Science Advisory Board's Statement regarding the PCAST Report

Introduction

On September 20, 2016, the US President's Council of Advisors on Science and Technology (PCAST) published a report on *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* in response to the President's question as to whether there are additional steps that could help ensure the validity of forensic evidence in the Nation's legal system. However, according to published reviews of this report [1-8], the PCAST report presents a flawed paradigm for forensic evaluation, misapplies statistics and the notion of probability, ignores existing data and literature in forensic science, and, as a result, this report is scientifically unsound.

Of interest to the Department of Forensic Sciences (DFS) are the sections on DNA, Latent Fingerprint Analysis, and Firearm Analysis. This is the third and final statement regarding the PCAST Report.

Firearm Analysis

Cognitive Bias

The report addresses concerns for the influence of bias on an examiner's decision. They break their concerns down into contextual bias and confirmation bias, as described by researchers i.e. Itiel Dror: [9]. While it focused on latent print analysis, there are clear parallel issues that are relevant to firearm analysis, and in pattern/impression evidence examination in general.

> *(1) Confirmation bias. Work by FBI scientists has shown that examiners often alter the features that they initially mark in a latent print based on comparison with an apparently matching exemplar. Such circular reasoning introduces a serious risk of confirmation bias. Examiners should be required to complete and document their analysis of a latent fingerprint before looking at any known fingerprint and should separately document any additional data used during their comparison and evaluation.*
>
> *(2) Contextual bias. Work by academic scholars has shown that examiners' judgments can be influenced by irrelevant information about the facts of a case. Efforts should be made to ensure that examiners are not exposed to potentially biasing information.*

We, the Science Advisory Board, state that at the time of this writing, the PCAST statement on bias has some validity and DFS has addressed these concerns in the best way possible. The DFS is revising the Standard Operating Procedures for firearm evidence examinations, which applies to both questioned evidence items and known firearms sources, to require examiners to analyze and compare the questioned evidence items prior to comparison to test fired ammunition components. This procedure limits the risk of confirmation bias. This analytical scheme has been employed in many crime laboratories for decades. The approach and its value were detailed in publications by Davis in the 1970's [10 & 11]. Regarding contextual bias, DFS has limited contact between the examiner and detective. This limits the examiners' exposure to task irrelevant case information that could be potentially biasing. However, recent research has demonstrated in firearm examinations that the actual extent of negative effects from contextual information is relatively small [12]. Examiners will not report or testify to the absolute identification of a firearm source to the exclusion of all firearms, 100% certainty, zero error rate or "a reasonable degree of scientific certainty". Therefore, jurors will receive testimony regarding firearm comparisons that is correct and consistent with accepted standards. The SAB is monitoring the activity of the Organization of Scientific Area Committees for Forensic Science (OSAC) including their

development of standards and published research and development gaps. The SAB will continue to advise DFS to make proper adjustments to advance and improve the firearms examination discipline when relevant standards are published by accredited forensic standard development organizations (ASTM, ASB, ISO) and approved for listing on the OSAC Registry of Approved Standards.

Proficiency Testing

The report addresses the need for proficiency testing. PCAST not only states proficiency testing should be required but at some point, blind proficiency testing should be employed. (p.10,57-59)

> *(3) Proficiency testing. Proficiency testing is essential for assessing an examiner's capability and performance in making accurate judgments. As discussed elsewhere in this report, proficiency testing needs to be improved by making it more rigorous, by incorporating it systematically within the flow of casework, and by disclosing tests for evaluation by the scientific community.*
>
> *Scientific validity as applied, then, requires that an expert: (1) has undergone relevant proficiency testing to test his or her accuracy and reports the results of the proficiency testing*
>
> *PCAST believes that test-blind proficiency testing of forensic examiners should be vigorously pursued, with the expectation that it should be in wide use, at least in large laboratories, within the next five years. However, PCAST believes that it is not yet realistic to require test-blind proficiency testing because the procedures for test-blind proficiency tests have not yet been designed and evaluated.*
>
> *While only non-test-blind proficiency tests are used to support validity as applied, it is scientifically important to report this limitation, including to juries—because, as noted above, non-blind proficiency tests are likely to overestimate the accuracy because the examiners knew they were being tested.*

