

RE: As Discussed

From: "Antell, Kira M. (OLP)" <(b) (6)>
To: "Mizelle, Chad (ODAG)" <(b) (6)>
Date: Wed, 12 Jul 2017 14:22:57 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics_KMA.docx (207.05 kB)

Looks good. Minor suggestions included. LMK if I can provide any other information. Are the article and your memo going to be the thrust of the briefing or will you do a briefing on those in advance of the OLP briefing next week?

From: Mizelle, Chad (ODAG)
Sent: Wednesday, July 12, 2017 2:06 PM
To: Antell, Kira M. (OLP) <(b) (6)>
Subject: As Discussed

RE: Forensics Memo

From: "Goldsmith, Andrew (ODAG)" <(b) (6)>
To: "Mizelle, Chad (ODAG)" <(b) (6)>
Date: Wed, 12 Jul 2017 15:26:16 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.v2 (ADG edits).docx (206.08 kB)

I made a few minor edits (main change is addition of sentence in first paragraph re: (b) (5) [REDACTED], which I think helps for context). If you're okay with changes, feel free to accept all and send to Jim. Thanks again.

From: Mizelle, Chad (ODAG)
Sent: Wednesday, July 12, 2017 2:00 PM
To: Goldsmith, Andrew (ODAG) <(b) (6)>
Subject: Forensics Memo

Andrew,

Attached is my draft memo. Looking forward to your thoughts/comments.

Best,
Chad

Forensics Speech Draft

From: "Ibrahim, Anitha (CRM)" <(b) (6)>
To: "Antell, Kira M. (OLP)" <(b) (6)>, "Mann, James (CRM)" <(b) (6)>, "Mizelle, Chad (ODAG)" <(b) (6)>, "Rybicki, David (CRM)" <(b) (6)>
Date: Thu, 13 Jul 2017 17:13:42 -0400
Attachments: Forensics Symposium Speech v2.docx (34.52 kB)

All –

Please find attached a draft for David's speech for the NAAG event next week. Please review and provide any feedback/edits.

Thank you!
Anitha

Anitha S. Ibrahim
Counsel to the Assistant Attorney General
Criminal Division | U.S. Department of Justice
(b) (6) | (b) (6)

Forensics Symposium Speech v2 (002)_KMA

From: "Antell, Kira M. (OLP)" <(b) (6)>
To: "Mizelle, Chad (ODAG)" <(b) (6)>
Date: Mon, 17 Jul 2017 17:21:11 -0400
Attachments: Forensics Symposium Speech v2 (002)_KMA.docx (36.74 kB)

See if you think this is sufficient. I think this threads the needle.

Position Proposal

From: "Hunt, Ted (ODAG)" <(b) (6)>
To: "O'Callaghan, Edward C. (ODAG)" <(b) (6)>, "Ellis, Corey F. (ODAG)" <(b) (6)>
Date: Thu, 31 Jan 2019 15:36:04 -0500
Attachments: Hunt DRAFT Memo-New Department Position.docx (31.41 kB)

Ed and Corey:

Attached is my draft proposal for creation of (b) (5) position, as discussed at today's meeting.

(One of the names that Andrew G. mentioned as an analogy to this proposal is (b) (5)

Let me know if you have any additional thoughts or ideas about the possible administrative logistics (or gymnastics) for such a position (b) (5).

In the meantime, I still need to (b) (6)

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)
(b) (6)

RE: Forensic Science Subcommittee Standing Meeting

From: "Antell, Kira M. (OLP)" <(b) (6)>
To: "Santos, Nelson A. (DEA)" <(b)(6), (b)(7)(C) per DEA>, "Isenberg, Alice R. (LD) (FBI)" <(b)(6), (7)(C), (7)(E) per FBI>, "Czarnopys, Greg P. (ATF)" <(b) (6)>, "Smith, David L. (USAEO)" <(b) (6)>, "(b)(6), (b)(7)(C) per BOP" (BOP)" <(b) (6)>, "Laporte, Gerald (OJP)" <(b) (6)>, "Downing, Richard (CRM)" <(b) (6)>, "(b)(6), (b)(7)(C) per USMS" (USMS)" <(b)(6), (b)(7)(C) per USMS, (b)(6), (b)(7)(C) per FBI" (OGC) (FBI)" <(b)(6), (b)(7)(C) per FBI>, "(b)(6), (b)(7)(C) per BOP" (BOP)" <(b) (6)>, "(b)(6), (b)(7)(C), (b)(7)(F) per DEA" <(b) (6)>, "Ibrahim, Anitha (CRM)" <(b) (6)>, "Williamson, Angela (OJP)" <(b) (6)>, "(b)(6), (7)(C), (7)(E) per FBI" (OTD) (FBI)" <(b)(6), (7)(C), (7)(E) per FBI>, "Dolan, Julia A. (ATF)" <(b) (6)>, "(b)(6), (b)(7)(C) per BOP" <(b) (6)>, "(b)(6), (b)(7)(C) per BOP" <(b) (6)>
Cc: "Thiemann, Robyn (OLP)" <(b) (6)>, "Beqian, Lernik (OLP)" <(b) (6)>, "Mizelle, Chad (ODAG)" <(b) (6)>, "(b)(6), (7)(C), (7)(E) per FBI" (OTD) (FBI)" <(b)(6), (7)(C), (7)(E) per FBI>
Date: Mon, 24 Jul 2017 15:10:01 -0400
Attachments: Fiber_pULTR_05252016.pdf (74.15 kB); LatentPrint_pULTR_05252016.pdf (73.58 kB); Agenda July 27, 2017_DISTRIBUTED.pdf (44.61 kB); latents_ULTR_FOR NCFS DISCUSSION_09092016.pdf (82.39 kB); Fiber_ULTR_FOR NCFS DISCUSSION_09092016.pdf (81.97 kB); Evett et al_FSI_278-16-23.pdf (363.88 kB); Budowle Response to PCAST Report 06-17-2017 (002).pdf (521.58 kB)

Good afternoon,

I am looking forward to our meeting on Thursday. This week I'd like to focus on the draft ULTR documents and how these drafts could be developed moving forward. I have attached documents here for your review.

1. ULTR 1.0 (latents and fiber) (put out for public comment in May 2016 and based on the FBI ASSTRs)
2. ULTR 2.0 (latents and fiber) (shared with the NCFS at meeting 11)
3. Evett Article. While this article primarily criticizes PCAST, it also suggests a different paradigm for reporting examiner results (match versus identification). If you are not familiar with this short commentary (8 page), please review it in advance of the meeting this week so you can share whether this distinction is meaningful or helpful from your perspective.
4. Agenda

If you'd like to add anything to the agenda, please let me know. If you can't make the meeting, please let me know who will be filling in for you. Finally, I think most of you have seen the Budowle affidavit on PCAST but for those of you who have not, I have attached it here.

Thanks,
Kira

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

From: Antell, Kira M. (OLP)

Sent: Friday, June 16, 2017 4:02 PM

To: Antell, Kira M. (OLP); Santos, Nelson A. (DEA); Isenberg, Alice R. (LD) (FBI); Czarnopys, Greg P. (ATF); Smith, David L. (USAEO); (b)(6), (b)(7)(C) per BOP (BOP); Laporte, Gerald (OJP); Downing, Richard (CRM); (b)(6), (b)(7)(C) per USMS (USMS); (b)(6), (b)(7)(C) per FBI (OGC) (FBI); (b)(6), (b)(7)(C) per BOP (BOP); (b)(6), (b)(7)(C), (b)(7)(F) per DEA (DEA); Ibrahim, Anitha (CRM); Williamson, Angela (OJP); (b)(6), (b)(7)(C) per FBI (OTD) (FBI); Mizelle, Chad (ODAG); Newman, Ryan (OLP); (b)(6) Julia Dolan; Scott, Kevin M. (OLP); Beqian, Lernik (OLP); Thiemann, Robyn (OLP)

Cc: (b)(6), (b)(7)(C) per BOP

Subject: Forensic Science Subcommittee Standing Meeting

When: Thursday, July 27, 2017 2:00 PM-3:30 PM (UTC-05:00) Eastern Time (US & Canada).

Where: OLP 4525 and Conference Line: (b) (6) /Passcode: (b) (6)

This document provides examples of the scientifically-supported conclusions and opinions that may be contained in Department of Justice reports and testimony. These examples are not intended to be all inclusive and may be dependent upon the precedent set by the judge or locality in which a testimony is provided. Further, these examples are not intended to serve as precedent for other forensic laboratories and do not imply that statements by other forensic laboratories are incorrect, indefensible, or erroneous. This document is not intended to, does not, and may not be relied upon to create any rights, substantive or procedural, enforceable by law by any party in any matter, civil or criminal, nor does it place any limitation on otherwise lawful investigative and litigative prerogatives of the Department.

**DEPARTMENT OF JUSTICE
PROPOSED UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS
FOR THE FORENSIC TEXTILE FIBER DISCIPLINE**

Purpose and Scope

If adopted, this document will apply to Department of Justice personnel who perform forensic examinations and/or provide expert witness testimony regarding the forensic examination of fiber evidence. This document does not imply that statements made or language used by Department personnel that differed from these proposed statements were incorrect, indefensible, or erroneous.

This document provides the acceptable range of opinions expressed in both laboratory reports and during expert witness testimony while acknowledging that this document cannot address every variable in every examination.

Statements Approved for Use in Fiber Examination Testimony and/or Laboratory Reports

Fiber Classification

The examiner may state or imply that a textile fiber is natural or manufactured (man-made).

Natural Fibers

1. The examiner may state or imply the type of natural fiber (e.g., cotton, wool, silk).

Manufactured Fibers

2. The examiner may state or imply the type of manufactured fiber (e.g., polyester, nylon). The examiner may further state or imply that the manufactured fiber is consistent with a particular sub-group (e.g., polyethylene terephthalate, nylon 6).

Comparisons

Inclusion

3. The examiner may state or imply that the questioned fiber exhibits the same microscopic characteristics and optical properties as the known sample and accordingly, the questioned fiber is consistent with originating from the source of the known sample or from another item comprised of fibers that exhibit the same microscopic characteristics and optical properties. A fiber association is not a means of positive identification and the number of possible sources for a specific fiber is unknown. However, due to the variability in manufacturing, dyeing, and consumer use, one would not expect to encounter a fiber selected at random to be consistent with a particular source.

Exclusion

4. The examiner may state or imply that the questioned fiber is dissimilar to the known fiber sample and accordingly, is not consistent with originating from the source of the known sample.

Statements Not Approved For Use in Fiber Examination Testimony and/or Laboratory Reports

Individualization

1. The examiner may not state or imply that a fiber came from a particular source to the exclusion of all other sources.

Statistical Weight

2. The examiner may not state or imply a statistical weight or probability to a conclusion or provide a likelihood that the questioned fiber originated from a particular source.

Zero Error Rate

3. The examiner may not state or imply that the method used in performing fiber examinations has a zero error rate or is infallible.

**DEPARTMENT OF JUSTICE PROPOSED UNIFORM LANGUAGE
FOR TESTIMONY AND REPORTS REVIEW SHEET**

Directions: This review sheet is designed to assist you in evaluating the attached Proposed Uniform Language for Testimony and Reports document against certain criteria while maintaining internal consistency in review and assessing comments.

Your use of this rating sheet is completely **optional**. While it is anticipated this review sheet will encourage comments on issues of particular importance, you are welcome to submit comments in any format that you believe appropriate. This review sheet is not intended to limit comments in any way.

If you elect to use the review sheet, you may find it helpful to frame your comments as suggested below.

Proposed Uniform Language Discipline Reviewed:

Reviewer Name:

Reviewer Organization:

Statements Approved for Use in Laboratory Reports and Expert Witness Testimony

Provide a summary of your assessment of the statements approved for use, including the most important highlights from the individual criteria comments.

- The statements approved for use are supported by scientific research.
- The statements approved for use accurately reflect consensus language.
- The statements approved for use are stated clearly.

Statements Not Approved for Use in Laboratory Reports and Expert Witness Testimony

Provide a summary of your assessment of the statements not approved for use, including the most important highlights from the individual criteria comments.

- The statements not approved for use are supported by scientific research.
- The statements not approved for use accurately reflect consensus language.
- The statements not approved for use are stated clearly.

This document provides examples of the scientifically-supported conclusions and opinions that may be contained in Department of Justice reports and testimony. These examples are not intended to be all inclusive and may be dependent upon the precedent set by the judge or locality in which a testimony is provided. Further, these examples are not intended to serve as precedent for other forensic laboratories and do not imply that statements by other forensic laboratories are incorrect, indefensible, or erroneous. This document is not intended to, does not, and may not be relied upon to create any rights, substantive or procedural, enforceable by law by any party in any matter, civil or criminal, nor does it place any limitation on otherwise lawful investigative and litigative prerogatives of the Department.

**DEPARTMENT OF JUSTICE
PROPOSED UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS
FOR THE FORENSIC LATENT PRINT DISCIPLINE**

Purpose and Scope

If adopted, this document will apply to Department of Justice personnel who perform forensic examinations and/or provide expert witness testimony regarding the forensic examination of latent print evidence. This document does not imply that statements made or language used by Department personnel that differed from these proposed statements were incorrect, indefensible, or erroneous.

This document provides the acceptable range of opinions expressed in both laboratory reports and during expert witness testimony while acknowledging that this document cannot address every variable in every examination.

Statements Approved for Use in Latent Print Examination Testimony and/or Laboratory Reports

Identification

1. The examiner may state or imply that an *identification* is the determination that two friction ridge prints originated from the same source because there is sufficient quality and quantity of corresponding information such that the examiner would not expect to see that same arrangement of features repeated in another source. While an *identification* to the absolute exclusion of all others is not supported by research, studies have shown that as more reliable features are found in agreement, it becomes less likely to find that same arrangement of features in a print from another source.

Inconclusive

2. An examiner may state or imply that an *inconclusive* result is the determination that there is insufficient quality and quantity of corresponding information such that the examiner is unable to identify or exclude the source of the print.

Exclusion

3. An examiner may state or imply that an *exclusion* is the determination that two friction ridge prints did not originate from the same source because there is sufficient quality and quantity of information in disagreement.

Statements Not Approved For Use in Latent Print Examination Testimony and/or Laboratory Reports

Exclusion of All Other Sources

1. An examiner may not state or imply that two friction ridge prints originated from the same source to the absolute exclusion of all other sources.

Absolute or Numerical Certainty

2. An examiner may not state or imply a level of certainty in his/her conclusion that is absolute or numerically calculated.

Zero Error Rate

3. An examiner may not state or imply that the method used in performing a friction ridge print comparison has a zero error rate or is infallible.

**DEPARTMENT OF JUSTICE PROPOSED UNIFORM LANGUAGE
FOR TESTIMONY AND REPORTS REVIEW SHEET**

Directions: This review sheet is designed to assist you in evaluating the attached Proposed Uniform Language for Testimony and Reports document against certain criteria while maintaining internal consistency in review and assessing comments.

Your use of this rating sheet is completely **optional**. While it is anticipated this review sheet will encourage comments on issues of particular importance, you are welcome to submit comments in any format that you believe appropriate. This review sheet is not intended to limit comments in any way.

If you elect to use the review sheet, you may find it helpful to frame your comments as suggested below.

Proposed Uniform Language Discipline Reviewed:

Reviewer Name:

Reviewer Organization:

Statements Approved for Use in Laboratory Reports and Expert Witness Testimony

Provide a summary of your assessment of the statements approved for use, including the most important highlights from the individual criteria comments.

- The statements approved for use are supported by scientific research.
- The statements approved for use accurately reflect consensus language.
- The statements approved for use are stated clearly.

Statements Not Approved for Use in Laboratory Reports and Expert Witness Testimony

Provide a summary of your assessment of the statements not approved for use, including the most important highlights from the individual criteria comments.

- The statements not approved for use are supported by scientific research.
- The statements not approved for use accurately reflect consensus language.
- The statements not approved for use are stated clearly.

**CRIME REDUCTION AND PUBLIC SAFETY TASK FORCE
FORENSIC SCIENCE SUBCOMMITTEE**

July 27, 2017 at 2:00 p.m.

Conference Line: (b) (6) /Passcode: (b) (6)

FSS Updates

- Crime Reduction and Public Safety Report Status
- Recent Events
 - Ninth Circuit Judicial Conference
 - National Association of Attorneys General Training and Research Institute Forensic Science Symposium (July 19-21)
- Needs Assessment Listening Session
- OSAC RFI
- Outstanding Projects

Component Updates

Uniform Language for Testimony and Reports

- ULTR 1.0
- ULTR 2.0
- Other Materials
 - Evett Article
 - OSAC Drafts

Upcoming Events

- International Association for Identification (August 7-12)
- International Conference on Forensic Inference and Statistics (September 6-8)
- Evidence Committee of the Judicial Conference (October 20)



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, Batohime – 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: ianevevt@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 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

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.

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

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. Brümmer, 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.

RE: FRE Conference on Forensics Materials

From: "Antell, Kira M. (OLP)" <(b) (6)>
To: "Goodhand, David (CRM)" <(b) (6)>, "Young, Cynthia (USAMA)" <(b) (6)>, "Wroblewski, Jonathan (CRM)" <(b) (6)>, "Hafer, Zachary (USAMA)" <(b) (6)>, "Hunt, Ted (ODAG)" <(b) (6)>, "Ibrahim, Anitha (CRM)" <(b) (6)>, "Smith, David L. (USAEO)" <(b) (6)>, "Shapiro, Elizabeth (CIV)" <(b) (6)>, "Isenberg, Alice R. (LD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>
Cc: "Morrissey, Brian (OAG)" <(b) (6)>, "Crytzer, Katherine (OLP)" <(b) (6)>, "Newman, Ryan (OLP)" <(b) (6)>, (b)(6), (b)(7)(C), (b)(7)(E) per FBI (OGC) (FBI) <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>, "Goldsmith, Andrew (ODAG)" <(b) (6)>, "Thiemann, Robyn (OLP)" <(b) (6)>, "Hur, Robert (ODAG)" <(b) (6)>
Date: Fri, 20 Oct 2017 11:52:57 -0400
Attachments: Lander Presentation_NCFS Meeting 12_EXCERPT.docx (35.03 kB);
4_lander_presentation_day1_session2.pdf (417.26 kB)

Good morning,

Some of you have asked what arguments I anticipate from the other FRE conference panelists. Those of you who are less familiar with these issues may find it helpful to review a presentation by Eric Lander, chair of PCAST, from a meeting of the National Commission on Forensic Science in January 2017. I have attached an edited transcript from that meeting (portions are highlighted to reflect Lander's likely arguments), his PPT slides, and a link to a video for those who want to see his presenting style (part 4 of the videos here: <https://www.nist.gov/topics/forensic-science/ncfs-meeting-12-webcast>).

