examiners. *See Otero*, 849 F. Supp. 2d at 435 (collecting cases). Furthermore, as Mr. Weller noted at the evidentiary hearing, the committee responsible for the PCAST Report did not include any firearm and toolmark examiners or researchers in the field, *see Evid. Hr’g Tr.* 47:18-23, thus raising the question of whether the PCAST Report criticism would even constitute a lack of acceptance from the “relevant scientific community.” For all of these reasons, this factor weighs in favor of admitting Mr. Monturo’s testimony.

6. The *Daubert* Analysis Urges Admission of Mr. Monturo’s Testimony

Balancing all five *Daubert* factors, the Court finds that the Government’s proposed expert testimony of Mr. Monturo is reliable and admissible, though subject to what the Court considers prudent limitations, discussed in detail below. The only factor that does not favor admissibility is the lack of objective criteria under the fourth *Daubert* factor, but as discussed, “the subjectivity of a methodology is not fatal under Rule 702 and *Daubert*.” *Ashburn*, 88 F. Supp. 3d at 246. And as other courts have also found, this deficiency “is countered by the method's relatively low rate of error, widespread acceptance in the scientific community, testability, and frequent publication in scientific journals.” *Romero-Lobato*, 379 F. Supp. 3d at 1122. Accordingly, the Court will allow the admission of Mr. Monturo’s expert testimony as to his firearm and toolmark identification analysis, subject to certain limitations.

D. Federal Rule of Evidence 702(d)

Federal Rule of Evidence 702(d) provides that qualified expert testimony is admissible only when “the expert has reliably applied the principles and methods to the facts of the case.” Fed. R. Evid. 702. Mr. Harris challenges the admission of Mr. Monturo’s testimony, asserting that he “has not applied the principles and methods reliably to the facts of the case.” Def.’s Mot. at 1. However, he provides no evidence or further analysis to flesh out this conclusory claim. Accordingly, the Court finds this argument to be without merit.

As previously described, Mr. Monturo detailed the firearm and toolmark examination he conducted in his report, providing both a description of his process and photo documentation. *See generally* Monturo Report. Mr. Monturo’s findings were then verified by another qualified examiner the same day. Monturo Report Notes at 2. In contrast, Mr. Harris has not put forth any evidence to suggest that Mr. Monturo applied the firearm and toolmarking methodology in an unreliable manner. Mr. Monturo also appears to be well-qualified, with the Government noting that he “has significant training and experience, has not failed any proficiency exams, and has designed consecutively manufactured firearms test kits for training other firearms examiners,” information that they plan to elicit at trial during qualification of his testimony and also set out in his curriculum vitae. Govt. Opp’n at 35. In light of his failure to identify any unreliability on Mr. Monturo’s part, and also because Mr. Harris will have the ability to question Mr. Harris regarding his analysis during cross examination, the Court is convinced exclusion on this ground is not warranted. *See Daubert*, 509 U.S. at 596 (“Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.”). If Mr. Harris has lingering concerns about Mr. Monturo’s application of the firearm and toolmark methodology in this case,

he is welcome to retain an independent expert to review Mr. Monturo's work, or have an independent examination of his own performed.

E. Federal Rule of Evidence 403

Next, Mr. Harris argues that even if the proposed testimony of Mr. Monturo is admissible pursuant to *Daubert* and Federal Rule of Evidence 702, it is inadmissible under Federal Rule of Evidence 403. Def. Mot. at 2. In support of this claim, Mr. Harris argues that Mr. Monturo's "conclusions appear to extend beyond his claimed expertise and are not reliable since they are not based on objective standards but rather his subjective observations and conclusions." *Id.* "The prejudice to Mr. Harris is simple, a connection to a firearm, a connection to a shell casing, all premised on analysis that at its best can only conclude that it 'may' be correct." Def. Supp. Mot. at 2.

Under Rule 403, a Court may exclude otherwise probative testimony if its value is substantially outweighed by unfair prejudice, confusing the issues, misleading the jury, undue delay, a waste of time, or cumulative evidence. Fed. R. Evid. 403. Mr. Harris's concern under Rule 403 appears to be that the value of Mr. Monturo's testimony will be substantially outweighed by the risk of him potentially misleading the jury through his reliance on a methodology Mr. Harris does not believe is sufficiently reliable. First, Mr. Harris's concerns about the reliability of the firearm and toolmarking methodology have already been analyzed, and the Court has found the underlying analysis sufficiently reliable such that Mr. Harris's concerns do not "substantially outweigh" the value of Mr. Monturo's testimony. Additionally, the Court believes that the risk of prejudice raised here can be alleviated through alternatives to exclusion. Cross-examination of Mr. Monturo's testimony, in conjunction with the appropriate limiting instruction governing the degree of certainty Mr. Monturo can express about his conclusions will sufficiently deter the risks of harm Mr. Harris has raised.

F. Limiting Instruction

In his final request, Mr. Harris asks that if the testimony of Mr. Monturo is not excluded, then the Court put in place limitations on his testimony. Def. Supp. Mot. at 6–7. Specifically, he requests that Mr. Monturo not “use the term ‘match’” but he “may be allowed to tell the jury that he could not exclude the gun as the weapon that produced a casing.” *Id.*

Limitations restricting the degree of certainty that may be expressed on firearm and toolmark expert testimony are not uncommon. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1117 (noting the “general consensus” of the courts “is that firearm examiners should not testify that their conclusions are infallible or not subject to any rate of error, nor should they arbitrarily give a statistical probability for the accuracy of their conclusions”); *Ashburn*, 88 F. Supp. 3d at 249 (limiting expressions of an expert’s conclusions to that of a “reasonable degree of ballistics certainty” or a “reasonable degree of certainty in the ballistics field.”); *Diaz*, 2007 WL 485967 at *1 (same).

With respect to Mr. Harris’s stated concerns, the Government has already agreed to a number of limitations on Mr. Monturo’s testimony, chief among them that he will not use terms such as “match,” he will “not state his expert opinion with any level of statistical certainty,” and he will not use the phrases when giving his opinion of “to the exclusion of all other firearms” or “to a reasonable degree of scientific certainty.” Govt. Opp’n at 12. These limitations are in accord with the Department of Justice Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmarks Discipline—Pattern Matching Examination. *See* Govt. Opp’n, Ex. 4 (“DOJ ULTR”), ECF No. 28-4. The DOJ ULTR permits firearms examiners to conclude that casings were fired from the same firearm when all class characteristics are in agreement, and “the quality and quantity of corresponding individual characteristics is such that the examiner

would not expect to find that same combination of individual characteristics repeated in another source and has found insufficient disagreement of individual characteristics to conclude they originated from different sources.” *Id.* at 2–3. This Court believes, as other courts have also concluded, *see Hunt*, 2020 WL 2842844, at *8, that the testimony limitations as codified in the DOJ ULTR are reasonable and should govern the testimony at issue here. Accordingly, the Court instructs Mr. Monturo to abide by the expert testimony limitations detailed in the DOJ ULTR.

III. CONCLUSION

For the foregoing reasons, Defendant’s Motion to Exclude Expert Testimony as to Firearm Examination Testing, ECF No. 22, is DENIED. An order consistent with this Memorandum Opinion is separately and contemporaneously issued.

Dated: November 4, 2020

RUDOLPH CONTRERAS
United States District Judge

**IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF OKLAHOMA**

UNITED STATES OF AMERICA,)
)
)
Plaintiff,)
)
v.)
)
DOMINIC EUGENE HUNT,)
)
Defendant.)

Case No. CR-19-073-R

ORDER

Before the Court is Defendant Dominic Hunt’s Motion in Limine to Exclude Ballistic Evidence, or Alternatively, for a *Daubert* Hearing. Doc. No. 67. The Government has responded in opposition to the motion. Doc. No. 81. Upon review of the parties’ submissions, the Court denies Defendant’s motion.

I. Background

On November 6, 2019, a federal grand jury returned a nine-count, third superseding indictment charging Defendant with, as relevant here, two counts of being a felon in possession of ammunition. Doc. No. 41. The two counts—Counts Eight and Nine—stem from two shootings: One in January of 2019 and another in February of 2019. *Id.* During the Oklahoma Police Department’s (OCPD) investigation at the scene of the first shooting, officers found a Blazer 9mm Luger cartridge casing—the basis for Count Eight. *Id.* at 5–6. During the OCPD’s investigation at the scene of the second shooting, officers found a Blazer 9mm Luger cartridge casing and two Winchester 9mm Luger cartridge casings—the basis for Count Nine. *Id.* at 6. Ronald Jones, a firearm and toolmark examiner for the

OCPD, examined the casings and concluded that all four casings were likely fired from the same unknown firearm, potentially a Smith & Wesson 9mm Luger caliber pistol. Doc. Nos. 81–1, 81–2. Howard Kong, a firearm and toolmark examiner for the Bureau of Alcohol, Tobacco, Firearms and Explosives’ (ATF) Forensic Science Laboratory, found the same. Doc. No. 81–4. The Government anticipates calling Mr. Jones and Mr. Kong at trial to “testify regarding their training, experience, and qualifications, the basis for firearms identification, their methods of examination in this case, their findings, and the basis for those findings.” Doc. No. 81, pp. 4–5. Specifically, the Government intends its experts to testify that:

(1) the ammunition charged in Count Eight was not fired from the Springfield Armory 9mm Luger caliber pistol [the Defendant’s brother] had on March 11, 2019; (2) the ammunition charged in Count Eight was not fired from the Smith & Wesson .40 caliber pistol [the Defendant’s cousin] was convicted of possessing on January 20, 2019; (3) the probability the ammunition charged in Count Nine were fired in different firearms is so small it is negligible; (4) the ammunition charged in Count Nine was not fired from [the] Smith & Wesson .40 caliber pistol . . . ; (5) the probability the ammunition charged in Counts Eight and Nine were fired in different firearms is so small it is negligible; and (6) the unknown firearm was likely a Smith & Wesson 9mm Luger caliber pistol.

Id. Defendant now moves to exclude the testimony of Mr. Jones and Mr. Kong, or alternatively, for a *Daubert* hearing. Doc. No. 67.

II. Legal Standard

When it comes to the admissibility of expert evidence, district courts maintain the role of gatekeeper. *Bitler v. A.O. Smith Corp.*, 400 F.3d 1227, 1232 (10th Cir. 2005). In that role, district courts must adhere to Federal Rule of Evidence 702, which demands that courts “assess proffered expert testimony to ensure it is both relevant and reliable.” *United*

States v. Avitia-Guillen, 680 F.3d 1253, 1256 (10th Cir. 2012). To do this, “the district court generally must first determine whether the expert is qualified” *United States v. Nacchio*, 555 F.3d 1234, 1241 (10th Cir.2009) (en banc). If the expert is sufficiently qualified, then “the court must determine whether the expert’s opinion is reliable” *Id.* “Although a district court has discretion in how it performs its gatekeeping function, ‘when faced with a party’s objection, [the court] must adequately demonstrate by specific findings on the record that it has performed its duty as gatekeeper.’” *Avitia-Guillen*, 680 F.3d at 1257 (quoting *Goebel v. Denver & Rio Grande W. R.R. Co.*, 215 F.3d 1083, 1088 (10th Cir. 2000)). “The proponent of expert testimony bears the burden of showing that its proffered expert’s testimony is admissible.” *Nacchio*, 555 F.3d at 1241.

Here, Defendant Hunt does not object to the relevancy of the experts’ testimony nor to the experts’ qualifications. Defendant objects only to the reliability of the experts’ testimony. Doc. No. 67, pp. 11–18. Therefore, the Court need only address whether the experts’ testimony is reliable. *See Avitia-Guillen*, 680 F.3d at 1257.

“To determine reliability, courts assess the reasoning and methodology underlying the [experts’] opinion” *Thompson v. APS of Oklahoma, LLC*, No. CIV-16-1257-R, 2018 WL 4608505, at *4 (W.D. Okla. Sept. 25, 2018) (internal quotation marks and citation omitted). “The reliability standard is lower than the merits standard of correctness, and plaintiffs need only show the Court that their experts’ opinions are reliable, not that they are substantively correct.” *Id.* (internal quotation marks and citation omitted). In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), the Supreme Court provided a non-exhaustive list of factors to aid in this determination:

(1) whether the particular theory can be and has been tested; (2) whether the theory has been subjected to peer review and publication; (3) the known or potential rate of error; (4) the existence and maintenance of standards controlling the technique's operation; and (5) whether the technique has achieved general acceptance in the relevant scientific or expert community.

United States v. Baines, 573 F.3d 979, 985 (10th Cir. 2009) (citing *Daubert*, 509 U.S. at 592–94).¹ The reliability inquiry, however, is fact- and case-specific: no one factor is dispositive or always applicable, and the goal remains “ensuring that an expert ‘employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.’” *Bitler*, 400 F.3d at 1233 (quoting *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152 (1999)).

III. Firearm Toolmark Identification

In his motion, Defendant challenges the Governments use of firearm toolmark identification. “Forensic toolmark identification is a discipline that is concerned with the matching of a toolmark to the specific tool that made it. Firearm identification is a specialized area of toolmark identification dealing with firearms, which involve a specific category of tools.” *United States v. McCluskey*, No. 10-2734, 2013 WL 12335325, at *3 (D.N.M. Feb. 7, 2013) (citation omitted). “Toolmark identification is based on the theory that tools used in the manufacture of a firearm leave distinct marks on various firearm components, such as the barrel, breech face, or firing pins . . . [and] that the marks are individualized to a particular firearm through changes the tool undergoes each time it cuts

¹ *Daubert* itself was limited to scientific evidence, see *United States v. Baines*, 573 F.3d 979, 985 (10th Cir. 2009), but in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), the Supreme Court made clear that the gatekeeping obligation of the district courts described in *Daubert* applies, not just to scientific testimony, but to all expert testimony. *Id.* at 141.

and scrapes metal to create an item in the production of the weapon.” *Id.* at 4. The field of firearm toolmark examination is based on the theory that some of these markings will be transferred to a bullet fired from the gun. *Id.* In conducting a firearm toolmark examination, a firearms examiner observes three types of characteristics:

(1) Class characteristics: i.e., the weight or caliber of the bullet, the number of lands and grooves, the twist of the lands and grooves, and the width of the lands and grooves, that appear on all bullet casings fired from the same type of weapon and are predetermined by the gun manufacturer;

(2) Individual characteristics: unique, microscopic, random imperfections in the barrel or firing mechanism created by the manufacturing process and/or damage to the gun post-manufacture, such as striated and/or impressed marks, unique to single gun; and

(3) Subclass characteristics: characteristics that exist, for example, within a particular batch of firearms due to imperfections in the manufacturing tool that persist during the manufacture of multiple firearm components mass-produced at the same time.

Ricks v. Pauch, No. 17-12784, 2020 WL 1491750, at *8–9 (E.D. Mich., 2020). Pursuant to the theory used by the Government’s experts in this case—the Association of Firearms and Toolmark Examiners (AFTE) method—“a qualified examiner can determine whether two bullets were fired by the same gun by comparatively examining bullets and determining whether ‘sufficient agreement’ of toolmarks exist,” meaning that there is significant similarity in the individual markings found on each bullet. *Id.* at 9.

IV. *Daubert* Analysis

The use of this type of firearm toolmark identification in criminal trials is “hardly novel.” *United States v. Taylor*, 663 F. Supp. 2d 1170, 1175 (D.N.M. 2009). “For decades . . . admission of the type of firearm identification testimony challenged by the defendant[] has been semi-automatic” *United States v. Monteiro*, 407 F. Supp. 2d 351, 364 (D. Mass. 2006); *see also, e.g., United States v. Hicks*, 389 F.3d 514 (5th Cir. 2004); *United States v. Johnson*, 875 F.3d 1265, 1281 (9th Cir. 2017). Indeed, no federal court has deemed such evidence wholly inadmissible. *See United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1117 (D. Nev. 2019). Having been routinely admitted, “[c]ourts [are] understandably . . . gun shy about questioning the reliability of [such] evidence,” *Monteiro*, 407 F. Supp. 2d at 364. However, because of the seriousness of the criticisms launched against the methodology underlying firearms identification by Defendant in this case, the Court will carefully assess the reliability of this methodology, using *Daubert* as a guide. *See, e.g., Taylor*, 663 F. Supp. 2d at 1176.²

The first *Daubert* factor asks whether the experts’ particular theory can be and has been tested. *Daubert*, 509 U.S. at 592–94. Defendant argues—without citation—that the theory of firearm toolmark identification rests on an assumption that has not been properly tested. Doc. No. 67, pp. 13–14. The Government responds that its experts’ testimony is based upon the theory and methodology developed by the Association of Firearms and

² Some Courts have analyzed whether firearm toolmark identification can fairly be called “science” before evaluating the *Daubert* factors. *See United States v. Glynn*, 578 F. Supp. 2d 567, 570 (S.D.N.Y. 2008). The Court need not conduct such an analysis here. Though Defendant argues firearm toolmark identification is not a science, Doc. No. 67, p. 14, it is clearly “technical or specialized, and therefore within the scope of Rule 702.” *United States v. Willock*, 696 F. Supp. 2d 536, 571 (D. Md. 2010), *aff’d sub nom. United States v. Mouzone*, 687 F.3d 207 (4th Cir. 2012).

Toolmark Examiners (AFTE), and that this theory has been well tested. Doc. No. 81, pp. 15–16. The Court agrees.

Put simply, the theory of firearm toolmark identification can be and has been tested. *See, e.g.,* The Association of Firearm and Tool Mark Examiners, *Testability of the Scientific Principle* (last visited May 14, 2020), <https://tinyurl.com/yal3ja4t> (collecting studies). This conclusion is supported by other courts within the Tenth Circuit that have already addressed the issue at length, *see, e.g., United States v. Taylor*, 663 F. Supp. 2d 1170, 1176 (D.N.M. 2009) (“[T]he methods underlying firearms identification can, at least to some degree, be tested and reproduced”), in addition to a number of other courts outside the Circuit, *see, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1118–19 (collecting cases where “federal courts have held that the AFTE method can be and has been frequently tested” and holding the same). Accordingly, this first *Daubert* factor weighs in favor of admissibility.

The second *Daubert* factor asks whether the technique has been subjected to peer review and publication. *Daubert*, 509 U.S. at 593–94. Defendant argues that there have not been enough studies done of firearm toolmark identification, and that the studies available have not been subject to peer review. Doc. No. 67, p. 14. The Government contends that analysis recently provided by federal courts tells a different story. The Court agrees.

In evaluating whether AFTE’s method of firearm toolmark identification satisfies the second *Daubert* factor, the United States District Court for the District of Nevada recently found that:

AFTE publishes its own journal, the appropriately named *ATFE Journal*, which is subject to peer review. According to AFTE’s website, the *ATFE Journal*, “is dedicated to the sharing of information, techniques, and

procedures,” and the papers published within “are reviewed for scientific validity, logical reasoning, and sound methodology.” [*What is the Journal?*, The Association of Firearm and Tool Mark Examiners, <https://afte.org/afte-journal/what-is-the-journal> (last visited May 1, 2019)]. Several published federal decisions have also commented on the *AFTE Journal*, with all finding that it meets the *Daubert* peer review element. See *U.S. v. Ashburn*, 88 F.Supp.3d 239, 245–46 (E.D.N.Y. 2015) (finding that the AFTE method has been subjected to peer review through the *AFTE Journal*); *U.S. v. Otero*, 849 F.Supp.2d 425, 433 (D.N.J. 2012) (describing the *AFTE Journal*'s peer reviewing process and finding that the methodology has been subjected to peer review); *U.S. v. Taylor*, 663 F.Supp.2d 1170, 1176 (D.N.M. 2009) (finding that the *AFTE* method has been subjected to peer review through the *AFTE Journal* and two articles submitted by the government in a peer-reviewed journal about the methodology); *U.S. v. Monteiro*, 407 F.Supp.2d 351, 366–67 (D. Mass. 2006) (describing the *AFTE Journal*'s peer reviewing process and finding that it meets the *Daubert* peer review element). And of course, the NAS and PCAST Reports themselves constitute peer review despite the unfavorable view the two reports have of the AFTE method.

Romero-Lobato, 379 F. Supp. 3d at 1119. The second *Daubert* factor thus weighs in favor of admissibility.

Defendant suggests that the studies mentioned above are insufficient because they were not “black-box” studies.³ Doc. No. 67, p. 14. Defendant then cites the PCAST Report, arguing that there has been only one black-box study on firearms identification and that this one study has never been subject to peer review. *Id.* The PCAST Report cited by Defendant “rejected studies that it did not consider to be blind, such as where the examiners knew that a bullet or spent casing matched one of the barrels included with the test kit” However, “The PCAST Report did not reach a conclusion as to whether the AFTE

³ A black-box study is a blind study where “many examiners are presented with many independent comparison problems—typically involving ‘questioned’ samples and one or more ‘known’ samples—and asked to declare whether the questioned samples came from the same sources as one of the known samples. The researchers then determine how often examiners reach erroneous conclusions.” President’s Council of Advisors on Science and Technology, Exec. Office of the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 49 (2016), available at <https://tinyurl.com/j29c5ua>.

method was reliable or not because there was only one study available that met its criteria.” *Id.* The Court does not similarly restrict its judicial review to techniques tested through black-box studies. The Court does, however, approve of the PCAST Report’s ultimate conclusion: “[W]hether firearms analysis should be deemed admissible based on the ‘current evidence’ is a decision that should be left to the courts.” *Id.*

The third *Daubert* factor asks whether the technique has a known or potential rate of error. *Daubert*, 509 U.S. at 594. Defendant contends that because there is only one black-box study, there is not enough information available to determine a known or potential rate of error in the field of firearm toolmark identification. Doc. No. 67, p. 14. The Government objects, citing federal cases discussing studies that evidence a low rate of error in firearms analysis. Doc. No. 81, pp. 17–18. Again, the Court agrees with the Government.

As noted above, the Court declines Defendant’s invitation to restrict judicial review to techniques tested through black-box studies. “*Daubert* does not mandate such a prerequisite for a technique to satisfy its error rate element.” *Romero-Lobato*, 379 F. Supp. 3d at 1120. Still, the Government bears the burden to demonstrate that its experts’ methodology is reliable. *See Nacchio*, 555 F.3d at 1241. To that end, the Government cites federal cases that discuss a number of studies which report a low error rate for the AFTE method. Doc. No. 81, p. 17 (citing *Romero-Lobato*, 379 F. Supp 3d at 1117–18 and *United States v. Otero*, 849 F. Supp. 2d 425, 433–34 (D.N.J. 2012)). Those cases discuss, for example, a Miami-Dade Study that reported a potential error rate of less than 1.2% and an error rate by the participants of 0.07%, in addition to an Ames Study that reported a false positive rate of 1.52%. *Id.*

Other federal courts examining the AFTE method's rate of error have likewise found it to be low. *See, e.g., v. Ashburn*, 88 F. Supp. 3d 239, 246 (E.D.N.Y. 2015) (“the error rate, to the extent it can be measured, appears to be low, weighing in favor of admission”); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1177 (D.N.M. 2009) (“this number [less than 1%] suggests that the error rate is quite low”). Even courts that have found it impossible to calculate an absolute error rate for firearm toolmark identification, have ultimately concluded that the known error rate is not “unacceptably high.” *United States v. Monteiro*, 407 F. Supp. 2d 351, 367–68 (D. Mass. 2006). Defendant does not introduce any contradictory studies. *See* Doc. No. 67, p. 14. Based on the record before the Court, this third *Daubert* factor weighs in favor of admissibility.

The fourth *Daubert* factor asks whether there are standards that control the technique's operation. *Daubert*, 509 U.S. 594. Defendant argues that there are no uniform standards controlling the AFTE method of firearm toolmark identification, and that instead, the AFTE method is based on subjective methodology. Doc. No. 67, p. 14. The Government argues that this subjectivity does not weigh against admissibility under the fourth *Daubert* factor. Doc. No. 81, p. 18. The Court disagrees.

A main criticism of the AFTE method is that firearm examiners do not reach their conclusions through objective criteria. *See Romero-Lobato*, 379 F. Supp. 3d at 1120-121. Instead, examiners use a high-powered microscope, in conjunction with their experience and training, to determine if there is “sufficient agreement” between the “unique surface contours” of two firearm toolmarks. *AFTE Theory of Identification*, The Association of Firearm and Tool Mark Examiners, available at <https://afte.org/about-us/what-is-afte/afte->

theory-of-identification (last visited May 14, 2020). “The statement that “sufficient agreement” exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.”⁴ *Id.* Ultimately, the AFTE itself recognizes that their method is “is subjective in nature.” *Id.* So too have other courts. *See Romero-Lobato*, 379 F. Supp. 3d at 1121 (collecting cases). This fourth factor, unlike the previous three, weighs against admissibility.

The fifth and final *Daubert* factor asks whether the theory or technique enjoys general acceptance within the relevant community. *Daubert*, 509 U.S. at 594. Defendant argues that the limitations of firearm toolmark identification is recent and growing, and that because courts have not seriously considered all aspects of the field or tested its reliability since the PCAST Report was published, the fifth *Daubert* factor is not satisfied here. Doc. No. 67, p. 15. The Government responds arguing that nearly every court to have addressed the issue has found that the AFTE method enjoys general acceptance within the

⁴ The AFTE further details their methodology in the following manner:

“[S]ufficient agreement” is related to the significant duplication of random toolmarks as evidence by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. Agreement is significant when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.

AFTE Theory of Identification, The Association of Firearm and Tool Mark Examiners, available at <https://afte.org/about-us/what-is-afte/afte-theory-of-identification> (last visited May 14, 2020).

relevant community—both before and after publication of the PCAST Report. Doc. No. 81, p. 19. The Court agrees.

The AFTE method easily satisfies this final factor. *See Romero-Lobato*, 379 F. Supp. 3d at 1122 (collecting cases finding the AFTE theory to be widely accepted in the relevant community and finding the same). In fact, the AFTE method used by the Government’s experts here, is “the field’s established standard.” *See Ashburn*, 88 F. Supp. 3d at 246. That the NAS and PCAST Reports criticize the method does not undermine the Court’s conclusion. “Techniques do not need to have universal acceptance before they are allowed to be presented before a court.” *Romero-Lobato*, 379 F. Supp. 3d at 1122 (citing *Daubert*, 509 U.S. at 588–99). Accordingly, this factor weighs in favor of admissibility.

Balancing the *Daubert* factors, the Court finds that the Government’s expert testimony, derived from the AFTE methodology, is reliable and therefore admissible—though subject to the limitations discussed below. The only factor that weighs against admissibility is the fourth *Daubert* factor, which highlights the AFTE’s subjective processes. But, “the subjectivity of a methodology is not fatal under Rule 702 and *Daubert*.” *United States v. Ashburn*, 88 F. Supp. 3d 239, 246 (E.D.N.Y. 2015). By its terms, Federal Rule of Evidence 702 permits an expert with sufficient knowledge, experience, or training to testify about a particular subject matter. *See Fed. R. Evid. 702; Romero-Lobato*, 379 F. Supp. 3d at 1120. *Daubert* does not impose a rigid requirement that the expert reach a conclusion through an entirely objective set of criteria. *See Daubert*, 509 U.S. at 594–595. Here, the lack of objective criteria is overcome by the Government’s introduction of evidence demonstrating that the method has been tested, reviewed by peers and subject to

publication, found to have a potential low rate of error, and widely accepted in the relevant community. Moreover, Defendant has not cited a single case where a federal court has completely prohibited firearms toolmark identification testimony under *Daubert*.

V. Federal Rules of Evidence 702(d)

Next, Defendant argues that even if the expert testimony is admissible under *Daubert*, the Government has not met its burden under Rule 702(d) to show that its experts reliably applied the AFTE method in this case. Under that Rule:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

...

(d) the expert has reliably applied the principles and methods to the facts of the case.

Fed. R. Evid. 702(d). Here, Defendant makes four specific objections. He argues that the Government has not complied with Rule 702(d) because its experts failed to document the basis for their findings, that a second examiner did not verify or review the experts' work, and that the experts failed to comply with two "validity" requirements discussed by the PCAST Report. Doc. No. 67, p. 17. The Government denies the validity of each objection. Doc. No. 81, pp. 21–23.

First, as the Government demonstrates, both Mr. Jones and Mr. Kong wrote detailed reports explaining their analysis. Doc. Nos. 81–9, 81–10. Second, those reports were reviewed by other examiners in the field. Doc. Nos. 81–1, 81–2, 81–3, 81–4. Finally, the two validity requirements discussed by the PCAST Report—that experts must provide evidence demonstrating their rigorous proficiency testing, in addition to whether they were aware of any facts of the case that might influence their conclusion—are not required under

Rule 702(d). Nevertheless, the Government has presented evidence demonstrating the experience, certifications, and continued training of both experts. *See* Doc. Nos. 81–6, 81–7, 81–8; *cf* Doc. No. 81–5. And both experts’ examination reports detail what case-specific facts they were aware of when drawing their conclusions. *See* Doc. Nos. 81–1, 81–2. Accordingly, the Court finds that Defendant’s objections are without merit.

VI. *Daubert* Hearing

As an alternative, Defendant requests a *Daubert* hearing to require the Government to prove that Mr. Jones’s and Mr. Kong’s testimony will be reliable before admitting their testimony. Doc. No. 17. Again, the Government objects. Doc. No. 81, pp. 24–25. Nothing requires the Court to hold a formal *Daubert* hearing in advance of qualifying an expert. *See Goebel v. Denver and Rio Grande Western RR Co.*, 215 F.3d 1083, 1087 (10th Cir. 2000); *see also Kumho Tire*, 526 U.S. at 152 (“The trial court must have the . . . latitude . . . to decide whether or when special briefing or other proceedings are needed to investigate reliability”). Considering the parties’ briefing, in addition to the *Daubert* and Rule 702 analysis above, the Court finds it unnecessary to conduct such a proceeding here. *See, e.g., Ashburn*, 88 F. Supp. 3d at 244 (finding *Daubert* hearing unnecessary). The reliability of the Government’s expert testimony has been sufficiently addressed on the briefs. *See Goebel*, 215 F.3d at 1087 (noting that a *Daubert* hearing “is not mandated” and that a district court may “satisfy its gatekeeper role when asked to rule on a motion in limine”).

VII. Expert Testimony Limitations

In his penultimate argument, Defendant asks the Court to place limitations on the Government’s firearm toolmark experts because the jury will be unduly swayed by the

experts if not made aware of the limitations on their methodology. Doc. No. 67, p. 18. The Government responds that no limitation is necessary because Department of Justice guidance sufficiently limits a firearm examiner's testimony. Doc. No. 81, pp. 23–24.

Some federal courts have imposed limitations on firearm and toolmark expert testimony. *See, e.g., Ashburn*, 88 F. Supp. 3d at 249. However, many courts have continued to allow unfettered testimony. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1117.

The general consensus is that firearm examiners should not testify that their conclusions are infallible or not subject to any rate of error, nor should they arbitrarily give a statistical probability for the accuracy of their conclusions. Several courts have also prohibited a firearm examiner from asserting that a particular bullet or shell casing could only have been discharged from a particular gun to the exclusion of all other guns in the world.

Id. (citing David H. Kaye, *Firearm-Mark Evidence: Looking Back and Looking Ahead*, 68 Case W. Res. L. Rev. 723, 734 (2018)).