The Science Advisory Board agrees that the PCAST statement on the need for proficiency testing is somewhat correct, but regarding the accuracy of examiners being overestimated in non-blind settings, this is not accurate. The members of the PCAST did not reference any published research or perform any studies to demonstrate that examiners are, or tend to be, more accurate in blind proficiency tests versus declared proficiency tests, but merely cited a psychological study on "Hawthorne Effect" and a pilot study. Recently a study by the Netherlands Forensic Science Institute in a double-blind test of firearm examiners resulted in a very high level of conclusion accuracy, and no instances of false identifications [13]. However, current available research is still insufficient to indicate that accuracy of examiners' results varies greatly from blind to non-blind test settings. DFS instead follows accreditation requirements, and via a quality assurance program, administers either internal or external proficiency tests to examiners on an annual basis. This complies with PCAST recommendations.

Under an accreditation setting, external proficiency tests should be supplied from not only an approved vendor, but one that has met the established criteria of ISO/IEC 17043:2010. In fact, DFS believes in testing the proficiency of the examiner in whichever discipline or sub discipline they are performing casework, firearms examination, or otherwise. Proficiency tests should be of a level comparable to standard laboratory casework, both in terms of quality and quantity of the samples. PCAST recommends that blind proficiency testing is the best way to move forward but does not offer ways to go about this or resources in which to achieve this goal. In fact, PCAST

points out *that it is not yet realistic to require test-blind proficiency testing because the procedures for test-blind proficiency tests have not yet been designed and evaluated.* SAB believes that without the addition of both monetary funding and employees to DFS, it would be impossible to set up a true blind proficiency testing system without creating a detrimental backlog of cases. This detrimental effect would be fully evident in firearm evidence examinations.

Foundational Validity

The report concludes that the only way to establish foundational validity, is through appropriately designed black box studies (p. 46):

> *Scientific validity and reliability require that a method has been subjected to empirical testing, under conditions appropriate to its intended use, that provides valid estimates of how often the method reaches an incorrect conclusion. For subjective feature-comparison methods, appropriately designed black-box studies are required, in which many examiners render decisions about many independent tests (typically, involving "questioned" samples and one or more "known" samples) and the error rates are determined. 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.*

In the spirit of the recommendations made within the PCAST report, the SAB does agree with the need for continued research in this area. However, PCAST subsequently fails to address how many studies of this type would be required to adequately establish the validity of firearm evidence comparisons. Additionally, PCAST does not describe what accepted scientific criteria would help determine such a number. Without the clear explanations of how, why, and the number required, the PCAST has not done its due diligence into the extensive firearm/toolmark validation studies performed and their recommendations are therefore unreliable. Under the current budgetary and administrative confines, continuation of this research is restrictive, however, with additional support from advisory bodies such as PCAST, appropriate research criteria could be identified and explored to add some additional scientific perspective regarding the examination method used. Federal funding is required for research in forensic science, not at just the federal level, but also at the state and local levels. Inherent variability with scientific research is a common occurrence because of variables present and research criteria. However, by allowing reproducibility of the research methods, a clear and relevant result may be obtained which would allow for a more accurate and reliable determination of false positive and negative rates amongst examiners.

As the OSAC Firearms and Toolmarks Subcommittee noted in its December 16, 2016 response to PCAST, there have been many research studies performed on the foundational validity of firearms and toolmark analysis [8]. They describe how PCAST reviewed several of these studies and discounted their results because they did not fit the PCAST definition of structured black-box studies. The OSAC Firearms and Toolmarks subcommittee continues by stating, "*we disagree with the premise that a structured black-box study is the only useful way to gain insight into both the foundations of firearm and toolmark identification and examiner error rates. Taken collectively, the published studies support the underlying principles of firearm and toolmark examination and the fact that examiner error rates are quite low* [8]." The discipline of firearms and toolmark analysis is built upon many decades of research that support the foundations of its practice [14].