I flag one thing in the transcript. (b)(5) per CIV

I look forward to our moot next week.

Thanks,
Kira

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

From: Antell, Kira M. (OLP)
Sent: Tuesday, October 17, 2017 4:06 PM
To: Antell, Kira M. (OLP); Goodhand, David (CRM); Young, Cynthia (USAMA); Wroblewski, Jonathan (CRM); Hafer, Zachary (USAMA); Hunt, Ted (ODAG); Ibrahim, Anitha (CRM); Smith, David L. (USAEO); Hur, Robert (ODAG); Shapiro, Elizabeth (CIV); Isenberg, Alice R. (LD) (FBI)
Cc: Morrissey, Brian (OAG); Crytzer, Katherine (OLP); Newman, Ryan (OLP); (b)(6), (b)(7)(C) per FBI (OGC) (FBI); Goldsmith, Andrew (ODAG); Thiemann, Robyn (OLP)
Subject: FRE Conference on Forensics Moot #2
When: Tuesday, October 24, 2017 2:00 PM-3:30 PM (UTC-05:00) Eastern Time (US & Canada)
Where: OLP Conference Room 4525 and Conference Line: (b) (6) /Passcode: (b) (6)

National Commission on Forensic Science
January 9-10, 2017
Transcript Excerpt: Scientific Foundations Panel

Presentation

ERIC LANDER: So, thank you very much. I'm delighted to come in response to your invitation to talk a little bit about the PCAST report. I'm just going to give an overview of it, and we'll have time for questions. I think we're going to do all three presentations, and then there will be Q&A afterwards, and I'm delighted to address them. But I co-chair PCAST. Just in case anyone is interested, PCAST is the sole science and technology advisory group to the White House.

There are many scientific advisory groups scattered through the federal government, but PCAST, in particular, works on those issues that involve policy and science, that crosscut science, technology, innovation, that may crosscut different departments, different fields. And we will work on things that are important to the economy, the defense. And so it's a very broad mandate.

And I've put up here, just for background, the range of things that PCAST does. We have produced, and I think we're done at this point, we approved our last three reports last Friday, so it will be a total of 39 reports, 37 open and two classified, across a wide range of topics in health, in energy and environment, the health of the U.S. Research Enterprise in general; several reports on advanced manufacturing that led to the creation of advanced manufacturing institutes we have in the country now, a report that just came out on ensuring the continued leadership of American semiconductors; a number of reports on information technology, including spectrum sharing that has had a very big impact, I'm proud of the work PCAST did there. A number of reports in education that are known amongst the PCAST folks as the Gates Report, because -- Gates Reports because your colleague, Jim Gates, and our colleague has really driven those reports. The report on forensic science about which I'll talk; multiple reports on nanotechnology, on agriculture, and there's a bunch of other ones on the website, but just to give a sense.

And so it is a group that consists of currently about 19 members, it's averaged about 20 members over the course of its life, 15 of whom have served for all eight years of the current administration. And they cover a wide range of topics here, expertise is all over the place. That's way too small to read, but you can look up the biographies on the website there. From academia and industry, and many, many, many different kinds of fields of study.

So, every report that PCAST undertakes is undertaken at the request of the President. So we'll have a conversation with the President. The President will say "I would like you to do a report on X." So, in this case, the President, as you know, in the creation of this very commission, is committed to your mission, which is ensuring the reliability of forensic science. And as another step in the same mission that created this commission, he asked PCAST to take a look at what else could be done to help support this mission. We came up with a plan of what we thought should be done. We sent it to the President. He said, "Yes, I like that work plan," and we proceeded to do that.

So my job today is to simply tell you about the nature of that report and then, when we get to the Q&A, we'll have a chance to -- you can pepper me with any questions you want. But briefly, we spent a year working on the report. We began -- well, we began talking with the President about it in early 2015. We actually launched our activities in September of 2015. We unanimously approved this report in September of 2016. Publically released it -- we usually take several weeks to do clean-up and proof-reading and all that. We publically released it about three weeks later. And then an addendum to the report was approved last Friday, which should be released today. I don't think it's up on the website yet, but it will probably be up by the end of the day or so, addressing remaining questions that had come up. So you'll have that addendum as well.

From the point of view of process, we tried to be, and we were in the case of this report compared to the other 38 reports we've done, maximally inclusive. This was the widest search for information and input that we had ever conducted. We spoke to 85 experts, the largest category amongst them are forensic

scientists, including eight from the FBI laboratory. And I should declare right at the beginning that I have enormous respect for the FBI laboratory. I have had the pleasure of working with people at the FBI laboratory going back to about 1990 when I had a chance to work with the director of the FBI laboratory on the early days of DNA fingerprinting. And I've collaborated with scientists there, and just have huge respect for the FBI laboratory. And so we were enormously grateful that a number of scientists, eight different scientists at the FBI laboratory were able to comment on specific details of our analysis of methods and things, and they were just great. Sent us all sorts of information about particular papers, "No, we disagree with that or that," and we made many, many, many improvements in response to these great scientists at the FBI laboratory.

Statisticians, we had a panel that included a number of judges who were enormously helpful. We really wanted to refer to them for the important legal context here. We also, for the first and only time in PCAST's work, put out a call for public comment or request for information, and received more than 70 detailed public comments. We're enormously grateful for those. And then we asked many agencies and individuals to suggest papers that we should review, and we ended up with a list of 2,100 papers which we reviewed. It was super. It was very, very helpful to get that list there, and it was not an easy thing to do, and I'm enormously grateful to our staff in this regard, one of whom is in the room, Tania Simoncelli, as well as Diana Pankevich and Kristen Zarrelli, who helped us in chewing up what papers we should look at more deeply. And then I and others on PCAST looked deeply at those papers. So, anyway, that was the process.

The report, you've all seen. It's about 173 pages, with 399 footnotes. It has a nine-page appendix that will come out today. And I'm sad to say we'll fall short of being the longest PCAST report by about six pages. There was another report that beats it. Oh, well. And it contains a variety of recommendations to a variety of agencies.

So let me just go to the main message of the report. The main message is, like, ludicrously simple. It is 173 pages. It does have 399 footnotes. It's a ludicrously simple message. First, and relevant to the discussion that you were just having, the report considers only forensic feature comparison methods. Lots of other things are interesting, but we didn't have time to look at them, so we've only looked at forensic feature comparison methods. And number two, it only concerns expert testimony in court. We make no statements whatsoever, the report pertains in no way whatsoever to what might be done in the course of investigation. I think investigation is a place where all sorts of inspiration may be helpful. We focused entirely on expert testimony in court.

Why? Because the federal law imposes a threshold requirement. It is the one place where hunches won't do. The law says one absolute thing, which is the basis of all of your work and our work. Expert testimony may only be admitted in court if it is based on methods that are reliable. More specifically 702C, "Expert testimony must be based on reliable principles and method," and 702D, that it "must be reliably applied." And Daubert makes very clear foundationally that in matters related to scientific evidence, evidentiary reliability rests on scientific validity.

So the reason why PCAST wrote a report is because that's a challenging thing. You have to know, what does it mean for a method to be reliable. So I want to just draw that strict distinction between all the wonderful conversations before about inspiration in investigations. I don't want to limit anybody from being inspired to find things that might get to the perpetrator in the course of investigation. But when you want to come to court, the law requires you need a reliable method.

So then the question is what's a reliable method? Well, if it's a feature comparison method, what you're saying is these features are so distinctive that it's reasonably likely that the evidence came from this source. It might be a class source. It might be an individual source. But that's a feature comparison method. I'm going to look at a set of features, I'm going to say I see them, I compare them to a potential source, they match within some degree, and they're distinctive enough. That's a really interesting scientific conclusion. And we took on the question, what does it mean for that method, this matches enough to that to draw a conclusion, to be reliable.

The entire PCAST report could be summarized as saying a forensic feature comparison method cannot be established as reliable unless you've empirically tested if it's reliable. That's it. You can't know otherwise. You have no business claiming a method is reliable if you haven't empirically tested if it's reliable. And then I guess a bunch of the rest of the report, chapter five, is devoted to the question of is that the case for many forensic feature comparison methods. Some yes. DNA simple mixtures, for example, DNA single source. The absolutely beautiful work that the FBI has undertaken with regard to latent fingerprints, yes. We can argue about what exactly the method is, but the -- what the accuracy is, but the FBI went out and measured it in their hands and they came up with an accuracy, a reliability.

You can't say a method is reliable unless you know its reliability, that is a number, something. It doesn't have to be perfect. Science is never perfect. There's always bounds on it. But the basic thing is if you know nothing about the reliability of a method, the method is not reliable. That's it. And it turned out that for a bunch of methods, we don't, because nobody's ever looked. We do -- we say that in greater length, with more footnotes, but that's basically what the report says is for a bunch of things, nobody's looked. Okay.

Now, I want to clear up a concern people might have. We have enormous respect for the professional practices within forensic science. They are very important to the practice of the field. They are valuable. They're important. Professional organizations, very important. Certification, very important. Accreditation, very important. Training programs, best practice manuals, extensive experience by examiners, papers in peer-reviewed journals, all those things I take my hat off to. They are very important to keeping a high quality field.

But it's very important to say none of them ever, no matter how much of it you do, can establish that a method is reliable unless you've empirically tested the method. The only relevant thing would be a peer-reviewed paper that tested the reliability. That would be very good. But peer-reviewed papers that describe other things count for nothing when we ask whether the method is reliable. Now, don't get me wrong, we do not fail to respect the importance of all of those things. They are very important, but they can never prove reliability because they don't test the only thing that matters. Does the method produce a result like it claims it produced? That's the heart of science.

We looked at seven areas. You know of them because you've, I'm sure, reviewed the report. I'm glad to address any of those. Basically, in two cases, it's clear that they're empirical tests that establish reliability and validity. In three cases, it's clear there's nothing there, just no tests. We couldn't find a thing that even would pass the laugh test. Bite marks, I can't imagine that anybody seriously, in this room, thinks that bite marks could pass the test of being a reliable method because there's no evidence to support it, and you can't support it without evidence. And the couple times people have tried, they're pathetically bad. I'd love to see great bite mark evidence, but you'd have to do an empirical test. And we all know that's not been done, so we know it fails the test of being a reliable method.

So we went through that, and we shouldn't be shy about it. Feel free to use an investigation. They have no business in federal court, obviously. In one case, we really did find a challenge, which was firearms. The issue is not that firearms people are not very smart and very careful and very thoughtful, and I have a lot of respect for the firearms community. They did a bunch of studies.

The big issue we ran into with firearms is many of those studies were designed in such a way, for example, these closed set tests, that they really did not come close to mimicking what really happens. And I mean, when you know that the answer is present in your possible known sources, you act very differently than when it's possible the answer isn't present there. And not surprisingly, those sort of tests produce error rates that are a hundred times lower than when you don't. And I realize that may not have been obvious to folks at the time, but now the evidence is super-duper clear, and you couldn't rely on the kind of "shooting fish in a barrel" kind of thing.

This is in no way to denigrate the forensic firearms examiners, because, in fact, there's been a study, a proper black box study, that has gone off and measured an error rate. And I believe you could do another one on the issue for us, and I'll admit this is the one that's exactly on the bubble is there's one such study.

From a scientific point of view, reproducibility matters, but that's a great one we could argue about. But I wouldn't argue very long because it's not hard to do that study, as demonstrated by the fact that somebody did it. So it's a great example.

And then, on complex DNA mixtures, I won't go into lots of detail other than to simply say this is an area that is ripe with potential. There are some things that are clearly solid, but it's clearly the case that if I said that's got a one-part-in-a-million contribution from you, you'd be pretty dubious about that and you'd want to know. And so the question there is not is DNA a good thing. I'm a big believer in DNA. It's what range -- within what range has reliability been demonstrated? We believe, reading the literature as best we can, it's been demonstrated within a certain range. And people can work to demonstrate it in larger and larger ranges. But again, it all comes down to this ridiculously simple notion of it's about empiricism. You actually have to do it empirically.

So what's needed? You know, a bunch of things. Threshold question of admissibility, should stuff come into court, you got to establish its reliability where it hasn't been established. You got to do it if you want to bring it into court. If you don't want to bring it into court, don't bother, that's fine. Now, some people, I think, have gotten the idea that PCAST somehow is not enthusiastic about the other things. By having focused on the reliability question as the threshold to admissibility, we may have given the impression we don't think it's also very important to do these fantastic white box studies. FBI has done white box studies. My colleague here, Dr. Champod, has looked at these things, namely what goes wrong, how do people make mistakes, how are conclusions drawn. Open up the black box and look inside. It's critical to improvement. We're wildly enthusiastic about that.

Moreover, we're sitting around talking about technologies that, frankly, ought to be converted in the next four or five years to objective methods. Companies can do facial recognition on the street really well. The technologies are becoming so good for matching, it would not at all be crazy to take -- yes, I know, latent fingerprints can be complicated, smeared, all sorts of things, but when I see what's happened with image recognition over the past three or four years, I believe this is an easy problem compared to many of the problems that are being solved.

And we can argue will we ever get rid of the forensic examiner; I'm not trying to ever get rid of the forensic examiner, but I have no doubt that a great deal of work could go to turn these things into objective methods. Firearms is actually somewhat easier because the patterns on bullets are much easier to digitize as a 3D image there and do matching on. And we're giving much too little attention to turning this into really objective science.

Why do I care? Because I think it will improve law enforcement. I think it will improve justice. I think it will decrease costs, increase accuracy. And we can argue all we want about should we do a black box method on this study, on the subjective method and all that. If we can turn it into an objective method, let's just do it. And then, finally, there's always the incremental improvements. Don't get me wrong, things can always get better, but what we were talking about is not could we get better, it's have we met the threshold test for this ability.

So those were the key things that we've talked about. We made a set of recommendations in the report. You've read them. And I won't go over those because you've got them all and we can always talk about them. And finally, the last point with regard to the appendix was the appendix was stimulated by the fact that the Department of Justice suggested in a statement following the PCAST report that "The report does not mention numerous published research studies which seem to meet PCAST criteria for appropriately designed studies providing support for foundational validity. That is, in short, we missed some important papers.

We take that very seriously, and so we reached out to the forensic community again, and to the Department of Justice, and said please tell us. We got back a bunch of responses of things we might look at additionally. And John Butler, God bless him, suggested we also look at the INTERPOL list of forensic papers. We actually reviewed the 8,000 papers you pointed us to. Happily, they are sorted by category, so we didn't have to read all 8,000 of them, but we did do that. And we could find no examples of

empirical studies that had been missed. We were wondering if the DOJ would write back to us about that. We didn't hear back from the DOJ. So I got in contact with Deputy Attorney General's Office and asked are there any other papers that you would like us to take a look at, and we were told -- this was about a few weeks ago -- that, no, in fact, upon reflection, there were no additional papers that DOJ thought PCAST should look at.

So, in any case, we've written up an eight-page appendix to the report that addresses the question of was there anything else missing. It readdresses this question of empiricism is necessary. There is no substitute for empiricism. And it makes clear our enthusiasm also for the white box in the other studies. I don't know exactly when that will go up, but it should be up sometime today. That's it on my end. And I think you guys have to wait for questions until we're done with the rest of the panel. Thank you.

[other presentations omitted]

Q&A

JOHN BUTLER: Thank you, Allen. So we have from now until about quarter to five to have a Q&A session with our three speakers. And we'll start with Jules.

JULES EPSTEIN: Thank you all. These questions are to Dr. Lander, although if Professor Champod wants to chime in, get it. I read the PCAST report. I'm not a scientist, but I have two follow-up questions from it. The part that talks about the foundational validity of latent prints, if I understand it, says there's enough there to show from black box studies that people can perform reliably. What it doesn't address, or if it does I missed it, so that's really question one, is what are the proper conclusions they are allowed to draw? In other words, I get it that it seems to say there's enough to say that, when given samples, people are really good at saying these two come from the same in a closed universe and these two don't. But that seemed different from any question of validity as to who else might have that, how many people in the world. So my first question is was I reading that right?

The flipside is taking a field like ballistics, firearms, where that said, okay, we're not there yet on the foundational support. I'm assuming, and this is where I'm, again, asking for enlightenment, that that still would permit some firearms comparison testimony. For example, this is a fire cartridge case at the scene. It's a 45-caliber. And the fire cartridge cases in the defendant's house are 45-caliber. And there are five lands and grooves, and there's this and that. So is there some demarcation on the latter discipline where you'd say, well, this much is okay, it's really after this line that we're concerned about? And thank you for all the work PCAST did.

ERIC LANDER: No, thank you. Let me do the last one first, because I think it will be most helpful. When we speak about a forensic feature comparison method, we're not speaking about a discipline. So, firearms is a discipline. A forensic feature comparison method is a way of doing a certain kind of comparison to reach a certain kind of conclusion. You might say can we tell whether these bullets were fired from the same class of gun; that's a method. And then you would test how well do you do in figuring out whether it was fired from the same class of gun. That wasn't actually what we did. We looked at the method and were very clear about can you associate it with a particular gun within a class. That's different. You can do a test for that.

So it might be the case that you can tell this bullet, I know what kind of bullet it is. That's good. It might be you can say I know what type of gun it was fired from. We didn't go look up papers on that. It turns out there aren't quite as many papers as you'd like, but it should be easier. You can tell which way the rifling was done by the way -- there are many things you can do to do class characteristics. We chose our method and we said associating with a specific gun within a class. For that, you can do a test. And I think there's confusion. There's not casting aspersions on a discipline, a field. There's "Is my method reliable."

So if we go back to your question about the proper conclusion about fingerprints, you might say, well, you know, in DNA, we're blessed with the distribution of alleles across a population, and we can make a database of the alleles. And we have the fact that, give or take, with very little bit of dependency, they distributed dependently and we can draw many conclusions about the frequency of each pattern. We

don't have a theory for the frequency of every fingerprint pattern, and therefore what we have to do is say, when I give you, Fingerprint Examiner, a bunch of fingerprints, latent prints taken from a particular place, and a bunch of possible knowns, as the FBI did running them against a large database to get the best match by the database, and then asked them whether, when they examined it, it was the right one. They looked at very large numbers by virtue of that process.