In accordance with recent guidance from the Department of Justice, *see* Doc. No. 81–11, the Government's firearm experts have already agreed to refrain from expressing their findings in terms of absolute certainty, and they will not state or imply that a particular bullet or shell casing could only have been discharged from a particular firearm to the exclusion of all other firearms in the world. Doc. No. 81, p. 24. The Government has also made clear that it will not elicit a statement that its experts' conclusions are held to a reasonable degree of scientific certainty. *Id.*

The Court finds that the limitations mentioned above and prescribed by the Department of Justice are reasonable, and that the Government's experts should abide by those limitations. *See* Doc. No. 81–11, p. 3. To that end, the Government's experts:

[S]hall not [1] assert that two toolmarks originated from the same source to the exclusion of all other sources. . . . [2] assert that examinations conducted in the forensic firearms/toolmarks discipline are infallible or have a zero error rate. . . . [3] provide a conclusion that includes a statistic or numerical degree of probability except when based on relevant and appropriate data. . . . [4] cite the number of examinations conducted in the forensic firearms/toolmarks discipline performed in his or her career as a direct measure for the accuracy of a proffered conclusion. . . . [5] use the expressions ‘reasonable degree of scientific certainty,’ ‘reasonable scientific certainty,’ or similar assertions of reasonable certainty in either reports or testimony unless required to do so by [the Court] or applicable law.

Id. As to the fifth limitation described above, the Court will permit the Government’s experts to testify that their conclusions were reached to a reasonable degree of ballistic certainty, a reasonable degree of certainty in the field of firearm toolmark identification, or any other version of that standard. *See, e.g., U.S. v. Ashburn*, 88 F. Supp. 3d 239, 249 (E.D.N.Y. 2015) (limiting testimony to a “reasonable degree of ballistics certainty” or a “reasonable degree of certainty in the ballistics field.”); *U.S. v. Taylor*, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009) (limiting testimony to a “reasonable degree of certainty in the firearms examination field.”). Accordingly, the Government’s experts should not testify, for example, that “the probability the ammunition charged in Counts Eight and Nine were fired in different firearms is so small it is negligible,” *see* Doc. No. 81, p. 5. To the extent Defendant wishes to question or clarify the experts’ findings, he may do so through cross examination or through direct examination of his own firearm toolmark expert.

VIII. Additional Expert Information

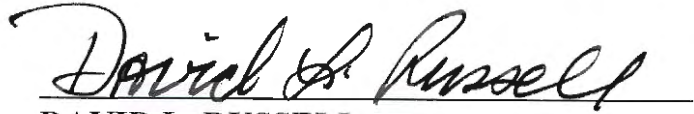
Defendant’s final objection is to the alleged lack of information relating to Mr. Jones’s expert testimony. Doc. No. 67, p. 19. Defendant claims that the Government should be required to provide “a significantly more detailed summary of what it expects Mr. Jones

will testify about.” *Id.* Notably, Defendant provides no support for his objection, and the Government has failed to respond in opposition. Upon review, the Court finds that the Government has provided sufficient information relating to Mr. Jones’s expert testimony. *See* Doc. No. 81, pp. 4–5; Doc. Nos. 81–1, 81–6, 81–7, 81–9.

IX. Conclusion

For the forgoing reasons, the Court denies Defendant Hunt’s Motion in Limine to Exclude Ballistic Evidence, or Alternatively, for a *Daubert* Hearing, Doc. No. 67.

IT IS SO ORDERED this 1st day of June 2020.



DAVID L. RUSSELL
UNITED STATES DISTRICT JUDGE

PAPER

Criminalistics

Beretta barrel fired bullet validation study

Jaimie A. Smith MS

Prince George's County Police
Department, Forensic Science Division,
Firearms Examination Unit, Landover, MD,
USA

Correspondence

Jaimie A. Smith MS, Prince George's
County Police Department, Forensic
Science Division, Firearms Examination
Unit, 7600 Barlowe Road, Landover, MD
20785, USA.

Email: (b)(6)

Funding information

Funding provided by Collaborative Testing
Service (CTS).

Abstract

A report published in 2016 by the President's Council of Advisors on Science and Technology (PCAST) criticized studies that have been published regarding the discipline of firearm identification. This study was designed to answer some of these criticisms and involved 30 consecutively manufactured Beretta brand 9 mm Luger caliber barrels. This study had an "open set" design to help the discipline of firearm identification establish "Foundational Validity" which is outlined in the PCAST report. Seventy-two qualified firearm examiners completed and submitted answers for this study that included 15 knowns and 20 unknowns. There were an additional 5 firearms with similar characteristics as the Beretta barrels that were also included as unknowns which provided "known non-match" comparisons. Test sets were created using the random function in Microsoft Excel. Collaborative Testing Services (CTS) funded, facilitated, distributed the tests, and collected the answers from qualified firearm examiners throughout the United States and the world. Firearm examiners were able to complete the test of fired bullets with a low error rate. The error rate for the corrected data was 0.08% (1 in 1250) with the lower confidence interval as low as 0.01% (1 in 10,000) and the upper confidence interval being as high as 0.4% (1 in 250).

KEYWORDS

barrels, Beretta, comparison, consecutively manufactured, error rate, firearm identification, fired bullets, foundational validity, microscopic examination, PCAST, validation study

2 | INTRODUCTION

Firearm and toolmark identification is a discipline within forensic science whose primary objective is to determine if a fired bullet or fired cartridge case was fired in a specific firearm or the same firearm by comparison to each other if a suspected firearm is not submitted. A firearm examiner can determine if a fired bullet from a victim or from a crime scene was fired from a specific firearm that was recovered at a scene or from a suspect. If no firearm is recovered, a firearm examiner can determine how many firearms were discharged at the scene. A firearm examiner microscopically evaluates fired evidence using an optical comparison microscope and observes the stria on the bearing surface of a fired bullet. These striae are marked on the bullet as it travels down the barrel of the firearm. They are accidental in nature and occur because of random imperfections within the barrel of the firearm. The patterns of these striations are considered by firearm examiners to be

unique. Many studies have been published supporting the idea that the striations on a bullet are unique (1–11). The striations are considered unique because the rifling tools during barrel manufacturing wear during their use and change microscopically. The greatest similarities between two barrels would be expected to occur in two barrels that were manufactured by the same rifling tool consecutively. There have also been many studies of a firearm examiners ability to differentiate evidence involving consecutively manufactured tools (2,3,5,11–38). Even though there is strong evidence supporting the discipline of firearm identification, there have been some expected criticisms considering the subjective nature of the analysis.

In 2009, the National Academy of Science Report (NAS) questioned the scientific validity of firearm and toolmark identification (39). Additional studies have been published after the NAS report that help support the scientific validity of firearm and toolmark comparisons (11,27–38,40–47). However, in September of 2016, the

Executive Office of the President President's Council of Advisors on Science and Technology (PCAST) published a Report to the President titled: *Forensic Science in the Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (48). It criticized several different forensic disciplines as well as the scientific validity of firearm and toolmark identification. In this report, PCAST outlines reasons they believed firearm/toolmark examinations did not meet the scientific criteria for "foundational validity". PCAST coined and defined the term "Foundational Validity". According to PCAST, since firearm identification is a feature-comparison method, its foundational validity can only be established through multiple independent black box studies ([48, p. 68]). In order to meet the scientific criteria for foundational validity, PCAST states that the following criteria must be met:

1. Studies must involve sufficiently a large number of examiners and be based on sufficiently large collections of known and representative samples from relevant populations to reflect the range of features or combination of features that will occur in the application.
2. Empirical studies should be conducted so that neither the examiner nor those with whom the examiner interacts have any information about the correct answer.
3. Study design and analysis framework should be specified in advance.
4. The empirical studies should be conducted or overseen by individuals or an organization that do not have a stake in the outcome of the studies.
5. Data, software, and results of the validation studies should be available to allow other scientists to review the conclusions.
6. To ensure that conclusions are reproducible and robust, there should be multiple studies by separate groups reaching similar conclusions. ([48, pp.52-53])

PCAST reviewed several studies that have been conducted in the field of firearm/toolmark identification in the past 15 years. They stated that many of the studies were not appropriate for assessing scientific validity and estimating the reliability because they employed artificial designs that differ in important ways from the problems faced in casework ([48, p.106]). These studies employed a "closed set" design where the source firearm is always present. They stated that the closed-set design is problematic in principle and underestimates the false positive rate in practice ([48, p.106]). Therefore, PCAST concluded that this design is not appropriate for assessing scientific validity and measuring reliability ([48, p.109]).

In order to address this criticism, more "open set" studies need to be conducted to have a black-box study that meets the scientific criteria for "foundational validity" set forth by PCAST as much as possible for firearm and toolmark identification.

With this goal in mind, the author's laboratory obtained 30 consecutively manufactured Beretta 9 mm Luger caliber barrels. These Beretta barrels were obtained by the laboratory in 1996 from Beretta

Highlights

- PCAST criticized firearm identification because of the few studies to support "Foundational Validity".
- A study of 30 consecutively manufactured Beretta barrels was created to address the concerns of PCAST.
- This test uses an "open set" design which was deemed appropriate by PCAST.
- CTS was used as a third party so that the participant did not communicate with the test designer.
- A low error rate was observed for firearm examiners when comparing fired bullets for this study.

U.S.A. Corp. of Accokeek, Maryland with the intent of performing a consecutively manufactured study. Given that the barrels were obtained in 1996, no one from the laboratory was present during the collection of the barrels and there is no formal documentation other than a packing list. The barrels are stamped numerically from 1 to 30 indicating the order of production. This experiment will provide participants in this study with a selection of known test standards from the 30 consecutively manufactured barrels and also provide them with 20 unknowns (a sample where the participant needs to determine if the bullet was fired from one of the barrels provided or some other barrel).

This experiment will be set up similarly to the Ten Consecutive Manufactured Ruger Barrel Study by James Hamby (49); however, instead of a "closed set", it will be an "open set". In an "open set", the participant should have no expectation that all questioned bullets should match one or more of the unknowns. Only firearm examiners who were qualified to do work by their laboratory were selected to participate in this experiment. There was an administrative section with several questions that each participant filled out, such as, years of experience in the field, type of lighting, type of scope, laboratory accreditation, certification, etc.

Two hundred tests were created for this study. Within the 200 tests, there were 20 different answer keys of 10 sets each. The 30 consecutive Beretta manufactured barrels and 5 "known non-matching" (in this study, "known non-match" refers to a bullet fired from a barrel that is not present in the provided knowns) 9 mm Luger caliber firearms with similar rifling characteristics as the 30 consecutive barrels from the laboratory's reference collection were included in the test sets. Each set of 10 was determined using the random number function present in Microsoft Excel. The random number function was generated and then repeated for the next 19 unknowns for each test set. Using this process for the 20 unknowns, it was possible to have multiple bullets from the same barrel. It was also possible for the unknown bullets to have been fired in a barrel which did not correspond to any of the knowns.

3 | MATERIALS AND METHODS

Thirty consecutively manufactured barrels were obtained from Beretta U.S.A. Corp in January of 1996 by a local laboratory. These barrels have been test fired many times, so there was no concern a "break-in" period would significantly affect the test samples. A "break-in" period is a short period after the barrel has been manufactured where several bullets have to be fired in the barrel before the striations mark in a reproducible manner (16,18). There were five additional pistols used in the test structure to provide "known non-match" fired bullets. All of the pistols have similar general rifling characteristics (GRC) to the known Beretta barrels that were provided. The general rifling characteristics (GRC) were six lands and grooves with a right hand twist where the land impression widths ranged from 0.072 to 0.076 inches and the groove impression widths ranged from 0.100 to 0.106 inches. The following pistols were used: Beretta model 92F, Ruger model P85 MKII, FEG model PJK-9HP, Fabrique Nationale model Hi-Power, and CZ model 75.

For this study, over 14,000 9 mm Luger caliber Federal FMJ cartridges with Lot# AE9AP were obtained and test fired through the barrels.

Figure 1 is a simplified flow chart to help visualize the procedure of how the test sets were created in this study. Each barrel/pistol was lubricated and cleaned prior to test firing the test set (there were approximately 400 bullets fired through each known barrel). Ten percent of the fired bullets were verified, by an AFTE certified firearm examiner, to display sufficient microscopic individual characteristics for identification. Prior to the firing process, every 10th bullet (1, 11, 21, 31, 41, etc) was marked with a sharpie for microscopic comparison to other fired bullets in that set of 100. The ten bullets from each set of 100 were intracompared. A bullet from each set of 100 was then microscopically intercompared to a bullet from each of the 4 sets of 100. Therefore, all of the bullets from 1 to the total number of bullets fired for that barrel should be identifiable; however, not all fired bullets were microscopically compared. A dry patch was run down the barrel after each set of 100 test fires.

After all 30 Beretta barrels were fired, the "known non-matching" pistols received from the laboratory's Firearms Reference Collection were fired using the same process outlined above; however, only about 100 bullets were fired through these pistols because the known exemplars did not need to be fired and therefore, lessened the number of test fires needed.

The Beretta barrels used in this study were manufactured using a broaching tool (50). Since the potential for subclass characteristics may be present, the procedure Ronald Nichols outlined in his journal article (51) was utilized. A cast was made from the muzzle to the chamber of the 30 Beretta barrels using Forensic Sil casting material. The cast was then cut in half and the muzzle end of the cast was compared to the chamber end of the cast. This comparison was conducted by an AFTE certified firearm examiner and no subclass characteristics were observed. Due to the exorbitant cost of making the cast, it was not possible to ship casts of the barrel to each examiner. If any participant asked about the potential for subclass

characteristics, they were told this method had been utilized to verify, there were no subclass characteristics.

Each test consisted of a set of three fired bullets each fired from 15 known standards (numbered 1 through 15) and 20 unknowns (labeled A through T). The random number generator feature on Microsoft Excel was used to determine the test sets. The function used to create the random number was RANDBETWEEN (x,y) where x is the lowest number and y is the highest number. Excel could select any number between x and y. This means that there could be multiple unknowns from the same barrel whether it is from a known barrel or an unknown non-matching barrel.

There were two sets of tests: the first set included barrels from 1 to 15, barrels 16 and 17 (not provided in this test as a known), the Beretta model 92F pistol, the Ruger model P85 MKII pistol, and the FEG model PJK-9HP pistol. The second set included barrels from 16 to 30, barrel 14 and 15 (not provided in this test as a known), the Beretta model 92F pistol, the FN model Hi-Power pistol, and the CZ model 75 pistol.

Once all of the test firing was completed, the bullets were scribed according to the Excel spreadsheet and packaged to be sent to Collaborative Testing Services (CTS). For each known of a particular test set, each bullet was scribed with the barrel number, and the set of standards were packaged into a coin envelope labeled with the barrel number. These knowns were placed in a large zip top plastic bag with the test set range (#1-#10, #11-#20, etc.) and the barrel number written on the bag. After all of the knowns were scribed for a particular barrel, the unknowns for that barrel were scribed with the appropriate letter, packaged in a coin envelope with that letter written on it, and put in a small zip top bag labeled with the test range and the appropriate letter. This procedure was performed for all 30 barrels.

For the fired bullets from barrels where a corresponding known was not present, the bullet was scribed with the appropriate letter and packaged in a coin envelope with the letter written on it and put in a small zip top bag with the test set range and the appropriate letter. For each test set, a large zip top plastic bag was labeled with the test set range and that it contained unknowns without a known present, incorporating "known non-match" in the test design.

Therefore, there were 15 large zip top plastic bags for each test set which contained the fifteen knowns (labeled 1-15 or 16-30) and unknowns (labeled A-T). In addition, there was one large zip top plastic bag labeled with the test set range and "unknowns without a known present" written on it.

These test sets were then sent to CTS for packaging and shipment. CTS assigned each test set a unique webcode. If more than one test set was ordered by a specific laboratory, different test sets were sent. This meant that no examiner in the same laboratory would have the same test. CTS managed communication with all of the participants in the study. At no time did the developer of the test know which particular tests were received by the participants.

The procedure outline below was the procedure that CTS used to package the test:

Knowns and unknown bullets

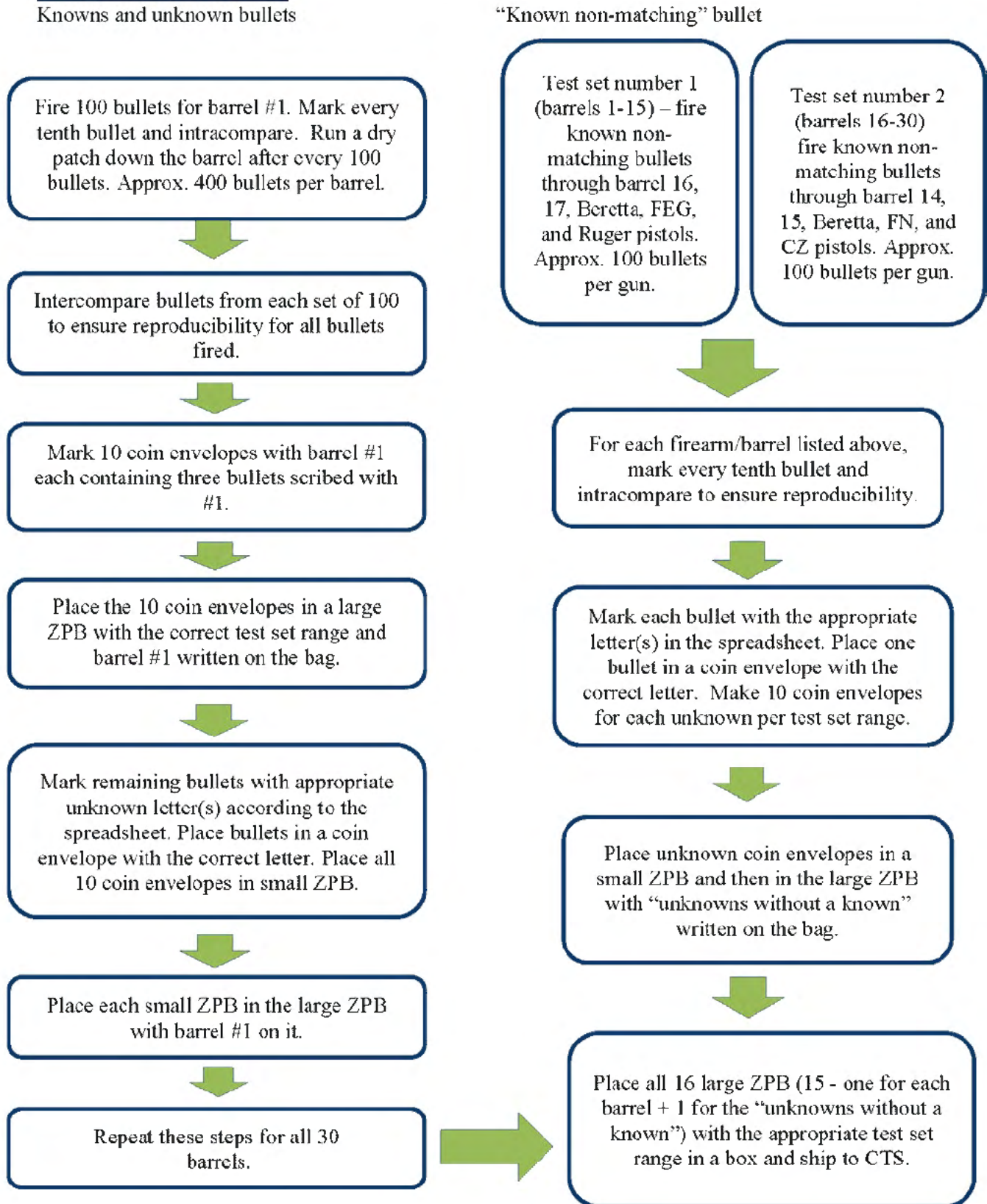


FIGURE 1 Simplified flow chart for procedure to create the test sets

1. With approximately 120 participants over a generated 200 Kits, the participants were spread out as evenly as possible, by utilizing up to 6 kits from each set of 10. Participants were

assigned a random alpha-numeric 6 digit code (WebCode). This was sorted alphabetically and the kits were assigned numerically to this sorted list.

2. CTS received boxes of the materials for the Kits in 10 kit ranges.
3. CTS unpacked the bags of Known and Questioned envelopes and laid them out on tables for the required number of kits to be used per range. As stated above this was approximately 6 per range.
4. The attached picture illustrates one of the multiple stations that were set up to lay out the envelopes as they were unpackaged from the provided bags. The known bullets were numerical, so no assistance in laying them out was used. However, to assist with the Questioned Bullets, paper with the alphabetical range was laid down so that no letter was missed during unpacking.
5. Once all the envelopes were laid out from the provided bags, it was verified that all items were present on the table for all of the necessary kits.
6. Then the full range of envelopes were picked up and packaged into the appropriately labeled zip top bag.
7. The kit ranges and their assigned webcodes were checked prior to laying out the samples, after they were packaged into the zip top bags, and again when the zip top bags were placed inside of a sample pack box.

Each participant would receive a box from CTS with a label containing the participant number and the appropriate webcode. Within that box, there would be 15 coin envelopes containing three bullets from each of the test standards (either labeled #1 through #15 or labeled #16 through #30) and 20 coin envelopes containing one bullet from an unknown (questioned) sample labeled letter A through T.

For test set number 1 (barrels #1-#15) and test set number 2 (barrels #16-#30), an average of 22% of the unknowns provided did not have a corresponding known provided. The first test set ranged from having three unknowns (15%) not provided to having seven unknowns (35%) not provided. While the second test set ranged from having three unknowns (15%) not provided to having six unknowns (30%) not provided. The number of duplicates for test set number 1 and number 2 range from two to five. The number of triplicates for test set number 1 and number 2 range from zero to two. Because of the importance of the consecutive nature of this study, the number of unknowns provided from consecutively produced barrels within each 15 barrel grouping was reviewed. For test set number 1, the number of unknowns from consecutively produced barrels ranged from 7 to 10 barrels and for test set number 2, it ranged from 6 to 13 consecutive barrels; however, the set with six (6) unknowns from consecutive barrels also had another set of 5 unknowns from another subgroup of consecutively produced barrels.

4 | RESULTS

After soliciting qualified examiners from the firearm examination community, there were a total of 110 participants who volunteered to receive the test and participate. All of the data was collected by CTS via their website; there were 74 participants (67.3%) who submitted results.

From the tests distributed, there were 1149 possible identifications to a known barrel, 151 possible identifications to another bullet present in the unknowns that are not present in the knowns, and 180 true eliminations (bullet where a known or another unknown is not present in the test). Therefore, there was a total of 1300 possible identifications and 6120 true eliminations ($180 * 34 [15 \text{ knowns} + 19 \text{ unknowns}] = 6120$).

Upon initial submission of the test results, there were 7 false identifications, 18 false eliminations, 23 missed identifications when the known was present and 22 missed identifications when only the unknown was present. See Table 1 for the data associated with results. In Table 1, the percentage of false identifications was calculated by dividing the number of false identifications by the number of correct identifications.

After looking at several of the false identification responses, it was realized that two examiners appeared to have incorrectly transferred the information into the wrong cell on the CTS website. One examiner made four false identifications because they had transposed the letters. On the answer sheet, the examiner had identified one of the unknowns to a specific barrel and then included other unknowns that had been identified to a different barrel. Another examiner made one false identification which was off by one letter; this would indicate that they read the wrong letter when filling in the answer sheet. A generic letter was sent by CTS to the participants who had incorrect responses stating that it was believed that they had made a typographical error and had ended up identifying one bullet to two different barrels. Below is the text of the email that was sent:

"It was noticed that there is an entry that appears to be a transcription error because there was an entry with more than one identification and your answers reference two different barrels. Any clarification that you could provide would be appreciated."

A response to the email was received from both examiners and their email response identified where the error was and what the correct answer should have been.

Another false identification was a typographical error. In the answer sheet, an unknown was identified as having been fired from barrel #1; however, barrel #1 was not one of the barrels provided for

TABLE 1 Error calculation based on original data submission

Type of error	Number	Total ^a	Error rate (%)
False identification	7	1251	0.56
False elimination	18	10935	0.16
Total (false identification and false elimination)	25	12186	0.21%
Missed identification (known present)	23	1251	1.84
Missed identification (unknown present)	22	1251	1.76

^aThe information for identifications was always filled in; however, for the false elimination data, some examiners left the area blank.

that test set. The barrels provided for that test set were barrels #16 through #30. Therefore, this had to be a typographical error. Two emails were sent to try and clarify what the correct response should have been; however, no response was received.

From the text described above, it is reasonable to determine that the five transferring errors and the one typographical error are administrative in nature and therefore, should not be counted as false positives. Since these tests were not technically or administratively reviewed, which is part of the normal process in most forensic laboratories, these errors would likely have been discovered during the administrative review process. For the results submitted, there was one false identification. Therefore, corrected responses from this test are in Table 2. In Table 2, the percentage of false identifications was calculated by dividing the number of false identifications by the number of correct identifications.

There were 18 false eliminations present in the study. These false eliminations were made by six examiners. Four examiners were responsible for 16 of the false eliminations (8, 3, 3, and 2), and two examiners made one false elimination each. The false elimination response in Tables 1 and 2 were calculated based upon the total number of eliminations present because not all examiners filled in the area designated for eliminations. This area was left blank by many examiners because most firearm examiners do not feel it is necessary to eliminate all other firearms if they have made an identification to a specific firearm.

After calculating the overall error rates of the examiners, the sensitivity and specificity were also calculated. Sensitivity is the number of identifications reported divided by the number of identifications present in the test. The number of identifications submitted in this test was 1251 and the identifications present in this test was 1300. Therefore, the sensitivity of this test is 96.2%. The specificity is the number of eliminations reported divided by the number of eliminations present in the test. The number of eliminations reported in this test was 10,935 and the number of eliminations present in this test was 47,876. Therefore, the specificity of this test is 22.8%. While the specificity of this test is on the low side, possible reasons are explained in the discussion.

TABLE 2 Error calculation based on corrected data from participants

Type of error	Number	Total ^a	Error rate (%)
False identification	1	1257	0.080
False elimination	18	10935	0.16
Total (false identification and false elimination)	19	12192	0.16
Missed identification (known present)	23	1257	1.83
Missed identification (unknown present)	22	1257	1.75

^aThe information for identifications was always filled in; however, for the false elimination data, some examiners left the area blank.

5 | DISCUSSION

The overall goal of a consecutive manufactured barrel study is to support the firearm identification community with scientific studies that show qualified firearm examiners can identify a fired bullet or fired cartridge case to a specific firearm within a small degree of error. The consecutively manufactured study is a "worst case scenario" where multiple barrels are manufactured consecutively (one after the other) with the same tool at the factory. In this and other consecutively manufactured studies, a firearm examiner can identify an unknown bullet to the correct barrel with a very low error rate. PCAST and other critics have found fault with many of the previous studies.

The first criterion that PCAST outlined: in order to establish foundational validity was the studies need to include a sufficiently large number of examiners and have large collections of representative samples that are typically found in casework. This is the largest consecutively manufactured barrel study known to date. Prior to this study, 10 consecutively manufactured barrels was the largest study that had been completed [3, 5, 6, 9, 11, 16, 20, 49]. Seventy-four examiners of the 110 that signed up completed the test (67.3%). This result is similar to other studies, such as the Ames Study where 218 out of 284 (76.8%) examiners participated (40) and the Smith study where 31 out of 47 (65.9%) examiners participated [41].

Since there are approximately 1200 firearm examiners (AFTE membership: Provisional [304], Regular [685] and Distinguished [174]) throughout the world, the number of participants in this study would have incorporated 6.3% of the firearm examiner in the world. This is obviously lower than desired; however, to be expected given the study had a large number of knowns and unknowns, it required a significant amount of time to complete the task. Since many firearm laboratories throughout the country and world have large backlogs and minimum manpower, it is reasonable to conclude participation could put an undue strain on their laboratories and participation would not be permitted by the employer in most cases. Also, examiners who would eagerly volunteer must manage time effectively and choose which studies to participate in because casework is still the priority.

In this study, Beretta barrels and pistols present in the Firearms Reference Collection were used. Many people purchase firearms chambered for the 9 mm Luger cartridge including the military, police departments, and civilian consumers for home defense. Since 1999, more than 44,000 firearms have been submitted to the firearm identification section of a local laboratory in a variety of different types of cases. Of those 44,000 firearms, more than 12% of those firearms have been chambered in 9 mm Luger caliber. Beretta is a popular manufacturer and they manufacture many different firearms chambered for the 9 mm Luger cartridge. For many years, the local police department used the 9 mm Luger Beretta model 92FS as their duty weapon. Beretta manufactured firearms are also commonly found in casework. Of the 5365 firearms chambered in 9 mm Luger submitted to the local police department since 1999, 515 of them were manufactured by Beretta. Therefore, Beretta accounted for approximately 9.6% of the 9 mm Luger submitted firearms. All of the pistols

selected for the unknowns were from the local laboratory's firearms reference collection. The local laboratory's firearms reference collection is a collection of firearms that have been seized during police investigations that occurred within the county. Therefore, all of the firearms used in this study are often seen in casework.

The second criterion for PCAST was: Empirical studies should be conducted so that neither the examiner nor those with whom the examiner interacts have any information about the correct answer. In this study, this criterion was met by a company called Collaborative Testing Services, Inc. (CTS). CTS is a company widely known throughout the forensic community as a proficiency test provider. All qualified firearm examiners filled out an application and submitted the application to CTS which served as the main point of contact for all of the participants in this study. CTS determined which tests were going to be shipped to the participant. In the event that a technical question needed to be answered, the test developer was contacted through CTS. In that event, the test developer did not know which specific test the participant was given because the webcode did not correlate to any information the test developer had.

The third criterion for PCAST was: Study design and analysis framework should be specified in advance. The study design and analysis framework were specified in advance. The local laboratory in collaboration with CTS specified in advance the design and analysis framework of the study. This was necessary so both parties knew and understood their responsibilities.

The fourth criterion for PCAST was: The empirical studies should be conducted or overseen by individuals or an organization that do not have a stake in the outcome of the studies. The study was conducted and overseen by CTS. In its capacity in this study, CTS served as the administrator of the test. CTS had no stake in the outcome of results of this study. CTS collected all of the answers submitted via their website and then forwarded the responses to the developer of the test.

The fifth criterion for PCAST was: Data, software, and results of the validation studies should be available to allow other scientists to

review the conclusions. The test materials and results of this validation study are available upon request.

The sixth criterion for PCAST was: To ensure that conclusions are reproducible and robust, there should be multiple studies by separate groups reaching similar conclusions. This study, along with many other studies that are currently being distributed, will help ensure that the conclusions are robust and reproducible. This study reaches similar conclusions previous studies have demonstrated which is that within a low error rate, firearm examiners are able to identify an unknown bullet to a specific firearm.

Along with the criterion described above, PCAST also found fault with previous studies because they did not incorporate an "open set". As described in the study design, this study incorporated an "open-set" concept. Known non-matching samples were included.

It was suggested in the PCAST report, that a 95% confidence interval be calculated for these studies using the Clopper-Pearson/Exact Binomial method, the Wilson Score interval, the Agresti-Coull (adjusted Wald) interval, and the Jeffreys interval. These calculations were done using the following website <https://epitools.ausvet.com.au/ciproportion>. The data is included in Table 3.

The 95% confidence interval for this study at the upper limit for the corrected results was an error between 0.24% and 0.5%. The 95% confidence interval at the upper limit for the reported results was a range of 0.97%–1.17%. According to sources (52,53), for a study this size, the best confidence interval method calculations would be either the Wilson Score, Agresti-Coull (adjusted Wald), or Jeffreys Interval.