Research does continue in this scientific area and the results of these studies will continue to add to the foundation and improve the performance over time.

Subjective to Objective Method

The report in its "Path Forward" section proposes the improvement of analysis from a subjective to a more objective based method (page 113-114):

> *A second—and more important—direction is (as with latent print analysis) to convert firearms analysis from a subjective method to an objective method.*
>
> *This would involve developing and testing image-analysis algorithms for comparing the similarity of tool marks on bullets. There have already been encouraging steps toward this goal. Recent efforts to characterize 3D images of bullets have used statistical and machine learning methods to construct a quantitative "signature" for each bullet that can be used for comparisons across samples. A recent review discusses the potential for surface topographic methods in ballistics and suggests approaches to use these methods in firearms examination. The authors note that the development of optical methods have improved the speed and accuracy of capturing surface topography, leading to improved quantification of the degree of similarity.*

The SAB supports the continued evaluation and testing of these methods in preparation for use in casework examination. There has been increased and accelerated global research in the measurement of firearm and toolmark similarity based primarily on 3D surface measurement topography, mathematically based measurement of surface similarity, and research databases used to statistically assess the chance agreement from a different firearm source. The SAB advises the Laboratory to maintain its current investigation in these emerging methods to ultimately use such validated methods to augment the current evidence examination protocols, and expert witness testimony.

References:

[1] I.W. Evett, C.E.H. Berger, J. Buckleton, C. Champod, G. Jackson, Finding the way forward for forensic science in the US - A commentary on the PCAST report, Forensic Science International 278 (2017) 16-23.

[2] FBI, Comments on: President's Council of Advisors on Science and Technology REPORT TO THE PRESIDENT Forensic Science in Federal Criminal Courts: Ensuring Scientific Validity of Pattern Comparison Methods, 2016.

[3] H. Ruslander, IAI PCAST Response, 2016.

[4] M. Ramos, NDAA PCAST Response, 2016.

[5] ASCLD-LAB, ASCLD PCAST Response, 2016.

[6] Association of Firearm and Tool Mark Examiners, "Response to PCAST Report on Forensic Science", October 31, 2016

[7] Bureau of Alcohol, Tobacco, Firearms and Explosives, "ATF Response to the President's Council of Advisors on Science and Technology Report," September 21, 2016

[8] OSAC, "OSAC Firearms and Toolmarks Subcommittee's Response to the PCAST Call for Additional References", December 14, 2016

[9] I.E. Dror, D. Charlton, A.E. Peron, "Contextual information renders experts vulnerable to making erroneous identifications," *Forensic Science International*, Vol. 156 (2006): 74-78.

[10] J.E. Davis, "Firearms Investigations: Miscellaneous Notes", *AFTE Journal*, Vol. 4, No. 3, August 1972

[11] J.E. Davis, "Firearms Training Notes", *AFTE Journal*, Vol. 9, No. 2, July 1977

[12] Jose Kerstholt, Aletta Eikelboom, Tjisse Dijkman, Reinoud Stoel, Rob Hermsen, Bert van Leuven, "Does suggestive information cause a confirmation bias in bullet comparisons?" *Forensic Science International,* 198 (2010) 138–142

[13] W. Kerkhoff, R.D. Stoel, E.J.A.T. Mattijssen, C.E.H. Berger, F.W. Didden, J.H. Kerstholt, "A part-declared blind testing program in firearms examination," *Science & Justice* 58 (2018) 258–263

[14] SWGGUN & AFTE "Joint submission by the Association of Firearm and Tool Mark Examiners and Scientific Working Group for Firearms and Toolmarks" to NSTC SOFS RDTE IWG- June 14, 2011. https://www.nist.gov/sites/default/files/documents/forensics/Annotated-Bibliography-Firearms-Toolmarks.pdf