The goal was to make a statement that is roughly it's pretty likely it comes from that person who we've identified in the database. The right statement to make is when people try to test that method to see if it was reliable, they found that the data said they might be making mistakes at a rate of one in three-hundred and change, but not more than that. I'd be totally comfortable with somebody going to court and saying there's very good evidence from a study done that says people can get this right under circumstances that reasonably resemble -- let's not fuss too much -- what goes on. And they do make mistakes. It's not like one in a billion and not one in a million. They do make mistakes, but when they measured it was on the order of one in 300. Jury factor that into your thinking. That's a reasonable conclusion to make.

So, A, a method is not a discipline. And B, the statement is, approximately, when we try to do that thing, we get it right pretty often, and occasionally we get it wrong. That's all PCAST really says, and it's just all of what science says. That's simply science. . . .

[conversation omitted]

ERIC LANDER: No, I want neither, because the law compels that the expert may not be there unless it's based on a reliable method. It's not my choice. It's Congress's decision.

JOHN BUTLER: I have Bill and then Suzanne.

WILLIAM THOMPSON: This is just a great panel. I'm enjoying this a lot. I want to pursue the issue Jules raised of sort of the demarcation of when -- at what point do we have a method that requires validation and when are we simply having an opinion based on expertise that we don't require validation for? And I think there's a lot of confusion about this that surrounds the PCAST report. We heard from Jonathan Wroblewski of the DOJ this morning that, in his opinion, the PCAST report would mean no testimony could be given about footwear comparisons whatsoever because there had been no black box studies. So what I thought I would do was -- let's imagine for a minute that I'm a footwear analyst and let me give you a series of conclusions that I think I might want to give in court as an expert. And Eric, maybe you could tell me which of those conclusions you think are the result of a method that would require validation.

ERIC LANDER: All right, so you want "standing on one foot" answers to these footwear questions. All right. Let's see.

WILLIAM THOMPSON: All right. So I'm looking at a mark at the scene, and I'd like to say that it is a shoe print. Do I have to validate that? I'd like to say it has the same pattern as the shoe found in the defendant's apartment. Do I have to validate the --

ERIC LANDER: Pattern, meaning [inaudible].

WILLIAM THOMPSON: Yeah, it has the same -- the pattern of the shoe is the same, the sole pattern. Do I have to -- I'd like to say I've measured it and found it to be exactly the same size. Do I have to validate that I can measure correctly? I'd like to say that the degree of wear, it appears from the print that this is a worn shoe and that the defendant's shoe is also worn. Do I have to validate that I can make that determination? I'd like to say on these shoe prints I see some accidental characteristics. I see some cuts and so on that appear to be in the same location in the print as in the defendant's shoe. And I'd like to express the opinion that the likelihood of seeing so many similarities if these shoes are from a different source is really rather low. In fact, I -- in following the Champod method, I would like to actually state a likelihood ratio that it's, you know 893 times more likely, in my opinion, that I would see these things --.

ERIC LANDER: Right, in footwear, it's sort of been said to be billions of times, according to the Bodziak book.

WILLIAM THOMPSON: Well, I got my number the same way he got his.

ERIC LANDER: Yeah.

WILLIAM THOMPSON: So -- so the question is which of those statements that I might like to make in court as an expert should I be allowed to make, and which ones would need validation through empirical research?

ERIC LANDER: Okay. So, for starters, the PCAST report, just to be very specific, looked at identifying, that is to say associating, a footwear impression with a particular source. It did not look at class characteristics. We looked at what was the ability to say it came not from a Nike size 12 running shoe of this given make, but whether it came from your Nike size 12 running shoe of this make. That was the method we looked at for which we looked for data.

WILLIAM THOMPSON: Do you think to give a statement about class characteristics, to state --

ERIC LANDER: So, we said with regard to identifying characteristics, we said there is no evidence whatsoever that anybody had done anything approaching a meaningful test. And we've since gone, and as part of the supplement to the report, talked to the president of the IAI and to PEPA [ph] and to the leading worker in this field, and they both agree there's never been a study, there needs to be a study, and one is being done in West Virginia today. It's great. West Virginia University is undertaking the right kind of study. I'm thrilled to hear they're going to do it.

With regard to class characteristics, we didn't look, but one could ask is there evidence that people can identify which shoe it is. But now let's not agonize to death over is it a shoe, is it a worn shoe, is it this large, because, remember, why did we care about any of this? We cared about any of this because in these forensic pattern comparison methods people are saying the characteristics are sufficiently distinctive that it would be very unexpected to see that. It's because people are coming in wearing the mantle of expertise, claiming that they can make statements, whether they give an actual number or they just imply, "hint hint, wink wink," "It's really rare." Such statements are the ones we care about. That's what we mean by pattern comparison method.

So if you ask me is a forensic footwear examiner entitled to make statements associated with a footprint -- a shoe print with a particular source shoe, the answer is no, of course not, because nobody has ever bothered to test whether they are any good at it. What they do is they've written papers that say, "Oh, dings and marks on shoes, they're probably kind of random and we can probably kind of multiply the probabilities." That's not science. So, no, of course they couldn't give that.

With regard to class characteristics, we didn't look. Maybe there's a set of papers. I'm not aware if there are actually, but we didn't actually look because we take seriously a method needs to be validated, and we chose that method. With regard to "is it a shoe print," go for it. I'm not very worried that that's a statement of the sort that is going to cause anybody to think we are saying this is a distinctive set of features. It's a shoe. No risk there. Even it's a worn shoe, not a particularly big risk there. Let's be real. We are here because people have misstated, for a long time now, claims that they can do things with high probative value when there's no evidence. So we don't have to say exactly where do we have to draw the line to know that there's a bunch of things on this side of the line we have to attend to.

WILLIAM THOMPSON: But I'm not sure I'm hearing a principled distinction.

ERIC LANDER: Yeah, it is.

WILLIAM THOMPSON: A category of things that require validation, in your view, and the category of things that --

ERIC LANDER: Anything that holds it out as a scientific method, offering evidence beyond the ordinary can of a juror is the basic point. If it's within the ordinary can of a juror -- "It's a red hair" -- go for it. You're not going to snow a juror.

WILLIAM THOMPSON: Okay.

ERIC LANDER: If you're saying, "Oh, my God, this DNA pattern is one in a gazillion," I can snow you on that. The clear line is if it is something within the ordinary ability of a juror, ordinary experience of a juror, then the truth is we don't need to qualify the guy as an expert. The cop can say it's a red hair. The lawyer can say "Isn't that red hair." But when we're talking about things where a lay jury is unable to interpret that statement and must, in fact, say, "Well, science says," then science damn well better know.

WILLIAM THOMPSON: What if the expert wants to say "In my experience, it's rare to see two different shoes that have exactly the same cuts in the sole"?

ERIC LANDER: Well, you know, A -- A, unfortunately for that poor expert, Congress doesn't allow it. The federal rules of evidence say that hunches don't do it. Your doctor, your general practitioner, they can have hunches. They can treat you based on hunches. There's not a guarantee that everything your doctor tells you to do is supported by detailed studies because Congress didn't pass that law. But, of course, the doctor's working on your behalf with your consent on your side. When the State is bringing a piece of evidence against a person, the Congress has said "reliable methods." That's it. You got to have -- it's there. So is it reliable to say, "In my opinion, I've looked at lots of shoes and cases." By the way, you didn't know if those prints matched those shoes, because that's just experience. No, the answer is, no, it fails the test, at least in the United States. Europe might be different.

JOHN BUTLER: All right. Suzanne, then Jerry.

SUZANNE BELL: Thank you. This is -- again, thank you for your hard work on the PCAST. I really appreciate it. My question is a little bit more general. It's directed primarily to you, Dr. Lander, but I think both other panelists can comment on this. Because you mentioned that you read so many thousands of papers or reviewed them, one of the things that we've been very concerned about is the quality of the forensic literature. And I understand that you have a problem being blunt, but what's your evaluation of the state of the forensic literature and is it coming up to the standards that we would expect of a discipline such as molecular biology, chemistry? I mean, what are your thoughts on that?

ERIC LANDER: So let me start by saying there are some first-rate people working in forensic science. And I want to associate myself with Dr. Champod's plea that our goal is not trained dogs. Our goal is real careful scientists looking at processes, measuring things, understanding processes. I totally buy it. I take my hat off to the collection of FBI studies on latent fingerprints, the one published in the National Academy of Science, gorgeous piece of work. The white box studies are even more beautiful because they examine such questions as, well, when there's a certain threshold number of minutiae that you have to reach to get somewhere, boy, there's an interesting spike just past the right of that. These are real good scientists doing really good work.

What I take away from these excellent papers in forensic science, Dr. Champod's, others of these papers, is it's totally possible. It says there's no excuse for crap. There's just no excuse for crap. Now, when we go back in the literature earlier, and when we look in other fields, there's a fair amount of crap in the literature. I cite -- you know, PCAST cites some of the hair papers where the results section of the paper consists of a single sentence. "My assistant gave me seven sets of ten hairs and asked if I could match them, and I got them all right," that's one of the papers cited by the Department of Justice in supporting the validity of hair analysis. I recited you the entirety of the data in that paper. That ain't science, and we know it. And I don't blame the DOJ, they probably didn't even read the paper. But had they read the paper, they'd be embarrassed to quote that paper.

Or another hair paper where people took a whole bunch of hairs and they characterized them by a bunch of characteristics. They found which ones had similar characteristics and then they did microscopic hair examination to see if they should match. And they almost never made a false association. But the study was designed where every single hair came from a different person and the examiner knew it. So, going into the study, if you know that the right answer is it's not matching, it's not a study.

What we are seeing, and I want to make it really clear, we are on a wonderful threshold. Forensic science stands at a crossroads led by people like you, by people at the FBI, that either it's going to turn into a serious science that says we can and will measure methods, we do make mistakes, we will report our mistakes, and we'll give up some of the past. We do not in any way disparage forensic science or forensic scientists. We want great forensic science. The only way you get great forensic science is the good scientists say this other stuff, if it ever was acceptable, is no longer acceptable.

Why is this a problem? The only reason forensic science is in the state it's in is not because there aren't good people or people don't know what to do, it's because everybody's gotten themselves in this little trap, admitting the problem about methods that are in ongoing use runs the risk of opening past cases or existing cases. If it wasn't for that, everybody in this room would be saying -- if we were only prospective, if we could pass a bill that says you have a four-year grace period, and after the four-year grace period this stuff has to be reliable, people like Dr. Champod and others would do the studies. But we must deny, right now, that there's a problem because it could reopen past cases. No other science labors under that, and we should call it as it is. It's not that people don't know. It's not that they're not smart enough to do it. PCAST, because we can be blunt, is simply saying that's the only thing that stops this from being a reliable science.

JOHN BUTLER: Jerry and then Julia.

GERALD LAPORTE: What an excellent panel. I wish we had more time for discussion The point being though is that experience -- and I don't want to sort of oversell experience, but experience does have some factor. So if I -- and just one more point that I want to make, it sounds, sometimes, like when you're pulling out impression and pattern evidence comparisons or you're not understanding that there are actually measurements, okay, it's a semantical measurement. So if I compare a shoe print from a scene and I have an actual shoe, and I have a -- there's a rock in one place, I got a nick in another, and I got a wear pattern in another place, you know, we are actually measuring the constellation of those.

ERIC LANDER: We know that. So let me address your question.

GERALD LAPORTE: Just one last thing. So we don't actually say, well, the nick is four millimeters from the tip and the rock is, you know, five millimeters from the base over on this side. What we do is we actually do a comparison and we overlay them, and we say, oh, look, they actually occur in the exact same place.

ERIC LANDER: So we're well aware of that. We had the benefit of a lot of forensic scientists who have walked us through these methods. There's no doubt, there's measurement, there's comparison. The bite mark guys, they actually measure things, they put in distortion factors. All those are great things. They don't tell you that you're getting the right answer, but they're good things to have.

And the fact that you've said all these examiners have done lots of practice problems and they get them right back in the lab convinces you that they're probably doing a good job. The only question is the fact that you, examiner, happen to be convinced, or you, examiner's supervisor, happens to be convinced does not constitute reliability until you share it and publish it. You must take one more step. And you're telling me you're very confident that most of these methods will be reliable, they can be shown to be reliable, and all I'm going to say is, "Good, show them."

Let me be very clear. When we are totally blunt about the problems, there's no disrespect to forensic science and forensic scientists. We're not saying people are pulling things out of their butts. We're not saying that people don't care. We're not saying they're evil in any way. We're saying that in order to bring

a method to court, you have an affirmative obligation . . . and that obligation is not to believe that we're reliable but to show that we are reliable.

The NIJ is in just the position to show we are reliable. I will be thrilled if every one of the methods that we discuss can be shown to be reliable within some degree of reliability. And if it's wrong five percent of the time, fine, just be able to say so. That's why I was disturbed when the Department of Justice says its examiners can't discuss error rates because the heart of science is "How reliable?"

So I want to apologize to any forensic scientists who, in any way, were insulted or put out by the fact that we say these things. We say them because we need your attention. We say them because all these other great things you're doing, your wonderful training, accreditations, all of which I respect, the reason we say they count for nothing is because, from a scientific point of view, as important as they are and as hard-working as you are about it, and as much as we want you to do it, they count for nothing with regard to actual demonstration of reliability. You, of all people, are in a position to ensure that we take the PCAST report and show that all of our concerns are unfounded. Good luck with the bite marks part.

GERALD LAPORTE: And for the record, I don't disagree with what you're saying. I mean, I think -- I love the idea of black box studies, white box studies, finding out more --.

ERIC LANDER: I actually prefer the stuff he's doing, if we can do it.

GERALD LAPORTE: So I -- I totally agree with that. I think just my comment would be let's be careful not to necessarily just throw everything out when maybe we don't have that empirical evidence, but we have a lot of anecdotal.

ERIC LANDER: No, no, no. We must throw things out if they fail to meet the standard. The thing that you hear often from people who are trying to have it both ways is, "Well, let's not throw it out because they still know stuff." Have that in investigations. The law requires a threshold. It's not that you know some stuff, it's have you met the threshold. Answer, if you have not done an empirical study, you just haven't met the threshold, no ifs, ands, buts. No amount of experience will actually tell us that your belief is correct. Anyway.

. . .

JOHN BUTLER: Right now I have Julia, Fred, and then Bonner. I don't know if we'll have time to get to all of them, so.

JULIA LEIGHTON: This question hasn't come up and I don't -- I raise it because it's out there in the public sphere. And I would characterize it so that it's understood the context in which I'm saying it is if you have a hard time defeating the message, you defeat the messenger. And one of the criticisms that came out very quickly of the PCAST report ran along these lines, that the working group was made up of scientists that were predisposed, biased against forensics, and that the report fails because the report had no forensic scientists or, more importantly I think, the forensic practitioners involved in the drafting of the report. I don't see those as going to the message, but I think that we're going to hear this in our day-to-day practices a lot, that the fault -- that you can fault the entire report by faulting the messengers.

ERIC LANDER: Oh, okay. Well, those are -- that's great questions. So the bias against forensic science I'm not sure I get. I've admitted my bias in favor of forensic science, having worked closely with the FBI, having -- you know, with Bruce Budowle, with John Hicks, with those at the FBI, and having put a lot of effort into the early days of DNA to get that on a sound footing. So I'm clearly biased in favor of forensic science. Jim Gates, I believe, is biased in favor of forensic science because he's here. Notwithstanding the fact that both of us are biased in favor of forensic science, I think it's okay that we're working on the PCAST report. I'm on the board of the Innocence Project, which collaborates with the FBI on forensic science, and they've done great work together.

So, you know, I think the question is are the people on the PCAST report who are knowledgeable about forensic science biased against it? No, certainly not. There are only two out of -- of course, a 19-person group, only two are familiar with forensic science as a discipline, and neither of us are practicing forensic scientists, as it should be. That is the case for the other 38 reports that we do, that the goal was not to get practitioners writing the report. That's not what the President's asked for. What the goal was was to hear from lots of practitioners.

So, in making a public RFI and getting input from 70, and having 85 people we spoke to in a variety of different settings, the largest category being forensic scientists, in having sent drafts of the sections in chapter five to the FBI laboratory that were kindly distributed by the head -- by the FBI laboratory to experts in the field, who wrote back with lots of comments, almost all of which we took. The question is not did we get -- did we outsource our writing to forensic scientists, but did we listen. I'd say, in this report, we sought the input of the forensic community far more than any of the other reports we've done precisely because we know how much information and value there is there.

So I don't know -- in any case, you know, even if you thought one or two people were biased against -- like I say, Jim and I are biased in favor, otherwise we wouldn't spend our time on it -- you got to remember the rest of the PCAST has no biases, they haven't been in the field, and they've done what the President has asked them to do with regard to many fields. Cybersecurity, we're not all experts in cybersecurity, one or two people are knowledgeable, but we talk to the whole IC and to many people at universities and go down the whole list. So, no, I think that's kind of if you don't like the message, complain that, oh, my -- and, of course, what's the message? The message is science requires empirical evidence. I told you, that's the message.

The President's Council of Advisory on Science and Technology is mighty qualified to say that science requires empirical evidence. That is not a statement that pertains to any particular field. And in the case of footwear examination to identify a particular shoe, I don't need much to know that if there's never been a single study to measure its reliability, it ain't science. So this is not a tough call in that sense. Your other question had to -- those were your questions. Thank you very much for those good questions. I hope I was blunt enough.

[omitted]

###

Forensic Science in the Criminal Courts: Ensuring Scientific Validity Of Feature-Comparison Methods





PCAST makes policy recommendations in the many areas where understanding of science, technology, and innovation is key to strengthening our economy and forming policy that works for the American people.