In the PCAST report, it was stated that closed-set studies have inconclusive and false-positive rates that are dramatically lower than those for an open designed study (p. 109). If one includes inconclusive results with false positive answers, the error rate will increase; however, it is inappropriate to include inconclusive results with false positive errors. An inconclusive result is reserved for an examiner when the class characteristics are the same and there are insufficient individual characteristics to reach a conclusion. If the firearm examiner believes that there is not enough

TABLE 3 Calculation of binomial confidence intervals for false identifications for both the original submission and the corrected data

Sample size	Positive number	Confidence	Proportion	Lower 95%	Upper 95%
1258	7	0.95			
Normal			0.0056	0.0015	0.0097
Clopper-Pearson			0.0056	0.0022	0.0114
Wilson			0.0056	0.0027	0.0114
Jeffreys			0.0056	0.0025	0.0109
Agresti-Coull			0.0056	0.0024	0.0117
1258	1	0.95			
Normal			0.0008	0.0008	0.0024
Clopper-Pearson			0.0008	0.0000	0.0044
Wilson			0.0008	0.0001	0.0045
Jeffreys			0.0008	0.0001	0.0037
Agresti-Coull			0.0008	0.0001	0.0050

information on the sample to come to a conclusion, then an inconclusive result is appropriate. Firearm examiners approach these tests as if they are casework; therefore, it would be inappropriate for an examiner to be forced to come to an identification or elimination if sufficient information is not observed on the items in question. A laboratory would not want to have a policy that forces a scientist to render an opinion if there is not enough information to make a determination. The same approach should be used for firearm examiners in this study. Also, inconclusive is neither a correct answer nor an incorrect answer. From the perspective of the defense attorney, this conclusion could be a benefit because it would allow for "reasonable doubt".

As stated above, it is not accurate to include inconclusive answers in the error rate because an inconclusive result is neither positive nor negative. These confidence interval calculations are based upon the theory that the result is either positive or negative, and an inconclusive result is not possible. However, in order to compare information that was published in the PCAST report, below the inconclusive result has been included in the error rate. For the submitted results, if one included false positive and inconclusive results, the results would be 52 out of the 1303 (4.0%) for the submitted result and 46 out of 1303 (3.5%) for the corrected result. When comparing the error rates of the submitted results, the false positive error was 0.56% and when the inconclusive results are included, the false positive and inconclusive error is 4.0% (7-fold increase). When comparing the errors rates of the corrected results, the false positive error was 0.08% and when the inconclusive results are included the false positive and inconclusive error is 3.5% (44-fold increase). This is by far much lower than the 100-fold error reported in the PCAST report ([48, p.11).

Some of the inconclusive results can be explained due to laboratory policies. In the additional questions that were provided with the answer sheet, one of the questions was whether there was a laboratory policy that did not allow examiners to eliminate two items based on differences in individual characteristics. There were 3 examiners who reported that their laboratory prohibited eliminating based on differences in individual characteristics because of a laboratory policy. Two examiners reported that they could only eliminate based on individual characteristics if it was verified by another qualified examiner. Since all of the fired bullets in this study have similar rifling characteristics, an examiner would have to eliminate based upon individual characteristics. For those two examiners who needed verification from another examiner to eliminate an item based on individual characteristics, it is unknown as to whether that examiner requested this procedure for the purposes of this test.

The number of inconclusive results for this study may be higher than other studies. This was a large test with many known samples. There were 15 knowns which typically represents far more knowns than an examiner would evaluate in routine casework. For a comparison of one unknown to the fifteen knowns, the examiner is comparing potentially conducting ninety ($90 = 15 \text{ knowns} * 6 \text{ per bullet}$) land impression comparisons. Therefore, there would be 1800 (90 land

impressions * 20 unknown bullets). In addition, with an average of more than 4 unknowns present per test, there would be potentially 24 comparisons (4 comparisons * 6 land impressions) per unknown for a total of about 1824 comparisons per test.

A sensitivity of 96.2% and specificity of 22.8% were calculated for this test. While the sensitivity is very good, the specificity was evaluated further. Of the 74 examiners who submitted results, many examiners either left the elimination area blank, put "N/A", or did not have a response. If an examiner left the elimination answer blank or put an "N/A", this meant that there were as many as 34 eliminations for one bullet that were missing (depending upon the test set). If the examiner left it blank for all of the bullets in a single test, this would mean that up to 680 ($34 * 20$) eliminations were potentially missing. There were several examiners who would eliminate the knowns, but did not eliminate the unknowns. Therefore, the number of eliminations went from 34 eliminations to 15 eliminations. This could be because the examiners did not realize that they were supposed to eliminate each unknown bullet from all of the unknowns. The normal process in most laboratories in casework is to compare all of the evidence to each other and to the tests, the directions for the study could have been more explicit. As discussed earlier, many firearms examiners did not fill in this area because they do not think it is necessary to eliminate all other firearms if they have made an identification to a specific firearm. Given this information, this perception has skewed the data for specificity for this study.

There were 16 examiners who had an inconclusive result for all of the eliminations in the test. The examiners in this study were asked to follow the AFTE Range of Conclusions and designate which inconclusive result that they were reporting. Below is the definition of inconclusive from the AFTE Range of Conclusions (54):

2. Inconclusive

- a. Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.
- b. Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.
- c. Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.

Of the 16 examiners who gave an inconclusive result, 9 examiners have a result of inconclusive (c), four have a result of inconclusive (b), and two have a result of inconclusive (a).

Therefore, the majority of the examiners who gave an inconclusive result, thought that it was inconclusive (c) and that there was disagreement of individual characteristics; however, just not enough disagreement of individual characteristics to come to a conclusion of an elimination.

There can be several reasons why an examiner would choose an inconclusive result over elimination. As discussed earlier, it may be a laboratory policy not to eliminate based on individual characteristics.

Therefore, five of sixteen examiners who have this result was due to a laboratory policy. Another reason that an examiner may give the result of inconclusive in a test like this is because they feel it was not feasible to determine the reproducibility of the marks. If there are two representations of the bullet fired from a specific barrel, then an examiner can determine what striations are reproducing and what striations are not. Often times in casework, an examiner will compare the tests to each other and the evidence to each other. If the evidence marks consistently and the tests mark consistently and the evidence and tests mark differently, then the examiner can come to the result that the evidence and the tests are from different firearms. However, if there is only one representative of the evidence, this decision becomes more complicated if some of the marks are similar. If this is the case, the conservative approach is for the result to be inconclusive.

For eliminations, there were 18 false eliminations and 10,935 correct eliminations for a false-elimination error rate of 0.16%. Of the 18 false eliminations, eight false eliminations occurred with one examiner (almost half of the errors). In recent journal publications [28, 40], false identifications and false eliminations are calculated separately. As a scientific discipline, it is important for the firearm examiners to pay attention to both false identifications and false eliminations. However, a false elimination is less problematic than a false identification because the subject of an investigation is not going to be imprisoned for a false elimination. After calculating both the false identifications and the false eliminations, total error rate was calculated for this study. The total error was calculated to be 0.21% for the original submission and 0.16% for the corrected results (Tables 1 and 2).

Besides what was discussed earlier, there were other additional questions that asked about the examiner such as, the years of experience, whether the examiner's laboratory was accredited, whether the examiner was certified, and the method the examiner used for the examination (pattern matching, QCMS, or both). All but two of the participants responded to these questions, so this information was based on 72 responses. From this information, the examiners had a range of experience that went from 1 year of experience to 50+ years of experience. The average years of experience was 12.3 years. 91.7% (66) of examiners were from accredited laboratories. 33.3% (24) of the examiners were certified firearm examiners. 92.9% (65) reported that they used pattern matching as the method for their comparison while 7.1% (5) reported that they used both pattern matching and QCMS (Quantifiable Consecutive Matching Stria) (2 of the responses were incomplete). While none of this information appeared to have an effect on the results of the test, it does represent the information pertaining to the background of the examiners in this test.

6 | CONCLUSIONS

This study was designed to respond to many of the criticisms presented in the PCAST report. It was modeled after the requirements outlined in the PCAST report to enable forensic disciplines which analyze impression evidence to establish Foundational Validity.

From the results of this study, trained and qualified firearm examiners throughout the United States and world are able to identify unknown samples to a known barrel in an "open set" format with a very low error rate. This test incorporated 30 consecutively manufactured Beretta barrels. It was divided into two different test sets, but combined results indicate, examiners are able to identify unknown bullets to the correct barrel from 30 consecutively manufactured barrels within a low error rate. Consecutively manufactured barrels are a firearm examiner's "worst case scenario" because a barrel manufactured by the same tool one after the next will have striations that are the most similar and it is more likely that an examiner could make an error. From the data submitted, the false identification error rate of the 74 examiners was 0.55% (1 in 182) with the result for the lower confidence interval as low as 0.2% (1 in 500) and with the upper confidence interval as high as 1.1% (1 in 91). The false identification error rate for the corrected data (data where the typographical errors were corrected) was 0.08% (1 in 1250) with the lower confidence interval being as low as 0.01% (1 in 10,000) and as high as 0.4% (1 in 250) for the upper confidence interval. These error rates are similar to previous studies (which may or may not have followed the model outlined in the PCAST Report) that have been published in the firearms examination discipline indicating that the specific requirements set up by PCAST have little effect on the overall error rates of firearm examiners.

ACKNOWLEDGEMENTS

The author would like to thank Collaborative Testing Service (CTS) for assisting in designing, administering, packaging, distributing, and compiling the answers for this study which includes Cathy Brown and Richard Hockensmith. I would like to acknowledge my fellow laboratory employees to include Director Kristen Lease, Firearms Manager Joseph Young, and all of the examiners in my laboratory for their help and assistance with this project. I would specifically like to thank Corporal Tara Mattingly for helping fire the 14,000 cartridges needed to complete this study. I would also like to thank all of the examiners and their agencies who took time out of their busy day to complete the many comparisons that were needed for this study.

REFERENCES

1. Biasotti AA. A statistical study of the individual characteristics of fired bullets. *J Forensic Sci.* 1959;4(1):34-50.
2. Goddard C. The Valentine's Day massacre: a study in ammunition tracing. *AFTE J.* 1980;12(1):44-59.
3. Lutz M. Consecutive revolver barrels. *AFTE Newslett.* 1970;9:24-48.
4. Murdock J. The effects of crowning on gun barrel individuality. *AFTE Newslett.* 1970;7:12-3.
5. Hall E. Bullet markings from consecutively rifled Shilen DGA barrels. *AFTE J.* 1983;15(1):35-53.
6. Miller J. An examination of two consecutively rifled barrels and a review of the literature. *AFTE J.* 2000;32(3):259-70.
7. Miller J. An examination of the conservative criteria for identification of striated toolmarks using bullet fired from ten consecutively rifled barrels. *AFTE J.* 2001;33(2):125-32.
8. Smith E. Cartridge case and bullet comparison validation study with firearms submitted in casework. *AFTE J.* 2005;37(4):130-5.

9. Brundage D. The identification of consecutively rifled gun barrels. *AFTE J.* 1998;30(3):438-44.
10. Hamby J. Identification of projectiles. *AFTE J.* 1974;6(5-6):22.
11. Monkres J, Luckie C, Petraco NDK, Milam A. Comparison and statistical analysis of land impressions from consecutively rifled barrels. *AFTE J.* 2013;45(1):3-20.
12. Flynn E. Toolmark identification. *J Forensic Sci.* 1957;2(1):95-106.
13. Butcher S, Pugh D. A study of marks made by bolt cutters. *J Forensic Sci Soc.* 1975;15(2):115-26. [https://doi.org/10.1016/S0015-7368\(75\)70965-9](https://doi.org/10.1016/S0015-7368(75)70965-9).
14. Watson D. The identification of toolmarks produced from consecutively manufactured knife blades in soft plastics. *AFTE J.* 1978;10(3):43-45.
15. Cassidy F. Examination of toolmarks from sequentially manufactured tongue and groove pliers. *J Forensic Sci.* 1980;25(4):796-809. <https://doi.org/10.1520/JFS11294J>.
16. Murdock J. A general discussion of gun barrel individuality and an empirical assessment of individuality of consecutively rifled 22 caliber rifles. *AFTE J.* 1981;13(3):84-111.
17. Tuira Y. Tire stabbing with consecutively manufactured knives. *AFTE J.* 1982;14(1):50-2.
18. Raven Matty W. 25 automatic pistol breech face tool marks. *AFTE J.* 1984;16(3):57-60.
19. Matty W. A comparison of three individual barrels produced from one button-rifled barrel blank. *AFTE J.* 1985;17(3):64-9.
20. Brown C, Bryant W. Consecutively rifled gun barrels present in most crime labs. *AFTE J.* 1995;27(3):254-8.
21. Lopez L, Grew S. Consecutively machined Ruger bolt faces. *AFTE J.* 2000;32(1):19-24.
22. Eckerman S. A study of consecutively manufactured chisels. *AFTE J.* 2002;34(4):379-90.
23. Lee S. Examination of consecutively manufactured slotted screwdrivers. *AFTE J.* 2003;35(1):66-70.
24. Thompson E, Wyant R. Knife identification project (KIP). *AFTE J.* 2003;35(4):366-70.
25. Bunch S, Murphy D. A comprehensive validity study for the forensic examination of cartridge cases. *AFTE J.* 2003;35(2):201-3.
26. Clow C. Cartilage stabbing with consecutively manufactured knives: A response to Ramirez v. State of Florida. *AFTE J.* 2005;37(2):86-116.
27. Weller T, Zheng A, Thompson R, Tulleners F. Confocal microscopy analysis of breech face marks of fired cartridge cases from 10 consecutively manufactured pistol slides. *J Forensic Sci.* 2012;57(4):912-7. <https://doi.org/10.1111/j.1556-4029.2012.02072.x>.
28. Giroux B. Empirical and validation study: consecutively manufactured screwdrivers. *AFTE J.* 2009;41(2):153-5.
29. Lancon D. Toolmarks in bone: continuing research with consecutively made knife blades. *AFTE J.* 2009;41(2):130-7.
30. Lyons D. The identification of consecutively manufactured extractors. *AFTE J.* 2009;41(3):246-56.
31. Mayland B, Tucker C. Validation of obturation marks in consecutively reamed chambers. *AFTE J.* 2012;44(2):167-9.
32. Cazes M, Goudeau J. Validation study results from Hi-Point consecutively manufactured slides. *AFTE J.* 2013;45(2):175-7.
33. Chu W, Tong M, Song J. Validation tests for the congruent matching cells (CMC) method using cartridge cases fired with consecutively manufactured pistol slides. *AFTE J.* 2013;45(4):361-6.
34. Zheng XA, Soones J, Thompson R, Villanova J, Kakal T. 2D and 3D topography comparisons of toolmarks produced from consecutively manufactured chisels and punches. *AFTE J.* 2014;46(2):143-7.
35. King E. Validation study of computer numerical control (CNC), consecutively manufactured screwdrivers. *AFTE J.* 2015;47(3):171-6.
36. Owens S. An examination of five consecutively rifled Hi-Point 9mm pistol barrels with three lands and grooves left twist rifling to assess identifiability and the presence of subclass characteristics. *AFTE J.* 2017;49(4):208-15.
37. Laporte D. An empirical and validation study of breechface marks on .380 ACP caliber cartridge cases fired from ten consecutively finished Hi-Point model C9 pistols. *AFTE J.* 2011;43(4):303-9.
38. Fadul T, Hernandez G, Stoiloff S, Gulati S. An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing 10 consecutively manufactured slides. *AFTE J.* 2013;45(4):376-93.
39. National Research Council, Committee on Identifying the Needs of the Forensic Science Community. Strengthening forensic science in the United States: A path forward. Washington, DC: National Academies Press; 2009. <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>. Accessed 2 Oct 2020.
40. Baldwin D, Bajic S, Morris M, Zamzow D. A study of false-positive and false-negative error rates in cartridge case comparisons; 2016. <https://www.ncjrs.gov/pdffiles1/nij/249874.pdf>. Accessed 12 Apr 2020.
41. Smith T, Smith G, Snipes J. A validation study of bullet and cartridge case comparisons using samples representative of actual casework. *J Forensic Sci.* 2016;61(4):939-46. <https://doi.org/10.1111/1556-4029.13093>.
42. Hamby J, Norric S, Petraco N. Evaluation of Glock 9 mm firing pin aperture shear mark individuality based on 1,632 different pistols by traditional pattern matching and IBIS pattern recognition. *J Forensic Sci.* 2016;61(1):170-6. <https://doi.org/10.1111/1556-4029.12940>.
43. Hamby J, Thorpe J. The examination, evaluation and identification of 9mm cartridge cases fired from 617 different Glock Model 17 & 19 semiautomatic pistols. *AFTE J.* 2009;41(4):310-24.
44. Stroman A. Empirically determined frequency of error in cartridge case examinations using a declared double-blind format. *AFTE J.* 2014;46(2):157-75.
45. McClarin D. Adding an objective component to routine casework: use of confocal microscopy for the analysis of 9 mm caliber bullets. *AFTE J.* 2015;47(3):161-70.
46. Song J. Proposed, "congruent matching cells (CMC)" method for ballistic identification and error rate estimation. *AFTE J.* 2015;47(3):177-85.
47. Keisler M, Hartman S, Kilmon A, Oberg M, Templeton M. Isolated pairs research study. *AFTE J.* 2018;50(1):56-8.
48. President's Council of Advisors on Science and Technology (PCAST). Report to the President - Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods. Washington, DC: President's Council of Advisors on Science and Technology; 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf. Accessed 2 Oct 2020.
49. Hamby J, Brundage D, Thorpe J. The identification of bullet fired from 10 consecutively rifled 9mm Luger Ruger pistol barrels: a research project involving 507 participants from 20 countries. *AFTE J.* 2009;41(2):99-110.
50. Smith J. Method of rifling by manufacturer. *AFTE J.* 2011;43(1):45-50.
51. Nichols R. Subclass characteristics: from origin to evaluation. *AFTE J.* 2018;50(2):68-88.
52. Calculate confidence limits for a sample proportion. <https://epitools.ausvet.com.au/ciproportion>. Accessed 9 Mar 2020.
53. Brown L, Cai T, Dasgupta A. Interval estimation for binomial proportion. *Stat Sci.* 2001;16(2):101-33.
54. AFTE Range of Conclusions. <https://afte.org/about-us/what-is-afte/afte-range-of-conclusions>. Accessed 22 May 2020.

How to cite this article: Smith JA. Beretta barrel fired bullet validation study. *J Forensic Sci.* 2020;00:1-10. <https://doi.org/10.1111/1556-4029.14604>

Couple Things

From: "Hunt, Ted (ODAG)" (b)(6)
To: "Donoghue, Richard (ODAG)" (b)(6)
Date: Tue, 17 Nov 2020 16:15:53 -0500

Rich,

I'm checking back in about two things:

1. The status of the request for DAG clearance of an OLC opinion on certain defined genetic genealogy legal issues that we need to address before a final DOJ GG policy can be released.
2. A few weeks ago i sent you a draft of a technical paper I wrote under the DOJ name addressing certain aspects of the PCAST Report for general release. It's been reviewed by the DOJ forensic working group that I chair. I'd like to send it out soon. Is any further review needed, or can I start the publication process?

Thanks,

Ted

Re: Couple Things

From: "Hunt, Ted (ODAG)" <(b)(6)>
To: "Donoghue, Richard (ODAG)" <(b)(6)>
Date: Wed, 18 Nov 2020 20:14:34 -0500

Thank you, Rich.

Sent from my iPhone

On Nov 18, 2020, at 5:33 PM, Donoghue, Richard (ODAG) <(b)(6)> wrote:

Ted,

With apologies for the delays, the approval for the OLC opinion request is attached. Furthermore, your PCAST response is good to go out.

Thanks for fighting the good fight on these challenging issues.

Rich

From: Hunt, Ted (ODAG) <(b)(6)>
Sent: Tue day, November 17, 2020 4:16 PM
To: Donoghue, Richard (ODAG) <(b)(6)>
Subject: Couple Things

Rich,

I'm checking back in about two things:

- 1) The status of the request for DAG clearance of an OLC opinion on certain defined genetic genealogy legal issues that we need to address before a final DOJ GG policy can be released.
- 2) A few weeks ago i sent you a draft of a technical paper I wrote under the DOJ name addressing certain aspects of the PCAST Report for general release. It's been reviewed by the DOJ forensic working group that I chair. I'd like to send it out soon. Is any further review needed, or can I start the publication process?

Thanks,

Ted

<Approval of DAG Memo Requesting OLC Opinion.pdf>

DAG Memo

From: "Hunt, Ted (ODAG)" (b)(6)
To: "Ellis, Corey F. (ODAG)" (b)(6) "Suero, Maya A. (ODAG)"
(b)(6)
Date: Wed, 20 Mar 2019 17:17:57 -0400
Attachments: Hunt Memo to the DAG.pdf (185.77 kB)

Corey and Maya

Attached is a memo for the DAG's review. He brought this up yesterday, so should be expecting it.

Thanks,

Ted

Senior Advisor on Forensic Science
Office of the Deputy Attorney General
United States Department of Justice
950 Pennsylvania Ave. NW
Washington, DC 20530

(b)(6)

RE: Advisory Committee on Evidence Rules

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "O'Callaghan, Edward C. (ODAG)" <(b) (6)>
Cc: "Hunt, Ted (ODAG)" <(b) (6)>, "Shapiro, Elizabeth (CIV)" <(b) (6)>
Date: Fri, 30 Nov 2018 15:41:50 -0500
Attachments: evidence meeting minutes.pdf (298.11 kB)

Ed – should you wish to see more, here are the minutes from the April 2018 Evidence meeting. Particularly noteworthy is the end of the discussion concerning Rule 801(d)(1)(A) at pp. 16-17:

However, following lengthy discussion by the Committee of potential amendments to Rules 807, 606, and 404(b) [detailed below], and after the lunch break, Rob Hur of the Department of Justice was recognized by the Chair. Mr. Hur stated that he was moved by the many good points made in opposition to the proposal to amend Rule 801(d)(1)(A), particularly those made by the Federal Public Defender. Having consulted with Betsy Shapiro and Andrew Goldsmith, Mr. Hur changed the Department of Justice vote on the proposed amendment from one in favor to one against, making the vote tally 5-4 *against* the proposed amendment, thus defeating it. Therefore, Rule 801(d)(1)(A) was not referred to the Standing Committee for release for public comment.

Consistent with our discussion earlier today, the (b)(5) per CIV
– Andrew

-----Original Appointment-----

From: Brown, Angela M (ODAG) **On Behalf Of** O'Callaghan, Edward C (ODAG)
Sent: Tuesday, November 13, 2018 1:23 PM
To: O'Callaghan, Edward C (ODAG); Goldsmith, Andrew (ODAG); Shapiro, Elizabeth (CIV)
Cc: Hunt, Ted (ODAG); Brown, Angela M. (ODAG)
Subject: Advisory Committee on Evidence Rules
When: Friday, November 30, 2018 11:00 AM-11:45 AM (UTC-05:00) Eastern Time (US & Canada).
Where: Room 4208

Advisory Committee on Evidence Rules
Minutes of the Meeting of April 26-27, 2018
Washington, D.C.

The Judicial Conference Advisory Committee on the Federal Rules of Evidence (the “Committee”) met on April 26-27, 2018 at the Thurgood Marshall Building in Washington, D.C.

The following members of the Committee were present:

Hon. Debra Ann Livingston, Chair
Hon. James P. Bassett
Hon. J. Thomas Marten
Hon. Shelly D. Dick
Hon. Thomas D. Schroeder
Daniel P. Collins, Esq.
Traci L. Lovitt, Esq.
A.J. Kramer, Esq., Federal Public Defender
Elizabeth J. Shapiro, Esq., Department of Justice

Also present were:

Hon. David G. Campbell, Chair of the Committee on Rules of Practice and Procedure
Hon. Jesse M. Furman, Liaison from the Committee on Rules of Practice and Procedure
Hon. Sara Lioi, Liaison from the Civil Rules Committee
Hon. James C. Dever III, Liaison from the Criminal Rules Committee
Robert K. Hur, Esq., United States Attorney for the District of Maryland
Dr. Joe S. Cecil, Esq.
Ted Hunt, Esq. (Department of Justice)
Andrew Goldsmith, Esq., (Department of Justice)
Professor Daniel J. Capra, Reporter to the Committee
Professor Daniel R. Coquillette, Reporter to the Standing Committee
Professor Catherine T. Struve, Assistant Reporter to the Standing Committee (by phone)
Professor Liesa L. Richter, Academic Consultant to the Committee
Dr. Timothy Lau, Esq., Federal Judicial Center
Rebecca A. Womeldorf, Esq., Secretary, Committee on Rules of Practice and Procedure
Bridget M. Healy, Esq., Attorney Advisor, Administrative Office of the Courts
Shelly Cox, Administrative Analyst, Committee on Rules of Practice and Procedure
Patrick Tighe, Esq., Rules Committee Law Clerk

I. Opening Business

Approval of Minutes

The Chair welcomed everyone to the meeting and solicited discussion of the minutes from the October 26, 2017 meeting of the Committee in Boston. A motion was made to approve the minutes, which was seconded and approved.

Standing Committee Meeting

The Chair reported on the Standing Committee meeting in January, 2018 during which she updated the Standing Committee concerning the projects and rules amendments being considered by the Evidence Advisory Committee. Judge Livingston noted that she received largely positive, albeit limited, feedback from the Standing Committee with respect to the projects being pursued by the Evidence Advisory Committee.

II. Symposium on Forensic Evidence, FRE 702, and *Daubert*

Judge Livingston then opened discussion of the first item on the agenda: the Committee's role in addressing challenges to forensic expert testimony, as well as more general problems under *Daubert* and Rule 702. Judge Livingston noted that this was the first opportunity the Advisory Committee had to discuss the vast array of information provided to the Committee at the fall symposium on expert forensic evidence and Rule 702, held at Boston College Law School. She further noted that the project began with recommendations from the President's Council of Advisors on Science and Technology ("PCAST") that the Advisory Committee draft a "Best Practices Manual" with respect to forensic evidence or alternatively prepare a new Committee note to Rule 702. Although no specific rule change was recommended, PCAST expressed interest in a revision to the detailed Committee note to FRE 702 to address special considerations associated with forensic evidence.

The Reporter made several observations about the PCAST recommendations. He noted that it is not statutorily permissible to revise a Committee note in the absence of any change to a rule. Although it might be possible that a relatively minor change to a rule would, after discussion, prove appropriate, the Committee has consistently followed the principle that it is not good rulemaking to amend a rule for the purpose of creating a note. In addition, there are problems with a Best Practices Manual emanating from the Advisory Committee. The Reporter noted that a Best Practices Manual for the authentication of electronic evidence was started under the auspices of the Committee, but ultimately had to be published under the names of the contributing authors because of concerns that a Best Practices Manual might be outside the Committee's rulemaking authority.

In light of these concerns, the Chair explained that the Advisory Committee would first discuss and consider the possibility for rule revisions that might assist courts and litigants in dealing with expert opinion evidence, particularly in the area of forensic feature comparison. Short of potential amendments to the Evidence Rules, the Committee could consider what role the Advisory Committee might play in the arena of expert forensic testimony.

The Chair thereafter recognized Dr. Joe Cecil, who had recently retired from the FJC and had served as the Liaison from the FJC to the Evidence Advisory Committee for many years, including in 2000 when the amendments to FRE 702 were enacted. Dr. Cecil is an author of the highly respected Reference Manual for Scientific Evidence relied upon by judges to better understand scientific evidence, and he contributed to the PCAST. The Chair explained that Dr. Cecil had been invited to share with the Committee how his work on scientific evidence might inform or assist in the Committee's inquiry into forensic expert testimony.

Mr. Cecil explained a bit of the background and focus of the Reference Manual for Scientific Evidence, noting that the first Manual was published in 1994 and that the most recent version came out in 2011 shortly after the National Academy of Sciences 2009 Report on forensic evidence. He noted that the Manual is now published in collaboration with the National Academy of Sciences and is extensively peer reviewed. He explained that the focus of the Manual is to give judges who may not have a science background the necessary scientific foundation to decide questions involving science in the courtroom. For example, the Manual includes chapters on statistics, toxicology, epidemiology, and forensic feature comparison. Dr. Cecil emphasized that the Manual is designed to impart scientific information, but is *not* designed to tell judges how to decide issues and cases. It is informative but not prescriptive. For those reasons, Dr. Cecil did not believe that the Reference Manual was a "substitute" for the Best Practices Manual envisioned by PCAST. Dr. Cecil stated that he was open to working with the Committee in the development of a Best Practices Manual should the Committee decide to sponsor such a project.

Judge Livingston inquired whether the FJC has education programs to further assist in addressing issues of forensic expert evidence. Dr. Lau remarked FJC currently does not sponsor many judicial programs on forensic evidence, but that programs could be developed if there is demand. He further noted that the European Union does have a Best Practices Manual on Forensic Evidence. The Reporter inquired of Dr. Cecil whether the FJC would be able to identify the scientists in the relevant fields that the Advisory Committee would need to consult in developing a Best Practices Manual. Dr. Cecil responded that the FJC was in contact with many noted scientists and could help the Committee in identifying those resources. He further noted that the National Academy of Sciences could also help identify experts. Judge Livingston inquired as to the timeline for the next edition of the Reference Manual. Dr. Cecil reported that no firm timeline exists, but that funds are currently being raised to support the publication of a new edition. The Reporter also inquired whether the Reference Manual would be able to resolve disputed issues identified in the PCAST report. Dr. Cecil stated that the Manual served to identify and explain such disputes, but does not provide resolution.

The Chair thereafter introduced a guest from the Department of Justice ("DOJ") who had been invited to the Committee meeting to explain the work being done by DOJ with respect to forensic investigation and testimony. Ted Hunt is the Senior Advisor on Forensic Evidence for DOJ. He began by stating that improving forensic investigation and evidence is a high priority for the Deputy Attorney General. He noted that his position as the Senior Advisor on Forensic Evidence was created last April and that a permanent working group on forensic evidence had been established to bring together all relevant stakeholders to improve and validate forensic

testing, and to provide guidelines for testimony by forensic experts. Mr. Hunt noted five key areas of focus:

1. Discontinuing statements by analysts and prosecutors expressing “a reasonable degree of scientific certainty” regarding findings. The Department directed prosecutors and analysts not to use this language in reporting results 18 months ago.
2. Establishing uniform terminology for examiners and analysts to employ in their reports and testimony to ensure that all terminology is scientifically based, appropriately qualified in scope, and not overstated. The first document on uniform terminology in latent print comparison was released in February of 2018 and additional directives for other disciplines will be forthcoming.
3. Monitoring expert forensic testimony for quality assurance to ensure that any mistakes are corrected immediately. This is a permanent program that evaluates testimony through real-time observation of testimonial presentations, as well as through transcript review. Feedback is promptly provided.
4. On-line posting of internal DOJ laboratory policies and procedures to enhance transparency. These documents are provided to defense counsel during discovery and also are being made publicly available, in order to provide greater insight and education into DOJ laboratory methodology, as well as to serve as a model for state crime labs.
5. Performing research and additional scientific study to strengthen the foundations of forensic science. The Department is conducting large-scale studies involving hundreds of examiners and thousands of forensic samples in a multi-year project in order to improve forensic methodologies.

Mr. Hunt concluded his remarks by emphasizing that each of the projects described was designed to enhance the reliability of forensics, to increase collaboration across federal and state laboratories, and to increase the capacity of forensic services.