39 Reports at the request of the President (2 classified)

- **Health**
 - Systems engineering for healthcare
 - Drug discovery and development
 - Health information technology
 - Pandemic flu vaccines
 - H1N1
 - Antibiotic resistance
 - Hearing technologies
- **Environment & Energy**
 - Climate change
 - Ecosystems and economy
 - Energy technologies
- **U.S. Research Enterprise**
- **Advanced Manufacturing**
- **Semiconductors**
- **Information Technology**
 - Privacy
 - Cybersecurity
 - Spectrum
 - Networking and IT R&D
- **Education**
 - Massively open online courseware
 - Tech and Training for middle skill workers
 - K-12 STEM education
 - Undergraduate STEM education
- **Forensic Science**
- **Nanotechnology**
- **Agriculture**

Current PCAST Members

Co-Chairs

John P. Holdren

Assistant to the President for Science and Technology

Eric S. Lander

President, Broad Institute of Harvard and MIT

Co-Vice Chairs

William Press

Univ Texas, Austin

Computer Science, Integrative Biology, Astrophysics

Maxine Savitz

Honeywell Corporation (retired)

Vice President National Academy of Engineering (former)

Members

Rosina Bierbaum

Univ Michigan, School of Natural Resources and Environment

Univ Maryland, Environment, Economics, Public Policy

S. James Gates, Jr.

Univ Maryland, College Park

Physics, String Theory, Particle Theory

Christine Cassel

Kaiser Permanente School of Medicine

Planning Dean, Gerontology

Mark Gorenberg

Zetta Venture Partners

Susan L. Graham

University of California, Berkeley

Electrical Engineering and Computer Science

Ed Penhoet

Alta Partners

Univ California, Berkeley (emeritus)

Biochemistry and Public Health

Chad Mirkin

Northwestern University

Chemistry, Nanotechnology

Eric Schmidt

Google (Alphabet)

Executive Chairman

Mario Molina

University of California, San Diego

Scripps Institution of Oceanography

Chemistry, Biochemistry, Atmospheric Sciences

Daniel Schrag

Harvard University Center for Environment

Geology, Environmental Science, Engineering

Michael McQuade

United Technologies Corporation

Senior VP for Science and Technology

Barbara Schaal

Washington University of St. Louis,

Dean of Faculty of Arts and Sciences, Biology

Vice-President, National Academy of Science (former)

Craig Mundie

Microsoft, Chief Strategy Officer (retired)

PCAST Report

Timeline:

Begun	Sept 2015
Unanimously Approved	Sept 1 2016
Publically Released	Sept 20, 2016
Addendum Approved	January 6, 2017

Process: Interviews and input from:

- ~85 experts (mostly forensic scientists (8 from FBI Lab), statisticians, judges, etc.)
- ~70 extensive public comments
- ~2100 scientific papers suggested and reviewed by PCAST

Report:

173 pages with 399 footnotes (plus 9-page addendum)

Recommendations to NIST, OSTP, FBI Lab, DOJ, Federal Judges

PCAST Report: Main Message

1. Report considers only (i) forensic feature-comparison methods and (ii) expert testimony in court.

Does not pertain to investigations

2. **Federal Law imposes a threshold requirement:** Expert testimony may only be admitted in court if it is based on methods that are **“reliable” and “scientifically valid”** (*F.R.E., Daubert*).

Requirement is not “flexible”

3. A forensic feature-comparison method cannot be established as **“reliable” unless the method itself has been empirically tested** to assess its degree of reliability.
4. *Some* important forensic feature-comparison methods have **never been subjected to meaningful empirical testing** to assess their reliability.

Good practices can't establish reliability of methods

Many practices are **valuable and important** in forensic disciplines

- professional organizations, certification, accreditation
- training programs
- best practices manuals
- extensive experience by examiners
- papers in peer-reviewed journals

However, *none* of these practices can establish in any way that a method is reliable or scientifically valid
-- because they don't actually test the method

Seven feature-comparison methods evaluated

1. DNA analysis of single-source and simple-mixture samples
2. DNA analysis of complex-mixture samples
3. Bitemark analysis
4. Latent fingerprint analysis
5. Firearms analysis
6. Footwear impression analysis
7. Microscopic hair comparison

Key issues

- In 2 cases, clear empirical tests establish reliability and validity
- In 3 cases, no empirical tests *whatsoever*
- In 1 case, only *one* empirical test properly designed to assess reliability
- In 1 case, issue is the *range* within which reliability has been established

What is needed

Threshold issue of admissibility: Establish Reliability

- **Black-box tests** for subjective methods not yet established as reliable and scientifically valid

Major improvement

- **White-box studies**, to understand and improve the methods
- **Technology development**, to convert subjective method to objective methods

Additional

- **Research aimed at incremental improvements**
- **Development of standards and best practices**

PCAST Recommendations

- 1. NIST should conduct ongoing evaluations** of validity and reliability of forensic science methods.
- 2. NIST (in partnership with others) should help move methods from subjective to objective** (e.g., fingerprints, firearms).
- 3. NIST should improve OSAC standards-development process** (forensic working groups) by adding a committee of independent scientists and statisticians.
- 4. OSTP should lead development of a national research strategy.**
- 5. FBI should undertake various scientific studies and receive increased funding.**
- 6. Attorney General should ensure that DOJ uses scientifically valid evidence.**
- 7. DOJ should withdraw and reissue its guidelines on testimony** (which forbid examiners from providing empirical evidence about accuracy).
- 8. Judges should “take account” of the scientific criteria for scientific validity.**

Forensic Science in the Criminal Courts: Ensuring Scientific Validity Of Feature-Comparison Methods



As Discussed

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Antell, Kira M. (OLP)" <(b) (6)>
Date: Wed, 12 Jul 2017 14:06:05 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.docx (203.71 kB)

Forensics Memo

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Goldsmith, Andrew (ODAG)" <(b) (6)>
Date: Wed, 12 Jul 2017 14:55:48 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.v2.docx (204.53 kB)

Andrew,

Attached is my draft memo. Looking forward to your thoughts/comments.

Best,
Chad

Fwd: Forensics

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Antell, Kira M. (OLP)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>
Date: Mon, 17 Jul 2017 08:48:56 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.v3.docx (204.84 kB); ATT00001.htm (216 bytes); 2017.07.12 OLP Memo on NCFS Recommendations.pdf (663.55 kB); ATT00002.htm (216 bytes); 2017.07.12 OLP Memo on ULTR and FSDR.pdf (1.48 MB); ATT00003.htm (168 bytes)

Begin forwarded message:

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Crowell, James (ODAG)" <(b) (6)>
Cc: "Hur, Robert (ODAG)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>
Subject: RE: Forensics

Jim,

Please find a draft of the memo attached. I've also attached two memos from OLP that provide a "deeper dive" into the issues discussed in the ODAG memo.

I strove to keep the ODAG memo very short—only two pages (including a header). While OLP's memos are a bit longer (4 pages and 6 pages, respectively), they do a good job of summarizing some tricky—and long running—issues.

Looking forward to your thoughts/edits. If the ODAG memo looks ready to go, I will print out a color copy for you to initial.

Best,

Chad

From: Crowell, James (ODAG)
Sent: Thursday, July 6, 2017 10:59 PM
To: Goldsmith, Andrew (ODAG) <(b) (6)>
Cc: Mizelle, Chad (ODAG) <(b) (6)>, Robert (ODAG) <(b) (6)>
Subject: Re: Forensics

Please send me a draft report before it goes to DAG. Thanks.

Sent from my iPhone

On Jul 6, 2017, at 7:46 PM, Goldsmith, Andrew (ODAG) <(b) (6)> wrote:

Chad - since you have forensics in your portfolio, let's work on this together. - Andrew

Sent from my iPhone - please excuse any typos.

On Jul 6, 2017, at 6:56 PM, Rosenstein, Rod (ODAG) <(b) (6)> wrote:

Please send me a brief memo summarizing the issues addressed in this article, and then let's meet to discuss:

https://www.washingtonpost.com/local/public-safety/science-organizations-renew-call-for-independent-us-committee-on-forensics/2017/06/28/3ab8cdea-5b6a-11e7-9b7d-14576dc0f39d_story.html?utm_term=.e52b48c4cf8e

Re: Forensic Science Subcommittee: Standing Meeting

From: "(b)(6), (b)(7)(C), (b)(7)(E) per FBI (OTD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>
To: "Antell, Kira M. (OLP)" <(b)(6)>, "Santos, Nelson A. (DEA)" <(b)(6), (b)(7)(C) per DEA>, "Isenberg, Alice R. (LD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>, "Czarnopys, Greg P. (ATF)" <(b)(6)>, "Smith, David L. (USAEO)" <(b)(6)>, "(b)(6), (b)(7)(C) per BOP (BOP)" <(b)(6)>, "Laporte, Gerald (OJP)" <(b)(6)>, "Downing, Richard (CRM)" <(b)(6)>, "(b)(6), (b)(7)(C) per USMS (USMS)" <(b)(6), (b)(7)(C) per USMS>, "(b)(6), (b)(7)(C), (b)(7)(E) per FBI (OGC) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>, "(b)(6), (b)(7)(C) per BOP (BOP)" <(b)(6)>, "(b)(6), (b)(7)(C), (b)(7)(F) per DEA", "Ibrahim, Anitha (CRM)" <(b)(6)>, "Williamson, Angela (OJP)" <(b)(6)>, "Dolan, Julia A. (ATF)" <(b)(6)>, "Scott, Kevin M. (OLP)" <(b)(6)>, "Begian, Lernik (OLP)" <(b)(6)>, "Thiemann, Robyn (OLP)" <(b)(6)>, "Mizelle, Chad (ODAG)" <(b)(6)>, "Newman, Ryan (OLP)" <(b)(6)>, "(b)(6), (b)(7)(C), (b)(7)(E) per FBI (OTD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>
Cc: "(b)(6), (b)(7)(C) per BOP" <(b)(6)>, "(b)(6), (b)(7)(C) per BOP" <(b)(6)>
Date: Fri, 14 Jul 2017 08:16:28 -0400

Kira,
Having reviewed (b)(6), (b)(7)(C) notes from yesterday I have one item that I would have added to the discussion yesterday (b)(5), (b)(6), (b)(7)(C) per FBI
your distro list as an FYI.

One follow up question regarding the September conference in Seattle on "Human Identification" - on which disciplines is that conference focused? DNA, physical anthropology?

Thank you,
(b)(6), (b)(7)(C), (b)(7)(E) per FBI
ologist
FBI-OTD-TODB
Building 27958A, Pod
Quantico, VA 22135
(b)(6), (b)(7)(C), (b)(7)(E) per FBI
(b)(6), (b)(7)(C), (b)(7)(E) per FBI

----- Original message -----
From: "Antell, Kira M. (OLP)" <(b)(6)>
Date: 7/11/17 5:10 PM (GMT-05:00)
To: "Santos, Nelson A. (DEA)" <(b)(6), (b)(7)(C) per DEA>, "Isenberg, Alice R. (LD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>, "Czarnopys, Greg P. (ATF)" <(b)(6)>, "Smith, David L. (USAEO)" <(b)(6)>, "Laporte, Gerald (OJP)" <(b)(6)>, "Downing, Richard (CRM)" <(b)(6)>, "(b)(6), (b)(7)(C) per USMS (USMS)" <(b)(6), (b)(7)(C) per USMS>, "(b)(6), (b)(7)(C), (b)(7)(E) per FBI (OGC) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>, "(b)(6), (b)(7)(C) per BOP (BOP)" <(b)(6)>, "(b)(6), (b)(7)(C), (b)(7)(F) per DEA", "Ibrahim, Anitha (CRM)" <(b)(6)>, "Dolan, Julia A. (ATF)" <(b)(6)>, "Scott, Kevin M. (OLP) (JMD)" <(b)(6)>, "Begian, Lernik (OLP) (JMD)" <(b)(6)>, "Thiemann, Robyn (OLP) (JMD)" <(b)(6)>, "Mizelle, Chad (ODAG)" <(b)(6)>, "Newman, Ryan (OLP) (JMD)" <(b)(6)>, "(b)(6), (b)(7)(C), (b)(7)(E) per FBI (OTD) (FBI)" <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>
Cc: "(b)(6), (b)(7)(C) per BOP" <(b)(6)>, "(b)(6), (b)(7)(C) per BOP" <(b)(6)>
Subject: Forensic Science Subcommittee: Standing Meeting

Attached is an agenda for Thursday's meeting. You'll note that BJA and NIJ will be sharing some information on their work on ensuring collection of lawfully owed DNA. They have asked that I distribute the attached short white paper.

I look forward to speaking with all of you this week. As always, don't hesitate to call or email me directly with questions or comments.

Thanks,
K

From: Antell, Kira M. (OLP)
Sent: Tuesday, July 11, 2017 9:21 AM
To: Santos, Nelson A. (DEA) <(b)(6), (b)(7)(C) per DEA>; Isenberg, Alice R. (LD) (FBI) <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>; Czarnopys, Greg P. (ATF) <(b)(6)>; Smith, David L. (USAFO) <(b)(6)>; (b)(6), (b)(7)(C) <(b)(6), (b)(7)(C) per BOP>; Laporte, Gerald (OIP) <(b)(6)>; Downing, Richard (CRM) <(b)(6), (b)(7)(C) per USMS>; (USMS) <(b)(6), (b)(7)(C) per USMS>; (OGC) (FBI) <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>; (b)(6), (b)(7)(C) per BOP <(b)(6), (b)(7)(C) per BOP>; Ibrahim, Anitha (CRM) <(b)(6)>; Williamson, Angela (OJP) <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>; (OTD) (FBI) <(b)(6), (b)(7)(C), (b)(7)(E) per FBI>; Julia.A.Dolan@usdoj.gov <(b)(6)>; Scott, Kevin M. (OLP) <(b)(6)>; Begian, Lernik (OIP) <(b)(6)>; Thiemann, Robyn (OIP) <(b)(6)>; Mizelle, Chad (ODAG) <(b)(6)>; Newman, Ryan (OIP) <(b)(6)>>
Cc: (b)(6), (b)(7)(C) per BOP <(b)(6)>; (b)(6), (b)(7)(C) per BOP <(b)(6)>>
Subject: RE: Forensic Science Subcommittee: Standing Meeting

Hello all!

I wanted to send a quick reminder that we will meet by phone only this week. I expect this meeting to be relatively short and I'd like to use our time to provide updates from our end on Department projects and receive updates from each of your components on issues or topics you'd like to share with the group. If you have any topics that may take a bit more time, please let me know and I will be sure to reserve it for you on our agenda. I will distribute an agenda on Thursday.

By Phone: Conference Line: (b)(6) /Passcode: (b)(6)

Thanks,
Kira

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

From: Antell, Kira M. (OLP)
Sent: Friday, June 16, 2017 4:00 PM
To: Antell, Kira M. (OLP); Santos, Nelson A. (DEA); Isenberg, Alice R. (LD) (FBI); Czarnopys, Greg P. (ATF); Smith, David L. (USAFO); (b)(6), (b)(7)(C) per BOP <(b)(6), (b)(7)(C) per BOP>; Laporte, Gerald (OIP); (b)(6), (b)(7)(C) per USMS <(b)(6), (b)(7)(C) per USMS>; (OGC) (FBI) <(b)(6), (b)(7)(C) per BOP>; (BOP); (b)(6), (b)(7)(C), (b)(7)(E) per DEA <(b)(6), (b)(7)(C) per BOP>; (DEA); Ibrahim, Anitha (CRM); Williamson, Angela (OJP); (b)(6), (b)(7)(C) per FBI <(b)(6), (b)(7)(C) per FBI>; (OTD) (FBI); (b)(6) Julia Dolan <(b)(6)>; Scott, Kevin M. (OLP); Begian, Lernik (OLP); Thiemann, Robyn (OLP); Mizelle, Chad (ODAG); Newman, Ryan (OLP)
Cc: (b)(6), (b)(7)(C) per BOP
Subject: Forensic Science Subcommittee: Standing Meeting
When: Thursday, July 13, 2017 2:00 PM-3:30 PM (UTC-05:00) Eastern Time (US & Canada).
Where: By Phone: Conference Line: (b)(6) /Passcode: (b)(6)

FW: OLP Revised Package_ (Original WF#: 3825829)

From: "Antell, Kira M. (OLP)" <(b) (6)>
To: "Eyler, Gustav (OAG)" <(b) (6)>
Cc: "Mizelle, Chad (ODAG)" <(b) (6)>
Date: Fri, 28 Jul 2017 09:20:24 -0400
Attachments: OLP Package on September NCFS Recommendations_06132017_DISTRIBUTED.pdf (8.55 MB)

Hi Gus,

This is the PDF exec sec package on the outstanding NCFS recommendations. As discussed yesterday, the current thinking is to permit Ted Hunt to start and review this package before the DAG signs it. I think there is a possibility that the public facing memo may shift but the consensus policy papers and the OLP recommendations are unlikely to move. Please let me know if you have any questions.

Thanks,
Kira

From: Begian, Lernik (OLP)
Sent: Tuesday, June 13, 2017 3:53 PM
To: Thomas, Sheaya (JMD) <(b) (6)>
Cc: Antell, Kira M. (OLP) <(b) (6)>
Subject: OLP Revised Package_ (Original WF#: 3825829)

Hi Sheaya,

OLP has a revised package that I will bring over shortly. The original WF # was 3825829. It will go to Chad Mizelle (ODAG) and Gustav Eyler (OAG). As always, I will appreciate it if you could share the new WF# when it is assigned.

Thank you,
Lernik

RE: Forensics

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Hur, Robert (ODAG)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>, "Crowell, James (ODAG)" <(b) (6)>
Date: Thu, 13 Jul 2017 11:03:14 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics_v6.docx (208.14 kB)

Rob,

Please find attached a revised draft, which incorporates your edits.

Best,
Chad

From: Hur, Robert (ODAG)
Sent: Thursday, July 13, 2017 9:17
To: Goldsmith, Andrew (ODAG) <(b) (6)>; Mizelle, Chad (ODAG) <(b) (6)>;
Crowell, James (ODAG) <(b) (6)>
Subject: RE: Forensics

Duplicative Information - See Bates Stamp 20220314-00992

Color Printing

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Simms, Donna Y. (ODAG)" <(b) (6)>
Date: Thu, 13 Jul 2017 14:18:51 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics_v6.docx (205.22 kB)

Donna,

Can you please print the attached in color? I will come by and initial and then give to Jim.

Best,
Chad



General

U.S. Department of Justice
Office of the Deputy Attorney General

July 13, 2017

MEMORANDUM FOR THE DEPUTY ATTORNEY GENERAL

THROUGH: Robert K. Hur
Principal Associate Deputy Attorney General

THROUGH: James A. Crowell
Chief of Staff and Associate Deputy Attorney General

FROM: Andrew D. Goldsmith
Associate Deputy Attorney General

Chad R. Mizelle
Counsel to the Deputy Attorney General

SUBJECT: Update on Forensics

On June 29, 2017, the Washington Post published an article entitled “Science Organizations Renew Call for Independent U.S. Committee on Forensics.” This memorandum briefly summarizes the major issues addressed in the article. We are also prepared to meet at your convenience to discuss the topics raised in the article, several of which will likely be addressed in separate briefings over the next few weeks.

President’s Council of Advisors on Science and Technology (PCAST) Report

Following a 2009 report from the National Academy of Sciences, which was critical of the state of forensic sciences, the Department instituted a number of measures (including the creation of the National Commission on Forensic Science) to improve the practice of forensics.