The Reporter asked Mr. Hunt about who it is that performs the testimonial monitoring function that he described. Mr. Hunt explained that a peer of comparable qualifications does the monitoring and immediately critiques in-court testimony of an examiner to prevent exaggeration or overstatement of results and to avoid deviation from uniform language tailored to each field of forensic study. The Chair asked Mr. Hunt how an expert testifying about a forensic method that had not been validated through black box studies was permitted to express confidence while testifying according to the Department’s program. In response, Mr. Hunt described international standards of accreditation established for various forensic disciplines based upon extensive literature and hundreds of training hours that demonstrated the reliability of those methods, though without the more rigorous black box studies emphasized in the PCAST Report. The Reporter followed up, asking Mr. Hunt whether a ballistics expert could say a shell casing was “a match” for a particular weapon. Mr. Hunt stated that pre-trial rulings by the court would determine exactly what the expert could say, but that a ballistics examiner should be able to say that a shell casing was fired from a particular gun. The Reporter again queried whether that meant that examiners could testify to a “match” according to the Department protocol described by Mr. Hunt, to which he responded that it depends upon the discipline.

Dr. Cecil offered that the DOJ efforts to improve research and quality control were commendable, but that difficult issues remain concerning identification of a match between a forensic sample and an exemplar. According to Dr. Cecil, DOJ guidelines continue to permit an examiner to state that she can identify the source of a particular sample and testimony to that level of certainty conflicts with the consensus in the scientific community that there is inadequate foundation for that specific attribution. Dr. Cecil noted that other groups, like the European Union, require more temperate terminology, involving a “likelihood” of attribution, in order to prevent overstatement. Mr. Hunt responded that the Department’s published documents on particular disciplines, such as the ULTR on Latent Prints, would list approved terms of art for the particular discipline, but then require explanations of those terms and a description of limitations. According to Mr. Hunt, it is impossible to craft a single term that accurately captures conclusions across forensic disciplines, and explanation of terminology is far more important than the particular term used.

A member of the Committee asked Dr. Cecil whether the concern of the scientific community is the failure of examiners to explain limitations or uncertainty surrounding a particular forensic methodology. Dr. Cecil explained that scientists prefer to express findings in confidence intervals that more accurately represent the likelihood of a match rather than in conclusions about a match. He stated that the concern of the scientific community is that there is inadequate foundation to make a specific attribution to a particular defendant for many disciplines. Scientists would prefer more discussion of confidence intervals in the legal arena. Mr. Hunt noted that the Department’s Latent Print document makes limitations on findings very transparent and that this publicly available document is accessible to defense counsel for purposes of cross-examination.

Another Committee member then asked Mr. Hunt what the remedy would be if an examiner did overstate conclusions during his testimony. Mr. Hunt stated that there would be a duty to notify the parties immediately of any misstatement by a testifying expert.

The Chair thanked Dr. Cecil and Mr. Hunt for their helpful contributions and explained that one possible response to the issues surrounding forensic testimony could be a change to the Rules. The Reporter directed the Committee’s attention to a draft of a new Rule 707 on Forensic Evidence on page 50 of the agenda materials. He noted that the draft rule was not a proposal, but more of a thought experiment drafted for the Symposium for purposes of discussion. The Reporter noted difficulties surrounding a definition of “forensic evidence” in a rule. In addition, the draft Rule 707 would overlap, problematically, with existing Rule 702. For that reason, amending Rule 702 might be a better solution.

The Reporter stated that one idea for amending Rule 702 would be a new subsection prohibiting an expert from overstating results. That more limited amendment was also prepared for discussion at the Symposium and was received favorably by a number of the panelists. An alternative would be a positive statement, such as that experts must accurately report the strength of their findings. The Reporter suggested that the Committee might review a formal proposal for such a textual change at a subsequent meeting.

Judge Dever, the Liaison from the Criminal Rules Committee, reported that Criminal Rules is addressing some of the concerns surrounding forensic expert evidence with potential amendments involving criminal discovery. Judge Dever stated that a subcommittee had been appointed to determine whether the expert disclosure obligations under Criminal Rule 16 should be broadened along the lines of Civil Rule 26. He suggested that more robust advance disclosure to criminal defendants could aid them in testing expert testimony through *Daubert* motions and could also help in avoiding overstatement by providing a meaningful opportunity for expert cross-examination. Given the wide array of subjects about which experts are testifying, a broader criminal discovery provision could give defendants better access to information to challenge experts in all fields. Professor Coquillette noted the importance of having the Criminal and Evidence Committees work together on the issue of expert testimony in criminal cases and also commended the Department of Justice for its efforts. Judge Dever noted that the Criminal Rules Committee was gathering information from all constituencies, the Department of Justice, the Federal Public Defender, as well as the scientific community to get a broad perspective on the issue of criminal discovery of expert opinion evidence.

The Chair thanked Judge Dever for his report and noted that it was very helpful to coordinate with the Criminal Rules Committee in thinking about potential amendments to the Evidence Rules. Of the possible amendments, the Chair noted that one preventing overstatement was one that seemed most plausible. She further noted the challenge presented by the disconnect between civil and criminal cases with respect to expert testimony that was highlighted at the Boston College symposium. Civil lawyers lamented the vast resources being needlessly consumed by *Daubert* challenges, while criminal lawyers expressed concern about the lack of attention being given to forensic expert testimony in criminal trials. The divergent experiences in civil and criminal cases present another challenge for rulemakers. She noted that a Best Practices Manual might be an alternative to rulemaking to address these matters.

The Reporter explained that it would not be possible to write a rule prohibiting overstatement by testifying experts on the criminal side only, because that would imply that overstatement is acceptable in civil cases, which of course it is not. He then provided an update on the case law regarding FRE 702 and forensic expert testimony and directed the Committee's attention to the case digest in the agenda materials. A review of recent cases revealed that courts are relying on precedent to support the admissibility of many forensic methods without conducting independent analysis of *Daubert* factors. The cases also showed significant overstatement by forensic experts, including testimony that a sample identification was "100% accurate." A Committee member asked what conclusion a testifying expert could make if testifying to a "reasonable degree of scientific certainty" constituted overstatement. Mr. Hunt responded that with sufficient foundation, an expert should be able to opine that a sample comes from a particular source, but stated that the Department of Justice did not believe that it was necessary to testify to a "reasonable degree of scientific certainty." Mr. Hunt stated that no "magic word" would be adequate in all cases and that explanation by the examiner of the meaning and limitations of her findings was more important.

The Reporter expressed concerns that the findings of both the National Academy of Sciences and of PCAST have been largely ignored by the courts in the recent opinions and that a

Best Practices Manual (that cannot emanate directly from the Evidence Advisory Committee) might also be ignored.

Judge Dever then asked Mr. Hunt whether the Department of Justice was working to monitor testimony by state examiners to the extent that state experts testify in federal cases. Mr. Hunt responded that federal prosecutors governed by Department policies would not elicit improper testimony from state examiners, and further noted that one of the goals of publishing Department of Justice best practices was to provide a model for state laboratories as well.

The Chair then noted that it might be advisable for the Evidence Advisory Committee to appoint a small subcommittee to do intense reading and study regarding the possible role of the Committee in addressing concerns with forensic evidence. She stated that she and the Reporter currently felt that an amendment to Rule 702 preventing overstatement of findings appeared to be the most promising possibility and that a potential amendment distinguishing between scientific and other types of expert opinion testimony appeared less viable.

Mr. Hur then thanked the Reporter for his detailed case digest and stated that the cases are the data that the Committee should be considering. He opined that the courts are grappling carefully and thoughtfully with *Daubert* issues and limiting expert testimony where necessary. He seconded Mr. Hunt's assertion that the Department of Justice was already working to prevent overstatement of expert conclusions. The Reporter emphasized the excessive reliance on precedent by the federal courts in place of detailed consideration of other *Daubert* factors, and the overstatement found in the cases. Mr. Hur noted the longstanding acceptance of certain scientific methods like latent fingerprint analysis. While he acknowledged that courts could start from the ground up in a *Daubert* analysis of such methodologies, he stated that the reliance on the longstanding precedent reaches the same result – the proper admissibility of such testimony. Mr. Hur further opined that the PCAST report is having an impact, noting that defense counsel have cited to it. He further emphasized that the PCAST report looked favorably on the black box studies conducted by the FBI in connection with fingerprint evidence. Mr. Hur stated that the courts need more time to absorb the PCAST report and for its findings to filter into *Daubert* analysis.

The Reporter then turned the Committee's attention to another concern about the application of Rule 702 raised by two members of the public in a law review article. Specifically, the article found that some federal courts treat the sufficiency of an expert's basis, and the application of the expert's methods, as questions of weight for the jury --- when in fact these matters are both questions of admissibility under Rule 702, as amended in 2000. The Reporter explained that the subdivisions of Rule 702 set forth admissibility requirements that a trial judge must find to be satisfied by a preponderance of the evidence *before* allowing the expert to testify before the jury. Therefore, federal courts that are treating these foundational requirements as matters of weight that may be given to a jury are indeed wrong. That said, the Reporter noted that FRE 104(a) clearly applies to the admissibility requirements of FRE 702, and that crafting an amendment that essentially tells federal courts to "apply the rule" may be challenging.

One member of the Committee remarked that the federal cases treating the requirements of FRE 702 as matters of weight are very troubling. Essentially, it is as if some courts are saying

that FRE 702 doesn't apply in their circuit. The Committee member suggested that it might be important to amend Rule 702 to prevent it from being ignored. Another Committee member also reported being taken aback by the federal courts blatantly ignoring Rule 702. That Committee member wondered whether a rule revision (that could also be ignored) would be the most fruitful solution or whether judicial education might be a better solution to the problem.

A Committee member reiterated the sharp divide between expert discovery in civil and criminal cases, noting that the adversarial process works out many issues with expert testimony on the civil side and that the failure of the adversarial process on the criminal side is placing greater burdens on trial judges to police the use of forensic experts. Judge Dever noted that the Department of Justice was training on this issue in an effort to get more information about testifying experts to defense counsel earlier in the process to allow for more adversarial testing. Andrew Goldsmith, the Criminal Discovery Coordinator in the Deputy Attorney General's office noted that a January, 2017 memo from Sally Yates on expert discovery was now part of the U.S. Attorney's Manual and that all federal prosecutors are receiving training on early disclosure. He opined that it was important for the Evidence Advisory Committee to collaborate with the Criminal Rules Committee and suggested that a rule change was unnecessary because prosecutors are giving defense counsel the information they need with respect to testifying experts. Professor Coquillette noted that issues regarding expert testimony are well resolved through adversarial testing in civil cases, but that has not historically been the case in criminal trials. He remarked that he was delighted to learn that the Department of Justice was working to rectify the imbalance.

Judge Livingston closed the discussion of the fall symposium and of Rule 702 and *Daubert*. She noted the sense of complexity of the issues raised and the need for further study by the Committee. She stated that proposals for rule amendments regarding overstatement of conclusions, and Rule 702 admissibility requirements, would be considered at a future meeting.

III. Proposed Amendment to Rule 807

The Reporter opened discussion of the proposed amendment to Rule 807 that was released for public comment. The public comment period closed on February 15, 2018. In order to facilitate discussion of revisions raised by the public comment and by the Standing Committee, the Reporter directed the Committee's attention to a supplementary memorandum prepared in advance of the meeting.

The Chair noted that the memo was designed to provide a draft of the amendment to Rule 807 that would make it easier to resolve issues raised during the public comment period. The Chair and the Reporter proceeded to walk the Committee through the following revisions to the proposed amendment as released for public comment:

- The language regarding the hearsay exceptions in Rules 803 and 804 was moved from an admissibility requirement back into the prefatory section of the rule. Both the American Association for Justice and Judge Furman recommended this change, noting concerns that a trial judge might find it necessary to test proffered hearsay against every exception

in Rules 803 and 804 before applying Rule 807 – which was never the intent of the proposal.

- In response to concerns that the term “substance” of the statement used in the amended notice provision could prove vague, a “See” cite to Rule 103(a)(2) governing offers of proof (in which the “substance” of the proffered evidence must be presented) was added to the Advisory Committee note.
- A reference to the use of corroborating evidence to determine the “accuracy” of a hearsay statement in the Advisory Committee note was replaced with language requiring the use of corroborating evidence to determine “whether a statement should be admissible under this exception.”
- In addition, language requiring a finding of “sufficient guarantees of trustworthiness” was retained over a requirement that a trial judge find the hearsay “trustworthy” to avoid any reading of the amendment that would make Rule 807 narrower and more difficult to satisfy.
- The language in the Rule text regarding Rules 803 and 804 was changed from “not specifically covered by a hearsay exception in Rule 803 or 804” to “not admissible under a hearsay exception in Rule 803 or 804” to reflect the “near-miss” interpretation given to the existing rule by the majority of courts. The near-miss issue was added to the Committee note as well.
- The word “limit” used in the proposed Committee note was changed to “guide” to better reflect the intent of the sufficient guarantees of trustworthiness requirement in informing the trial court’s exercise of discretion.
- A reference to Rule 104(a) was added to the Note, in response to a suggestion from a member of the Standing Committee.
- A reference to the Confrontation Clause was added to the Note, in response to a suggestion from a member of the Standing Committee.

The Committee discussed the revised draft of the proposed amendment to Rule 807 and the accompanying Committee note. Judge Furman suggested replacing omitted language in the Committee note clarifying that a trial judge need not make a finding that the hearsay is not admissible under any Rule 803 or 804 exception before employing the residual exception. The language was removed from the Committee note when the Rule 803/804 language was eliminated as an admissibility requirement and moved back into the preface. Judge Furman expressed concern that a trial judge might still think that such findings were necessary and advocated retaining the clarifying language. He also proposed deleting language in the note that rule 807 should be “invoked only when necessary” as unduly limiting. Committee members agreed with these suggestions.

Another Committee member argued that if the intent of Rule 807 is not to allow parties to use the residual exception unless they need it, then inadmissibility under Rules 803 and 804 should be required. The Chair responded that making it an admissibility requirement would risk forcing trial judges to make a threshold examination of every Rule 803 and 804 hearsay exception before applying Rule 807 – which was not intended, and which would unnecessarily constrain the use of the rule. Judge Campbell raised the concern that the Committee Note would say that a party could *not* use Rule 807 to admit hearsay admissible through Rules 803 and 804 (suggesting that a party could not proceed directly to Rule 807 to admit hearsay) when nothing in the text of Rule 807 would prevent a party from doing just that. The Reporter noted that case law interpreting existing Rule 807 does prohibit parties from proceeding directly to Rule 807. Judge Campbell proposed altering the Committee note to provide that nothing in the amendment is intended to “alter the case law holding that parties may not proceed directly to the residual exception, without considering the admissibility of the hearsay under Rules 803 and 804.” Committee members agreed with that suggestion. Another Committee member noted that Rule 807 is always the last exception argued by parties and the Reporter highlighted litigants’ natural incentives to start with the Rule 803 and 804 hearsay exceptions because Rule 807 is ordinarily more difficult to satisfy.

The Reporter then explained that revised language in the Committee note had been added to deal with the “near-miss” precedent and the new rule text stating that hearsay not “admissible” through a Rule 803 or 804 exception (as opposed to “not specifically covered by” an exception) could be admissible under Rule 807. He noted that the language was designed to suggest that courts employing a near-miss analysis of hearsay offered through Rule 807 should think about how nearly a proffered hearsay statement misses a standard exception, as well as about the importance of the requirement of a Rule 803 or 804 exception that the hearsay statement fails to satisfy. One Committee member expressed concern that the near-miss language in the Committee note might lead some to believe that near-miss analysis was a substitute for considering sufficient guarantees of trustworthiness. The proposed Committee note was revised to clarify that a near-miss analysis may be part of an inquiry into guarantees of trustworthiness, but is not a replacement for that inquiry. Judge Furman also expressed concern that litigants and judges might not appreciate which requirements of the Rule 803 and 804 hearsay exceptions are the “important ones.” The reference to the importance of the admissibility requirements was removed from the Committee note to accommodate that concern.

The Reporter next explained that a member of the Standing Committee suggested adding a sentence to the Committee note clarifying that testimonial hearsay satisfying the requirements of Rule 807 would nonetheless be excluded under the Sixth Amendment Confrontation Clause in a criminal case. Given that the Constitution prohibits the admission of uncross-examined testimonial hearsay through *any* of the hearsay exceptions, the Chair queried why this reference to the Sixth Amendment was needed in the note to Rule 807 when the notes to the other hearsay exceptions contain no such caveat. The Reporter responded that the categorical exceptions generally avoid the admissibility of testimonial hearsay, because the admissibility requirements require a showing that would be inconsistent with primary motivation for use in a criminal prosecution. For example, a record that satisfies the requirements of the business records exception in Rule 803(6) would, by definition, not be testimonial, because it would have to be made in the course of regularly conducted activity. And a statement admissible as an excited

utterance will not be testimonial because it must be made under the influence of a startling event, which is inconsistent with preparing a statement for a criminal prosecution. In contrast, Rule 807 presents the greatest risk of admitting testimonial hearsay due to its “sufficient guarantees of trustworthiness” standard. So there is some justification for adding the language about the right to confrontation in the Committee Note. No further objections were made to its inclusion.

The Committee then discussed changes to the notice provision and the Committee Note regarding notice. The Reporter noted that the “See” cite to Rule 103(a)(2) in the Committee Note was designed to inform the court’s inquiry into whether the “substance” of the statement had been disclosed. He also noted that language in the note regarding case law under the former requirement that “particulars” be disclosed had been removed as unhelpful. The Reporter also explained that conflicting statements about the rigor or flexibility of the good cause exception to the notice requirement had been removed. The suggestions were a provision that good cause should not be easily found (provided by a Standing Committee member) and a provision that good cause should be easily found as to criminal defendants (provided by the National Association of Criminal Defense Lawyers). The Committee decided to leave the interpretation of good cause to trial judges and the extensive pre-existing case law from courts that had applied a good cause exception even though it was not specifically provided for in the rule.

At the conclusion of the Committee’s discussion, the Chair explained that the Reporter would provide a clean copy of the revised Rule 807 and accompanying Committee note reflecting all changes made during the discussion and that the Committee would vote on sending the proposed amendment to the Standing Committee, with the recommendation that it be released for public comment, on the following day. Thereafter, the Committee adjourned.

The Committee meeting resumed Friday, April 27

Mr. Hur served as the representative of the Department of Justice, as Ms. Shapiro could not be present.

IV. Rule 702 and Rule 104(a) Admissibility Requirements (Revisited)

Judge Livingston explained that the Committee would take Rule 807 back up later in the day after all Committee members had a chance to review the latest version of the proposed amendment prepared by the Reporter. She then asked the Reporter to share an idea for resolving the misapplication of Rule 702 by federal courts who are treating the Rule’s admissibility requirements as matters of weight. The Reporter suggested that the preface to Rule 702 that precedes the admissibility requirements could be modified to address this concern by stating that a qualified expert may testify if “the court finds the following by a preponderance of the evidence.” The Reporter explained that adding this language would emphasize that the Rule 702 requirements are admissibility requirements governed by Rule 104(a). He explained that a Committee Note could accompany such a revision, explaining that it was a needed clarification to address confusion in the courts. While the new language would basically state the existing rule --- that Rule 104(a) applies to the Rule 702 requirements --- it has the benefit of making the principle explicit, thus hard to ignore. And it might be justified in light of the disregard of the admissibility requirements by many courts.

Judge Campbell then opened the discussion with an example from a hypothetical trial in which an expert testifies in a *Daubert* hearing that he rejects 7 of 10 seminal studies in an area and is relying on the 2 or 3 minority studies in the field as the basis for his opinion. Judge Campbell queried, if the judge is not persuaded that the three minority studies are reliable and sufficient, but the jury might be, does the judge exclude? The Reporter responded that the trial judge must make a finding by a preponderance of the evidence on the admissibility requirements before allowing the expert to testify, and that it would be error to permit the testimony if the judge is not satisfied that the expert's basis is sufficient, as would be the case in Judge Campbell's hypothetical. Another Committee member stated that the question is whether Rule 702 works under a Rule 104(b) analysis, and the Reporter responded that this was indeed the issue that some courts were struggling with, but that the admissibility requirements in Rule 702 are clearly governed by Rule 104(a) --- as also stated in *Daubert* itself. The Reporter then asked whether the Committee members would be interested in reviewing a draft with revised prefatory language requiring a finding of each of the Rule 702 requirements by a preponderance of the evidence. Committee members expressed interest in reviewing such a draft and the Chair suggested that such a proposal might be part of the broader conversation the Committee would continue to have about its role in helping trial judges apply Rule 702.

V. Prior Inconsistent Statements: Possible Amendment to Rule 801(d)(1)(A)

Judge Livingston next opened the discussion of a potential amendment to Rule 801(d)(1)(A) that would allow for substantive admissibility of prior inconsistent statements of witnesses that were recorded audio-visually and available for presentation at trial. She acknowledged that the Committee had been considering the proposal for a long time. She traced the history of Rule 801(d)(1)(A), noting that the original Advisory Committee had favored a wide open approach allowing substantive admissibility of all prior inconsistent statements by testifying witnesses --- an approach that is now employed in a number of states, including California and Wisconsin. She noted that Congress pushed back on this proposal, expressing concern that a criminal defendant might be convicted solely on the basis of out of court statements of a witness who did not implicate the defendant at trial. This concern resulted in the compromise rule embodied in existing Rule 801(d)(1)(A) requiring prior inconsistent statements to be made under oath and in a prior proceeding if they are to be used substantively.

The Chair noted that this Advisory Committee began reviewing prior inconsistent statements due to concern that the limiting instructions provided to jurors when such statements are admitted for impeachment purposes only are difficult to comprehend and follow. In addition, the Committee noted Wigmore's opinion that cross-examination is the greatest engine for the discovery of truth in exploring the possibility of broader admissibility of hearsay statements made by testifying witnesses. Some expansion of the admissibility of prior inconsistent statements was also thought to be consistent with the basic thrust of the Federal Rules of Evidence to make more information admissible and available to the fact-finder. With the caveat that evidence rulemaking should focus on the process of deriving the truth at trial, some value was also seen in the likelihood that a rule allowing substantive admissibility of audio-visually recorded statements would encourage more recording and greater documentation of witness

statements. On the other hand, concerns had been expressed about the reliability of prior inconsistent statements and the ways in which the oath and the grand jury process contribute to reliability. Other potential downsides to an amendment could be added litigation costs needed to determine whether statements were recorded “audio-visually” or were made “off camera.” And questions had arisen about the impact of the amendment at a time when recording technology was exploding to include dash-cam and body-cam footage, as well as cellphone and social media recordings. There were also lingering concerns over the impact on summary judgment practice in civil cases. The Chair noted that every straw vote taken on the proposal in the Committee resulted in 2/3 of the Committee in favor of exploring the amendment and 1/3 opposing it.

After this introduction, the Reporter noted that the Department of Justice had proposed allowing substantive admissibility of prior inconsistent statements acknowledged by a witness at trial, in addition to audio-visual witness statements. Committee members inquired about the interaction between the audio-visual and acknowledgement proposals. The Chair explained that the Department’s proposal would be more liberal because it would allow substantive admissibility of any prior inconsistent a witness would acknowledge while on the stand – whether recorded or not. Judge Campbell asked whether case law had developed over how a witness “acknowledges” a prior statement. The Reporter noted that there was case law in jurisdictions with an acknowledgement rule and that the acknowledgement provision had sometimes resulted in problematic inquiries at trial, but that this was not an inevitable outcome.

Dr. Lau noted that technologies making it relatively easy to create fake video content were proliferating and that the Committee should consider that falsifying video material might become extremely easy 5-10 years from now. The Reporter responded that if this was a problem, then it was a problem for *all* electronic evidence, not just the narrow band of audiovisual statements that would be admissible under the amendment. The Federal Public Defender noted that defendants and witnesses already deny making statements that appear on video and that experts are employed to determine whether a defendant actually made a statement reflected in a recording.

The Chair asked Dr. Lau to report on the survey performed by the Federal Judicial Center on the proposed admissibility of audio-visual inconsistent witness statements. Dr. Lau noted that federal judges seemed to be split along lines similar to those in the Committee, with little appetite for the adoption of wide-open substantive admissibility of prior inconsistent statements and some support for a compromise approach to expanding admissibility. Judges expressed few concerns about expanded use of prior inconsistent statements in civil cases. In criminal cases, judges reported encountering oral prior inconsistent statements more frequently than they encounter audio-visual statements. Judge Livingston noted the bottom line in the survey that 58% of judges supported or strongly supported the proposal, while 29% opposed or strongly opposed it.

The Reporter thanked the FJC for the survey and the report and noted appreciation for feedback received from the American Association of Justice (“AAJ”), the National Association of Criminal Defense Lawyers (“NACDL”), and the Innocence Project on the proposal as well. He noted that the feedback from AAJ was largely favorable. The AAJ suggested adding a reference to future recording technologies in the Committee note. The Innocence Project suggested a pilot project to further explore the proposal in action due to two primary concerns: 1)

the possibility that a recorded statement may be the last in a long series of statements taken from the witness that may not reflect all of what the witness has said and 2) the concern that a defendant could be convicted solely on the basis of a prior inconsistent statement. The Reporter first noted that it would be wonderful to be able to conduct million dollar pilot projects in connection with rulemaking efforts, but that no Committee had ever done such a project prior to rulemaking and that it would be impossible. He also responded to the substantive concerns raised by the Innocence Project. He noted that a Federal Rule of Evidence could not mandate the recording of all of a witness's statements because that would exceed the Advisory Committee's statutory mandate. He explained that an evidence rule might condition admissibility of one recorded statement on the availability of all other statements in recorded form to the opponent, but questioned whether that would be advisable. With respect to the concern that a defendant could be convicted on the basis of a prior inconsistent statement alone, the Reporter reiterated that Rule 801(d)(1)(A) makes statements admissible for their truth, but does *not* deal with the sufficiency of the evidence to convict. He noted that Congress rejected the same objection to Rule 801(d)(1)(C) dealing with prior statements of identification and that a Committee note could clarify that the amendment does not speak to sufficiency.

Judge Furman noted that the issue of admissibility is intertwined with sufficiency because a prior inconsistent statement that could not be used to get a case to the jury under the existing rule could support submission to the jury under the proposal. He queried whether the Committee has solicited feedback from the defense bar in states where there is wide-open substantive admissibility of prior inconsistent statements. The Reporter responded that the Committee had received such feedback and described research by Professor Dan Blinka into the practice in Wisconsin that solicited input from all constituencies, the defense bar included. That report suggested that there is very little controversy over substantive admissibility of prior inconsistent statements in that jurisdiction. The Reporter also obtained input from noted Evidence expert Professor Ed Imwinkelried, who reported little activity in the California cases concerning the substantive admissibility of prior inconsistent statements in California. The Chair stated that it is not surprising that there is little controversy over the admissibility of prior inconsistent statements in Wisconsin and California because the wide-open rule that makes all such statements substantively admissible is straightforward. She expressed concern, however, that a compromise position that allows only audio-visual or acknowledged prior inconsistent statements could generate significant litigation over the scope of those limitations.

Another Committee member reminded the Committee of the symposium at Pepperdine in 2016 in which California prosecutors talked about the impact of substantive admissibility of prior inconsistent statements in obtaining plea agreements in domestic violence cases, and in proving up gang-related prosecutions, where witnesses often recant. He noted the report that defendants would accept a plea knowing that a prosecution could proceed even without the cooperation of the victim. The Chair noted that one of the concerns of the Innocence Project is that innocent defendants might plead guilty if witness statements taken in the aftermath of an incident, that have since been recanted, can form the basis of a prosecution. The Federal Public Defender also noted situations in which a domestic partner calls police out of anger at a partner and recants later because there was no abuse. He explained that there are times when the initial report is not accurate, even in the domestic violence context, and that the proposal would allow substantive use of these recanted early reports. He also reiterated the concerns of the Innocence

Project about a series of interviews that lead up to the final audio-visual statement and the inability of the jury to view the entire back and forth that created the prior inconsistent statement. Finally, he expressed concern that the government might claim that a prior inconsistent statement was substantively admissible under the proposed rule even if the *defense* sought to offer the statement only for impeachment purposes. The Reporter noted that an Advisory Committee note had been included to prevent that possibility. The Federal Public Defender further expressed concern about unreliable body-cam or cell phone recordings, noting that defense lawyers could record witnesses exonerating defendants and substantively admit those statements if the witness shows up and testifies favorably for the prosecution. He suggested that the proposal could create abuses and litigation on both sides of criminal cases.

Another Committee member noted that any prior inconsistent statement may already be used to impeach a testifying witness and that juries don't understand the limiting instruction accompanying such statements. This Committee member suggested that the proposal would be an improvement because it would impose more rigor with respect to the prior inconsistent statements admitted substantively than is currently required of prior inconsistent statements already allowed to impeach. Judge Lioi remarked that it does matter a great deal in criminal cases if the prior inconsistencies are allowed fuller use because substantive admissibility may be enough to defeat a defendant's otherwise valid Rule 29 motion for acquittal. The Chair also noted potential impact on summary judgment practice in civil cases if plaintiffs produce audio-visual statements that are inconsistent with a witness's deposition testimony. Judge Campbell noted that such a recorded statement may allow a civil case to go to trial under the proposal where summary judgment could be granted under the existing rule. The Reporter noted that if the recorded statement were a sham designed to defeat summary judgment, existing case law would permit a judge to disregard the statement even after an amendment. He further queried whether an audio-visually recorded statement by a witness expected to testify at trial that supported the plaintiff's case shouldn't mean that the case *should* proceed to trial.

Another Committee member questioned the absence of an oath requirement for statements that would be admissible under the proposal, indicating that the statements would lack the gravity of the statements admissible under existing Rule 801(d)(1)(A). The Reporter noted that the trial cross-examination before the jury required by the Rule was designed to reveal any weaknesses in the statement. Another Committee member remarked that the effect on Rule 29 practice in criminal cases should drive the result on the proposal, especially in light of evidence suggesting that jurors do not follow instructions with respect to prior inconsistent statements offered only for impeachment once they get a case. This Committee member suggested that audio-visually recorded statements of a testifying witness who is subject to cross-examination at trial -- that the jury can view for itself -- might be worthy of substantive effect and justifiably affect Rule 29 practice. The Committee member expressed some uncertainty regarding the Department of Justice proposal to include acknowledged witness statements in an amendment. The Reporter suggested that the Department's acknowledgement proposal should be included in the rule, if it were released for public comment, in brackets to signal that the Committee had not endorsed the acknowledgement option, but was seeking input from the public concerning it. He noted that this was done with the selective waiver provision of Rule 502 that did not ultimately find its way into the rule as enacted.

Another Committee member asked whether there is data suggesting that jurors do not understand limiting instructions regarding prior inconsistent statements offered for impeachment only. The Reporter noted that there was such data, involving mock juries, as well as judicial experience. The Committee member suggested that jurors do understand when instructed clearly. Another Committee member expressed concern about the voluminous dockets of the federal trial courts and the possibility that the proposed rule could increase the volume of cases requiring evidentiary hearings or trial. The Committee member noted the high volume of prisoner cases that could be impacted by an amended rule. The Reporter suggested that recordings submitted by plaintiffs in prisoner litigation would reflect anticipated testimony at a new trial that might necessitate evidentiary hearings, even without Rule 801(d)(1)(A).