In 2016, PCAST issued a report concluding that additional changes were needed to further improve the practice of forensic science. Many of the current forensic issues facing the Department stem from the Report’s criticisms. Specifically, the report concluded that unless a forensic discipline has been scientifically validated – i.e., unless a discipline has a known error rate – then no testimony associating evidence to a source in that discipline should be admitted in court. The report also concluded that firearm, shoeprint, tire tread, and complex-source DNA analyses were not sufficiently validated.

When PCAST issued the report, the Department publicly declined to adopt the recommendations, stating that “the current legal standards regarding the admissibility of forensic evidence are based on sound science and sound legal reasoning.”

National Commission on Forensic Science (NCFS)

In April 2017, the Department declined to renew the charter of the NCFS, a federal advisory committee created to advise the Department on issues related to forensics. Simultaneous with this decision, the Department issued a request for public comment, seeking input on ways the Department can advance the practice of forensic science.

The Department received and is reviewing more than 250 comments in response. Some of the commenters advocate for reinstating the NCFS. Others argue that a body independent of the Department should be responsible for determining the validity of various forensic disciplines. Some simply request that outside scientists retain a role in Department decision-making. We expect to have a full analysis of the comments in the coming weeks.

NCFS's Recommendations

NCFS made six recommendations before its charter expired that are still pending before the Department. A separate four-page memo from OLP outlining the recommendations and proposed responses is attached. The recommendations related to digital accreditation, proficiency testing, and documentation are largely noncontroversial and all of the relevant components (ATF, BOP, DEA, FBI, USMS, CRM, EOUSA, and OJP) believe the Department should adopt these three recommendations with minor tweaks. The components also agree that the two recommendations related to medicolegal death investigation should be declined; because the Department does not conduct medicolegal death investigations, our equities in this area are limited.

The most controversial recommendation relates to “technical merit.” NCFS recommended that the National Institute of Standards and Technology (NIST) independently evaluate the technical merit of certain forensic practices. The components strongly oppose this recommendation on the ground that giving NIST this sort of oversight will likely have legal admissibility implications and result in the exclusion from court of reliable, accurate, and probative evidence.

Uniform Language for Testimony and Reports (ULTR) and Forensic Science Discipline Review (FSDR)

The ULTR is a prospective project aimed at providing consensus language for Department examiners to use in their testimony. This project began in the previous administration. Department components and external stakeholders unanimously support it. As noted in the Washington Post article, this project was put on hold to give the new administration an opportunity to review the project.

FSDR is a retrospective review of testimony provided by Department examiners to identify the circumstances that lead to testimonial overstatement, i.e., instances where an examiner's testimony is inconsistent with the underlying forensic report. Importantly, if this review identifies any inconsistencies in a case where there was a conviction, the reviewer would notify the prosecutors, defense attorneys, and defendants, thus potentially giving rise to fresh challenges. This project, too, is currently on hold. Although the previous administration publicly announced its intent to conduct the FSDR, it never began the project. The U.S. Attorney community strongly opposes the FSDR, primarily on the ground that it will force U.S. Attorneys to expend valuable resources re-litigating cases.

A separate six-page memo from OLP on the ULTR and FSDR is attached. OLP recommends that the Department continue with ULTR, while discontinuing FSDR. OLP notes, however, that discontinuing the FSDR will likely present some risks, including negative press, and

suggests an alternative means to ensure forensic examiners are testifying consistent with scientific principles and just outcomes.

Fwd: Forensics

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Antell, Kira M. (OLP)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>
Date: Mon, 17 Jul 2017 08:48:56 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.v3.docx (204.84 kB); ATT00001.htm (216 bytes); 2017.07.12 OLP Memo on NCFS Recommendations.pdf (663.55 kB); ATT00002.htm (216 bytes); 2017.07.12 OLP Memo on ULTR and FSDR.pdf (1.48 MB); ATT00003.htm (168 bytes)

Begin forwarded message:

From: "Mizelle, Chad (ODAG)" <(b) (6)>
To: "Crowell, James (ODAG)" <(b) (6)>
Cc: "Hur, Robert (ODAG)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>
Subject: RE: Forensics

Duplicative Information - See Bates Stamp 20220314-02134

Fwd: Forensics

From: (b)(6) Chad Mizelle
To: "Antell, Kira M. (OLP)" <(b)(6)>, "Goldsmith, Andrew (ODAG)" <(b)(6)>
Date: Mon, 17 Jul 2017 08:48:51 -0400
Attachments: 2017.07.12 Memo to DAG on Forensics.v3.docx (204.84 kB); ATT00001.htm (216 bytes); 2017.07.12 OLP Memo on NCFS Recommendations.pdf (663.55 kB); ATT00002.htm (216 bytes); 2017.07.12 OLP Memo on ULTR and FSDR.pdf (1.48 MB); ATT00003.htm (168 bytes)

Begin forwarded message:

From: "Mizelle, Chad (ODAG)" <(b)(6)>
To: "Crowell, James (ODAG)" <(b)(6)>
Cc: "Hur, Robert (ODAG)" <(b)(6)>, "Goldsmith, Andrew (ODAG)" <(b)(6)>
Subject: RE: Forensics

Duplicative Information - See Bates Stamp 20220314-02134


Upcoming Evidence Committee Meeting -- Please Respond by April 22.

From: "Shapiro, Elizabeth (CIV)" <(b) (6)>
To: "O'Callaghan, Edward C. (ODAG)" <(b) (6)>, "Fountain, Dorothy (ATR)" <(b) (6)>, "Goldberg, Stuart M. (TAX)" <(b) (6)>, "Himmelhoch, Sarah (ENRD)" <(b) (6)>, "Stemler, Patty (CRM)" <(b) (6)>, "Wroblewski, Jonathan (CRM)" <(b) (6)>, "Gardner, Joshua E (CIV)" <(b) (6)>, "Goldsmith, Andrew (ODAG)" <(b) (6)>, "Smith, David L. (USAEO)" <(b) (6)>
Cc: "Hunt, Ted (ODAG)" <(b) (6)>
Date: Thu, 11 Apr 2019 15:46:12 -0400
Attachments: FRE 615 - 2019 agenda.pdf (1.79 MB); FRE 106 - 2019 agenda.pdf (5.66 MB); FRE 702 - 2019 agenda.pdf (2.48 MB); Rule 404(b) - 2019 agenda.pdf (1.49 MB)

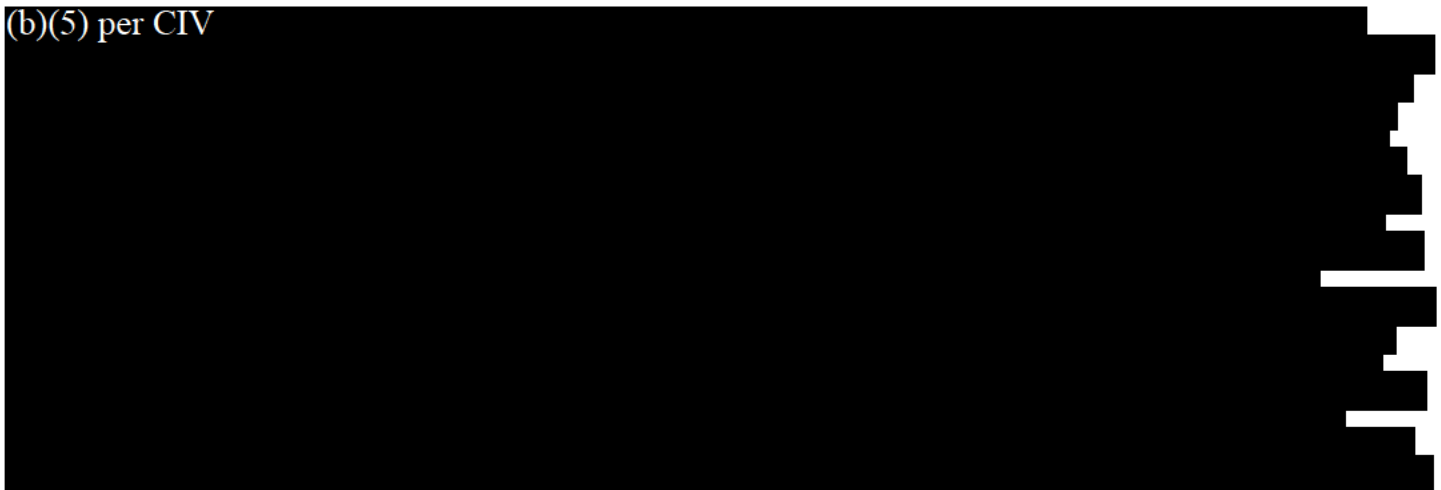
Dear Colleagues:

Our next Evidence Committee meeting is on May 3, and there are a number agenda items on which I would appreciate your input. Most of these proposals you have seen before, but they have evolved.

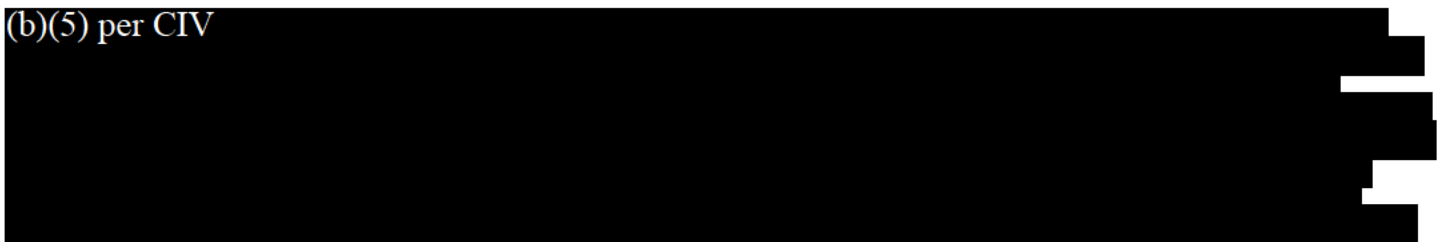
(b)(5) per CIV



(b)(5) per CIV



(b)(5) per CIV



(b)(5) per CIV

(b)(5) per CIV

(b)(5) per CIV

If you have questions or want to discuss, feel free to call or email me.

Best,
Betsy

ORDHAM

University School of Law

Lincoln Center, 150 West 62nd Street, New York, NY 10023-7485

Daniel J. Capra
Philip Reed Professor of Law

Phone: (b) (6)
e-mail: (b) (6)

Memorandum To: Advisory Committee on Evidence Rules
From: Daniel J. Capra, Reporter
Re: Possible Amendments to Rule 702
Date: April 1, 2019

The Advisory Committee has been considering possible amendments to Rule 702 for the last two years. By the time of the last meeting, the Committee's focus had narrowed to two possible changes:

1. An amendment that would prevent an expert from overstating the results that could be reliably obtained from the method used by the expert.
2. An amendment clarifying that the questions of sufficiency of facts of data and reliable application of method are questions for the court, and must be proved to the court by a preponderance of the evidence under Rule 104(a).

At the last meeting – after a miniconference that was devoted mostly to these two possible amendments --- the Committee requested that drafting alternatives be prepared to capture the concept of overstatement. As to the weight/admissibility issue, the Committee made no final determination, but interest was expressed in addressing the problem in a Committee Note should the amendment regarding overstatement be approved.

This memorandum further develops the matters that the Committee wished to further consider, based on discussion at the last meeting. It is divided into three parts. Part One is a discussion of the overstatement problem and whether an amendment might be useful. Part Two is a short discussion of the admissibility/weight problem. Part Three sets forth two drafting alternatives, and accompanying draft Committee Notes.

In addition, an extensive digest on recent case law on forensic evidence is set forth in the agenda book immediately after this memo. It was previously part of the memo but it got so lengthy that I thought it would be better accessed as a freestanding document .

I. The Problem of Overstatement

A. *Overstatement of Results in Forensics*

Many speakers at the Boston College Symposium in 2018 argued that one of the major problems with forensic experts is that they overstate their conclusions --- examples include testimony of a “zero error rate” or a “practical impossibility” that a bullet could have been fired from a different gun; or that the witness is a “scientist” when the forensic method is not scientific. Expert overstatement was a significant focus of the PCAST report. And a report from the National Commission on Forensic Sciences addresses overstatement with its proposal that courts should forbid experts from stating their conclusion to a “reasonable degree of [field of expertise] certainty,” because that term is an overstatement, has no scientific meaning and serves only to confuse the jury. The DOJ has weighed in with a prohibition on use of the “reasonable degree of certainty” language, as well as important limitations on testimony regarding rates of error (as discussed below).

Both the National Academy of Science and PCAST reports emphasize that forensic experts have overstated results and that the courts have done little to prevent this practice --- the courts are often relying on precedent rather than undertaking an inquiry into whether an expert’s opinion overstates the results of the forensic test.

Judge Rakoff, at the Boston Symposium, suggested that a provision prohibiting an expert from overstating results should be added to Rule 702 --- and that this would be a meaningful change because the courts have not relied on any language in the existing rule to control the problem of overstatement. And Judge Browning, at the Denver Symposium, suggested that while he does prevent overstatement by pruning an expert’s conclusions, textual language on overstatement might be useful to provide a specific source of authority.¹

It goes without saying that most of the problems of forensic overstatement occur at the state level --- and especially this may be so going forward, given the DOJ’s attempts at quality control at the federal level. But the case law digest on federal cases, set forth in the agenda book after this memo, supports the notion that overstatement of forensic results is a problem. There are many reported cases in which experts’ conclusions went well beyond what their basis and methodology could support --- claims such as zero rate of error, or opinions to a reasonable degree of scientific certainty. And, as discusses below, there is an argument that problems remain with forensic “identification” testimony even under the DOJ protocols. Thus, it would seem that there is good reason to seek to control overstatement, especially in forensic evidence cases. Such a venture would surely be more straightforward, and less science-dependent, than a rule that seeks to regulate forensic expert testimony from top-to-bottom.

¹ Though to be fair, Judge Browning also, in the context of an opinion about something else, appended long footnotes that generally came out against: 1) Amendments to Rule 702 of any kind, and 2) the Reporter, who was accused of pushing Federal judges around in order to justify his existence.

B. Can Overstatement by Forensic Experts be Controlled Without an Amendment?

Assuming that overstatement by forensic experts is a problem --- a pretty good assumption looking at the case law digest --- are there other sources of regulation that might make an amendment unnecessary? Three possible sources might exist: 1) Court regulation under existing law; 2) Education efforts; and 3) DOJ efforts to regulate forensic experts. These are discussed in turn.

1. Court Regulation: The case digest demonstrates that some courts are making efforts to control overstatement. But it is only a handful that are really doing so. Many courts *think* they are doing so by prohibiting experts from testifying to a zero error rate. But those courts as an alternative are allowing experts to testify to a reasonable degree of scientific or professional certainty, which is a meaningless and yet misleading standard. Given that most courts rely on precedent in this area, and that the *best* precedent is to allow testimony to a reasonable degree of scientific or professional certainty, there seems to be little hope for meaningful regulation by the courts any time soon.

2. Education: It might be thought that the NAS report, the PCAST report, and other sources would lead to more regulation of overstatement of forensic experts. But the case digest indicates that these reports have made very little practical impact on the courts. The National Commission on Forensic Science report attacking the “reasonable degree of certainty” standard was issued several years ago² and has been widely distributed, but courts are still happily using that standard as if it has solved the problem of overstatement. Judicial training through FJC may well be useful, but will it be as impactful as a rule amendment? Given the fact that courts rely heavily on precedent in evaluating forensic testimony, it would seem that for a *court* to act, a change of law is at least an important means of effectuating change in accompaniment with judicial education.

3. DOJ: The Department is making extensive efforts in trying to control some of the prior problems that were evident in the testimony of forensic experts. Apropos of overstatement, a DOJ directive instructs Department scientists working in federal laboratories, and United States attorneys, to refrain from using the phrase “reasonable degree of scientific certainty” when testifying, and to disclose other limitations on their results. There are a number of directives, each targeted toward a specific forensic discipline, but they all provide regulation on overstatement of results. An example is the directive regarding toolmark testimony, in pertinent part as follows:

² See <https://www.justice.gov/ncfs/file/795146/download>

- An examiner shall not assert that two or more fractured items were once part of the same object unless they physically fit together or when a microscopic comparison of the surfaces of the fractured items reveals a fit.
- When offering a fracture match conclusion, an examiner shall not assert that two or more fractured items originated from the same source to the exclusion of all other sources. This may wrongly imply that a fracture match conclusion is based upon statistically-derived or verified measurement or an actual comparison to all other fractured items in the world, rather than an examiner's expert opinion.
- An examiner shall not assert that examinations conducted in the forensic firearms/toolmarks discipline are infallible or have a zero error rate.
- An examiner shall not provide a conclusion that includes a statistic or numerical degree of probability except when based on relevant and appropriate data.
- An examiner shall not 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. An examiner may cite the number of examinations conducted in the forensic firearms/toolmarks discipline performed in his or her career for the purpose of establishing, defending, or describing his or her qualifications or experience.
- An examiner shall not 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 a judge or applicable law.

These standards addressed directly to overstatement obviously represent an important advance and they are an excellent development. *But despite these efforts there remains an argument that an amendment limiting overstatement will be useful and even necessary.* This is so for a number of reasons:

- There are questions of implementation of the DOJ protocols, as the edict has been in effect since 2016 and experts are still using the "reasonable degree" standard in many courts, according to the case digest. A case from 2018, discussed in the case digest, indicates that a ballistics expert was prepared to testify that it was a "practical impossibility" for the bullet to be fired from a different gun. Also there are questions about the impact of the DOJ standards on witnesses from state labs. This is not at all to understate the DOJ efforts. It is just to say that there may be room for court regulation as a supplement to these efforts.

- Even if the "reasonable degree" language is eradicated --- and it may not be because judges may require it --- there remains debate about what an expert *can* testify to as an alternative. One can argue that courts should be controlling such an important debate, the outcome of which can literally be the difference between freedom and a prison sentence.

- Leaving protections up to the DOJ means that any failure in compliance is not actionable—even though the result might be an unjust conviction, or more likely a guilty plea that would not otherwise have been entered.

- Adding something to Rule 702 that the Department is already doing should not be burdensome on the Department. Indeed there is precedent for such an approach --- the proposed amendments to the notice provisions of Rule 404(b), according to the Department, impose no obligations on U.S. attorneys that they are not already doing. Yet there is definite value to the system in codifying those obligations, as the Committee unanimously determined.

- The Department’s reforms, as salutary as they are, would not affect overstatement by experts called by any litigants other than the government in a criminal case.

- There is no guarantee that the Department’s protocols will remain in place --- administrations change, objectives change, and nobody has a right to enforce an existing DOJ protection. With an amendment to Rule 702, there is a pretty strong guarantee that limitations on overstatement will remain in place.

- Finally, Joe Cecil, an expert on forensic evidence, who is preparing the new FJC Manual on the subject, has provided a statement in response to the Reporter’s question about the DOJ standards. That statement indicates that the standards are a big step forward *but do not answer all concerns about overstatement*. Joe writes as follows:

Hi Dan,

You asked “If the DOJ standards on what forensic experts say is perfectly executed, are there still concerns about overstatement? If yes, please explain.”

The answer is yes, there are still concerns, especially regarding fingerprints and toolmarks.

First, it is important to note that the DOJ initiative will help to resolve some of the most important problems that arise in forensic science testimony. The DOJ standards will improve current practice by: 1) eliminating the use of the terms “reasonable degree of scientific certainty” and similar statements that have no scientific foundation; 2) eliminating claims that forensic techniques are free of error; 3) prohibiting forensic examiners from citing the number of examinations conducted as an indication of the accuracy of their conclusion; and, 4) offering statistical estimates without relevant and appropriate data. Monitoring forensic science testimony also will bring about greater consistency and allow early identification of emerging problems. These are important steps in strengthening the accuracy of forensic science testimony.

Nevertheless, concerns about overstatement of findings will persist. Based on the scientific assessments I have seen of forensic research on pattern matching evidence e.g., fingerprints, toolmarks I am confident that distinguished members of the science community will conclude that the current research does not provide a sufficient factual

foundation to support a statement by a forensic examiner that a comparison of two or more specific patterns indicate that they *originated from the same source* --- a conclusion that is permitted under the DOJ standards.

The courts may encounter this issue when there is a *Daubert* challenge to the proffered report and testimony of a forensic examiner that concludes that a comparison of two or more patterns indicate that they originated from the same source. For example, a forensic examiner may wish to testify that the correspondence between a fingerprint found at a crime scene and the fingerprint of a suspect indicates that the suspect is the source of the fingerprint, or that toolmarks found at a crime scene indicates that a specific tool in the possession of the suspect is the source of the crime scene toolmarks. The DOJ Uniform Language for Testimony and Reports for fingerprints and toolmarks would allow such testimony.

* * *

*The DOJ Uniform Language for Testimony and Reports attempts to walk a fine line between allowing the forensic expert to testify to identity of the source of a crime scene sample and disavowing any certainty that this is in fact the case.*³ * * * [T]he forensic examiner is allowed to conclude that the fingerprints or toolmarks originated from the same source. However, this conclusion is then subject to qualifications that make clear that such a conclusion should not be interpreted as indicating that the examiner has in fact identified the source of the crime scene pattern. According to the Uniform Language, a “source identification” of a toolmark means only that the examiner has seen sufficient pattern agreement to “provide extremely strong support for the proposition that the two toolmarks came from the same source and extremely weak support for the proposition that the two toolmarks came from different sources.” While this sounds as though the strength of the evidence is based on a statistical assessment, the Uniform Language makes clear that this is *merely the examiner’s opinion*, and has no statistical foundation * * *. The same tension is found in the Uniform Language for fingerprint identification.

[F]or these two types of pattern matching evidence, the Uniform Language permits the forensic examiner to testify that the crime scene sample came from the suspect, based only on the examiner’s subjective opinion that there is strong support for a match and weak support for no match. The Uniform Language offers no guidance on how to interpret what constitutes strong support and weak support, and disavows any suggestion that the conclusion is based on any knowledge of the frequency of different patterns in the population. Here is the relevant qualification from the Uniform Language for fingerprint examiners:

³ Reporter’s Note: This fine line (or fuzzy line) was evident in the explanations provided by the DOJ at the Denver Miniconference: See 87 Fordham L.Rev. at 1370-71 (explaining that a statement of identification is permissible because “it is not an empirical claim on the external world. . . “The claim is simply based on identification, and identification is different than individualization and uniqueness.”).

An examiner shall not assert that two friction ridge skin impressions originated from the same source to the exclusion of all other sources or use the terms ‘individualize’ or ‘individualization.’ This may wrongly imply that a ‘source identification’ conclusion is based upon a statistically-derived or verified measurement or actual comparison to all other friction ridge skin impression features in the world’s population, rather than an examiner’s expert opinion.

So under the Uniform Language forensic examiners may testify two prints originated from the same source, but not to the exclusion of all other sources since that would imply a scientific basis for the opinion. What am I missing? It is sufficient to say that this is just the examiner’s opinion with no additional support? Isn’t that the type of “ipse dixit” justification that the Supreme Court rejected in *GE v Joiner*?

Forensic examiners’ untethered opinion testimony that declares a match with no empirical basis is exactly what has raised the ire of the scientific community. The President’s Council of Advisors on Science and Technology (PCAST) questioned whether such a subjective conclusion would meet the FRE 702(c) standard of reliable principles and methods which it termed “foundational validity” . PCAST summarized its conclusion regarding pattern matching testimony as follows:

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

So, I believe it is fair to say that those scientists who prepared the PCAST report will still be concerned about overstatement, even if the DOJ standards are perfectly executed.

Similarly, the scientists who participated in the fingerprint identification study by the American Association for the Advancement of Science (AAAS) are likely to continue to be concerned about overstatement. The AAAS report noted that presently there is no basis “for assessing the rarity of any particular feature, or set of features, that might be found in a fingerprint. Examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and there are no scientific criteria for determining when the pool of possible sources is limited to a single person.” The AAAS scientists are unlikely to be swayed by DOJ standards that specifically reject the need for such statistical information as a basis for fingerprint testimony.

In fact, after the DOJ released the Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline, Rush Holt, the Chief Executive Officer for the AAAS wrote to Deputy Attorney General Rod Rosenstein, expressing concern about the Uniform Language for fingerprint examiners. Holt was particularly concerned about the

lack of scientific basis for the Uniform Language that allows an examiner to conclude that latent prints have a common source. The letter expressed the following concern:

There is an aspect of your Uniform Language, however, that is not in agreement with the scientific conclusions of the AAAS report. Although the Uniform Language you put forward forbids an examiner from making the unsupported claim that the pattern of features in two prints come from the same source to the exclusion of all others, it does allow examiners to say they “would not expect to see that same arrangement of features repeated in an impression that came from a different source.”

There is no scientific basis for estimating the number of individuals who might have a particular pattern of features; therefore, there is no scientific basis on which an examiner might form an expectation of whether an arrangement comes from the same source. The proposed language fails to acknowledge the uncertainty that exists regarding the rarity of particular fingerprint patterns. Any such expectations that an examiner asserts necessarily rest on speculation, rather than scientific evidence.

As there is no empirical basis for examiners to estimate the frequency of any particular pattern observable in a print, the term identification or, in your proposed language source identification, should not be used.

So concerns regarding overstatement will continue, even if the DOJ Uniform Testimony guidelines are perfectly implemented. The core problem is the decision to allow forensic examiners in some areas to testify that he or she can determine that the defendant is the source of the crime scene evidence i.e., source identification. There are a number of alternative forms of testimony that avoids these concerns. The AAAS report suggests the following testimony by a fingerprint examiner:

The latent print on Exhibit ## and the record print bearing the name XXX have a great deal of corresponding ridge detail with no differences that would indicate they were made by different fingers. There is no way to determine how many other people might have a finger with a corresponding set of ridge features, but it is my opinion that this set of features would be unusual.

Other forensic science agencies have disavowed the source identification standard. The Department of the Army Defense Forensic Science Center allows its fingerprint examiners to testify as follows:

The latent print on Exhibit ## and the record finger/palm prints bearing the name XXXX have corresponding ridge detail. The likelihood of observing this amount of correspondence when two impressions are made by different sources is considered extremely low.

While the subjective nature of the assessment is still a problem, this does represent a more measured statement than claiming to having identified the source of a crime scene print.

The 2018 Report of the American Statistical Association on Statistical Statements for Forensic Evidence supports Joe Cecil’s conclusion that the DOJ-sanctioned statement of “identification” raises the possibility of a problematic overstatement of an expert’s conclusions. The Association states as follows:

The ASA strongly discourages statements to the effect that a specific individual or object is the source of the forensic science evidence. Instead, the ASA recommends that reports and testimony make clear that, even in circumstances involving extremely strong statistical evidence, it is possible that other individuals or objects may possess or have left a similar set of observed features. We also strongly advise forensic science practitioners to confine their evaluative statements to expressions of support for stated hypotheses: e.g., the support for the hypothesis that the samples originate from a common source and support for the hypothesis that they originate from different sources.

The ASA report is addressing, in the above passage, the very concerns that support an amendment prohibiting overstatement. The ASA further states that “a comprehensive report by the forensic scientist should report the limitations and uncertainty associated with measurements, and the inferences that could be drawn from them” --- again, directed straight to the concerns that animate an amendment prohibiting overstatement.

In sum, even if the DOJ Guidelines are perfectly implemented, an argument remains for an amendment to Rule 702 that would specifically preclude an expert from overstating a conclusion.

C. Support for a Proposal to Regulate Overstatement

At the Chair’s suggestion, the Reporter contacted some individuals involved with the PCAST report to determine whether the working draft addressed to overstatement --- developed over the last few meetings --- was on the right track. They were asked their thoughts whether the proposed amendment will effectively address at least some of the concerns expressed about forensic expert testimony. There was no attempt to be comprehensive, because broader input is part of the public comment process .

Professor Brandon Garrett, an expert on forensic evidence at Duke Law School, reviewed the proposed amendment on overstatement and submitted this opinion:

I write to strongly endorse the revision presently under consideration to Rule 702, regarding the testimony of expert witnesses. My research includes work in law and in psychology, as well as collaborations with statisticians, and with forensic crime laboratories, regarding scientific evidence. I should note that the views expressed in this letter do not reflect those of Duke University or Duke School of Law, where I work, or that of the Center for Statistics and Applications to Forensic Evidence (CSAFE), a research center that I participate in.

The proposed revision would add a new subsection e , providing that an expert may not overstate the conclusions that may reasonably be drawn from the principles and methods used. I strongly favor this proposal. The central problem that this proposal addresses is that experts may reach conclusions that are not supported by the facts or by the method employed and that there has been a tendency in many disciplines to overstate conclusions.

Testimonial overstatement has contributed to large numbers of wrongful convictions. Experts have made such claims of infallibility, together with other unscientific and invalid claims, in a disturbing number of cases in which persons were ~~later~~ exonerated by post-conviction DNA testing. Brandon L. Garrett Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 Va. L. Rev. 1, 1 (2009) exploring “the forensic science testimony by prosecution experts in the trials of innocent persons, all convicted of serious crimes, who were later exonerated by post-conviction DNA testing” .

Nor is it an isolated problem. Entire disciplines have been plagued by testimonial overstatement. A massive FBI review of almost 3,000 cases involving microscopic hair comparison found that over 96% involved testimony flawed by overstatement of several different types. FBI/DOJ Microscopic Hair Comparison Analysis Review, at <https://www.fbi.gov/services/laboratory/scientific-analysis/fbidoj-microscopic-hair-comparison-analysis-review>. Indeed, 33 of those cases involving testimonial overstatement had been death penalty cases; in nine of those cases, the defendants had already been executed and five died of natural causes, as of March 2015.

Moreover, when such testimonial overstatement has occurred and has been brought to the attention of judges, in response, judges have often viewed their responsibility to regulate expert testimony as limited to the methods used and the admissibility of the type of expertise. Judges have sometimes viewed incorrectly, in my view the conclusions reached and how those conclusions are expressed as a matter for the jury to assess, rather than an integral feature of the expert’s work. In my view, the ultimate conclusion reached is an integral feature of the expert’s work and it must be reviewed as part of the judge’s

gatekeeping responsibilities. This proposal valuably addresses what has become, in practice, a very important and troubling gap in the coverage of Rule 702.

Obviously more could be done to address the problem that experts may draw conclusions that are overstated and do not follow from the facts or their methods. However, I also want to highlight the importance of the notes accompanying this proposal, which help to explain the concept of non-overstatement of conclusions. Perhaps most important is what the Committee Note says regarding failure to mention error rates. No 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. No technique that involves human interpretation or judgment is error free. And if a type of analysis was so reliable that no human judgment was involved, one would likely not need an expert to explain it and reach conclusions about it. The entire purpose of an expert is to contribute judgment, experience, and use of sound scientific methods to analysis of facts relevant in a case. In research conducted in collaboration with Greg Mitchell, we have found that error-rate information is highly salient to lay jurors. See, e.g. Brandon L. Garrett and Gregory Mitchell, *How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information and Error Acknowledgement*, 10 J. Empirical Legal Stud. 484 (2013).

In the past, unfortunately, experts have made false and startling statements, like that there was a “zero error” rate in their type of expert work. See, e.g. Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. Crim. L. Criminology 985, 1043, 1048 (2005). For example, the American Association for the Advance of Science AAAS report described “decades of overstatement by latent print examiners.” Am. Ass’n for the Advancement of Sci., *Latent Fingerprint Examination: A Quality and Gap Analysis* 11 (2017). Zero error rates do not exist but asserting infallibility would predictably impact the jury powerfully.

Not only should experts be barred from claiming infallibility, but they must disclose the actual error rates, if they have been adequately measured. If error rates for a method have not been adequately measured using sound “black box” studies under realistic conditions, then experts must disclose that their technique is of unknown validity and reliability and in such situations, other prongs of Rule 703 and Rule 403 may each bar admissibility of the expert testimony).

Expert evidence should never be presented in court without evidence of its error rates and of the proficiency or reliability of not just the method, but the particular examiner using the method. See President’s Council of Advisors on Sci. Tech., Exec. Office of the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* 9–11 (2016). Such proficiency testing should involve tests of realistic difficulty and such testing should be done blind, so that the participant does not know that it is a test. Jonathan J. Koehler, *Proficiency Tests to Estimate Error Rates in the Forensic Sciences*, 12 Law, Prob. Risk 89, 94 (2013) “Blind proficiency testing has been used in some forensic science areas, including the Department of Defence’s forensic

urine drug testing programme and the HIV testing programme.” ; Joseph L. Peterson et al., *The Feasibility Of External Blind DNA Proficiency Testing. II. Experience With Actual Blind Tests*, 48 J. Forensic Sci. 1, 8 2003).

Jurors should hear about the proficiency of the particular expert, and of that person’s reliability in reaching conclusions using a method. Brandon L. Garrett and Gregory Mitchell, *The Proficiency of Experts*, 166 U. Penn. L. Rev. 901 2018); see also Gary Edmond, *Forensic Science Evidence and the Conditions for Rational Jury Evaluation*, 39 Melb. U. L. Rev. 77, 85-86 2015) “[R]egardless of qualifications and experience, rigorous proficiency testing tells us whether the forensic analyst performs a task or set of tasks better than non-experts or chance. A significantly enhanced level of performance is precisely what it means to be an expert.” .

In the past, scientific experts have also used vague terminology like “identification” or “match” – and the Committee Note could valuably note that there are additional types of problematic conclusion testimony apart from the use of terms like “reasonable scientific certainty.” The AAAS report, for example, noted that terms like “match,” “identification,” “individualization,” and other synonyms should not be used by examiners, nor should they make any conclusions that “claim or imply” that only a “single person” could be the source of a print. AAAS Report at 11.

The Committee Note could also address claims of experience – which can be used to bolster statements that something the expert observes is rare or common based on one’s experience, without citing to any empirically valid support. The Department of Justice’s Model Uniform Language on Latent Fingerprint Evidence, for example, explicitly cautions against the use of such experience-based claims to suggest probabilities connected with a conclusion, as does the protocol for the FBI’s review of microscopic hair evidence. FBI/DOJ Microscopic Hair Comparison Analysis Review, at <https://www.fbi.gov/services/laboratory/scientific-analysis/fbidoj-microscopic-hair-comparison-analysis-review>.

I also note that some experts testify about general research, and are therefore cautious about connecting general research to the facts in a case, and therefore may be much less likely to risk overstatement. For example, experts may also testify about more general scientific research to provide a “framework” to educate factfinders, and they may explain industry or professional norms as well. See Laurens Walker John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559, 570 1987).

I hope that these views are of use as you consider this important proposal. Please feel free to contact me at your convenience if I can be of further assistance.

In addition, a number of experts involved in the PCAST report have reported that the amendment, and especially the Committee Note, would be useful in regulating what that PCAST found to be a significant problem of overstatement. Among those who have reviewed the draft amendment are Dr. Eric Lander who provided some suggestions on the Committee Note, Judge Patti Saris, and Dr. Karen Kafadar. All thought that the amendment and the Note would be an important tool in addressing a real problem.

D. Trial Court Evaluations of an Expert’s “Credibility”

At the last meeting, during the discussion of the proposed amendment on overstatement, the thought was expressed that the amendment might lead to the court assessing the “credibility” of an expert, and that this was inappropriate. The example discussed was an expert testifying that he was “certain” of his opinion; under the amendment, the trial judge might have to exclude the testimony if she found that the testimony of “certainty” was an overstatement given the underlying data and method that the expert used. The thought was expressed that such an exclusion would amount to a credibility determination, and the credibility of the expert is to be left to the jury.

But the process that the judge used in this hypothetical would be no different than that used to judge any of the other admissibility requirements currently in Rule 702. For example, if an expert states that he relied on sufficient data, and the judge finds that the data is not sufficient to support the opinion, the judge must exclude the evidence. Is the judge in that case wrong because she does not believe the expert’s assertion? If “credibility” assessments are prohibited in that circumstance, then logically the judge cannot disagree with any of the expert’s assertions, because to do so would challenge the expert’s credibility.

In fact a *Daubert* hearing today is rife with “credibility” determinations. If an expert states that he relied on a report, but the adversary shows to the judge’s satisfaction that the expert could not have so relied and come to the opinion he did, then the judge should disregard the expert’s assertion and review the expert’s basis accordingly. Similarly, under the proposed amendment, if the expert states that there is a zero rate of error when a forensic methodology applies, that assertion is demonstrably untrue --- incredible --- and the expert should be prohibited from testifying to that overstatement.

The role of “credibility” determinations at a *Daubert* hearing is complicated, but credibility determinations are clearly not always barred. If the expert says that he employed a reliable method, or that his conclusion is not an overstatement, it may be that the expert did not in fact employ reliable methods, or did in fact overstate the conclusion. If the trial judge does not intervene, this would mean that the jury would hear unreliable expert testimony, contrary to the principle of *Daubert*.