The Chair again expressed skepticism about the proposal, noting concerns about Rule 29 practice in criminal cases and summary judgment practice in civil cases, concerns about plea bargaining impact and increased litigation costs surrounding the Rule. Although she doubted whether a change was worth the candle, she noted that social science has shown that jurors do not understand limiting instructions and noted the results of the Federal Judicial Center survey revealing that the majority of trial judges favored the change. The Chair noted that the Committee could send it out for public comment or table the idea for two years. Another Committee member queried what the standard for releasing a proposal for public comment should be. Judge Campbell noted that there are many potential standards, but that the consensus on the Standing Committee was that the public comment process should not be used as a research tool. On the other hand, if the Advisory Committee thinks the Rule is probably a good idea depending upon what public comment reveals, that is a sound basis for forwarding a proposal. The Reporter noted that the Rule 801(d)(1)(A) proposal certainly had not been rushed to public comment given several years of research, an FJC survey, two symposia, and Committee consideration at six consecutive meetings. Professor Coquillette noted that the risk of sending something forward to the Standing Committee improvidently was a loss of credibility for the Advisory Committee. The Reporter observed that negative public comment has been a catalyst for effective rule changes; in 2006 a proposal to amend Rule 408 to allow civil settlements to be admissible in criminal cases was released at the urging of the Department of Justice. The Reporter noted that very negative commentary fostered a compromise rule, which is now in effect. The Chair opined that tabling the proposal would provide the Committee more time to see how body and dash cameras, as well as cell phone recordings affect trials in the future.

The Reporter explained that the question for the Committee was whether to send the proposal forward to the Standing Committee to be released for public comment or to remove it from the Committee's agenda. A Committee member made a motion to refer the proposed amendment to the Standing Committee with the acknowledgement provision included in brackets for release for public comment. The Committee voted 5-4 *in favor* of sending the proposed amendment to the Standing Committee. The Committee then proceeded through the proposed Committee note to determine which portions of that note would advance with the proposed rule, and reached agreement on a Committee Note.

However, following lengthy discussion by the Committee of potential amendments to Rules 807, 606, and 404(b) [detailed below], and after the lunch break, Rob Hur of the Department of Justice was recognized by the Chair. Mr. Hur stated that he was moved by the many good points

made in opposition to the proposal to amend Rule 801(d)(1)(A), particularly those made by the Federal Public Defender. Having consulted with Betsy Shapiro and Andrew Goldsmith, Mr. Hur changed the Department of Justice vote on the proposed amendment from one in favor to one against, making the vote tally 5-4 *against* the proposed amendment, thus defeating it. Therefore, Rule 801(d)(1)(A) was not referred to the Standing Committee for release for public comment.

VI. Rule 807 Approved

After the Committee reviewed all revisions to the proposed amendment to Rule 807, it was unanimously approved for transmission to the Standing Committee, with the recommendation that it be sent to the Judicial Conference for approval.

The text and Note of the Rule, a GAP report, and a summary of public comment, are attached to these Minutes.

VII. Rule 606(b) and *Pena-Rodriguez*

The Chair next raised the Rule 606(b) ban on juror testimony about deliberations, and the impact of the Supreme Court's 2017 decision in *Pena-Rodriguez v. Colorado*. The Court in *Pena-Rodriguez* held that Rule 606(b) could not be applied to bar testimony of racist statements about the defendant made in juror deliberations --- such a bar violated the defendant's Sixth Amendment right to a fair trial. The Chair noted that the Committee had discussed three potential amendments to Rule 606(b) to bring the rule text in line with *Pena-Rodriguez* at its spring 2017 meeting, and had tabled the issue after discussion. Rule 606(b) was back on the Committee's agenda again to consider the need for an amendment to reflect the holding. The Chair explained that if the Committee decided not to take action on Rule 606(b) at this meeting, the topic would be tabled for at least a year to observe the case law developing in the wake of *Pena-Rodriguez*.

The Reporter directed the Committee's attention to a digest of federal cases interpreting *Pena-Rodriguez*, and observed that courts have declined to expand the exception to the no-impeachment rule beyond that holding --- which was limited to statements of racial bias toward the defendant in jury deliberations. He then briefly outlined the potential amendments previously considered by the Committee, including an amendment that would expand an exception beyond that required by *Pena-Rodriguez*, one that would seek to codify the racial animus exception from *Pena-Rodriguez* narrowly in rule text, and a generic amendment that would create an exception to the no-impeachment rule for evidence required by the Constitution. The Committee previously rejected both the expansive and narrowly-tailored potential amendments as problematic, and at the meeting it focused on the more generic constitutional exception in the rule that would flag the *Pena-Rodriguez* issue for litigators consulting only rule text.

Two possibilities have been considered. First, an amendment that makes an exception to the no-impeachment rule "when excluding the testimony would violate a party's constitutional rights." This generic constitutional exception would be modeled upon the one that currently exists in Rule 412(b)(1)(c). Due to concern in the Committee at the spring 2017 meeting that a generic constitutional exception in Rule 606(b) could be read to expand upon *Pena-Rodriguez* and to permit post-verdict juror testimony in any case where a defendant claims violation of a

“constitutional right” by the jury, a Committee member suggested using the restrictive language of the AEDPA in a Rule 606(b) amendment to avoid such an expansive reading. Such an amendment would allow juror testimony about deliberations when “excluding the testimony would violate clearly established constitutional law as determined by the Supreme Court of the United States.” This proposal was suggested as a way to send up a red flag or at least a yellow light for courts considering using Rule 606(b) to expand beyond the holding in *Pena-Rodriguez*. The Reporter explained that the use of the AEDPA language would be problematic due to its substantive restriction on lower courts and suggested that a generic constitutional exception like the one in Rule 412 was a better solution for the Committee to consider. The Chair and the Committee agreed that the AEDPA alternative would not work, and proceeded to reconsider the generic constitutional exception.

The Reporter also brought to the attention of the Committee a law review note to be published in the *Columbia Law Review* on *Pena-Rodriguez* that chronicled the Advisory Committee’s inaction on Rule 606(b). The note advocated expansion of the *Pena-Rodriguez* exception to the no-impeachment rule beyond racist statements and favored a general constitutional exception in Rule 606(b) that would accommodate such future expansions. The Chair reiterated that the goal of the Committee was to raise the *Pena-Rodriguez* issue for the trial lawyer consulting only the text of evidence rules, without suggesting expansion.

Judge Campbell expressed concern that even a generic constitutional exception would invite lawyers to seek expansion of the *Pena-Rodriguez* holding. He posited a case in which a defendant claims that the jury violated his constitutional rights and points to a constitutional exception to Rule 606(b) to show that the court must hear juror testimony. Judge Campbell suggested that the lack of an exception in Rule 606(b) currently helps courts hold the line on *Pena-Rodriguez* because courts can point to the prohibition in the Rule as support for the idea that no other exceptions exist. If the Committee removes that constraint, he suggested that courts might feel compelled to expand to create exceptions to Rule 606(b) for other constitutional violations. The Reporter noted that the Committee note accompanying an amendment would explain that no expansion was intended. The Reporter also reiterated that courts are finding that *Pena-Rodriguez* did not create constitutional rights outside the narrow circumstance it recognized, meaning there is no other constitutional right to introduce post-verdict juror testimony.

Judge Furman noted that there is a recognized constitutional right not to have the jury draw an adverse inference from a defendant’s silence. If a defendant claims that right was violated in the jury room, Judge Furman queried why an amended Rule 606(b) wouldn’t also allow juror testimony on that point. The Reporter responded that courts had already rejected such arguments after *Pena-Rodriguez* and that nothing in any Evidence Rule could determine substantive constitutionality.

A Committee member suggested that Judges Campbell and Furman made compelling points and that it would be difficult for a court to refuse to take juror testimony about other constitutional violations with an amended Rule 606(b) containing a generic constitutional exception. The Committee member stated that the proposal to amend Rule 606(b) was rightly tabled by the Committee in the spring of 2017 to avoid potential expansion by rule.

The Reporter emphasized that it is not optimal to have an evidence rule that could be applied unconstitutionally, and queried whether the language of an amendment might be tweaked to provide some signal in rule text without suggesting any expansion of *Pena-Rodriguez*. Another Committee member suggested that the only way to truly prevent expansion would be to reference *Pena-Rodriguez* in rule text. The Reporter suggested that it would not be appropriate rulemaking to have an amendment that specifically referenced a case, and moreover that to do so would be to risk the possibility that another amendment would be required should the Supreme Court expand upon the *Pena-Rodriguez* exception.

Other Committee members, after this discussion, agreed that a potential constitutional exception was problematic and that tabling the issue was appropriate. The Chair wrapped up the discussion by noting that the issue would be tabled for one to two years to allow more time for case law to develop before the Committee reconsidered action on Rule 606(b).

VIII. Possible Amendment to Rule 404(b)

The Chair next turned the Committee's attention to potential amendments to Rule 404(b) that had been considered in light of recent Seventh and Third Circuit cases limiting admissibility of evidence of uncharged misconduct in criminal cases. The Chair explained that four different proposals remained on the Committee's agenda: 1) a proposal to restrict use of the "inextricably intertwined" doctrine that takes prior act evidence outside the protections of Rule 404(b); 2) a substantive amendment requiring judges to exclude bad act evidence offered for a proper purpose, where the probative value as to that purpose proceeds through a propensity inference; 3) a proposal to add the balancing test from Rule 609(a)(1)(B) to Rule 404(b) to require that the probative value of prior act evidence offered against a criminal defendant outweigh unfair prejudice; and 4) a proposal to expand the prosecution's notice obligation in criminal cases. The Chair explained that she met with the Reporter prior to the meeting in an effort to streamline the Committee's consideration by subjecting each proposal to an independent determination and vote by the Committee.

The Chair first addressed the "inextricably intertwined" proposals. She stated that the inextricably intertwined doctrine in the courts is problematic, partly due to the variable terminology adopted by courts employing it (including acts that "pertain" to the charged crime, those that are "integral" to the charged crime, those which "complete" the story of the charged crime, or are "intrinsic" to the charged crime). The proposal before the Committee to limit the inextricably intertwined doctrine was an amendment requiring all acts "indirectly" proving the charged crime to proceed through Rule 404(b). The Chair concluded that such an amendment would not be workable or helpful in applying Rule 404(b), particularly because it might sweep any and all conduct apart from the act specifically charged into a Rule 404(b) analysis. The Chair gave an example of a defendant fleeing the scene of the charged crime as indirect evidence that would have to proceed through Rule 404(b) if such an amendment were adopted. One Committee member noted that the inextricably intertwined doctrine is important in determining which acts of a defendant are "other" acts for purposes of Rule 404(b) and opined that the restyling project was wrong to move the word "other" (to read "crimes, wrongs or other acts" instead of "other, crimes, wrongs or acts"). That Committee member suggested that if any other

amendments to Rule 404(b) are proposed, the word “other” should be relocated to its former position. The Reporter agreed that a change might be made if other amendments were proposed, but noted that such a change would not affect the case law on inextricably intertwined acts, because courts would still need to decide which acts were “other” regardless of the placement of the term. The Reporter also noted that the style change did not result in any change in the courts in the application of the inextricably intertwined doctrine.

The Committee determined that it would no longer proceed with any attempt to rectify the “inextricably intertwined” doctrine through an amendment to Rule 404(b).

The Chair then recommended that the Committee remove from the agenda the proposal to bar admission of uncharged misconduct unless the court found the evidence probative of a proper purpose by a chain of reasoning that did not rely on any propensity inferences. She noted that the proposal came from the Seventh Circuit’s opinion in *United States v. Gomez*. She expressed skepticism that a required “chain of non-propensity inferences” could be a workable requirement. She suggested that requiring a trial judge to find a chain of non-propensity inferences sounded more like taking an evidence exam than managing a trial. She further suggested that the original Advisory Committee had rejected “mechanical solutions” in drafting Rule 404(b) and had rejected the notion that there was a truly binary distinction between a “propensity use” and use for a proper purpose -- to show “intent” for example. The line between intent and propensity is often difficult if not impossible to draw. The Chair concluded that *Gomez* made the exercise in eliminating propensity inferences sound easy and straightforward when it often is not.

One Committee member suggested that Rule 404(b) is the most critical rule of evidence in a criminal case and that the real reason that other acts are offered is in fact to suggest the defendant’s propensity to commit crimes. In this Committee member’s opinion, this evidence improperly tips the scales significantly against the defendant, and so the prosecution ought to bear a heavier burden in establishing admissibility. The member concluded that incorporating the *Gomez* test would not be too burdensome on judges, and that the amendment should be adopted. The Federal Public Defender agreed, stating that Rule 404(b) evidence is by far the most prejudicial evidence offered in criminal trials. He noted that proof of Rule 404(b) acts often consumes far more time at trial than proof of the charged offense. He further contended that the instruction given to jurors regarding the use of Rule 404(b) evidence is incomprehensible and offers defendants no protection.

Rob Hur noted that the Department shared the Chair’s concerns that requiring articulation of the chain of reasoning would be unworkable. He opined that a review of pre-trial transcripts reveals that trial courts are already putting the burden on prosecutors to demonstrate the admissibility of this evidence and that Rule 404(b) issues are thoroughly flushed out at the trial level. Mr. Hur further stated that the recent shift in Circuit precedent was having an effect on prosecutorial behavior vis a vis Rule 404(b). Prosecutors know they need to follow the Rule and defend the admissibility of the evidence on appeal. Therefore, he argued that the courts are resolving these issues appropriately and no amendment is necessary. The DOJ did concede that an amendment to the notice provision of Rule 404(b), to codify what the Department is already doing to ensure that defendants receive timely and proper notice, might be viable.

In response to the suggestion that further development in the courts would resolve any problems with Rule 404(b), the Reporter pointed to a recent opinion in the Tenth Circuit, *United States v. Banks*. In that case, the court acknowledged recent efforts to analyze other acts carefully in other circuits, but rejected this trend and held summarily that drug crimes are admissible in the Tenth Circuit to show knowledge. The Reporter suggested that cases like *Gomez* might arguably go too far in preventing use of other act evidence through Rule 404(b), but that other circuits may continue to do too little to prevent misuse. He suggested that an amendment that falls somewhere in between these divergent approaches may be optimal. Mr. Hur cautioned that Congress may get involved if the Committee chose to pursue an amendment limiting admissibility of Rule 404(b) evidence.

The Chair highlighted another recent Tenth Circuit opinion, *United States v. Henthorn*, in which the government was permitted to offer evidence to show that the defendant's first wife died alone in his presence in very suspicious circumstances, to rebut the defendant's argument that his second wife's death while alone with him in suspicious circumstances was an unfortunate accident. She noted that the relevance of the prior accident turned to some degree on the doctrine of chances --- it is highly unlikely that one husband would lose two wives in such similar and tragic circumstances by accident. But she also explained that some suggestion of the defendant's propensity to kill his wives might be found in the evidence. She noted that Wigmore opined that there should be room for a difference of opinion. The Chair explained that the propensity ban in *Gomez* failed to account for that difference of opinion and could confuse trial judges.

A motion to remove the non-propensity inference requirement from discussion passed by a vote of 6-3.

The next amendment alternative discussed was a proposal to add a new balancing test to Rule 404(b) requiring the probative value of other acts evidence offered against a criminal defendant to outweigh unfair prejudice. The Reporter explained that this alternative would offer a more flexible solution that avoids the mechanical tests rejected by the Advisory Committee Note to the current rule, and would avoid any rigid requirement of a chain of non-propensity inferences. He noted that the proposed balancing test would not be a true "reverse" balancing because it would not require probative value to "substantially" outweigh prejudice. Instead, it would be the same balancing test found currently in Rule 609(a)(1)(B), that protects criminal defendants from similar character prejudice. He suggested that it made good sense to have similar balancing tests governing Rule 404(b) and Rule 609(a)(1)(B) evidence offered against criminal defendants because the two rules deal with similar character concerns. He further explained that Congress crafted the protective test in Rule 609(a)(1)(B) that could be usefully applied to Rule 404(b) evidence as well. The Reporter explained that making the balancing test slightly more protective would eliminate the characterization of Rule 404(b) as a rule of inclusion --- a characterization that has resulted in almost per se admission of prior offenses in many federal drug prosecutions. Still, the balancing test would continue to permit probative other acts to be admitted. The Reporter noted that there is support for such a balancing test in pre-Rules cases and that the Uniform Rules of Evidence and some states employ the more protective standard.

Rob Hur from the Department of Justice noted that the applicable balancing represents a policy choice about Rule 404(b) evidence and that Congressional adoption of Rule 404(b), limited only by the standard Rule 403 balancing test, is reason enough to reject a balancing amendment. Another Committee member expressed concern that a balancing amendment would not help courts deal with the issue of what counts as prejudice and whether propensity uses are permissible. That Committee member suggested that no change be made unless it is one to fix the concern about other acts offered for propensity purposes. The Reporter responded that a balancing test requiring the prosecution to demonstrate that probative value outweighs bad character prejudice would do a better job of protecting defendants from improper uses of Rule 404(b) evidence. Another Committee member questioned whether having the same test for Rules 404(b) and 609(a)(1)(B) was appropriate, given that the past convictions are offered for impeachment only under Rule 609, but can be offered on the merits under Rule 404(b). The Reporter responded that the prejudice in both instances is the same, and that the different goals in admitting the evidence is factored in as part of the consideration of probative value --- so that there is no reason not to apply the same test for both situations.

The Chair asked for a straw vote on whether to continue discussing a balancing amendment or whether to remove it from the agenda. The Committee voted 5-4 to continue discussing the balancing alternative.

One Committee member queried why the test to protect criminal defendants from character prejudice in Rule 609(a)(1)(B) should differ from the balancing test in Rule 404(b), apart from historical practice. The Chair noted that Rule 404(b) helps the prosecution sustain its burden of proof, while Rule 609 pertains to impeachment only. The Reporter then noted that decisions about balancing and protections are indeed policy decisions commonly underlying rules of evidence like Rule 412. The policy underlying the balancing amendment of Rule 404(b) would be living up to our commitment to try cases and not people. Judge Lioi commented that the Rule 403 factors serve that purpose well and put the government through its paces, to which the Reporter responded that the proposed balancing test would utilize the identical factors but would simply replace the Rule 403 balance favoring inclusion with one requiring probative value to outweigh prejudice. Another Committee member noted that an amended balancing test would ensure that Rule 404(b) is a rule of exclusion and not inclusion. The Reporter noted that it would be a rule of “mild exclusion” where it would simply require probative value to overcome prejudice to even a slight degree to be admitted.

The Chair then stated that Rule 404(b) is not a rule of exclusion. Instead, it prohibits one inference that a defendant is a bad person due to past misdeeds. She opined that other act evidence relevant to anything other than that bad character inference is admissible subject to Rule 403. She further argued that young prosecutors are so nervous about overstepping with Rule 404(b) evidence that they often limit comments on such evidence in closing argument to brief statements that the evidence is admissible to prove “intent” for example. The Chair concluded that the balancing test should not be made more protective because it might limit the admissibility of evidence prosecutors need to prove a case.

The Reporter noted that the courts permissively admitting other act evidence under the Rule 403 standard are not necessarily ruling incorrectly because that standard favors admissibility so heavily. The question raised by a balancing alternative is whether Rule 404(b) should allow evidence of other acts to come in as freely as it does. Although the drafters of Rule 404(b) limited it only with Rule 403, the Reporter emphasized that there is much less legislative history regarding Congressional intent for Rule 404(b) than there is regarding the proposed balancing test found in Rule 609. Therefore there should not be substantial concern about overriding congressional intent.

At the conclusion of these remarks, another straw vote was taken on whether to proceed with consideration of a balancing amendment. The Committee vote was 7-2 against continuing consideration of a balancing amendment.

The Committee then discussed the final potential amendment to Rule 404(b) – changes to the notice provision in criminal cases. The Reporter explained that a proposal to eliminate the requirement that the defense request notice in criminal cases had already been unanimously approved by the Committee. The Reporter also called the Committee’s attention to a proposed amendment to the notice provision circulated to the Committee by the DOJ prior to the meeting. This provision would require a prosecutor to “provide reasonable notice of the general nature of any such evidence.” It would also require a prosecutor to “articulate in the notice the non-propensity purpose for which the prosecutor intends to offer the evidence and the reasoning supporting the purpose.” Finally, it would require the prosecution to provide notice “in writing” before trial or during trial “if the court, for good cause, excuses lack of pretrial notice.”

Committee members raised concerns about requiring the prosecution to provide notice of only the “general nature” of Rule 404(b) evidence. Some discussion was had about requiring the government to disclose “the substance” of the evidence to make the Rule 404(b) notice provision consistent with the notice provision in the proposed amendment to Rule 807. Concern was also raised about the lack of any timing requirement for the notice. Some suggested that requiring notice 14 days in advance of trial could be superior, although Mr. Hur thought a timing requirement could prove rigid and unworkable. The Reporter suggested that the language used in the proposed amendment to Rule 807 requiring disclosure sufficiently before trial to allow the opponent to meet the evidence could be a useful solution to the timing issue, and would promote uniformity in the Rules. Other Committee members agreed that trial judges set deadlines in pre-trial orders and that including a 14-day limit in rule text was unnecessary.

The Federal Public Defender commented that prosecutors commonly provide the minimum notice possible and resist all efforts by the defense to obtain more information. He noted that there is a great deal of needless litigation over who the Rule 404(b) witness will be and what act will be proved and that prosecutors rely on the terms “general nature” in Rule 404(b) to defend minimal notice. The Reporter queried whether use of the term “substance” would represent an improvement over “general nature.” The Department of Justice suggested that the articulation requirement in the proposed notice provision would resolve the existing concerns over the quality of the notice. The Federal Public Defender did not think the articulation of reasoning requirements would necessarily help in identifying the specific act to be proved and thought that a “particulars” or “specific details” requirement would be superior.

Judge Furman suggested putting the term “substance” together with the “fair opportunity to meet the evidence” qualification to address the problem. Judge Campbell suggested deleting the required description of the act in the notice and simply stating that the prosecutor must provide “reasonable notice of any such evidence” --- which all agreed was workable. Committee members agreed that requiring notice in writing sufficiently in advance of trial “to give the defendant a fair opportunity to meet the evidence” would be a good solution to the timing issue as well. The DOJ noted that the good cause exception to the notice requirement should apply to all of the prosecutor’s obligations (including articulation). The Reporter explained that the good cause exception was made applicable to all notice obligations due to its placement at the conclusion of all notice requirements, and that the Committee Note could emphasize that the good cause exception would go to articulation as well as timing.

The Committee voted unanimously to approve the amendment to the notice provision of Rule 404(b).

The Reporter then took the Committee through the text of Rule 404(b) and a proposed Committee Note that was set forth in the agenda book. During that discussion, one Committee member proposed moving the word “other” in the heading of Rule 404(b) and in the text of Rule 404(b)(1) to return the word to its correct pre-restyling position: “Other crimes, wrongs, or acts.” The Committee unanimously agreed with this proposal. The Reporter also recommended changing “Permitted Uses” in the heading of Rule 404(b)(2) to “Other Uses.” He explained that headings were added to the Rule as part of the restyling and that “Other Uses” more accurately reflects the operation of Rule 404(b)(2). The Committee tentatively agreed with this proposal.

The Committee generally approved the proposed Committee Note, subject to further wordsmithing after the meeting. After discussion by email, the following changes were made to the proposal:

- “Permitted uses” in the heading of Rule 404(b)(2) would be retained.
- Two changes proposed by the Style Subcommittee to the Standing Committee would be implemented.
- The good cause provision would be amended to provide, consistently with Rule 807, that if the court finds good cause to allow notice during the trial, that notice can be given in any form.
- Minor changes to the Committee Note were made to clarify that the good cause exception as to articulation would apply to additional proper purposes that became evident after notice was provided.

The Committee, by email, unanimously approved the text and the Committee Note of the proposed amendment to Rule 404(b). The proposed amendment will be submitted to the Standing Committee with the recommendation that it be released for public comment.

The Committee resolved that it would revisit certain questions during public comment, such as whether notice provided after trial has begun (upon a showing of good cause) must be made in writing, and whether the Committee Note should be changed with respect to good cause and the articulation requirements.

The text and Committee Note of the proposed amendment to Rule 404(b) is attached to these Minutes.

IX. Possible Amendment to Rule 106

The next item on the agenda for the Committee's consideration was a potential amendment to the Rule 106 rule of completion. The amendment would rectify a conflict in the courts over the admissibility of otherwise inadmissible hearsay to complete misleading statements, and would include oral statements within the coverage of Rule 106. The Reporter reminded the Committee that Judge Paul Grimm had raised these problems about Rule 106 for the Committee's consideration, and directed the Committee's attention to Judge Grimm's thoughtful opinion on the issues in the agenda materials.

The Reporter explained that the hearsay issue relates to a very narrow circumstance in which the government offers a portion of a defendant's statement that is misleading (as a statement of a party opponent under Rule 801(d)(2)(A)) and the remainder of the statement is necessary for completion --- but is hearsay. Some courts find that the hearsay rule bars the defendant's attempt to admit the remainder of his own hearsay statement through Rule 106 to correct the distortion, because a defendant may not admit his own hearsay statement under Rule 801(d)(2). In those cases, the unfairness created by the government's misleading presentation of a partial statement goes uncorrected. The question for the Committee is whether this result is appropriate under the traditional "door-opening" approach of the evidence rules that seeks to ensure that adversaries are not prejudiced by a misleading presentation of evidence.

The Reporter explained that Rule 502(a), regarding subject matter waiver of privilege, borrowed the language of Rule 106 exactly and embodies the same principle: that a misleading use of privileged information by one side allows the opponent full access to privileged materials on the same subject to correct any distortion. He argued that it was difficult to understand why the government should be permitted to lodge a hearsay objection to prevent needed completion of a misleading statement, when similar behavior by a litigant is sufficient to waive privilege. An amendment would be necessary to address the cases in which courts prevent defendants from correcting a misleading partial statement due to the rule against hearsay.

One option previously discussed by the Committee would be to amend Rule 106 to allow the completing statement to be admitted solely for its not-for-truth purpose in showing the full "context" of the partial statement already admitted. The Reporter suggested, however, that the "context" option would be problematic in that the parties would not be left on equal footing: the government could argue the truth of the misleading portion of the statement, while the defendant could not argue the truth of the completing portion. The only way to a fair result would be to allow the completing statements to be admissible for their truth. Otherwise the proponent is given an advantage from a misleading presentation.

The Reporter also noted that, prior to a style amendment designed to make Rule 106 gender neutral, the language of Rule 106 required the *proponent* of the original partial and misleading statement to admit the completing portion of the statement at the same time the misleading portion was admitted. If the government were required to admit the completing statement itself,

the hearsay objection would be eliminated because the government would be offering the defendant's entire statement through Rule 801(d)(2)(A), as a statement by a party-opponent. That prior version of the Rule suggests that Congress did not intend to have the hearsay rule prevent completion of a misleading partial statement. Moreover, the legislative history indicates that Congress rejected a DOJ request to provide in Rule 106 that the completing statement had to be independently admissible.

Judge Furman suggested that a return to the language requiring the original proponent to do the completing would be a good alternative to an amendment that would allow the opponent's completion over a hearsay objection. This would avoid establishing a hearsay exception outside the context of Article 8 of the Federal Rules. The Reporter expressed concern that a return to the old provision might be too subtle to correct the unfair result in some of the recent cases. A Committee member stated that requiring the proponent to do its own completing would not be too subtle and would represent a more surgical solution to the problem than a broader hearsay exception would.

Another Committee member noted a footnote in Judge Grimm's opinion on Rule 106 stating that the Advisory Committee had voted unanimously against an amendment to address these issues in 2002-2003, finding that the costs of an amendment exceed its benefits due to judicial handling of the issues. The Reporter explained that amendments to Rule 106 had come up in 2002 and again in 2006, but were rejected due to other more pressing rulemaking priorities at the time. He noted that recent cases allowing misleading partial statements to go uncorrected present a more significant conflict and concern in the case law. The Chair queried whether the conflict is confined to the Sixth and Ninth Circuit, and whether everyone else is basically getting it right. The Reporter noted prior amendments designed to correct even lesser conflicts and concluded that an amendment would be the only way to correct the unfairness in the Circuits that allow a misleading partial statement to go uncorrected, given the many years in which this conflict has gone uncorrected.

The Chair agreed that the function of the Advisory Committee is to resolve conflicts, but advocated proceeding slowly. She expressed reluctance to propose a hearsay exception for completing statements and more interest in a housekeeping amendment that would require the party offering a misleading portion to also offer the completing remainder --- without creating a broader hearsay exception. The Chair noted that the Department of Justice had proposed limiting completion to circumstances in which the original portion is "misleading." The Reporter noted that Judge Grimm thought that limiting the rule to "misleading" statements would be workable.

Judge Furman reiterated his proposal to return to the language of Rule 106 requiring the original proponent to complete the proffered statement, to be accompanied by Advisory Committee notes explaining that hearsay is not a bar to completion and that the Committee was returning to the original language to resolve the split in the cases. Judge Campbell expressed the concern that opponents would use such a requirement as a tactical advantage to interrupt the proponent of a statement repeatedly to demand completion. Judge Furman noted that the Rule 106 existing requirement that completion is required only in narrow circumstances would limit

such interruptions. The Reporter stated that limiting Rule 106 to “misleading” statements expressly might further clarify that the Rule is limited in scope.

The Chair asked the Committee whether it was interested in considering an amendment requiring the proponent to do its own completion, with a “misleading” limitation added to the rule text. The Committee voted to consider such a proposal for the next meeting with a Committee note explaining that there “can be no hearsay objection because the proponent is required to introduce the completing portion.”

The discussion then moved to whether oral statements should be covered by Rule 106. The Chair noted that Rule 106 currently applies only to written or recorded statements and that Judge Grimm advocates extending Rule 106 to cover oral statements needed to complete misleading statements. She noted that many courts allow completion of oral statements through their inherent Rule 611(a) authority, but that the question was whether to bring oral statements under the umbrella of Rule 106. The Reporter noted that one concern that had been raised about completing oral statements was the difficulty in proving the content of an oral statement. He noted that Judge Grimm thought that extensive and distracting inquiries into the content of an oral statement could be prevented by the trial judge through Rule 403 --- and that courts have done so. The Reporter further questioned why the difficulty in proving the content of completing oral statements should foreclose their use, when the difficulty in proving the content of the oral statement originally offered by the proponent poses no obstacle to its proof.

Committee members discussed practical problems in the completion of oral statements testified to by a witness and how they might be handled at trial. Judge Lioi noted that the most common statements sought to be corrected at trial appear in depositions or in transcripts of wiretap recordings. In those cases, she explained, the trial judge knows exactly what was said, can see whether a proffered portion is misleading, and decide how much of the remainder is necessary to complete. Extending Rule 106 to oral statements might open up a can of worms because it would allow completion without providing the judge access to this crucial information needed to rule on this issue. The Reporter stated that an Advisory Committee Note would be useful in giving the court guidance that trial judges should decline to consider completion of oral statements if problems of proof become too complicated and time-consuming. Andrew Goldsmith from the DOJ noted that Criminal Rule 16 ensures pre-trial notice of any oral statements of the defendant that will be offered at trial, meaning that disputes about completion should not arise on the fly in the heat of trial. The Reporter remarked that such pre-trial disclosures should make completion issues surrounding a defendant’s oral statements easier to resolve.

The Committee voted to continue consideration of an amendment to Rule 106 that would add oral statements to the rule at its next meeting. The Reporter agreed to write up amendment alternatives for the fall meeting including a hearsay exception proposal, a requirement that the proponent complete to avoid the hearsay issue, the addition of the limiting term “misleading,” and the addition of oral statements to Rule 106.

X. Proposed Amendments to Rule 609(a)(1)

The Chair explained that there were multiple proposals on the table concerning Rule 609(a)(1) and the use of a criminal defendant's non-dishonesty felony convictions to impeach his trial testimony. She noted that there are only a small number of states with greater protections for criminal defendants, and that the vast majority of states are following the federal approach. The Reporter noted that the first alternative to an amendment was to prohibit non-dishonesty felony impeachment of criminal defendants --- or even more broadly to abrogate Rule 609(a)(1) entirely. The Committee at the previous meeting, however, expressed reluctance about such bans, as in tension with the hard-fought compromise in Congress that resulted in Rule 609(a).

The Chair asked whether Committee members wished to discuss an abrogation alternative. No interest was expressed in pursuing abrogation and no further discussion about an amendment abrogating Rule 609(a)(1)(B) impeachment was had.

The Reporter noted another potential amendment, suggested by Professor Ric Simmons, to limit Rule 609(a)(1) impeachment to theft convictions. Michigan follows this approach. The Reporter explained that such an amendment would allow impeachment with the non-dishonesty felony convictions most probative of untruthfulness --- like theft and receipt of stolen property -- - while eliminating impeachment with less probative felonies like assault and sex crimes. The Reporter recognized that there could be some difficulty in defining the crimes to be included in a theft-related amendment (such as receipt of stolen property) but a Committee Note might be useful in defining such crimes. A Committee member opined that crimes such as drug distribution should not be absolutely barred, because they are often indicative of a life of underhandedness that could be probative for impeachment. The Chair noted that defense counsel in criminal cases frequently impeach prosecution witnesses with felony convictions that are not theft-related, and suggested that defendants it would not be advisable to abrogate impeachment for these witnesses, or solely for the criminal defendant. The Committee thereafter rejected a potential amendment to Rule 609(a)(1) that would limit felony impeachment to theft-related offenses.

The Reporter then raised the possibility of an amendment to the balancing test in Rule 609(a)(1)(B) suggested by Professor Jeff Bellin. A small adjustment to the balancing test could restore congressional intent to protect defendants from routine felony impeachment and provide defendants with prior convictions a more meaningful opportunity to testify. This revision would require courts to consider the marginal impeaching value of prior felony convictions in light of the inherent bias of a criminal defendant testifying to evade conviction. Professor Bellin notes that a defendant is already significantly impeached by his desire to avoid punishment and that the probative value of prior felony convictions is reduced by this alternative impeaching factor. A balancing test that expressly requires courts to take the defendant's bias into account would result in a more accurate assessment of probative value. Professor Bellin has also suggested that courts should be strongly cautioned against admitting prior felonies similar to the current charges for the purpose of impeachment. The Reporter noted that the extensive digest compiled in the agenda materials on Rule 609(a)(1)(B) rulings demonstrates that courts frequently admit similar crimes for impeachment purposes. The Reporter described data compiled by Professor Bellin indicating that jurors do not limit consideration of prior felonies to impeachment, do not follow limiting instructions as to impeachment, and that jurors punish

defendants who choose to remain off the stand to avoid impeachment with a silence penalty notwithstanding instructions not to do so.

Judge Campbell contended that the suggested modifications to the Rule 609(a)(1)(B) balancing test seemed pretty prescriptive and would micromanage a trial judge's balancing process unduly. Further, Judge Campbell thought that including some specific factors for consideration might suggest the omission of others, making the amended test underinclusive. In the end, he did not see why it would be advisable to mandate specifics for trial judges applying this balancing test. The Reporter agreed that it may not have been necessary to include such specifics in the initial rule, but that evidence from the cases shows that judges are not properly accounting for these factors such that spelling them out now may be necessary. Moreover, the proposed amendment focuses on marginal probative value and the similarity of the conviction to the crime charged, but does not purport to limit the court's use of other factors.

The Chair stated that trial judges don't think in terms of "marginal probative value," but evaluate impeachment in light of the defendant's position on the stand and in the hurly burly of the courtroom. The Reporter responded that the reported cases belie that notion --- they indicate that the courts do take account of other matters affecting marginal probative value (such as other convictions) but not the self-interest of the defendant.

The Chair expressed her view that it was inadvisable to micromanage trial judges in their assessments of probative value and prejudicial effect. No Committee member provided further discussion or moved for the adoption of a proposed amendment to the balancing test. In the absence of any further comment, the Chair stated that the proposed amendment to the balancing test would be tabled. The Reporter noted that he had hoped for a more robust Committee exchange on potential amendments to Rule 609(a)(1)(B), particularly with regard to the balancing test.

XI. Rule 611 and Illustrative Evidence

The final item on the agenda originated with a proposal from a law review article suggesting that the Committee should adopt a rule on the use of illustrative evidence at trial. The line between "demonstrative" evidence, used substantively to prove disputed issues at trial, and "illustrative" evidence, offered solely as a pedagogical aid to assist the jury in understanding other evidence, is a difficult one to draw. An idea for a draft of an amendment to Rule 611 was included in the agenda materials to govern the use of truly "illustrative" evidence at trial. This draft rule was not designed as a proposal for the Committee, but was included to give the Committee an idea of what might be done if it wished to consider the matter further. The draft amendment was placed in Rule 611 because courts typically find authority to regulate illustrative evidence in Rule 611(a). The draft would not cover demonstrative evidence at all, but would regulate the use of illustrative aids. It would prohibit a judge from sending an illustrative aid to the jury during deliberations absent the consent of all parties.

Judge Campbell asked whether there is any indication that courts are confused about these issues. The Reporter noted that there is some confusion in the cases regarding the distinction between demonstrative and illustrative evidence, and also between pedagogical summaries and

those substantively admissible under Rule 1006. The Reporter opined that there was not a crying need for an amendment, but that there could be value in providing organizing principles around illustrative evidence. The Chair asked for the experience of the trial judges in the room with respect to illustrative aids. There was a consensus among judges that illustrative aids present no significant difficulty and that there is no need for a rule covering their use. Several members of the Committee noted, however, that they found the Maine rule on illustrative evidence and the thoughtful accompanying legislative notes, which were included in the agenda materials, to be extremely valuable.

XII. Closing Matters

The Committee thanked the Reporter for the immense amount of work he put into the excellent agenda materials and the meeting was adjourned.

XIII. Next Meeting

The fall meeting of the Evidence Rules Committee will be held at the University of Denver in Colorado on Friday, October 19, 2018.

Respectfully submitted,

Liesa L. Richter
Daniel J. Capra

FW: Proposed Presentations for FRE Conference

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Crowell, James (ODAG)" <(b) (6)>
Date: Mon, 28 Aug 2017 09:08:39 -0400

Jim – this is another important development in the FRE/Forensics world. I just spoke to Ted and Kira, and they agree with my suggestion that (b) (5) on the various Advisory Committees, (b) attend the Boston FRE/Forensics conference/symposium in October, and (c) call (b) (5). Given our schedules, it looks like Friday 9/15 is the earliest we'll all be available – will that work for you (and possibly Rob)? Thanks – Andrew

From: Antell, Kira M. (OLP)
Sent: Friday, August 25, 2017 5:33 PM
To: Goldsmith, Andrew (ODAG) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

(b) (5)
■ (b) (5); (b) (5) per CIV

From: Goldsmith, Andrew (ODAG)
Sent: Friday, August 25, 2017 11:28 AM
To: Antell, Kira M. (OLP) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Can you give me a quick call? Thanks.

From: Antell, Kira M. (OLP)
Sent: Friday, August 25, 2017 10:18 AM
To: Shapiro, Elizabeth (CIV) <(b) (6)>
Cc: Begian, Lerrick (OLP) <(b) (6)>; Smith, David L. (USAEO) <(b) (6)>; Ibrahim, Anitha (CRM) <(b) (6)>; Young, Cynthia (USAMA) <(b) (6)>; Hunt, Ted (ODAG) <(b) (6)>; Goldsmith, Andrew (ODAG) <(b) (6)>; Hafer, Zachary (USAMA) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Thank everyone! Bet y, please go ahead and share our proposal with Capra

From: Young, Cynthia (USAMA) <(b) (6)>
Sent: Friday, August 25, 2017 9:39 AM
To: Goldsmith, Andrew (ODAG) <(b) (6)>; Antell, Kira M. (OLP) <(b) (6)>; Hafer, Zachary (USAMA) <(b) (6)>; Hunt, Ted (ODAG) <(b) (6)>
Cc: Shapiro, Elizabeth (CIV) <(b) (6)>; Begian, Lerrick (OLP) <(b) (6)>; Smith, David L. (USAEO) <(b) (6)>; Ibrahim, Anitha (CRM) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Thanks, Kira. Fine with me

From: Goldsmith, Andrew (ODAG) <(b) (6)>
Sent: Friday, August 25, 2017 9:27 AM
To: Antell, Kira M. (OLP) (JMD) <(b) (6)>; Young, Cynthia (USAMA) <(b) (6)>; Hafer, Zachary (USAMA) <(b) (6)>; Hunt, Ted (ODAG) (JMD) <(b) (6)>
Cc: Shapiro, Elizabeth (CIV) <(b) (6)>; Begian, Lerrick (OLP) (JMD) <(b) (6)>; Smith, David L. (USAEO) <(b) (6)>; Ibrahim, Anitha (CRM) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Looks good to me.

From: Antell, Kira M. (OLP)
Sent: Friday, August 25, 2017 8:59 AM
To: Young, Cynthia (USAMA) <[REDACTED]>; Goldsmith, Andrew (ODAG) <[REDACTED]>;
Hafer, Zachary (USAMA) <[REDACTED]>; Hunt, Ted (ODAG) <[REDACTED]>;
Cc: Shapiro, Elizabeth (CIV) <[REDACTED]>; Begian, Lernik (OLP) <[REDACTED]>; Smith, David
L. (USAE0) <[REDACTED]>; Ibrahim, Anitha (CRM) <[REDACTED]>
Subject: RE: Proposed Presentations for FRE Conference

I am writing to follow up and make sure no one has any issues with the proposed presentations. Betsy will need to email Dan Capra with this information no later than this afternoon so we can verify [REDACTED]. If you do have issues, please let us know by noon today so we can make any needed edits.

Thanks,
Kira

From: Antell, Kira M. (OLP)
Sent: Wednesday, August 23, 2017 10:45 AM
To: Young, Cynthia (USAMA) <[REDACTED]>; Goldsmith, Andrew (ODAG) <[REDACTED]>;
Hafer, Zachary (USAMA) <[REDACTED]>; Hunt, Ted (ODAG) <[REDACTED]>;
Cc: Shapiro, Elizabeth (CIV) <[REDACTED]>; Begian, Lernik (OLP) <[REDACTED]>; Smith, David
L. (USAE0) <[REDACTED]>; Ibrahim, Anitha (CRM) <[REDACTED]>
Subject: Proposed Presentations for FRE Conference

Hello all,

Yesterday, Andrew Goldsmith, Betsy Shapiro, Ted Hunt, and I met to discuss presentations at the October FRE Mini Conference on Forensics. Dan Capra, the reporter, has suggested that we have too many Department representatives on the first panel and proposed that we reduce our representatives from three to two. The four of us agreed [REDACTED].
[REDACTED] (b)(5); (b)(5) per CIV [REDACTED]

I propose the following general topics for each presenter. I welcome your input. After we make any necessary edits to the below, [REDACTED].

(b)(5)

(b)(5); (b)(5) per CIV

Kira Antell
Senior Counsel
Office of Legal Policy
U.S. Department of Justice
950 Pennsylvania Avenue, NW
Washington, DC 20530

[REDACTED]
[REDACTED]

RE: Proposed Presentations for FRE Conference

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Simms, Donna Y. (ODAG)" <(b) (6)>
Cc: "Crowell, James (ODAG)" <(b) (6)>
Date: Mon, 28 Aug 2017 09:13:56 -0400

Great, thanks. The others would be Ted Hunt, Kira Antell, and Betsy Shapiro. 30 minutes is more than sufficient.

From: Crowell, James (ODAG)
Sent: Monday, August 28, 2017 9:09 AM
To: Goldsmith, Andrew (ODAG) <(b) (6)>; Simms, Donna Y. (ODAG) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Donna

Please schedule with Andrew so that Rob and I can attend a meeting with Andrew and whoever else he deems appropriate for 9/15.

Thank you,

Jim

From: Goldsmith, Andrew (ODAG)
Sent: Monday, August 28, 2017 9:09 AM
To: Crowell, James (ODAG) <(b) (6)>
Subject: FW: Proposed Presentations for FRE Conference

Duplicative Material

RE: Proposed Presentations for FRE Conference

From: "Simms, Donna Y. (ODAG)" <(b) (6)>
To: "Crowell, James (ODAG)" <(b) (6)>
Date: Mon, 28 Aug 2017 09:19:29 -0400

Will do.

From: Crowell, James (ODAG)
Sent: Monday, August 28, 2017 9:10 AM
To: Goldsmith, Andrew (ODAG) <(b) (6)>; Simms, Donna Y. (ODAG) <(b) (6)>
Subject: RE: Proposed Presentations for FRE Conference

Duplicative Material

DAG Meeting-Feb 12 9:10 a.m.

From: "Hunt, Ted (ODAG)" <(b) (6)>
To: "Murphy, Marcia (ODAG)" <(b) (6)>
Date: Fri, 09 Feb 2018 13:30:34 -0500
Attachment Briefing Document for Forensic Bitemark Discussion doc (24.5 kB)

Marcy,

Attached is a short briefing document for the DAG's review in preparation for our 9:10 a.m. meeting on Monday.

Thanks,

Ted

Ted R. Hunt
Senior Advisor to the Attorney General on Forensic Science
Office of the Deputy Attorney General
United States Department of Justice
950 Pennsylvania Ave. NW
Washington, DC 20530

(b) (6)

Fwd: Rules Summary DAG Standing Comm

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Suero, Maya A. (ODAG)" <(b) (6)>
Cc: "Hovakimian, Patrick (ODAG)" <(b) (6)>
Date: Fri, 06 Sep 2019 16:37:34 -0400
Attachments: Rules Summary DAG Standing Comm jjw edits 090619.docx (27.46 kB); ATT00001.htm (1.25 kB)

Here's the memo for the DAG in preparation for next week's meeting on the Rules Committees. Please let me know if you have any questions.

Sent from my iPhone please excuse any typos

|
.
|

RE: PCAST Talkers

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Antell, Kira M. (OLP)" <(b) (6)>
Date: Thu, 16 Feb 2017 12:52:16 -0500

Thanks!

From: Antell, Kira M. (OLP)
Sent: Thursday, February 16, 2017 12:50 PM
To: Goldsmith, Andrew (ODAG) <(b) (6)>
Subject: RE: PCAST Talkers

Duplicative Material

DRAFT Remarks

From: "Hunt, Ted (ODAG)" <(b) (6)>
To: "Crowell, James (ODAG)" <(b) (6)>, "Hur, Robert (ODAG)" <(b) (6)>
Date: Fri, 15 Sep 2017 11:22:58 -0400
Attachment DRAFT Remark KBI KS AG Meeting Sept 20 doc (46 27 kB)

Jim/Rob:

Attached above are draft remarks that I've prepared for a talk I'm giving next week (Wednesday) to the Kansas Bureau of Investigation and the Kansas Attorney General's Office at the KBI's new lab in Topeka, KS.

These remarks have already been reviewed by OLP. Note that they contain a couple references to the PCAST Report, and that these are ODAG's first public comments on that Report.

Please let me know if you have any questions or comments.

Thanks,

Ted

Ted R. Hunt
Senior Advisor to the Attorney General on Forensic Science
Office of the Deputy Attorney General
United States Department of Justice
950 Pennsylvania Ave, NW
Washington, D.C. 20530
(b) (6)

RE: DRAFT Remarks

From: "Crowell, James (ODAG)" <(b) (6)>
To: "Hunt, Ted (ODAG)" <(b) (6)>
Cc: "Hur, Robert (ODAG)" <(b) (6)>
Date: Mon, 18 Sep 2017 14:24:19 -0400

Good by me

From: Hunt, Ted (ODAG)
Sent: Friday, September 15, 2017 11:23 AM
To: Crowell, James (ODAG) <(b) (6)>; Hur, Robert (ODAG) <(b) (6)>
Subject: DRAFT Remarks

Duplicative Material

702/PCAST TPs

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Hur, Robert (ODAG)" <(b) (6)>
Cc: "Antell, Kira M. (OLP)" <(b) (6)>, "Shapiro, Elizabeth (CIV)" <(b) (6)>, "Hunt, Ted (ODAG)" <(b) (6)>
Date: Thu, 21 Sep 2017 13:59:30 -0400
Attachment Propo ed Talker for Call with Judge Living ton on 702 09202017 doc (25 2 kB); ATT00001 t t (2 bytes)

Rob - here are the talkers for R 702/PCAST. (b) (5)

- Andrew

FW: Advisory Committee on Rules of Evidence, agenda materials for October 26-27, 2017 meeting

From: "Brown, Angela M. (ODAG)" <(b) (6)>
To: "Hur, Robert (ODAG)" <(b) (6)>
Date: Tue, 03 Oct 2017 10:44:45 -0400
Attachments: Evett et al, Finding the Way Forward, FS International (2017).pdf (418.04 kB); UNT Center for Human Identification, Response to PCAST Report June 2017 pdf (521.58 kB)

Printed and placed in your in box.

Angela M. Brown
Office of the Deputy Attorney General
Department of Justice

(b) (6)
(b) (6)

From: (b) (6) Bridget Healy <(b) (6) Bridget Healy [mailto:(b) (6)]>
Sent: Tuesday, October 03, 2017 10:40 AM
To: (b) (6) Debra Livingston <(b) (6) Debra Livingston [mailto:(b) (6)]; dcapra@law.fordham.edu; (b) (6) James Bassett <(b) (6) James Bassett [mailto:(b) (6)]>; (b) (6) Daniel Collins <(b) (6) Daniel Collins [mailto:(b) (6)]>; Hur, Robert (ODAG) <(b) (6) Hur, Robert (ODAG) [mailto:(b) (6)]>; (b) (6) AJ Kramer <(b) (6) AJ Kramer [mailto:(b) (6)]>; (b) (6) Traci Lovitt <(b) (6) Traci Lovitt [mailto:(b) (6)]>; (b) (6) J. Thomas Marten <(b) (6) J. Thomas Marten [mailto:(b) (6)]>; (b) (6) Shelly Dick <(b) (6) Shelly Dick [mailto:(b) (6)]>; (b) (6) Thomas Schroeder <(b) (6) Thomas Schroeder [mailto:(b) (6)]>; (b) (6) Liesa Richter <(b) (6) Liesa Richter [mailto:(b) (6)]>; (b) (6) William Sessions <(b) (6) William Sessions [mailto:(b) (6)]>
CC: (b) (6) James Dever <(b) (6) James Dever [mailto:(b) (6)]>; (b) (6) Lyndsav Hayes <(b) (6) Lyndsav Hayes [mailto:(b) (6)]>; (b) (6) Sara Lioi <(b) (6) Sara Lioi [mailto:(b) (6)]>; Shapiro, Elizabeth (CIV) <(b) (6) Elizabeth Shapiro (CIV) [mailto:(b) (6)]>; (b) (6) David Campbell <(b) (6) David Campbell [mailto:(b) (6)]>; (b) (6) Nancy Outley <(b) (6) Nancy Outley [mailto:(b) (6)]>; (b) (6) Daniel Coquillette <(b) (6) Daniel Coquillette [mailto:(b) (6)]>; (b) (6) Irene Dalbec <(b) (6) Irene Dalbec [mailto:(b) (6)]>; (b) (6) Barbara Alcon <(b) (6) Barbara Alcon [mailto:(b) (6)]>; (b) (6) Kathy Stephenson <(b) (6) Kathy Stephenson [mailto:(b) (6)]>; (b) (6) Jeanette Santos <(b) (6) Jeanette Santos [mailto:(b) (6)]>; (b) (6) Krystle Dalke <(b) (6) Krystle Dalke [mailto:(b) (6)]>; (b) (6) Timothy Lau <(b) (6) Timothy Lau [mailto:(b) (6)]>; Brown, Angela M. (ODAG) <(b) (6) Brown, Angela M. (ODAG) [mailto:(b) (6)]>; (b) (6) Rebecca Womeldorf <(b) (6) Rebecca Womeldorf [mailto:(b) (6)]>; (b) (6) Patrick Tighe <(b) (6) Patrick Tighe [mailto:(b) (6)]>
Subject: Advisory Committee on Rules of Evidence, agenda materials for October 26-27, 2017 meeting

Hi everyone,

Please find attached two additional articles that relate to the report included at Tab 9C of the agenda book. They have been added to the online version of the agenda materials as well.

Sincerely,
Bridget

Bridget Healy
Attorney Advisor
Office of General Counsel, Rules Committee Staff
(b) (6)
(b) (6)

----- Forwarded by Bridget Healy/DCA/AO/USCOURTS on 10/03/2017 09:31 AM -----

From: Bridget Healy/DCA/AO/USCOURTS
(b) (6) Debra Livingston <(b) (6) Debra Livingston [mailto:(b) (6)]>; (b) (6) Daniel Capra <(b) (6) Daniel Capra [mailto:(b) (6)]>; (b) (6) James Bassett <(b) (6) James Bassett [mailto:(b) (6)]>; (b) (6) Daniel Collins <(b) (6) Daniel Collins [mailto:(b) (6)]>; (b) (6) Robert Hur <(b) (6) Robert Hur [mailto:(b) (6)]>; (b) (6) AJ Kramer <(b) (6) AJ Kramer [mailto:(b) (6)]>; (b) (6) Traci Lovitt <(b) (6) Traci Lovitt [mailto:(b) (6)]>; (b) (6) J. Thomas Marten <(b) (6) J. Thomas Marten [mailto:(b) (6)]>; (b) (6) Shelly Dick <(b) (6) Shelly Dick [mailto:(b) (6)]>; (b) (6) Thomas Schroeder <(b) (6) Thomas Schroeder [mailto:(b) (6)]>; (b) (6) Liesa Richter <(b) (6) Liesa Richter [mailto:(b) (6)]>; (b) (6) William Sessions <(b) (6) William Sessions [mailto:(b) (6)]>; (b) (6) James Dever <(b) (6) James Dever [mailto:(b) (6)]>; (b) (6) Lyndsav Hayes <(b) (6) Lyndsav Hayes [mailto:(b) (6)]>; (b) (6) Sara Lioi <(b) (6) Sara Lioi [mailto:(b) (6)]>; (b) (6) Elizabeth Shapiro <(b) (6) Elizabeth Shapiro [mailto:(b) (6)]>; (b) (6) David Campbell <(b) (6) David Campbell [mailto:(b) (6)]>; (b) (6) Nancy Outley <(b) (6) Nancy Outley [mailto:(b) (6)]>; (b) (6) Daniel Coquillette <(b) (6) Daniel Coquillette [mailto:(b) (6)]>; (b) (6) Irene Dalbec <(b) (6) Irene Dalbec [mailto:(b) (6)]>; (b) (6) Barbara Alcon <(b) (6) Barbara Alcon [mailto:(b) (6)]>; (b) (6) Kathy Stephenson <(b) (6) Kathy Stephenson [mailto:(b) (6)]>; (b) (6) Jeanette Santos <(b) (6) Jeanette Santos [mailto:(b) (6)]>; (b) (6) Krystle Dalke <(b) (6) Krystle Dalke [mailto:(b) (6)]>; (b) (6) Timothy Lau <(b) (6) Timothy Lau [mailto:(b) (6)]>; (b) (6) Angela Brown <(b) (6) Angela Brown [mailto:(b) (6)]>; (b) (6) Rebecca Womeldorf <(b) (6) Rebecca Womeldorf [mailto:(b) (6)]>; Patrick Tighe/AZD/09/USCOURTS@USCOURTS

Dear Committee members and invited guests,

The agenda materials are now available on [uscourts.gov](http://www.uscourts.gov/rules-policies/archives/agenda-books/advisory-committee-rules-evidence-october-2017) at the following link: <http://www.uscourts.gov/rules-policies/archives/agenda-books/advisory-committee-rules-evidence-october-2017>. Please let our office know if you have any issues accessing or downloading the materials. We look forward to seeing you in Boston!

Sincerely,

Bridget Healy
Attorney Advi or
Office of General Counsel, Rules Committee Staff

(b) (6)
(b) (6)



Review Article

Finding the way forward for forensic science in the US—A commentary on the PCAST report

I.W. Evett^{a,*}, C.E.H. Berger^b, J.S. Buckleton^{c,d}, C. Champod^e, G. Jackson^f^a Principal Forensic Services Ltd., 34 Southborough Road, Bickley, Bromley, Kent, BR1 2EB, United Kingdom^b Institute for Criminal Law and Criminology, Faculty of Law, Leiden University, PO Box 9520, 2300 RA Leiden, The Netherlands^c Environmental Science & Research Ltd, Private Bag 92021, Auckland 1142, New Zealand^d Department of Statistical Genetics, University of Washington, Box 357232 Seattle, WA 98195-7232, United States^e Ecole des Sciences Criminelles, Faculty of Law, Criminal Justice and Public Administration, Université de Lausanne, Batochime – quartier Sorge, CH-1015 Lausanne-Dorigny, Switzerland^f Abertay University, Dundee, DD1 1HG, United Kingdom

ARTICLE INFO

Article history:

Received 16 March 2017

Received in revised form 30 April 2017

Accepted 18 June 2017

Available online 26 June 2017

Keywords:

Forensic inference

Evidence

Comparison methods

Probability

Likelihood ratio

ABSTRACT

A recent report by the US President's Council of Advisors on Science and Technology (PCAST), (2016) has made a number of recommendations for the future development of forensic science. Whereas we all agree that there is much need for change, we find that the PCAST report recommendations are founded on serious misunderstandings. We explain the traditional forensic paradigms of *match* and *identification* and the more recent foundation of the logical approach to evidence evaluation. This forms the groundwork for exposing many sources of confusion in the PCAST report. We explain how the notion of treating the scientist as a black box and the assignment of evidential weight through error rates is overly restrictive and misconceived. Our own view sees inferential logic, the development of calibrated knowledge and understanding of scientists as the core of the advance of the profession.

© 2017 Elsevier B.V. All rights reserved.

Contents

In Memoriam	17
1. Introduction	17
2. The logical approach	17
2.1. Framework of circumstances	17
2.2. Propositions	17
2.3. Probability of the observations	17
3. The match paradigm	18
4. The identification paradigm	18
5. Misconceptions, fallacies and confusions in the PCAST report	18
5.1. Confusion between the match and identification paradigms	18
5.2. Judgement	19
5.3. Subjective versus Objective	19
5.4. Transposed conditional	19
5.5. "Probable match"	20
5.6. Foundational validity and accuracy	20
5.7. The PCAST paradigm	21
5.8. The scientist as a "black box"	21
5.9. Black box studies	21
5.10. Governance	22

* Corresponding author.

E-mail address: ianevett@btinternet.com (I.W. Evett).

6.	Our view of the future	22
6.1.	Logical inference	22
6.2.	Calibration	22
6.3.	Knowledge and data	22
7.	Conclusion	23
	References	23

In Memoriam

This paper is dedicated to the memory of Bryan Found who did so much to advance the profession of forensic scientist through his work on calibrating and enhancing the performance of experts under controlled conditions. He will be sorely missed.

1. Introduction

This paper is written in response to a recent report on forensic science of the US President's Council of Advisors on Science and Technology (PCAST) [1]. There have already been several responses to the report from the forensic community [2–7] which have resulted in an addendum to the report [8]. Our main concern is that the report (and its addendum) fails to recognise the advances in the logic of forensic inference that have taken place over the last 50 years or so. This is a serious omission which has led PCAST to a narrowly focussed and unhelpful view of the future of forensic science.

The structure of our paper is as follows. In Section 2 we briefly outline our view of the requirements imposed by logic on the assessment of the probative value of evidence. This allows us to set up a framework against which we can contrast some of the suggestions of the report. In Sections 3 and 4 we briefly explain the notions of “match” and “identification” paradigms that have underpinned much of forensic inference over the last century or so. Section 5 will point out misconceptions, fallacies, sources of confusion and improper terminology in the PCAST report. Our contrasting view of the future path for forensic science follows in Section 6.

2. The logical approach

Much has been written over the past 40 years on inference in forensic science. The frequency of appearance of articles, papers and books on the topic has increased markedly in recent years. Practically all of this material is founded on a logical, probabilistic approach to the assessment of the probative value of scientific observations [9,10]. The PCAST report mentions this body of work only briefly and pays scant attention to its principles [11], which we list and explain briefly as follows.

2.1. Framework of circumstances

It is necessary to consider the evidence within a framework of circumstances.

A simple example will illustrate this. Imagine that a sample¹ has been obtained from a crime scene which yielded a DNA profile from which the genotype of the originator of the sample has been inferred. A suspect for the crime is known to have the same genotype. Because the alleles revealed by a DNA profile will be found in different proportions in different ethnic groups, it is relevant to the assessment of the probative value of this

¹ The term “sample” is used generically to describe what is available for forensic examination. The term is not used here to suggest any statistical sampling process.

correspondence of genotypes that a credible eyewitness of the crime said that the offender was of a particular ethnic appearance.

It follows that, when presenting an evaluation, the scientist should clearly state the framework of circumstances that are relevant to their assessment of the probative value of the observations, with a caveat that, if details of the circumstances change, the evaluation must be revisited.

2.2. Propositions

The probative value of the observations cannot be assessed unless two propositions are addressed.

In a criminal trial, these will represent what the scientist believes the prosecution may allege and a sensible alternative that represents the defence position.² In taking account of both sides of the argument, the scientist is able to assess the evidence in a balanced, justifiable way and display to the court an unbiased approach, irrespective of which side calls the witness.

Propositions may be formed at any of at least four levels in a hierarchy of propositions [12–14]. These levels are termed offence, activity, source and sub source. We do not discuss these in any depth here. Most of the PCAST report appears to address questions at the source or sub source level. Examples of these would be:

1. Sub source: The DNA came from the person of interest (POI),³ or
2. Source: This fingerprint was made by the POI.

2.3. Probability of the observations

It is necessary for the scientist to consider the probability⁴ of the observations given the truth of each of the two propositions in turn.

The ratio of these two probabilities is widely known as the *likelihood ratio* (LR) and this is a measure of the weight of evidence that the observations provide in addressing the issue of which of the propositions is true. A likelihood ratio greater than one provides support for the truth of the prosecution proposition. A likelihood ratio less than one provides support for the truth of the defence proposition.

It cannot be sufficiently emphasized that it is the scientist's role to provide expert opinion on the probability of the *observations* given the proposition. The role of assigning a value to the probability of the *proposition* given the observations is that of the jury in a criminal trial. This probability will take account, not just of the scientific observations, but also of all of the other evidence presented at court.

² We recognise that the scientist, particularly at an early stage of proceedings, may not know the position that defence will take. It is common practice for the scientist to adopt what appears to be a reasonable proposition, given what is known of the circumstances—making it clear that this is provisional and subject to change at any time.

³ A source level DNA proposition would specify the nature of the recovered material, e.g. “the semen came from the POI”.

⁴ This could be a probability density, depending on the nature of the observations. But the principle remains unchanged.