Judge Becker considered the complex relationship between expert credibility and reliability in *Elcock v. Kmart Corp.*, 233 F.3d 734, 750–751 (3d Cir. 2000). The trial judge in

Elcock held a *Daubert* hearing and determined that one of the plaintiff's experts did not pass the reliability threshold. The judge relied in part on the fact that the expert had engaged in criminal acts involving fraud, and so was not a credible witness; the fraudulent activity was not in any way related to the expert's professional life, however. Judge Becker found the trial court's reliance on these bad acts to be error, and stated that on remand "the district court should not consider Copemann's likely credibility as a witness when assessing the reliability of his methods." Judge Becker added, however, the following important elaboration:

We do not hold ... that a district court can never consider an expert witness's credibility in assessing the reliability of that expert's methodology under Rule 702. ***Such a general prohibition would be foreclosed by the language of Rule 104(a), which delineates the district court's fact-finding responsibilities in the context of an in limine hearing on the Daubert reliability issue. Indeed, consider a case in which an expert witness, during a Daubert hearing, claims to have looked at the key data that informed his proffered methodology, while the opponent offers testimony suggesting that the expert had not in fact conducted such an examination. Under such a scenario, a district court would necessarily have to address and resolve the credibility issue raised by the conflicting testimony in order to arrive at a conclusion regarding the reliability of the methodology at issue.*** We therefore recognize that, under certain circumstances, a district court, in order to discharge its fact-finding responsibility under Rule 104(a), may need to evaluate an expert's general credibility as part of the Rule 702 reliability inquiry.

While Judge Becker properly concluded that credibility determinations would have to be made at a *Daubert* hearing, he emphasized that those determinations are limited to testimony about how the expert reached her opinion, as opposed to witness-credibility more generally:

Although *Daubert* assigns to the district court a preliminary gatekeeping function—requiring the court to act as a specialized fact-finder in determining whether the methodology relied upon by an expert witness is reliable—it does not necessarily follow that the court should be given free rein to employ its assessment of an expert witness's *general credibility* in making the Rule 702 reliability determination. To conclude otherwise would be to permit the district court, acting in its capacity as a *Daubert* gatekeeper, to improperly impinge on the province of the ultimate fact-finder, to whom issues concerning the general credibility of witnesses are ordinarily reserved.

Thus the distinction as articulated by Judge Becker is between credibility determinations bearing *directly* on the expert's methods and application, and general credibility issues that apply to all witnesses. Judge Becker posited the following example:

For instance, in situations involving an attempt to attack an expert witness's credibility on the basis of prior bad acts or convictions, at least one prominent evidence commentator has noted that an expert's prior dishonesty or misconduct should not qualify as an appropriate factor in assessing methodological reliability *when the acts are wholly unrelated to the expert's use of a particular methodology*, but that a court should take such

dishonesty or misconduct into account when the nexus between the acts and the expert's methodology is more direct, e.g., when the prior dishonest acts involve fraud committed in connection with the earlier phases of a research project that serves as the foundation for the expert's proffered opinion. See Edward J. Imwinkelreid, *Trial Judges—Gatekeepers or Usurpers? Can the Trial Judge Critically Assess the Admissibility of Expert Testimony Without Invading the Jury's Province to Evaluate the Credibility and Weight of the Testimony*, 84 Marq. L. Rev. 1, 39 (2000). Under this approach, for instance, the fact that an expert witness falsely reported his salary on an income tax return has little if any bearing on the reliability of a diagnostic test he frequently employs, but the fact that the expert lied about whether his methodology had been subjected to peer review, or *intentionally understated the test's known rates of error, is a different matter entirely*.

It would seem that the Becker quote above is spot-on for answering concerns about “credibility” determinations made by a judge ruling on possible overstatement of an expert's conclusions. If the expert overstates the certainty of a conclusion (understates the rate of error) then *Daubert* obligates the judge to prohibit such an unreliable assertion from being made at trial.

Thus, if the attack on credibility has nothing to do with the expert's methods, but only with a general character for truthfulness, the issue of credibility should be left to the jury—the opponent can bring impeachment evidence before the jury by way of cross-examination as with any witness. As applied to the facts of *Elcock*, the credibility evidence should not have been used by the trial court, because it related to acts of dishonesty and fraud completely outside the expert's work in the particular case.⁴ On the other hand, if the expert in *Elcock* were found to have misstated or even lied about doing a test in this particular case, the trial court must disregard the expert's conclusion that is purportedly based on the test. If that is a “credibility” determination, then so be it.

It should be noted that while a trial court is considering credibility when evaluating an admissibility requirement under Rule 702 (such as sufficiency of basis), the addition of an overstatement requirement would not, and should not, be a vehicle allowing the trial judge to nitpick an expert into oblivion. Nothing in an amendment limiting overstatement requires the judge to get into the difference between “highly likely” and “very likely” for example. The preponderance standard of Rule 702 does not require that the expert be absolutely correct or completely precise. The draft Committee Notes, *infra*, emphasize this point.

In sum, the proposed amendment limiting overstatement is no different from any of the existing admissibility requirements of 702 insofar as there is concern that trial judges will improperly make “credibility” determinations. If the judge finds that the expert overstated the opinion, then the trial judge should prohibit the opinion.

⁴ See also *Cruz-Vazquez v. Menmonite Gen. Hosp. Inc.*, 613 F.3d 54 (1st Cir. 2010) (error to exclude expert because he was biased in favor of plaintiffs in medical cases and was generally affiliated with plaintiffs' lawyers; those considerations are for the jury in assessing the weight of the expert's testimony).

E. Should a Rule on Overstatement Apply Beyond Forensics?

While overstatement by experts in areas other than forensics is less publicized, there are arguments for any amendment regulating overstatement to apply to all expert testimony. Those arguments are:

1) a limit to “forensic” experts would skew Rule 702, because all current parts of the rule apply to all experts;

2) the term “forensic” is hard to define in rule text, as it goes beyond feature-comparison (for example to arson investigations) and there are disputes about just which disciplines are forensic;

3) there is no other rule of evidence that focuses specifically on a subset of witnesses;

4) if it is a good idea to require a court to regulate overstatement, it certainly can’t hurt to have that tool available outside the forensic disciplines; and

5) *Most importantly, there are a number of reported cases in which an expert appears to have gotten away with a conclusion that is not fairly supported by the data, methodology and application. And there are many cases in which the courts have required an expert outside of forensics to testify to a “reasonable degree of [field of expertise] certainty.”*

That is, there is a problem of overstatement outside the forensic area. And while it is not as evident as in the forensic area, overstatement does exist. What follows is a case digest:

Case Digest on Overstatement by Non-Forensic Experts⁵

1. Expert Overstatement Permitted

In some federal cases, non-forensic expert opinion testimony is admitted that appears to overstate the conclusions that reliably flow from the expert’s methodology. See, e.g.:

- * *United States v. Machado-Erazo*, 901 F.3d 326 (D.C. 2018): The government offered an expert on cellphone location. The disclosure under Rule 16 was deficient, because the “report” was nothing but pictures of cellphone towers. (!) At a hearing the government assured the trial judge that the expert would offer testimony about only the “general location” of cell phones, rather than precise locations. At trial, before a different judge, the

⁵ This digest is not intended to be comprehensive. It collects a representative example of cases decided within the last five years. The digest was prepared with the substantial help of Professor Liesa Richter.

expert testified to precise locations. The court of appeals found that it was error to admit this testimony --- and that there was a violation of Rule 16 --- but found the error to be harmless.

- ***United States v. Chikvashvili***, 859 F.3d 285, 292-93 4th Cir. 2017) government expert in healthcare fraud resulting in death prosecution was permitted to testify that the misreading of patient x-rays was the “but-for cause” of two patients’ deaths and that standard medical procedures “would have averted” their deaths. Doctor also opined that one patient’s elective surgery “would have been postponed” with an accurate reading of his x-ray).
- ***United States v. Tingle***, 880 F.3d 850, 855 7th Cir. 2018) rejecting defendant’s argument that DEA agent’s expert testimony violated FRE 704(b) where agent testified that the amount of drugs found in defendant’s residence was “definitely for distribution” and that the gun found in residence “was utilized by [the defendant] to protect himself and/or the methamphetamine and the currency.” .
- ***Adams v. Toyota***, 867 F.3d 903, 916 8th Cir. 2017) affirming admission of expert testimony in which an engineer “ruled out” pedal misapplication as a potential cause of sudden acceleration accident .
- ***United States v. Lopez***, 880 F.3d 974 8th Cir. 2018) affirming admission of DEA agent’s expert testimony that appellate court characterized as opining that “illegal drugs entering the market are of such high purity that it has become physically impossible even for seasoned addicts to consume large amounts of methamphetamine” .
- ***Wendell v. Glaxo Smith Kline, LLC***, 858 F.3d 1227 9th Cir. 2017) district court erred in excluding medical experts’ opinions that prescription drug caused the plaintiff’s rare cancer where one expert testified to “a one in six million chance” that the plaintiff would have developed the cancer without exposure to the drug).
- ***United States v. Wells***, 879 F.3d 900 9th Cir. 2018) affirming admission of expert testimony by a tire expert to refute a murder defendant’s alibi that he was not at work at time of murders because he got a flat tire; the expert concluded that the nail in the tire “had been inserted” in the tire “manually” rather than picked up while driving .
- ***United States v. Lozano***, 711 Fed. App’x 934 11th Cir. 2017) permitting government’s drug trafficking expert to testify that “blind mule theory” has “no factual basis” .
- ***U.S. Information Systems, Inc. v. International Broth. of Elec. Workers Local Union No. 3, AFL-CIO***, 313 F.Supp.2d 213 S.D.N.Y. 2004): An expert in antitrust economics testified to damages, and the opponent argued that the claims were overstated, because he used a discounting factor that was unsupported. The court held that the expert could testify, concluding that while “the accuracy of Dr. Dunbar s figures may be open to dispute, his methodology with respect to damages is sound.”

- *Flavel v. Svedala Indus.*, 875 F.Supp. 550 E.D.Wi. 1994) in an age discrimination action, the fact that a statistics expert artificially inflated his findings by using employee ages as of a certain date raised a question for the jury, not the court .
- *Etherton v. Owners Ins. Co.*, 35 F. Supp.3d 1360, 1364, 1368 D. Colo. 2014), aff'd 829 F.3d 1209 10th Cir. 2016) rejecting challenge to admission of expert testimony that the plaintiff's many injuries "were entirely caused" by collision and that "every single rear-end collision that has ever occurred" is a plausible mechanism for causing lumbar disc injury .

2. Expert Overstatement Regulated

There are a number of reported cases in which it appears that courts are regulating expert attempts to overstate their results (sometimes by appellate court correction):

- *United States v. Machado-Erazo*, 2018 WL 4000472 D.C. Cir.) district court erred in admitting FBI agent's expert testimony about "precise location" of cell phones "within a half mile" of a particular cell tower, but the error was harmless .
- *United States v. Naranjo-Rosaro*, 871 F.3d 86, 96 1st Cir. 2017) trial court erred in allowing agent handling drug-sniffing dog to testify as a lay witness, but error was harmless where agent's testimony would have been admissible expert opinion and where the agent conceded that the dog's alerts to drugs "did not establish the presence of drugs in the house" .
- *In re Vivendi Sec. Litig.*, 838 F.3d 223, 256 2nd Cir. 2016) affirming admissibility of expert testimony based upon an event study about artificial inflation in a company's stock price due to misapprehension of a company's liquidity risk; emphasizing that the expert did not purport to establish that the company's fraud *caused* the misapprehension).
- *Nease v. Ford Motor Co.*, 848 F.3d 219, 225 4th Cir. 2017 reversing a verdict for the plaintiff in a product liability action due to the district court's erroneous admission of testimony by the plaintiff's expert "to a reasonable degree of engineering certainty" that the throttle on the plaintiff's truck contained a design defect that caused an acceleration accident; the expert's opinion was not supported by the information he had and the methodology he used).
- *Rheinfrank v. Abbott Labs, Inc.*, 680 Fed. App'x 369, 376 6th Cir. 2017) finding no error in the district court's ruling refusing to allow the plaintiff's regulatory expert to testify that "DepoKote was known to be the most teratogenic drug"; the expert was not in a position to evaluate the relative risks of epilepsy drugs .
- *Abrams v. Nucor Steel Marion, Inc.*, 694 Fed. App'x 974 6th Cir. 2017) affirming exclusion of an opinion by a toxicological expert that persons who reside ".25 to .50 miles"

from the defendant's plant "for a period of ten years or more" will suffer harm from chronic exposure to manganese; the opinion was an overstatement .

- ***United States v. Pembroke***, 876 F.3d 812 6th Cir. 2017) affirming admission of expert testimony regarding cell tower location analysis because the government did not attempt to put defendant's cell phone in a very "specific" or "precise" location, but rather attempted to show the general geographical proximity to the locations of the robberies at the pertinent times; the court stated that the disclaimers about the limits of the methodology would have been good fodder for cross-examination of the expert .
- ***United States v. Reynolds***, 626 Fed. App'x 610 6th Cir. 2015) affirming admission of expert testimony concerning cell tower location analysis because the agent did not purport to rely on data to place the defendant *in* the home when child pornography was downloaded, but rather used data to *exclude* the presence of other members of household during relevant times because cell phones of other individuals connected to cell towers were far away from home during downloads.
- ***Krik v. Exxon Mobile Corp.***, 870 F.3d 669, 675 7th Cir. 2017) affirming exclusion of a toxicological expert's testimony that asbestos exposure is "either zero or it's substantial; there's no such thing as not substantial exposure," as unsupported by dose-dependent causation of cancer .
- ***United States v. Lewisbey***, 843 F.3d 653, 659-60 7th Cir. 2016) affirming admission of expert testimony about the general location of the defendant's cell phone based on call records and cell tower data, where the district court appropriately barred the agent "from couching his testimony in terms that would suggest that he could pinpoint the exact location of Lewisbey's phones." .
- ***United States v. Hill***, 818 F.3d 289, 295 7th Cir. 2016): The court held that cell site analysis expert testimony should include a "disclaimer" regarding accuracy. The expert should not "overpromise on the technique's precision or fail to account for its flaws." The court affirmed the admission of cell site analysis testimony by an FBI agent where the agent made it clear that the defendant's phone records were "consistent" with him being at or near relevant locations at relevant times, but clarified that he could not state whether a phone was "absolutely at a specific address."
- ***Murray v. Southern Route Maritime, S.A., et al.***, 870 F.3d 915 9th Cir. 2017) affirming the district court's admission of expert testimony about the theory of low-voltage diffuse electrical injury, where the district court highlighted the narrow nature of the expert's opinion about the theory, and did not permit the expert to testify that the plaintiff's injuries were *caused* by low-voltage shock).

3. The "Reasonable Degree of Certainty" Standard in Civil Cases

A rule prohibiting overstatement in forensic evidence cases would likely result in prohibiting an expert from testifying to a “reasonable degree of [field] certainty” of a feature-comparison match. As stated above, the DOJ has abandoned the standard, it has been rejected by scientific panels, and it is a classic example of overstatement. But in civil cases, there is a complication in rejecting the reasonable degree of certainty standard. In federal civil cases, litigants frequently object that the expert testimony offered by their opponents is unreliable and insufficient due to the experts’ *failure* to opine “to a reasonable degree of certainty.” Moreover, some states appear to require a reasonable certainty standard as a matter of state substantive law -- which is controlling in diversity cases, assuming that in fact it is substantive. *See, e.g., Antrim Pharmaceutical LLC v. Bio-Pharm., Inc.*, 310 F. Supp.3d 934 (N.D. Ill. 2018) (explaining that Illinois law permits plaintiffs to recover lost profits only if they can establish them “to a reasonable degree of certainty”; finding expert testimony sufficient to establish lost profits to the requisite degree of certainty); *Miranda v. Count of Lake*, 900 F.3d 335 (7th Cir. 2018) (“In Illinois, proximate cause must be established by expert testimony to a reasonable degree of medical certainty.”); *Day v. United States*, 865 F.3d 1082 (8th Cir. 2017) (Under Arkansas law, a medical expert must testify that “the damages would not have occurred” without the defendant’s negligence; expert’s opinion “must be stated within a reasonable degree of medical certainty or probability.”).

It is arguable whether a state’s requirement of a “reasonable degree of certainty” standard is in fact a matter of substantive law, if what it means is that an expert’s testimony to a lesser standard is inadmissible. A state that requires experts to testify to a reasonable degree of medical certainty is enforcing that “law” through a rule of evidentiary exclusion --- you can’t testify unless you say those magic words. Under Federal Rule 402, state rules of evidence cannot be used to exclude relevant evidence in a Federal Court --- the only possible sources of exclusion are the federal constitution, federal statutes, and national rules of procedure.

But assuming that a state rule imposing the reasonable degree of certainty standard is a substantive requirement, even if a misguided one, then nothing in an evidence rule can change it. So it may be that a Committee Note supporting any change should flag the issue of the possibility of substantive law requiring such a statement from an expert --- the draft Committee Note at the end of this memo does exactly that.⁶

Beyond the substantive limitations that might be imposed by state law, some federal courts go further and find that an expert’s opinion fails *Daubert* due to its lack of certainty, while others uphold the admissibility of expert opinions *because* they are stated with the requisite degree of certainty. Other courts hold that the “magic words” of reasonable degree of certainty are not required by *Daubert* and Rule 702. A sampling of recent cases is immediately below.

Here are some recent cases on “reasonable degree of certainty” and Daubert:

⁶ The DOJ standards prohibiting testimony to a reasonable degree of certainty, set forth above, contain an exception for cases in which the law requires such testimony.