3. The match paradigm

In most forensic comparisons, one of the items will be from a known origin (such as: a reference sample for DNA profiling from a particular individual; a pair of shoes from a suspect; a set of control fragments of glass from a broken window). The other will be from an unknown, or disputed origin (such as: DNA recovered from a crime scene; a footwear mark from the point of entry at a burglary; or a few small fragments of glass recovered from the clothing of a suspect). It is convenient to refer to these as the *reference* and *questioned* samples, respectively. The matter of interest to the court relates to the origin of the questioned sample. This question will be addressed scientifically by carrying out observations on both samples. These observations may be purely qualitative: such as, for example, the shapes of the loops of letters such as “y” and “g” in a passage of handwriting. They may be quantitative and discrete, such as the alleles in a DNA STR profile. Or they may be quantitative and continuous, such as the refractive index of glass fragments. The match paradigm calls for a judgement, by the scientist, as to whether or not the two sets of observations agree within the range of what would be expected if the questioned sample had come from the same origin as the reference sample. The basis for that judgement may, in the case of quantitative observations, be based on a set of pre-determined criteria; but where the observations are qualitative such criteria may be vague or purely judgemental.

If the two sets of observations are considered to be outside the range of what may have been expected if the two samples had come from the same source then the result may be reported as a “non match”. Depending on the nature of the observations, this provides the basis for a strong implication that the questioned and reference samples came from different sources. In many instances this conclusion will be non-controversial in the sense that prosecution and defence will be content to accept it.

However, when the result of the comparison is a “match” it does not logically follow that the two samples do share the same source or even that they are likely to be from the same source. It is possible that the two samples came from two different sources that, by coincidence, have similar properties. Throughout the history of forensic science there has been the notion – often imperfectly expressed – that the smaller the probability of such a coincidence, the greater the evidential value to be associated with the observed match. In DNA profiling, for example, we encounter the notion of a “match probability”. The implication of this approach is that the jury should assign an evidential weight that is related to the inverse of the match probability.

The logical approach has done much to clarify the rather woolly inference that historically has been associated with the match paradigm but it has also demonstrated the considerable advantages of the single stage approach implied by the assignment of weight through the calculation of the likelihood ratio, over the rather clumsy and inefficient two stage approach implied by the match paradigm. This has already been pointed out by Morrison et al. [4].

4. The identification paradigm

Historically, fingerprint comparison was seen to be the gold standard by which the power of any other forensic technique could be judged. The paradigm here was the notion of “identification”⁵ or

⁵ Kirk [15] defined the term identification as only placing an object in a restricted class. The criminalist would, for example, identify a particular mark as a fingerprint. Individualization was defined by Kirk as establishing which finger left the mark. An opinion of the kind “this latent mark was made by the finger which made this reference print” is an individualization.

“individualization” (the terms are used synonymously here). Provided that sufficient corresponding detail was observed, the outcome of a comparison between a fingerprint of questioned origin and a print taken from a known person would be reported as a categorical opinion: the two were definitely made by the same person.

So, the match and identification paradigms are related with the difference that in the latter the scientist is allowed to state that the match probability is so infinitesimally small that it is reasonable to conclude that the two items came from the same source. Historically, many examiners would have claimed that the source was established with certainty to the exclusion of all others.

The identification paradigm went largely unchallenged for many years until later in the 20th century when its logical basis was questioned (see, for example, [16] or more recently [17,18]) and also when, in a number of high profile cases, misidentifications with serious consequences were exposed.

An example of the paradigm is given in box 6, p. 137 of the PCAST report (DOJ proposed uniform language) (emphasis added).

The examiner may state that it is his/her opinion that the shoe/tire *is the source of the impression* because there is sufficient quality and quantity of corresponding features such that the examiner would not expect to find that same combination of features repeated in another source. This is the highest degree of association between a questioned impression and a known source.

The PCAST report rightly indicates that the conclusions conveying “100 percent certainty” or “zero or negligible error rates” are not scientifically defensible. Such conclusions tend to overestimate the weight to be assigned to the forensic observations.

5. Misconceptions, fallacies and confusions in the PCAST report

The most serious weakness in the PCAST report is their flawed paradigm for forensic evaluation. Unfortunately, the report contains more misconceptions, fallacies, confusions and improper wording. In this section we will discuss the main problems with the report.

5.1. Confusion between the match and identification paradigms

This is the first source of confusion in the report. For example, from p. 90 of the report (emphasis added):

An FBI examiner concluded with “100 percent certainty” that the fingerprint *matched* Brandon Mayfield . . . even though Spanish authorities were unable to confirm the *identification*.

On p. 48 we find (emphasis added):

To meet the scientific criteria of foundational validity, two key elements are required:

(1) a reproducible and consistent procedure for (a) identifying features within evidence samples; (b) comparing the features in two samples; and (c) determining based on the similarity between the features in two samples, whether the samples should be declared to be a proposed *identification* (“*matching rule*”).

We have seen that declaring a match and declaring an identification are not the same thing. Declaring a match implies nothing about evidential weight whereas declaring an identification implies evidential weight amounting to complete certainty.

The PCAST report proposes an approach that is fusion of the match and identification paradigms. See, from p. 45/46:

Because the term “match” is likely to imply an inappropriately high probative value, a more neutral term should be used for an examiner’s belief that two samples came from the same source. We suggest the term “proposed identification” to appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong. We will use this term throughout the report.

If a scientist says that the questioned and reference samples match, the immediate inference to be drawn from this (as we have explained) is that they might have come from the same source but it is also true that they might not have come from the same source. These two statements make no implication with regard to evidential weight. Weight only comes from the second stage of the paradigm which entails coming up with some impression of rarity. The identification paradigm, on the other hand, is different in that it implies a statement of certainty: the two samples certainly came from the same source.

The PCAST paradigm requires that the scientist should make a categorical statement (an identification) that cannot be justified on logical grounds as we have already explained. Most scientists would be comfortable with the notion of observing that two samples *matched* but would, rightly, refuse to take the logically unsupportable step of inferring that this observation amounts to an *identification*.

5.2. Judgement

The report emphasises the value of empirical data (emphasis added):

The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, *is not a matter of ‘judgment’*. It is an empirical matter *for which only empirical evidence is relevant.* ([1], p. 6)

This denial of the importance of judgement betrays a poor understanding of the nature of forensic science. We offer a simple example.

Mr POI is the suspect for a crime who was arrested at time T in location Z . Some questioned material has been found on the clothing of Mr POI which is to be compared with reference material taken from the crime scene. Denote the observations on the two samples by y and x respectively. Whichever paradigm we follow, we are interested in the probability of finding material with observations y on the clothing of Mr POI if he had nothing to do with the crime. Ideally, of course, we would like a survey carried out near to time T and in the general region of Z and of people of a socio economic group Q that would include Mr POI. But this is, of course unrealistic. What we do have is a survey of materials on clothing carried out at some earlier time T' and at another location Z' and of a slightly different socio economic group Q' . Who is to make a judgement on the relevance of this survey data to the case at hand? We would argue that this is where the knowledge and understanding of the forensic scientist is of crucial importance.

The reality is, of course, that the perfect database never exists. The council is wrong: it is most certainly *not* the case that “only empirical evidence” is relevant. Without downplaying the importance of data collections, they can only inform judgement it is judgement that is paramount and informed judgement is founded in reliable knowledge.

5.3. Subjective versus Objective

PCAST give their definition of the distinction between “objectivity” and “subjectivity” p. 5 footnote 3.

Feature comparison methods may be classified as either objective or subjective. By objective feature comparison methods, we mean methods consisting of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising little or no judgment. By subjective methods, we mean methods including key procedures that involve significant human judgment . . .

What is suggested is that many of the decisions be moved from the examiner to the procedure and/or software. The procedure or software will have been written by one or more people and the decisions about what models are used or how decisions are made are now enshrined in paper or code. Hence all the subjective judgements are now made by this person or group of people via the paper or code. Whereas this approach could be viewed as repeatable and reproducible, the objectivity is illusory.

In the US environment, subjectivity has been associated with bias and sloppy thinking, and objectivity with an absence of bias and rigorous thinking. It is worthwhile examining whence the fear of subjectivity arises. There is considerable proof that humans are susceptible to quite a number of cognitive effects many of which can affect judgement. We suspect that the fear is that these effects bias the decisions in ways that are detrimental to justice. Hence, it is bias arising from cognitive effects that is the enemy, not subjectivity.

If we return to the concept of enforced precision, we could assume that trials could be conducted on such a system and that the outputs could be calibrated. Such a system could be of low susceptibility to bias arising from cognitive effects. We suspect that these are the goals sought by PCAST. We certainly could support calibrating subjective judgements but we see little value in pretending that writing them down or coding them makes them objective.

5.4. Transposed conditional

We are concerned by the report’s poor use of the notion of probability. In particular we note in the report many instances where the fallacy of the transposed conditional either occurs explicitly or is implied. We have seen that the logic of forensic inference directs us to assign a value to the probability of the observations given the truth of a proposition. The probability of the truth of a proposition is for the jury *not* the scientist. Confusion between these two different probabilities has been called the “prosecutor’s fallacy” [19]. We prefer the term *transposed conditional* because, in our experience, the fallacy is regularly committed by prosecutors, defence attorneys, the judiciary and the media alike.

The fallacy is widespread, even though it can be grounds for a retrial if given in testimony by an expert witness. The document [20] that attempts to explain DNA statistics to defence attorneys in the US describes incorrectly a likelihood ratio for a mixture profile as:

4.73 quadrillion times more likely⁶ to have originated from [suspect] and [victim/complainant] than from an unknown individual in the U.S. Caucasian population and [victim/complainant].” ([20], p. 52)

⁶ We are fully aware of the distinction made in statistical theory between “likelihood” and “probability”. We believe that attempting to explain that distinction in this paper would cause more confusion than the worth of it. It is our experience that in courts of law the two terms are taken to be synonymous.

This is a classic example of the transposed conditional. It is a transposition of the likelihood ratio, which would be more correctly presented as follows:

The DNA profile is 4.73 quadrillion times more likely to be obtained if the DNA had originated from the suspect and the victim/complainant rather than if it had originated from an unknown individual in the U.S. Caucasian population and the victim/complainant.

The contrast between these two statements, though apparently subtle, is profound. The first is an expression of the probability (or odds) that a particular proposition is true – this, we have seen, is the probability that the jury must address, not the scientist.⁷ The second considers the probability of the *observations*, given the truth of one proposition then the other, which is the appropriate domain for the expertise of the scientist. It is important to realise that the first statement is not a simple rephrasing of the second statement. Whereas the second may be a valid representation of the scientist's evaluation in a given case, the first most definitely cannot be.

Consider the following quote from the first paragraph on footwear methodology in the PCAST report ([1], p. 114):

Footwear analysis is a process that typically involves comparing a known object, such as a shoe, to a complete or partial impression found at a crime scene, to assess whether the object is likely to be the source of the impression.

This is wrong. We state again that it is not for the scientist to present a probability for the truth of the proposition that the object was the source of the impression. The scientist addresses the probability of the outcome of the comparison *if* the object were the source of the impression: this probability forms the numerator of the likelihood ratio. Just as important, of course, is the probability of the outcome of the comparison *if* some other object were the source of the impression. The latter forms the denominator of the likelihood ratio. It is the two probabilities, taken together, that determine the evidential weight in relation to the two propositions of interest to the court.

The PCAST report sentence clearly states that the objective of the footwear analysis is to present a probability for the proposition given the observations, and not for the observations given the proposition. This is clearly a transposition of the conditional.

Similarly, the scientist is not in a position to consider the probability addressed in the following ([1], p. 65 and repeated on p. 146):

... determining, based on the similarity between the features in two sets of features, whether the samples should be declared to be likely to come from the same source ...

We have seen that it is not for the scientist to consider the probability that the samples came from the same source given the observation of a “match”. It is another example of the fallacy of the transposed conditional.

This confusion is systematic in the original report and we note that it continues into the addendum ([8], p. 1) (emphasis added):

These methods seek to determine whether a questioned sample *is likely to come* from a known source based on shared features in certain types of evidence.

We have seen that this is most certainly *not* what a feature comparison should aspire to. It is not the role of the forensic

scientist to offer a probability for the proposition that a questioned sample came from a given source since this would require the scientist to take account of all of the non scientific information which properly lies within the domain of the jury.

The need for precision of language when presenting probabilities is exemplified by two quotations from the report. First, from p. 8 when talking about the interpretation of a DNA profile:

Could a suspect's DNA profile be present within the mixture profile? And, what is the probability that such an observation might occur by chance?

As we read it, this second sentence can be taken to mean:

What is the probability that such an observation would be made if the suspect's DNA were not present in the mixture?

Within the logical paradigm, this is a legitimate question to ask – it is the probability of the observations given that one of the propositions were true.

However, later in the report we find (p. 52):

the random match probability – that is, the probability that the match occurred by chance”.

There is an economy of phrasing here that obscures meaning and the reader could be forgiven for believing that the question implied by the second phrase is:

What is the probability that the two samples had come from different sources and matched by chance?

This is a probability of a proposition (the two samples came from different sources) given the observation (a match) and would imply a transposed conditional. We are aware that the council may respond that this is not at all what they meant – to which we would respond that the council should have been far more careful in its phraseology.

5.5. “Probable match”

In giving their definition of the distinction between “objectivity” and “subjectivity” p. 5 – see footnote 3 the report states:

how to determine whether the features are sufficiently similar to be called a probable match.

The council do not say what they mean by a “probable match” but it seems to us that it is another example of confusion between the match and identification paradigms. Following the match paradigm there is no such thing as a probable match – the two samples either match or they do not.

5.6. Foundational validity and accuracy

The report distinguishes two types of scientific validity: “foundational validity” and “validity as applied”. We confine ourselves to the first of these (p. 4):

Foundational validity for a forensic science method requires that it be shown based on empirical studies to be *repeatable, reproducible, and accurate*, at levels that have been measured and are appropriate to the intended application. Foundational validity, then, means that a method can, *in principle*, be reliable.

Repeatability refers to the ability of the same operator with the same equipment to obtain the same (or closely similar) results when repeating analysis of the same material. Reproducibility refers to the ability of the equipment to obtain the same (or closely similar) results with different operators. As such, both are

⁷ In Bayesian terms, the first statement is one of posterior odds. This can be derived from the second statement either by assigning prior odds of one (which would be highly prejudicial in most criminal trials) or by making the mistake of transposing the conditional. Neither is acceptable behaviour for a scientist.

expressions of precision, which is how close each measurement or result is to the others.

Accuracy is a measure of how close one or a set of measurements is to the true answer. This has an obvious meaning when we know or could know the true answer. We could imagine some measurement such as the weight of an object where that object has been weighed by some very advanced technique and we can accept that as the “true” weight. We wish then to consider the accuracy of some other, perhaps cheaper, technique. We could assess the accuracy of this second technique by using it to weigh the object multiple times and observing the deviation of the results from the “true” weight of the object.

For some questions in forensic science, such as “How much heroin is in this seized sample?” or “How much ethanol is in this blood sample?”, the notion of the accuracy of an applied analytical technique is relevant because it is possible to assess a technique’s accuracy using trials with known quantities of heroin or ethanol. However, when it comes to answering a question such as “What is the probability that there would have been a match with a suspect’s shoe if it did not make the mark at the scene of crime?”, then there is no sense in which there is a “true answer”. The values that experts assign for such probabilities will vary depending on the specific knowledge of the experts and the nature of any databases that experts may use to inform their probabilities.

We could use a weather forecaster as an illustration. If she says that there is a 0.8 probability of a sunny day tomorrow, there can be no sense in which this is a “true” statement. Equally, if tomorrow brings rain, she is not “wrong” in any sense. Nor is she “inaccurate”. A probabilistic statement of this nature may be unhelpful or misleading, in the sense that it may lead us to make a poor decision, but it cannot be either true or false.

Once we abandon the idea of a true answer for probabilities, we are left with the difficult question of what we mean by accuracy. We suggest that the report does a disservice to the important task of calibrating probabilities by a simplistic allusion to accuracy.

The PCAST report says (p. 46):

Without appropriate estimates of accuracy, an examiner’s statement that two samples are similar or even indistinguishable is scientifically meaningless; it has no probative value, and considerable potential for prejudicial impact. Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy.

We have seen that the report is wrong here—it is not a matter of “accuracy” but of evidential weight.

5.7. The PCAST paradigm

The PCAST report proposes an approach that is fusion of the match and identification paradigms. See, from p. 45/46:

Because the term “match” is likely to imply an inappropriately high probative value, a more neutral term should be used for an examiner’s belief that two samples came from the same source. We suggest the term “proposed identification” to appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong. We will use this term throughout the report.

First, we have seen that the term “match”, if used properly, makes no implication of probative value: it implies that the two samples might have come from the same source but also might have come from different sources. This is evidentially neutral. Second, we have seen that there is no place for the “examiner’s

belief that two samples came from the same source”: it is not for the scientist to assign a probability to the proposition that the two samples came from the same source.

Next we must consider what the council understand the phrase “proposed identification” to mean. Do they mean that, because it is an identification, it is a categorical opinion? Note that the qualifier “proposed” does not make the identification less than categorical—if it were probabilistic it could not be “wrong”.⁸ If it is not probabilistic then the scientist is to provide a categorical opinion while telling the court that he/she might be wrong! It is difficult to believe that any professional forensic scientist would be happy to be put in this position.

5.8. The scientist as a “black box”

On page 49 we find:

For subjective methods, procedures must still be carefully defined but they involve substantial human judgment. For example, different examiners may recognize or focus on different features, may attach different importance to the same features, and may have different criteria for declaring proposed identifications. Because the procedures for feature identification, the matching rule, and frequency determinations about features are not objectively specified, the overall procedure must be treated as a kind of “black box” inside the examiner’s head.

The report justifiably emphasises weaknesses of qualitative opinions. The intuitive “black box” view of the scientist will certainly have been true in many instances in the past and, indeed, in certain quarters in the present day. But for us the solution is emphatically not to continue to treat this as an acceptable state of affairs for the future. The PCAST view appears to be “it’s a black box, so let’s treat it like a black box”. Our approach has been, and will continue, to break down intuitive mental barriers by expanding transparency, knowledge and understanding. We do not see the future forensic scientist as an *ipse dixit* machine whatever the opinion, we expect the scientist to be able to explain it in whatever detail is necessary for the jury to comprehend the mental processes that led to it.

5.9. Black box studies

That the council intend the proposed identification to be categorical is clarified in the following from page 49 (emphasis added):

In black box studies, many examiners are presented with many independent comparison problems—typically, involving “questioned” samples and one or more “known” samples—and asked to declare whether the questioned samples came from the same source as one of the known samples.⁹ The researchers then determine how often examiners reach erroneous conclusions.

PCAST proposes that the error rates from such experiments would be used to assign evidential value at court.

We are strongly against the notion that the scientist should be forced into the position of giving categorical opinions in this way. Whereas, we are strongly in favour of the notion of calibrating the

⁸ Though, of course, it would be logically incorrect because it would imply a transposed conditional.

⁹ In footnote 111 the report says: “Answers may be expressed in such terms as “match/no match/inconclusive” or “identification/exclusion/inconclusive”. This strengthens our belief that the council see match and identification as interchangeable”.

opinions of forensic scientists under controlled conditions we see those opinions expressed in terms of statements of evidential weight. We return to the subject of calibration later.

5.10. Governance

PCAST suggests that forensic science should be governed by those, such as metrologists, from outside the profession. This speaks to the view, reinforced by a very selective reference list, that the forensic science discipline is not to be trusted with developing procedures, testing them, and self governance. We do not reject input from outside the profession: we welcome it. But our own observations are that those outside may be engaged to different extents, varying from a passing interest to years of study. They may be unduly influenced by headlines in newspapers highlighting or exaggerating deficiencies. On occasion, these same commentators from outside the profession may not recognise the limitations in their own knowledge base where it concerns specifically forensic aspects, may be reticent to consult subject matter experts from amongst practising scientists and may give well intentioned, but erroneous, advice [1,21].

6. Our view of the future

6.1. Logical inference

The recommendations of the PCAST report are founded on a conflation of two classical forensic paradigms: match and identification. These paradigms are as old as forensic science but their inadequacies and illogicalities have been comprehensively exposed over the last 50 years or so. All of us maintain, and have done so in our writings, that the future of forensic science should be founded first on the notion of logical inference and second on the notion of calibrated knowledge. The former leads to a framework of principles (which have been adopted by ENFSI) and we are disappointed that PCAST has apparently chosen to ignore, or at most pay lip service to, this fundamental change. The second is a deeper and far richer concept than the profoundly limited notion of false positive and false negative error rates: this is the notion of *calibration*.

6.2. Calibration

We are most definitely in favour of the studying of expert opinion under controlled circumstances, see for example Evett [22] but proficiency testing is far more than the counting of errors. The PCAST black box approach calls for a categorical opinion that is recorded as right or wrong but we have seen that forensic interpretation is far richer and more informative than simple yes/no answers. In a source level proficiency test we expect the participants to respond with a statement of evidential weight in relation to one of two clearly stated propositions. Support thus expressed for a proposition that is, in fact, false is undesirable because it is misleading not “wrong”. Obviously, the desirable outcome of the proficiency test is a small value for the expected weight of evidence in relation to a false proposition. But whatever the outcome, the study must be seen as a learning exercise for all participants: the pool of knowledge has grown. The notion of an error rate to be presented to courts is misconceived because it fails to recognise that the science moves on as a result of proficiency tests. The work led by Found and Rogers [23] has shown how the profession of handwriting comparison in Australia and New Zealand has grown in stature because of the culture of advancing knowledge through repeated study under controlled conditions. To repeat then, our vision is not of the black box/error rate but of continuous development through calibration and feedback of opinions.

A striking example of forensic calibration is the evolution of fingerprints evidence from the identification paradigm to the logical paradigm via mathematical modelling [24,25]. Instead of the categorical identification, we have a mathematical approach that leads to a likelihood ratio. The validation of such approaches is founded on two desiderata: we require large likelihood ratios in cases in which the prosecution proposition is true; and small likelihood ratios in cases in which the defence proposition is true. Investigation of performance in relation to these two desiderata is undertaken by considering two sets of comparisons: one set in which it is known that the two samples came from the same source; and one set in which it is known that the two samples came from different sources. There have been major advances over recent years in how the likelihood ratio distributions from such experiments may be compared and evaluated (Ramos [26], Brümmer [27] see also Robertson et al. [28] for a layman’s introduction to calibration). The elegance and performance of such methods far transcends the crude PCAST notion of “false positive” and “false negative” error rates.

6.3. Knowledge and data

The PCAST report focuses on “feature comparison” methods and, as we have explained, this has meant that it is concerned with inference relating to source level propositions. At this level, the report sees data as the sole means for assigning probabilities. An important part of the role of the forensic scientist is concerned with inference with regard to activity level propositions. Consider, for example, a question of the form “what is the probability of finding this number of fragments of glass on Mr POI’s jacket if he is the person who smashed the window at the crime scene?” The answer is heavily dependent on circumstantial information (how large is the window? where was the person who smashed the window standing? was any implement used? how much time elapsed between the breaking of the window and the seizure of the jacket from Mr POI? etc.) and the variation in this between cases is vast. There is no single database to inform such probabilities. The scientist will, it is hoped, be thoroughly familiar with all of the published literature on glass transfer in crime cases [29] and may, if resources permit, carry out experiments that reproduce the current case circumstances. The knowledge and judgement of other scientists who have encountered similar questions is also relevant. We agree with PCAST that length of experience is not a measure of reliability of scientific opinion: the foundation is reliable knowledge. Too little effort has been devoted within the forensic sphere thus far to the harnessing of knowledge through knowledge based systems but see [29] for examples of how such a system was created for glass evidence interpretation.

We do not deny the importance of data collections but the view that data may replace judgement is misconceived. A data collection should be used to inform reliable knowledge not replace it.

We have explained that our view of the scientist is the antithesis of the PCAST “black box” automaton. Although there is a need for data, PCAST are mistaken in seeing it as the be all and end all: qualitative judgement will always be at the centre of forensic science evidence evaluation. We reject the PCAST vision of the scientist who gives a categorical opinion and a statement about the probability that the opinion is wrong. We see the model scientist as deeply knowledgeable about her domain of expertise and able to rationalise the opinion in terms that the jury will understand. The principles have been expressed elsewhere [11] as balance, logic, robustness and transparency. There is no place for the black box. We agree that the scientist should be able to provide the court with evidence of performance under controlled conditions. Found and Rogers [23] have provided a model for handwriting comparison

and we see such approaches as extending into other areas: the emphasis is on calibration of probabilistic assessments.

7. Conclusion

The 44th US president's request was "to consider whether there are additional steps that could usefully be taken on the scientific side to strengthen the forensic science disciplines and ensure the validity of forensic evidence used in the Nation's legal system" ([1], p. 1). We suggest that the report has very little emphasis on positive steps and does much to reinforce poor thinking and terminology.

Our own view of the future of forensic science is based on the principle that forensic inference should be founded on a logical framework for reasoning in the face of uncertainty. That framework is provided by probability theory coupled with the recognition that probability is necessarily subjective and conditioned by knowledge and judgement. It follows that our view of the forensic scientist is a knowledgeable, logical and reasonable person. Whereas data collections are valuable they should be viewed within the context of reliable knowledge. The overarching paradigm of reliable knowledge should be founded on the notion of knowledge management, including comprehensive systems for the calibration of expert opinion.

References

- [1] President's Council of Advisors on Science and Technology, Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, Washington DC, 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.
- [2] Federal Bureau of Investigation—FBI, Comments on: President's Council of Advisors on Science and Technology Report to the President on Forensic Science in Federal Criminal Courts: Ensuring Scientific Validity of Pattern Comparison Methods. September 20, 2016. www.fbi.gov/file-repository/fbi-pcast-response.pdf/view.
- [3] National District Attorneys Association—NDAA, Report Entitled Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. November 16, 2016. <http://www.ciclt.net/ul/ndaajustice/PCAST/NDAAPCASTResponseFINAL.pdf>.
- [4] G.S. Morrison, D.H. Kaye, D.J. Balding, D. Taylor, P. Dawid, C.G.G. Aitken, S. Gittelson, G. Zadora, B. Robertson, S. Willis, S. Pope, M. Neil, K.A. Martire, A. Hepler, R.D. Gill, A. Jamieson, J. de Zoete, R.B. Ostrum, A. Caliebe, A comment on the PCAST report: skip the match/non-match stage, *Forensic Sci. Int.* 272 (2017) e7–e9, doi:<http://dx.doi.org/10.1016/j.forsciint.2016.10.018>.
- [5] Association of Firearm and Tool Mark Examiners AFTE, Response to PCAST Report on Forensic Science. October 31, 2016. https://afte.org/uploads/documents/AFTE_PCAST_Response.pdf.
- [6] Bureau of Alcohol Tobacco Firearms and Explosives—ATF, ATF Response to the President's Council of Advisors on Science and Technology Report. September 21, 2016. https://www.theiai.org/president/20160921_ATF_PCAST_Response.pdf.
- [7] The International Association for Identification (IAI), IAI Response to the President's Council of Advisors on Science and Technology Report, 2016. https://www.theiai.org/president/IAI_PCAST_Response.pdf.
- [8] President's Council of Advisors on Science and Technology, An addendum to the PCAST report on forensic science in criminal courts, Washington DC, 2017. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf.
- [9] C.G.G. Aitken, F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd ed, John Wiley & Sons Ltd., Chichester, 2004.
- [10] C. Aitken, P. Roberts, G. Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*, London, 2011. <http://www.rss.org.uk/uploadedfiles/userfiles/files/Aitken-Roberts-Jackson-Practitioner-Guide-1-WEB.pdf>.
- [11] Expressing evaluative opinions: a position statement, *Sci. Justice* 51 (1) (2011) 1–2, doi:<http://dx.doi.org/10.1016/j.scijus.2011.01.002>.
- [12] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, *Sci. Justice* 38 (3) (1998) 151–156, doi:[http://dx.doi.org/10.1016/S1355-0306\(98\)72099-4](http://dx.doi.org/10.1016/S1355-0306(98)72099-4).
- [13] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, *Sci. Justice* 38 (4) (1998) 231–240, doi:[http://dx.doi.org/10.1016/S1355-0306\(98\)72117-3](http://dx.doi.org/10.1016/S1355-0306(98)72117-3).
- [14] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, Case pre-assessment and review in a two-way transfer case, *Sci. Justice* 39 (2) (1999) 103–111, doi:[http://dx.doi.org/10.1016/S1355-0306\(99\)72028-9](http://dx.doi.org/10.1016/S1355-0306(99)72028-9).
- [15] P.L. Kirk, The ontogeny of criminalistics, *J. Crim. Law Criminol. Police Sci.* 54 (1963) 235–238.
- [16] D.A. Stoney, What made us ever think we could individualize using statistics, *J. Forensic Sci. Soc.* 31 (2) (1991) 197–199, doi:[http://dx.doi.org/10.1016/S0015-7368\(91\)73138-1](http://dx.doi.org/10.1016/S0015-7368(91)73138-1).
- [17] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, *Forensic Sci. Int.* 177 (2–3) (2008) 120–132, doi:<http://dx.doi.org/10.1016/j.forsciint.2007.11.008>.
- [18] A. Biedermann, S. Bozza, F. Taroni, The decisionalization of individualization, *Forensic Sci. Int.* 266 (2016) 29–38, doi:<http://dx.doi.org/10.1016/j.forsciint.2016.04.029>.
- [19] W.C. Thompson, E.L. Schumann, Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy, *Law Hum. Behav.* 11 (3) (1987) 167–187, doi:<http://dx.doi.org/10.1007/BF01044641>.
- [20] E.H. Holder, M.L. Leary, J.H. Laub, DNA for the Defense Bar, U.S. Department of Justice Office of Justice Programs, Washington, DC, 2012.
- [21] National Research Council - Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science*, National Academy Press, Washington, D.C, 1992.
- [22] I. Evett, The logical foundations of forensic science: towards reliable knowledge, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370 (1674) (2015), doi:<http://dx.doi.org/10.1098/rstb.2014.0263>.
- [23] B. Found, D. Rogers, The initial profiling trial of a program to characterize forensic handwriting examiners' skill, *J. Am. Society of Questioned Document Examiners* 6 (2) (2003) 72–81.
- [24] C. Champod, C.J. Lennard, P.A. Margot, M. Stoilovic, *Fingerprints and other Ridge Skin Impressions*, CRC Press, Boca Raton, 2016.
- [25] C. Neumann, I.W. Evett, J. Skerrett, Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, *J. Roy. Stat. Soc. Ser. A. (Stat. Soc.)* 175 (Part 2) (2012).
- [26] D. Ramos, J. Gonzalez-Rodriguez, G. Zadora, C. Aitken, Information-theoretical assessment of the performance of likelihood ratio computation methods, *J. Forensic Sci.* 58 (6) (2013) 1503–1518, doi:<http://dx.doi.org/10.1111/1556-4029.12233>.
- [27] N. Brümmner, J. du Preez, Application-independent evaluation of speaker detection, *Comput. Speech Language* 20 (2006) 230–275, doi:<http://dx.doi.org/10.1016/j.csl.2005.08.001>.
- [28] G.A. Robertson, C.E.H. Vignaux, *Interpreting Evidence—Evaluating Forensic Science in the Courtroom*, 2nd ed., John Wiley & Sons, Ltd., Chichester, 2016.
- [29] J.M. Curran, T.N. Hicks, J.S. Buckleton, *Forensic Interpretation of Glass Evidence*, CRC Press LLC, Boca Raton, 2000.



June 17, 2017

To whom it may concern:

When the President's Council of Advisors on Science and Technology (PCAST) Report first was published in 2016, it was obvious that the report was not particularly helpful from a scientific perspective as it was myopic, full of error, and did not provide data to support its contentions. A more significant concern regarding the failings of the PCAST Report was that it claimed its focus was on science, but obviously was dedicated substantially to policy. Initially I considered writing a critique about the failings of the PCAST Report to assist the community. But the problems with this report were so obvious that I did not think it would be necessary to devote time to such an effort. Indeed my prediction was correct in that the report would be (and has been) rejected by the scientific community as well as overwhelmingly by the courts. However, the PCAST Report is being relied on by the Public Defender Service in U.S. v. Benito Valdez (Motion to Exclude the Testimony of the Government's proposed expert witness in Firearms Examination and Memorandum of Points and Authorities in Support, dated June 2, 2017) as a scientifically sound review of the state of the forensic sciences. Therefore, it has become necessary to address the serious limitations of the PCAST Report and convey that it is an unsound, unsubstantiated, non-peer-reviewed document that should not be relied upon for supporting or refuting the state of the forensic sciences.

My credentials to be able to opine on the failings of the PCAST Report are based on my work of more than 30 years in research, development, validation, and implementation of DNA typing methodologies for forensic applications (my CV is attached). I received a Ph.D. in Genetics in 1979 from Virginia Polytechnic Institute and State University. From 1979-1982, I was a postdoctoral fellow at the University of Alabama at Birmingham and carried out research predominately on genetic risk factors for such diseases as insulin dependent diabetes mellitus, melanoma, and acute lymphocytic leukemia. In 1983, I joined the research unit at the FBI Laboratory Division to carry out research, development, and validation of methods for forensic biological analyses. The positions I held at the FBI include: research chemist, program manager for DNA research, Chief of the Forensic Science Research Unit, and the Senior Scientist for the Laboratory Division of the FBI. I have contributed to the fundamental sciences as they apply to forensics in analytical development, population genetics, statistical interpretation of evidence, and in quality assurance. Some of my technical efforts have been: 1) development of analytical assays for typing myriad protein genetic marker systems, 2) designing electrophoretic instrumentation, 3) developing molecular biology analytical systems to include RFLP typing of VNTR loci and PCR-based SNP, VNTR and STR assays, and direct sequencing methods for mitochondrial DNA, 4) new technologies such as use of massively parallel sequencing; and 5) designing image analysis systems. I worked on laying some of the foundations for the current

statistical analyses in forensic biology and defining the parameters of relevant population groups. I have published approximately 600 articles (more than any other scientist in the area of forensic genetics), made more than 730 presentations (many of which were as an invited speaker at national and international meetings), and testified in well over 250 criminal cases in the areas of molecular biology, population genetics, statistics, quality assurance, validation, and forensic biology. In addition, I have authored or co-authored books on molecular biology techniques, electrophoresis, protein detection, forensic genetics, and microbial forensics. I was directly involved in developing the quality assurance standards for the forensic DNA field in the United States. I have been a chair and member of the Scientific Working Group on DNA Methods, Chair of the DNA Commission of the International Society of Forensic Genetics, and a member of the DNA Advisory Board. I was one of the original architects of the CODIS National DNA database, which maintains DNA profiles from convicted felons, from evidence in unsolved cases, and from missing persons.

Some of my efforts over the last 16 years also are in counter terrorism, including identification of victims from mass disasters, microbial forensics and bioterrorism. I was an advisor to New York State in the effort to identify the victims from the WTC attack. In the area of microbial forensics, I was the chair of the Scientific Working Group on Microbial Genetics and Forensics, whose mission was to set QA guidelines, develop criteria for biologic and user databases, set criteria for a National Repository, and develop forensic genomic applications. I also have served on the Steering Committee for the Colloquium on Microbial Forensics sponsored by American Society of Microbiology, was an organizer of four Microbial Forensics Meetings held at The Banbury Center in the Cold Spring Harbor Laboratory, and participated on several steering committees for NAS sponsored meetings.

In 2009 I became Executive Director of the Institute of Applied Genetics and Professor at the University of North Texas Health Science Center at Fort Worth, Texas. I currently direct the Center for Human Identification. I also direct an active research program in the areas of human forensic identification, microbial forensics, emerging infectious disease, human microbiome, molecular biology technologies, and pharmacogenetics (or molecular autopsy). I also currently am an appointed member of the Texas Forensic Science Commission.

Of note, the PCAST Committee relied on my work and as a noted expert which is supported by the report's citation of my work several times all in a favorable manner. Indeed, I am the scientist at the FBI that is mentioned as Dr. Lander's co-author to bolster his credentials in the forensic sciences (see footnotes 17 and 20). My work is cited in footnotes 33, 149, 183, 185, 187, and 209.

The report lacks scientific substance. It is cloaked with a veneer of science but in actuality is an attempt to set policy. The report discusses and advocates validation (a topic all should agree is important). Yet the topic is only addressed superficially providing definitions that already are well known with generalizations and terms it calls criteria. Nothing novel was provided by the report (see examples in references 1-7 that already have discussed the same criteria but to a greater degree than in the report). Moreover, the report does not provide any substantial guidance on how to perform validation studies for any of the disciplines it addresses. There are basic validation criteria such as sample size, power analyses, types of samples, sensitivity, specificity, dynamic range, purity of analyte, etc. that the report does not address per se or only touches upon (and instead uses black box studies for its only endeavor into sampling uncertainty and for a

misguided attempt at addressing the potential for error). The PCAST Committee could have done a service to the community if it had selected some validation studies that it claims to have reviewed (although such claims are suspect as there is no documentation supporting the claims) and described specifically those studies that the PCAST Committee deemed inappropriate and/or inadequate. Then, the PCAST Committee could have laid out how those studies should have been performed with the real substantive criteria and examples that are necessary to perform a validation study. Leading by example would have been helpful; instead the report just dismisses most of the work performed in 2000 plus articles that it claims (sic) to have reviewed. The report criticizes the forensic community for a lack of validation studies but does not describe what is lacking in any substantive way.

The Report does not describe data from each of the disciplines that could be relied upon. It is difficult to believe that in 2000 papers, the PCAST Committee claims to have relied upon, that there are no data of value. There are no indications that the PCAST Committee actually assessed the data in the literature. There is little if any documentation in this regard which should be extremely troubling to all given the PCAST Committee's strong positions of the importance of validation, documentation, and peer-reviewed publication for the forensic science community. The PCAST Committee clearly takes a "do as I say, not as I do" position. The report contains no discussion on the criteria that were used to assess the literature, the criteria that were used to dismiss the literature as inadequate, and no documentation that any data (if existing) are readily available to support that the PCAST committee performed a sound, full and complete review. Again, these issues are most disconcerting because it is apparent that the PCAST Committee in its undertaking did not hold itself up to the same standards of validation, documentation, and peer-review that it espouses the forensic community should embrace (compounded as a number of the criticisms in the report are unfounded). The report provides some guidance on basic statistics, such as estimating false positive rates (which are not novel). However, this lecturing on proper statistics is troubling to say the least as the report misuses statistics in its own cursory efforts.

The following are examples from the report to support my above claims. They are not comprehensive as it is unnecessary to go page-by-page to indicate the serious problems with the PCAST Report. A few examples should suffice to demonstrate why this report has been so underwhelming and been ignored by most scientists and the courts. In pointing out the failings of the report I will focus on topics that transcend the disciplines and specifically on my area of expertise, i.e., DNA; I could not adequately address the other disciplines and what data do or do not exist in those forensic science areas. I leave specifics of other disciplines to those with requisite expertise. However, I stress that since the report misinforms on forensic DNA applications, which is considered the "gold standard" and well-documented in the scientific literature (even the report acknowledges that), then there is a strong indication that perhaps the report missed the mark on the other disciplines as well.

I take the position that improvements in forensic sciences are needed. Indeed, all science continues to improve. It is never static. In my field of DNA typing, I and others have been and currently are working on developing better/improved methods, such as the use of next generation sequencing and new software tools. It would be improper to say that any method is perfect and cannot be made better. That position, though, is not a wholesale condemnation of the forensic sciences. Each discipline, or better yet each application, should be assessed in context as a holistic system (not solely based on validation as the report seemingly myopically espouses) and

the types/quality of samples encountered in specific cases. The report's generalization of issues avoids addressing an extremely important question – was the analysis/interpretation in this case performed correctly?

The first two examples presented below are particularly egregious and point to the dearth of substance in the report. The report states on page 2

“In the course of its study, PCAST compiled and reviewed a set of more than 2,000 papers from various sources—including bibliographies prepared by the Subcommittee on Forensic Science of the National Science and Technology Council and the relevant Working Groups organized by the National Institute of Standards and Technology (NIST); submissions in response to PCAST's request for information from the forensic-science stakeholder community; and PCAST's own literature searches.”

On page 67 of the report it is stated

“PCAST compiled a list of 2019 papers from various sources—including bibliographies prepared by the National Science and Technology Council's Subcommittee on Forensic Science, the relevant Scientific Working Groups (predecessors to the current OSAC), and the relevant OSAC committees; submissions in response to PCAST's request for information from the forensic-science stakeholder community; and our own literature searches.”

There were two citations to support the review of the 2000 or so papers that the PCAST relied upon:

www.nist.gov/forensics/workgroups.cfm.

www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf.

Neither of these sites appear to show (or allow for ready identification) what those articles were that the PCAST Committee reviewed and then relied upon. More so, there are no criteria and no data in the report or at these sites on what the PCAST Committee actually read, noted, reviewed, quantified, calculated, accepted, rejected, and/or debated. The report advocates emphatically and repeatedly the virtues of validation, documentation, and peer-review. Yet the report does not contain such information and thus does not meet as a minimum the requirements that it lambasted the forensic science community for lacking. This inconsistency between recommended requirements and lack of performance by the PCAST Committee is most noted as there is substantial documentation in the forensic science community (in many disciplines) but not in this report.

This lack of documentation should be considered in light of the report's statements on pages 1 and 22

“PCAST concluded that there are two important gaps: (1) the need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) the need to

evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.”

The report also states on pages 4 and 21

“It is the proper province of the scientific community to provide guidance concerning scientific standards for scientific validity, and it is on those *scientific* standards that PCAST focuses here.”

Yet the PCAST Committee did not provide its data to support the validity of its own work. There simply is no accounting of the PCAST Committee’s work to demonstrate it assessed the 2000 papers and how it came to the conclusions it rendered.

This evident failing is exacerbated by the reports statement on page 6

“The forensic examiner must have been shown to be *capable* of reliably applying the method and must *actually* have done so. Demonstrating that an expert is *capable* of reliably applying the method is crucial—especially for subjective methods, in which human judgment plays a central role. From a scientific standpoint, the ability to apply a method reliably can be demonstrated only through empirical testing that measures how often the expert reaches the correct answer. Determining whether an examiner has *actually* reliably applied the method requires that the procedures actually used in the case, the results obtained, and the laboratory notes be made available for scientific review by others.”

No one knows what method(s) the PCAST Committee used; but it is clear that it did not hold itself to the same standard either by *capability* or *actually* performing. This report cannot be held up for scientific review (as indicated on page 6 of the report – see immediately above). There are no notes or results available.

As the report says repeatedly (see pages 6 and 32)

“We note, finally, that neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual evidence of foundational validity and reliability.”

The academic and professional standings of the PCAST Committee members are not a substitute for good practices (none of which are documented). No one should take seriously this report because it has little substance to support its contentions.

The second most egregious example is the misuse and disregard for statistics. It may appear to the casual observer that the PCAST Committee is steeped in statistics and thus all statistics presented must be meaningful. For example, the report dedicates Appendix A for some discussion on statistics. But this guidance is rather basic and not particularly helpful to guide the community for any specific discipline or application. Yet when it comes to substance the PCAST Committee fails again which is evident in its own use of statistics. Consider the statements in the report on page 3

“Reviews by the National Institute of Justice and others have found that DNA testing during the course of investigations has cleared tens of thousands of suspects and that DNA-based re-examination of past cases has led so far to the exonerations of 342 defendants. Independent reviews of these cases have revealed that many relied in part on faulty expert testimony from forensic scientists who had told juries incorrectly that similar features in a pair of samples taken from a suspect and from a crime scene (hair, bullets, bitemarks, tire or shoe treads, or other items) implicated defendants in a crime with a high degree of certainty.”

Then on page 26

“DNA-based re-examination of past cases, moreover, has led so far to the exonerations of 342 defendants, including 20 who had been sentenced to death, and to the identification of 147 real perpetrators.”

A similar statement is found on page 44 (footnote 94). These findings appear to support the assertion on page 44 of the report

“It is *important* because it has become apparent, over the past decade, that faulty forensic feature comparison has led to numerous miscarriages of justice.”

I do not dispute that there have been 342 post-conviction exonerations. I am not sure what the number of exonerations is when the report says “many relied in part on faulty expert testimony” – because the report does not quantify what is meant by many. However, one wrongful analysis or testimony is one too many, and every effort should be made to minimize forensic science errors. The exoneration of 342 convicted felons is serious and topic in its own right (and again way too many). But this number is statistically meaningless and out of context. The PCAST Committee should have recognized this obvious aspect of the use of numbers. The PCAST Committee did not perform any statistical analyses or even appear to collect the data necessary to put these numbers in proper perspective. The PCAST Committee should have identified how many cases in total that have been reviewed to date (especially given that the report discusses the proper way to calculate a false positive rate, the Committee does not follow through with the same verve). This number of 342 may be and is likely a very small percentage of the total number of cases reviewed, especially since the innocence project has been around for 25 years (see <https://25years.innocenceproject.org/>). Moreover, the PCAST Committee did not convey how many post-conviction analyses that have been performed over the past 25 years in which there was no evidence of improper scientific performance, findings or faulty testimony. It would seem that such obvious basic information eluded the PCAST Committee. Those cases that were reviewed over the past 25 years in which no misuse of forensic science analyses were detected would indicate that perhaps the forensic science field is not so scientifically corrupt as the report implies. More so it would indicate that proper results can be obtained (at least most of the time).

The report discusses error rates substantially using statements such as on page 6

“Similarly, an expert’s expression of *confidence* based on personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.”

The PCAST Report also recommends

“For subjective feature-comparison methods, because the individual steps are not objectively specified, the method must be evaluated as if it were a “black box.”

Smrz et al (8) (a paper of which I am a co-author) recommended the black box approach after the review of the FBI Laboratory’s latent print misidentification related to the Madrid bombing incident, and the PCAST Report advocates the use of such black box studies. I concur that a black box approach has some value but strongly caution that one must consider the proper utility of such studies. The authors of the PCAST Report calculated upper bound error rates based on the results of the very few black box studies they discuss; the PCAST Committee seemingly implies that these upper bound error rates are somehow meaningful to report in every case analysis. A black box study can demonstrate generally whether or not a method can yield reliable results where a human is substantially involved in the interpretation of results. But it does not necessarily help address error that may or may not have occurred during a specific case analysis.

There are several problems with such a simplistic generalization that the authors of the PCAST Report have taken regarding use of black box studies. A black box study only tests those individuals involved in the study. Therefore, the performance of the rest of the analysts of the forensic science community is not covered by the study, and the results of the study may not apply to those analysts. Some individuals perform better than others in black box studies. The average rate inflates the performance of the poorer analysts and deflates the performance of the better analysts tested in the study. Therefore, the error rate values calculated by the PCAST authors likely do not apply to most analysts. Moreover, the information content and quality of results from a forensic science analysis vary from sample to sample. Treating all sample results equally and applying a single error rate does not convey the chance for error in a particular analysis. As the PCAST Report states (see below) DNA mixture interpretation is more challenging than interpretation of single source DNA profiles. If the PCAST Committee recognizes that differences in the quality of DNA evidence affect difficulty of interpretation, then the PCAST Committee should have been able to realize that the same holds for black box study results and different quality evidence (another obvious inconsistency in the report).

A known error rate or proficiency test mistake is at best some indirect measure of the verity of the proposed results in any given sample analysis, but can never be a direct measure of the reliability of the specific result(s) in question (9). Consider a hypothetical crossing of a street where there is a 1% error (arbitrary for sake of discussion) of being hit by a car. At the beginning of the journey crossing the road there is a 1% error of being hit. While crossing the road the chance can increase or decrease depending on circumstances (possibly being greater at the center of the road and less within lanes). If the individual successfully crosses the road, then the error drops to zero. Of course, different roads (such as a busy interstate vs a rural back road) have different *a priori* chances of error (i.e., similar to the quality of evidence affects the degree of difficulty). Ultimately the issue of crossing the road is did the individual successfully cross the road or get hit. The same holds for casework, i.e., is there an error or is there not an error in the performance or analysis. Given that the black box studies mentioned in the report did have a good degree of success, there is support that a process can generate a reliable result. Thus it still comes back to determining if an error of consequence was committed in a specific case. Oddly not mentioned in the PCAST Report is that most of the forensic disciplines addressed carry out non-consumptive forms of examination. Therefore, the most direct way to measure the truth of

the purported results is to have another expert conduct his/her own review, as is advocated by the National Research Council Report II for DNA analyses (10). Re-analysis would be more meaningful instead of espousing hypothetical error rates, which may not apply to the actual results and/or analysts involved. Indeed, the above mentioned black box studies and the missing data on total number of cases from innocence project case reviews do support that tests can yield reliable results but that most of the problems (as discussed below for DNA mixtures) have been due to misapplication. Therefore, case peer-review can be an effective approach to identify misapplications. However, the PCAST Report seems to ignore the value of this practice which demonstrates the reports myopic assessment of the forensic sciences and lack of consideration of a holistic systems approach.

The PCAST Report singles out validation as essentially the sole basis for reliability. Instead under a systems approach there are several components that impact an outcome, and the reliance on these several features increases validity and reliability in any one case. Quality performance is an essential component for obtaining reliable results and for reducing the chance of error. Quality assurance provides an infrastructure to promote high performance, address errors that arise, and improve processes. In addition to validation studies, there are other mechanisms such as technical review of a case that reduce error. This technical review is performed within the laboratory before issuing a report and also outside the laboratory when an expert witness is acquired by the opposing side to assess results and interpretations. The PCAST Report seems to ignore the value of these additional quality measures and the strength of the adversary system. Error rates are difficult to calculate; they are fluid. When an error of consequence (i.e., a false “match”) occurs, under a sound quality assurance program corrective action is taken (to include review of cases analyzed by the examiner prior to and post the discovery of the error). When the corrective action is such that the individual will no longer commit that error, it no longer impacts negatively on the individual’s future performance. In fact, he/she is better educated and less likely to err. The calculation of a current error rate then should not include past error(s). Having said that, past error should not be ignored; if desired, it could be raised in court or other deliberations. The defense (or prosecution), if it believes it useful, should make use of such information during a cross-examination of an expert. But the PCAST Report does not address the shortcomings of the calculated error rate as it uses it; it treats the upper bound error rate calculation from black box studies as if they are robust and specific (which they are not).

Notably the PCAST Report tends to dismiss experience and judgment, implying it has little value. I agree that experience and judgment standing alone should be considered with caution. However, the vast majority of forensic science disciplines work in a systems approach, i.e., many facets to the process; experience is but one factor among several to effect a quality result. Even though the PCAST Report dismisses experience it again shows its inconsistencies about the province of experience. Consider the following statements on page 55 of the report

“In some settings, an expert may be scientifically capable of rendering judgments based primarily on his or her “experience” and “judgment.” Based on experience, a surgeon might be scientifically qualified to offer a judgment about whether another doctor acted appropriately in the operating theater or a psychiatrist might be scientifically qualified to offer a judgment about whether a defendant is mentally competent to assist in his or her defense.”

“By contrast, “experience” or “judgment” cannot be used to establish the scientific validity and reliability of a metrological method, such as a forensic feature-comparison method. The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of “judgment.” It is an empirical matter for which only empirical evidence is relevant. Moreover, a forensic examiner’s “experience” from extensive casework is not informative—because the “right answers” are not typically known in casework and thus examiners cannot accurately know how often they erroneously declare matches and cannot readily hone their accuracy by learning from their mistakes in the course of casework.”

Even to a lay person these statements should be obviously inconsistent, troubling and point to the inadequacy of the PCAST Committee addressing the topic of forensic science reliability. I fail to see why the medical and psychology fields can have another expert review another’s work (on what may be life and death decisions) and opine on the analyses/interpretations; yet a qualified forensic science analyst cannot perform a technical review of forensic work to assess analyses/interpretations (especially since the report has ignored data that support that at some level forensic testing is reliable). The logic of the PCAST Committee escapes me.

The PCAST Report discusses DNA typing and the limitations that have been encountered with mixture interpretation. For example on page 75 the report states

“DNA analysis of complex mixtures—defined as mixtures with more than two contributors—is inherently difficult and even more for small amounts of DNA.”

I concur that it is more challenging to interpret DNA mixtures compared with single-source DNA profiles. But the report fails to add that difficult does not necessarily translate into impossible or that proper interpretations can be made. The difficulties with mixture interpretation were not due to a lack of good, valid approaches to employ as there were valid approaches and also not due to the fact that there is some subjective judgment with interpretations. The issue, and it is a serious one, was that many of the practitioners in the forensic DNA community were inadequately trained, did not seek out solutions, or instead chose to wait for guidance (see pages 77-78 of the PCAST report and discussion on Texas and mixture interpretation). These issues were similar to the mixture interpretation problems at the Department of Forensic Sciences in Washington, DC (in which I was the scientist who identified the problems).

The PCAST Report assails the use of the Combined Probability of Inclusion (CPI) which is one of the methods used by the community and endorsed by the DNA Advisory Board (11) 17 years ago. However, the discussion of the Texas Forensic Science Commission (TFSC) (of which I was deeply involved in the review of mixture interpretation for the State) and how it pursued and addressed inappropriate interpretation of mixtures actually implies that valid methods do exist; otherwise how could a group of international experts (of which I was one of the experts) assess the situation, determine that there are problems in the application of interpretation guidelines, and provide guidance to the community to implement sound procedures?

The PCAST Committee on page 78 of the report states

“The TFSC also convened an international panel of scientific experts—from the Harvard Medical School, the University of North Texas Health Science Center, New Zealand’s

forensic research unit, and NIST—to clarify the proper use of CPI. These scientists presented observations at a public meeting, where many attorneys learned for the first time the extent to which DNA-mixture analysis involved subjective interpretation. Many of the problems with the CPI statistic arose because existing guidelines did not clearly, adequately, or correctly specify the proper use or limitations of the approach.”

The report properly focuses on lack of detailed guidelines on interpretation and does not suggest that the principles of how to calculate the CPI are erroneous. Indeed, nowhere in the report are there any data to indicate that the CPI is foundationally erroneous.

Yet, the report then states on page 78

“In summary, the interpretation of complex DNA mixtures with the CPI statistic has been an inadequately specified—and thus inappropriately subjective—method. As such, the method is clearly not foundationally valid.”

The allegation that the CPI is not foundationally valid demonstrates the lack of understanding (and again the lack of documentation of review) by the PCAST Committee. In fact, these statements also demonstrate another report inconsistency – this time about the principles of statistical calculations related to DNA profiles. On page 72 the report states

“The process for calculating the random match probability (that is, the probability of a match occurring by chance) is based on well-established principles of population genetics and statistics.”

The random match probability is one approach to calculating a statistic for single-source samples and appears to be endorsed by the PCAST Committee as well-established and thus valid. Yet, the PCAST Committee takes the opposite position for the CPI stating it is not foundationally valid. If one reads my colleagues and my most recent paper on the CPI (12), cited in the PCAST Report, it is clear that the principles of the foundational validity of the CPI are the same as those for the random match probability. Consider a similar situation which is the chance of drawing four aces in a row from a standard deck of cards is estimated to be 1 in 270,275. This value is based on probability theory and does not require an empirical testing to be published in the peer reviewed literature to support its validity. The CPI and random match probability use the same population frequency data and the same well-established principles of population genetics and statistics. While this is another example of myopia by the PCAST Committee, it borders on the bizarre that the PCAST Committee failed to understand the foundations of DNA statistics.

All know the PCAST Committee had access to the most recent paper on the use of the CPI (and the references within that paper) as it is stated on page 78 of the report

“Because the paper appeared just as this report was being finalized, PCAST has not had adequate time to assess whether the rules are also *sufficient* to define an objective and scientifically valid method for the application of CPI.”

I note that the CPI is a rather simple concept and its foundations are basic. It is surprising that the PCAST Committee, which touts its vast expertise, could not readily assess the paper. Given the importance of their report and this topic it also is surprising that they would not have done so before finalizing their report.

The PCAST Report recognizes that probabilistic genotyping is an advancement to improve or reduce subjectivity in DNA mixtures (see page 79). I concur. But the report states on page 79

“Appropriate evaluation of the proposed methods should consist of studies by multiple groups, *not associated with the software developers*, that investigate the performance and define the limitations of programs by testing them on a wide range of mixtures with different properties.”

Also the report states on page 81

“Because empirical evidence is essential for establishing the foundational validity of a method, PCAST urges forensic scientists to submit and leading scientific journals to publish high-quality validation studies that properly establish the range of reliability of methods for the analysis of complex DNA mixtures.”

Publication is part of the peer-review process and I support publication by the developers and others who adopt the method. But the PCAST Committee has placed a requirement that is unrealistic to meet which is publication by the user laboratories. It is likely that a few at most laboratories will be able to publish their validation testing of the software. Anyone who serves on editorial boards of scientific journals should know that journals are unlikely to publish additional studies because they are not considered novel. Yet, the PCAST Committee failed to recognize this fact.

It is important to stress that the report contains no criticisms of probabilistic genotyping and still there are no data contained in the report that demonstrate that the PCAST Committee actually reviewed (or better yet tested) the current probabilistic genotyping software programs (even though it claims to have done extensive review, such as the undocumented 2000 papers).

Forensic laboratories are required to perform validation studies, and there are substantial data on mixtures that support the validity of mixture interpretation and use of probabilistic genotyping. Mixture studies are required to be performed by every laboratory engaged in analyzing such evidence as part of their validation studies. Many of these studies lack novelty and thus will never be published in peer-review journals. However, the PCAST Committee could have contacted a number of forensic DNA laboratories who have implemented one of the probabilistic genotyping software programs (as there were laboratories operating or near implementation of the tools at the time of the report’s publication) to gain access to the validation data to determine whether there are sufficient data to support the already peer-reviewed published work. There is no indication that the PCAST Committee made any effort to become informed to opine on the reliability and validity of probabilistic genotyping.

The PCAST Committee simply ignored a wealth of validation data residing in crime laboratories. If the PCAST Committee had taken a holistic approach, they would have considered the totality of data in determining whether there is support for the validity and reliability of probabilistic genotyping. Peer-review publications by the developers and validation data by the users combined clearly support the software and its applications. Indeed, this failure of the PCAST Committee of not considering all available data is reminiscent of a similar situation that occurred 25 years ago with another report – the National Research Council I Report (NRC I) (13). The NRCI Report proposed a non-scientific, *ad hoc* way to calculate statistics called the ceiling principle. The ceiling principle had no genetics foundation or validity and was roundly rejected. One of the bases for the proposed ceiling principle approach (espoused by the NRC I Committee) was a lack of population data. There were substantial population data in crime

laboratories world-wide at the time the NRC I Report was published; but the NRC I Committee did not seek out the data. As soon as the NRC I Report was published, I reached out to my colleagues around the world and gathered the existing data which were then compiled into a five volume compendium (14). If the NRC I Committee had chosen to consider extant population data, they might have prepared a more informed Report. The outcome was that the National Academy of Sciences convened a second committee and produced the sound NRC II Report (10), which was steeped in fundamental population genetics and statistical applications. The findings of the NRC II Report in part were based on the data I compiled in the five volume compendium which were available prior to the publication of the rejected NRC I Report. The PCAST Report has taken the same blinded approach and ignored extant data with a similar outcome as 25 years ago – a report that provides little value for assessing the state-of-the-art and even less value for providing guidance to improve the forensic sciences.

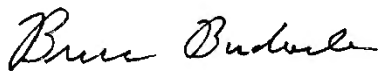
In conclusion, the few examples above demonstrate that the PCAST Report 1) is **not** scientifically sound, 2) is **not** based on data, 3) is **not** well-documented, 4) misapplies statistics, 5) is full of inconsistencies, and 6) does **not** provide helpful guidance to obtain valid results in forensic analyses.

References

1. Budowle, B. and SWGMGF Members: Quality assurance guidelines for laboratories performing microbial forensic work. *Forens. Science Communications* October 5(4): 2003, At: www.fbi.gov/hq/lab/fsc/current/2003_10_guide01.htm.
2. Budowle, B., Schutzer, S.E., Einseln, A., Kelley, L.C., Walsh, A.C., Smith, J.A.L., Marrone, B.L., Robertson, J., and Campos, J.: Building microbial forensics as a response to Bioterrorism. *Science* 301: 1852-1853, 2003.
3. Budowle, B., Schutzer, S.E., Morse, S.A., Martinez, K.F., Chakraborty, R., Marrone, B.L., Murch, R.S., Jackson, P.J., Williamson, P., Harmon, R., and Velsko, S.P.: Guidance for validation in microbial forensics. *Applied and Environmental Microbiology* 74:5599-5607, 2008.
4. Budowle, B., Connell, N.D., Bielecka-Oder, A., Colwell, R.R., Corbett, C.R., Fletcher, J., Forsman, M., Kadavy, D.R., Markotic, A., Morse, S.A., Murch, R.S., Sajantila, A., Schmedes, S.E., Ternus, K.L., Turner, S.D., Minot, S.: Validation of high throughput sequencing and microbial forensics applications. *BMC Invest. Genet.* 5:9, 2014.
5. The FBI Director's Forensic Quality Assurance Standards for DNA Testing Laboratories, http://media.wix.com/ugd/4344b0_4a22824ce56f43d4b1a4d2486409f95d.pdf
6. SWGDAM Validation Guidelines for Forensic DNA Analysis Methods, https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf
7. SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems, https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf

8. Smrz, M.A., Burmeister, S.G., Einseln, A., Fisher, C.L., Fram, R., Stacey, R.B., Theisen, C.E., and Budowle, B.: Review of FBI latent print unit processes and recommendations to improve practices and quality. J. Forens. Ident. 56(3):402-434, 2006.
9. Harmon R. and Budowle, B.: Questions about forensic science. Science 311:607, 2006.
10. NRC II Report, The Evaluation of Forensic DNA Evidence, 1996, <https://www.nap.edu/catalog/5141/the-evaluation-of-forensic-dna-evidence>
11. Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated From Pertinent Population Database(s), DNA Advisory Board, Forensic Science Communications 2(3), 2000, <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm>
12. Bieber, F.R., Buckleton, J., Budowle, B., Butler, J., and Coble, M.D.: Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genetics 17(1):125, 2016.
13. NRC I Report, The Evaluation of Forensic DNA Evidence, 1992, <https://www.nap.edu/read/5141/chapter/1#ii>
14. Federal Bureau of Investigation: VNTR Population Data: A Worldwide Study, Volumes I-IV, Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia, 1993.

I declare under penalty of perjury that the forgoing is true and correct to the best of my knowledge.



Bruce Budowle, Ph.D.
Director
Center for Human Identification
University of North Texas Health Science Center
Fort Worth, Texas 76107
Email: (b) (6)
Tel: (b) (6)