- ***Johnson v. Memphis Light, Gas & Water Div.***, 695 Fed. App'x 131 6th Cir. 2017): The trial court excluded the expert opinion of a medical examiner that the decedent's cause of death was "probable heat stroke," after the defendant objected that the opinion was not stated to the requisite "reasonable degree of medical certainty." The Sixth Circuit found that exclusion was error, in light of the medical examiner's testimony that "probable" did not mean "possible or maybe" but instead meant "reasonable to think" and "more likely than not." In finding the medical examiner's testimony admissible under *Daubert*, the appellate court noted that, although lawyers and judges routinely use the phrase "reasonable degree of certainty", there is no "consensus" as to its precise meaning. The court noted that "reasonable degree of certainty" is a term of art in the law that has no analog for practicing physicians carrying out their professional duties. The court concluded that there is "no magic words test" for an expert's testimony in the Sixth Circuit and that experts need not attach such language to an opinion to make it admissible, nor can the phrase save an otherwise unreliable opinion from exclusion.
- ***Wendell v. Glaxo Smith Kline, LLC***, 858 F.3d 1227 9th Cir. 2017) exclusion of medical experts' opinions was error where both experts testified that their opinions were "based on a reasonable degree of medical certainty" even though they "would not satisfy the standards required for publication in peer-reviewed medical journals." .
- ***Murray v. Southern Route Maritime, S.A., et al.***, 870 F.3d 915 9th Cir. 2017) rejecting the defendant's argument that medical experts should have been excluded because they failed to provide "more probable than not" testimony, reasoning that the experts confirmed their opinions "to a reasonable degree of certainty on a more-probable-than-not basis" .
- ***West v. Bayer Healthcare Pharm., Inc.***, 293 F. Supp.3d 82 D.D.C. 2018) rejecting the defendant's motion to exclude the plaintiff's causation experts, as to a claim based upon bacterial contamination of a pharmaceutical product, due to the experts' alleged inability to "conclusively rule out" every other possible cause of plaintiff's injuries; the experts' opinions that the plaintiff's symptoms were "more likely than not" caused by contamination were adequate; in support of its holding, the court quoted a case finding that testimony that defendant's negligence "more likely than not" caused plaintiff's harm "based on a reasonable degree of medical certainty" was adequate .
- ***Guzman-Fonalledas v. Hospital Espanol Auxilio Mutuo***, 308 F. Supp.3d 604 D.P.R. 2018) approving admission of expert testimony to a "reasonable degree of medical and surgical pathology certainty" that the plaintiff's mistaken diagnosis constituted a significant deviation from the usual standards of medical care .
- ***Hewitt v. Metro-North Commuter Railroad***, 244 F. Supp.3d 379 S.D.N.Y. 2017) in the plaintiff's suit against a railroad alleging shoulder injury suffered as a result of the requirements of his job as a coach cleaner, the court approved testimony by an ergonomics expert about the ergonomic risks in the plaintiff's job and measures that could have been taken to avoid those risk, "to a reasonable degree of ergonomic certainty" .

- ***Jordan v. Iverson Mall Ltd. Ptsp.***, 2018 WL 2391999 D.Md.): The defendants argued that the plaintiffs’ medical expert should not have been allowed to testify because she never stated that her opinion was to a “reasonable degree of medical certainty.” The court reviewed Fourth Circuit case law, which requires the expert to have a reasonable degree of medical certainty for an opinion on causation to be admissible. But the court concluded that the Fourth Circuit case law does not require the expert to say the magic words “reasonable degree of certainty.” In this case, the court found that the expert was testifying to a reasonable degree of certainty even though she never used that term.
- ***Ernst v. City of Chicago***, 39 F. Supp.3d 1005 N.D. Ill. 2014) expert’s use of uncertain qualifiers, such as “might”, “possible”, “potentially”, “appear to be”, and “likely” were not a reason to exclude opinion as speculative .
- ***Bullock v. Volkswagen Group of Amer., Inc.***, 160 F. Supp.3d 1365 M.D. Ga. 2016) rejecting defendants’ challenge to the admission of the plaintiff’s expert in automobile mechanics, based on the expert’s failure to express his opinions about acceleration to a “reasonable degree of scientific certainty or probability”; the court found that the expert’s trial testimony established that he held his opinions “to the requisite degree of certainty required under the law” even though he failed to use the “magic words” .
- ***Rangel v. Anderson***, 202 F. Supp.3d 1361 S.D. Ga. 2016) doctor’s testimony using terms like “possible” and “likely” interchangeably in describing cause of plaintiff’s injuries highlighted his lack of certainty; testimony failed to establish a reasonable degree of medical certainty and thus failed to satisfy *Daubert* .

Reporter’s Comment: A movement toward abrogating the “reasonable degree of certainty” standard in civil cases could be a salutary development. The National Commission on Forensic Sciences pointed out that such a standard is “not required by *Daubert*.” The question under *Daubert* is whether an opinion is reliable and helpful, and surely an opinion can so qualify without the meaningless and confusing buzzwords of “reasonable degree of medical certainty.”

Moreover, the courts that require an expert to testify to a reasonable degree of certainty appear to be confusing admissibility of the opinion and the weight of the evidence. Assuming reliable methodology, if an expert testifies that something is possible, why would that not be admissible under Rule 702? It would certainly seem relevant and helpful. Such an opinion would be unlikely to constitute *sufficient* evidence of causation, but that is not the question to be answered on a *Daubert* motion.

All in all, an amendment to address expert overstatement on the civil side might be valuable in drawing the courts away from the reasonable degree of certainty standard.

II. A Short Discussion of the Admissibility/Weight Problem

As stated above, the Committee has been considering the possibility of an amendment to Rule 702 that would emphasize that the questions of sufficiency of basis subdivision b) and reliability of application subdivision d) are questions of admissibility and not weight. The Chair appointed a Rule 702 Subcommittee to study this matter and report to the Committee. That report was submitted to the Committee at the last meeting.

The Committee's inquiry was in response to a law review article highlighting a number of cases that appear not to have read the Rule as it is intended. The Rule provides that the requirements of sufficient basis and reliable application must be treated as questions of admissibility, and so must be established by a preponderance of the evidence under Rule 104(a). But the cases cited in the law review article appeared to be treating these admissibility requirements as questions of weight.

The last memo to the Committee on this subject took a deep dive into the cases that have been cited as the leading examples of courts ignoring the Rule 104(a) standard for questions of sufficiency of basis and reliability of application. The takeaway points from the case law survey were as follows:

- A court's declaration that sufficiency of basis and reliability of application are "questions of weight" is not necessarily a misapplication of Rule 702/104(a). That is because even under 104(a) there are disputes that will go to weight and not admissibility. When the proponent has met the preponderance standard and the opponent responds with some deficiency that does not drive the proponent's showing of a preponderance, then that deficiency is a question of weight and not admissibility --- under the preponderance standard.
- Because there remain questions of weight under Rule 104(a), one must be cautious in jumping to the conclusion that a court is ignoring Rule 702/104(a) when it states something like "the defendant's challenges to the expert's opinion present questions of weight and not admissibility." That is a different statement than a *broader* one such as "challenges to the sufficiency of an expert's basis raise questions of weight and not admissibility" a misstatement made by circuit courts in a disturbing number of cases. But even where that broader statement is made, the focus must be on what the challenges are and what the court has found in terms of the expert's basis, methodology and application. That is to say, a court that makes the broader statement might actually have found that basis and application were more likely than not satisfied in the specific case. The fact that the court makes an overbroad, generalized statement is not ideal, but it's only dictum if the court actually ended up finding the standards met by a preponderance.

- There is no doubt that in some circuits the courts routinely state the misguided notion that arguments about sufficiency of basis and reliability of application almost always go to weight and not admissibility. But in many of the reviewed cases, the expert arguably satisfied the Rule 104(a) standard anyway, so the court's cavalier treatment of Rule 702(b) and (d) appears to make no difference to the result. In other cases, it cannot be determined whether the court used the 104(a) or the 104(b) standard in assessing sufficiency of basis and application. Evaluation of the cases is muddled by two complications: 1) courts rarely specifically articulate the standard of proof that they are employing; and, more importantly, 2) there will be a line to draw for admissibility and weight no matter what standard of proof is employed.⁷

Discussion at the last Committee Meeting:

At the last meeting a number of Committee members observed that it would be useful to educate the courts that it is incorrect to make broad statements that sufficiency of basis and reliable application are questions of weight and not admissibility. Members also stated that it would be useful if courts articulated the standard of proof that they were actually applying. But Committee members did not conclude that the proper remedy was to amend the text of the Rule to emphasize that the Rule 104(a) standard applies to all admissibility requirements of Rule 702. The confounding problem of amending the text is that the Rule 104(a) standard *already* applies to these admissibility requirements --- as the court itself makes clear in *Daubert* and *Bourjaily*. Adding the preponderance standard to the text of the rule may raise questions about its applicability to all the other rules --- the Rule 104(a) standard applies to almost all the admissibility requirements in the Federal Rules, but it is not specifically stated in the text of any of them.

The Committee's reaction at the last meeting to a proposed amendment to the text of Rule 702 that would add a Rule 104(a) standard was, it was fair to say, not wildly enthusiastic. But no vote was taken to drop the proposal. Therefore, one of the drafting alternatives below sets forth such an amendment.

The Committee seemed more receptive to an alternative: if a proposal to amend Rule 702 to prevent overstatement were approved by the Committee, the Committee Note to that amendment could provide instruction on the Rule 104(a) question --- including encouraging courts to specify that they are applying that standard. Accordingly, one of the drafting alternatives below adds Rule

⁷ A rough count of the cases highlighted in the law review article as being problematic (along with a number of recent cases decided after its publication) found the following: 1. Five circuit court opinions in which the court appeared to apply a Rule 104(b) standard to the questions of sufficiency of basis and reliable application; 2. Six circuit opinions in which the court used inappropriate Rule 104(b) language, but actually appeared to apply the Rule 104(a) standard to those questions; 3. Three district court opinions that wrongly applied the Rule 104(b) standard; 4. Four district court opinions that used Rule 104(b) language but actually appeared to review under Rule 104(a); and 5. Three district court opinions in which Rule 104(b) language was used and there is not enough to determine from the opinion which standard was actually applied.

104(a) -related instructions to the Committee Note that would accompany an amendment regarding overstatement.

III. Drafting Alternatives

This section presents two drafting alternatives. Alternative 1 adds an admissibility requirement to address overstatement of conclusions, and includes comment on the Rule 104(a) /104 b) question. Alternative 2 combines the first alternative with the addition of the Rule 104(a) standard to the text.

Note: The “overstatement” language has been tweaked in response to comments and suggestions made at the last meeting. The changes were worked on by Judge Schroeder Chair of the Rule 702 Subcommittee , Dan Collins, and the Reporter.

A. Alternative 1 --- Overstatement Regulation.

Rule 702. Testimony by Expert Witnesses

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:

- 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; and
- e) the testimony is limited to the opinions that may reasonably be drawn from the reliable application of the principles and methods.]

Or: “ e) the expert does not overstate the opinions that result from the expert’s reliable application of the principles and methods.”

Draft Committee Note

Rule 702 has been amended to provide that an expert may testify only to opinions that can reasonably be drawn from the principles and methods used by the expert. Experience shows that even when experts use reliable methodology and apply it reliably, some experts state the opinion in terms that overstate or exaggerate the results that the expert could reliably reach. For example, an expert may testify that something is a fact even though it is only the expert's opinion. Or an expert may express a degree of certainty that the methodology does not support. Even when experts reliably apply reliable principles and methods to arrive at opinions, testimony that inaccurately states their conclusions undermines the purposes of the Rule. 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 claims of an expert concerning the strength of the evidence produced by a method.

The amendment applies to all experts but it has special relevance to testimony of forensic experts. Forensic experts often explicitly or implicitly express opinions about probabilities – for example, when comparing features to assess the possible origin of an evidence sample. It is important that the expert accurately inform the factfinder of the meanings of the results that are reached. A forensic expert who states or implies that a method or conclusion is “infallible,” “certain,” or “error-free” will by definition be stating an opinion that cannot reasonably be drawn, because such statements cannot be empirically supported. Also, many forensic processes do not comport with the scientific method, so testimony that such a process is “scientific” is not supported --- and is prohibited under this amendment. The amendment requires the expert to accurately inform the factfinder of the meaning of the results found by the expert. 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 limits inherent in the methodology. Claims of identification or probabilities based only on the expert's experience, without empirically valid support, would not be admissible because they are not reasonably drawn from the method used.

Claims that an expert expresses an opinion to a “reasonable degree of [scientific/medical/forensic] certainty” should be prohibited under the amendment. That phrase has no scientific meaning and is misleading. See National Commission on Forensic Science, *Testimony Using the Term “Reasonable Scientific Certainty”*, <https://www.justice.gov/ncfs/file/795146/download> “Rather than use ‘reasonable...certainty’ terminology, experts should make a statement about the examination itself, including an expression of the uncertainty in the measurement or in the data. The expert should state the bases for that opinion e.g., the underlying information, studies, observations and the limitations relating to the results of the examination.” . Examples of properly verified conclusions, when supported by the data and methodology, include statements such as “cannot be ruled out” or “more likely than not.” Of course this amendment does not bar testimony that is required by substantive law.

Nothing in the amendment requires the court to nitpick an expert's opinion so that it is perfect expression of what the basis and methodology can support. The Rule 104(a) standard does

not require perfection. On the other hand, it does not permit the expert to express a conclusion that is clearly unsupported by the expert's basis and methodology.

A requirement of an accurate conclusion derived from the methodology is integrally related to the admissibility requirements of Rule 702(b)- d), all of which are intended to assure that an expert's opinion is helpful. Those admissibility requirements, like the requirement of an accurately stated conclusion, are evaluated by the court under Rule 104(a), under which the proponent must establish that the admissibility standards are met by a preponderance of the evidence. *See Bourjaily v. United States*, 483 U.S. 171 (1987). Unfortunately many courts have held or declared that the critical questions of the sufficiency of an expert's basis, and the application of the expert's methodology, are generally questions of weight and not admissibility. These rulings are an incorrect application of Rules 702 and 104(a).

Of course some challenges to expert testimony will raise matters of weight rather than admissibility even under the Rule 104(a) standard. For example, if the court finds by a preponderance of the evidence that an expert has relied on sufficient studies to support an opinion, the fact that the expert has not read every single study that exists will likely raise a question of weight and not admissibility. But this does not mean, as certain courts have held, that arguments about the sufficiency of an expert's basis *generally* go to weight and not admissibility. Rather it means that once the court has found the admissibility requirement to be met by a preponderance of the evidence, any remaining attack by the opponent will go only to the weight of the evidence. In order to avoid confusion on this subject, it is useful for the trial court to specify that it is applying the Rule 104(a) preponderance standard to all the admissibility requirements of Rule 702.

B. Alternative B --- Combining Overstatement Regulation With Articulation of the Preponderance Standard of Proof.

Rule 702. Testimony by Expert Witnesses.

For a witness to testify as an expert in the form of an opinion or otherwise, the court must find the following requirements to be established by a preponderance of the evidence: 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:

a) ~~the expert's~~ witness'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~~ the testimony is the product of reliable principles and methods; and—

d) ~~the expert-~~ witness has reliably applied the principles and methods to the facts of the case- ;

e) the witness is qualified as an expert by knowledge, skill, experience, training, or education; and

f) the testimony is limited to the opinions that may reasonably be drawn from the reliable application of the principles and methods.

Or: “ f) the expert does not overstate the opinions that result from the expert’s reliable application of the principles and methods.”

Draft Committee Note

Rule 702 has been amended in two respects. First, the rule now clarifies and emphasizes that the admissibility requirements set forth in the Rule must be established by a preponderance of the evidence --- which may include evidence other than privileged information that would not be

admissible at trial. *See* Rule 104(a). Of course the Rule 104(a) standard applies to most of the admissibility requirements set forth in the Evidence Rules. *See Bourjaily v. United States*, 483 U.S. 171 (1987). But unfortunately many courts have held that the critical questions of the sufficiency of an expert's basis, and the application of the expert's methodology, are generally questions of weight and not admissibility. These rulings are an incorrect application of Rules 702 and 104(a), and are rejected by this amendment. There is no intent to raise any negative inference as to the applicability of the Rule 104(a) standard of proof for other rules. The Committee concluded that emphasizing the preponderance standard in Rule 702, specifically, was made necessary by the courts that have ignored it when applying that Rule.

Of course some challenges to expert testimony will raise matters of weight rather than admissibility even under the Rule 104(a) standard. For example, if the court finds by a preponderance of the evidence that an expert has relied on sufficient studies to support an opinion, the fact that the expert has not read every single study that exists will likely raise a question of weight and not admissibility. But this does not mean, as certain courts have held, that arguments about the sufficiency of an expert's basis *generally* go to weight and not admissibility. Rather it means that once the court has found the admissibility requirement to be met by a preponderance of the evidence, any remaining attack by the opponent will go only to the weight of the evidence. In order to avoid confusion on this subject, it is useful for the trial court to specify that it is applying the Rule 104(a) preponderance standard to all the admissibility requirements of Rule 702.

Second, Rule 702 has been amended to provide that an expert may testify only to opinions that can reasonably be drawn from the principles and methods used by the expert. Experience shows that even when experts use reliable methodology and apply it reliably, some experts state the opinion in terms that overstate or exaggerate the results that the expert could reliably reach. For example, an expert may testify that something is a fact even though it is only the expert's opinion. Or an expert may express a degree of certainty that the methodology does not support. Even when experts reliably apply reliable principles and methods to arrive at opinions, testimony that inaccurately states their conclusions undermines the purposes of the Rule. 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 claims of an expert concerning the strength of the evidence produced by a method.

The amendment applies to all experts but it has special relevance to testimony of forensic experts. Forensic experts often explicitly or implicitly express opinions about probabilities – for example, when comparing features to assess the possible origin of an evidence sample. It is important that the expert accurately inform the factfinder of the meanings of the results that are reached. A forensic expert who states or implies that a method or conclusion is “infallible,” “certain,” or “error-free” will by definition be stating an opinion that cannot reasonably be drawn, because such statements cannot be empirically supported. Also, many forensic processes do not comport with the scientific method, so testimony that such a process is “scientific” is not supported --- and is prohibited under this amendment. The amendment requires the expert to accurately inform the factfinder of the meaning of the results found by the expert. 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 limits inherent in the methodology. Claims of identification or probabilities based on the only on the expert's experience, without empirically valid support, would not be admissible because they are not reasonably drawn from the method used.

Claims that an expert expresses an opinion to a “reasonable degree of [scientific/medical/forensic] certainty” should be prohibited under the amendment. That phrase has no scientific meaning and is misleading. See National Commission on Forensic Science, *Testimony Using the Term “Reasonable Scientific Certainty”*, <https://www.justice.gov/ncfs/file/795146/download> “Rather than use ‘reasonable...certainty’ terminology, experts should make a statement about the examination itself, including an expression of the uncertainty in the measurement or in the data. The expert should state the bases for that opinion e.g., the underlying information, studies, observations and the limitations relating to the results of the examination.” . Examples of properly verified conclusions, when supported by the data and methodology, include statements such as “cannot be ruled out” or “more likely than not.” Of course this amendment does not bar testimony that is required by substantive law.

Nothing in the amendment requires the court to nitpick an expert's opinion so that it is perfect expression of what the basis and methodology can support. The Rule 104(a standard does not require perfection. On the other hand, it does not permit the expert to express a conclusion that is clearly unsupported by the expert's basis and methodology.