Department of Forensic Sciences Science Advisory Board's Statement with regard to 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.

As appropriate to the disciplines offered by the Department of Forensic Sciences, the Advisory Board will address the disciplines of Forensic Biology (DNA), Latent Fingerprint Analysis, and Firearms Analysis. The Board has decided to address these disciplines separately, beginning with Forensic Biology. The other disciplines will be addressed in the next few meetings.

DNA

According to published reviews of this report (e.g., [1-4]), 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, state that the PCAST report is scientifically unsound.

The PCAST report concludes that the DNA analysis of single-source specimen and simple mixtures of two contributors is a foundationally valid and reliable method, yet raises several concerns about the interpretation of complex DNA mixtures (pp. 75-83). Regarding the latter, the report concludes (page 82):[1]

> *Objective analysis of complex DNA mixtures with probabilistic genotyping software is relatively new and promising approach. Empirical evidence is required to establish the foundational validity of each such method within specified ranges. At present, published evidence supports the foundational validity of analysis, with some programs, of DNA mixtures of 3 individuals in which the minor contributor constitutes at least 20 percent of the intact DNA in the mixture and in which the DNA amount exceeds the minimum required level for the method. The range in which foundational validity has been established is likely to grow as adequate evidence for more complex mixtures is obtained and published.*

We, the Science Advisory Board, state that at the time of this writing, the range in which foundational validity has been established for the interpretation of complex mixtures at DFS using

[1] Note that an addendum to the report that appeared in January 2017 reached a slightly different conclusion (page 8):

> *PCAST found that empirical testing of PG* [probabilistic genotyping] *had largely been limited to a narrow range of parameters (number and ratios of contributors). We judged that the available literature supported the validity and reliability of PG for samples with three contributors where the person of interest comprises at least 20% of the sample. Beyond this approximate range (i.e. with a larger number of contributors or where the person of interest makes a lower than 20% contribution to the sample), however, there has been little empirical validation.*

probabilistic genotyping[2] extends from DNA mixtures of 2 individuals up to DNA mixtures of 5 individuals. The PCAST notion of a lower limit percentage of the minor contributor as a criterion for deciding whether a DNA profile is interpretable or uninterpretable is scientifically unsound. The scientific criterion for making this decision is the quantity of information in the electropherogram(s) for a particular contributor.[3] DFS has a valid pre-evaluation phase in place for making this decision.

More specifically, an internal validation study conducted by the DNA analysts at DFS[4] consisting of over 10,000 comparisons to 100 DNA mixtures ranging from 2 contributors to 5 contributors has addressed each of the PCAST concerns listed below (PCAST, pp. 79-80).

> *These probabilistic genotyping software programs clearly represent a major improvement over purely subjective interpretation. However, they still require careful scrutiny to determine (1) whether the methods are scientifically valid, including defining the limitations on their reliability (that is, the circumstances in which they may yield unreliable results) and (2) whether the software correctly implements the methods. This is particularly important because the programs employ different mathematical algorithms and can yield different results for the same mixture profile.* (PCAST, page 79)

The internal validation study conducted at DFS demonstrates that the interpretation of complex mixtures using STRmix™ version 2.4 in conjunction with GlobalFiler™ PCR Amplification Kit and 3500/3500xL Genetic Analyzer is scientifically valid for mixtures of 2 to 5 individuals.

To test the correctness of the software's implementation of the method, the DFS internal validation study reproduced the likelihood ratio values for each locus of a single-source profile in quadruple, once for each of four allele frequency databases. These results confirm that the software correctly implements the method.

> *Appropriate evaluation of the proposed methods should consist of studies by multiple groups, not associated with the software developers, that investigate the performance and define the limitations of programs by testing them on a wide range of mixtures with different properties. In particular, it is important to address the following issues:*
>
> *(1) How well does the method perform as a function of the number of contributors to the mixture? How well does it perform when the number of contributors to the mixture is unknown?* (PCAST, page 79)

---

[2] Note that probabilistic genotyping does not identify contributors with 100% certainty. Instead it applies mathematical models and probability theory to assign probabilities to the observed peak heights given different sets of potential contributors. The conclusion is therefore probabilistic, taking the form of a likelihood ratio.

[3] The quantity of information in the electropherogram(s) for a particular contributor depends on the quantity of data and the information known about the mixture.

[4] The DFS internal validation study strictly follows the FBI approved SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems available at https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf (accessed January 2, 2018). It was approved by the Technical Leader on 1/7/2016 for the Idenfiler Plus PCR Amplification kit and on 2/24/2017 for the GlobalFiler PCR Amplification kit. A summary of the results is available at https://dfs.dc.gov/page/fbu-validation-studiesperformance-checks (accessed January 5, 2018), and these results have been published in a peer-reviewed journal as part of a larger compilation of results from STRmix™ internal validation studies [5].

The DFS internal validation study tested the performance of the method for 40 mixtures with 2 contributors, and 20 mixtures each for 3, 4 and 5 contributors. These mixtures varied in DNA quantity and mixture proportions to represent the typical profiles[5] encountered by the laboratory. The method correctly and reliably produced the expected results for each of the different number of contributors tested.

In addition, the results of the FBI internal validation study on the performance of STRmix™ version 2.3.06 contains a total of 290 mixtures with 2, 3, 4, and 5 contributors, for each of which the software proved to be appropriately sensitive and specific [6].

In casework, the number of contributors is always unknown (e.g., [7]). The DNA analyst assigns the number of contributors based on the number of peaks and the peak height information in the electropherogram.

To test the effect of an incorrect assignment of the number of contributors, the DFS internal validation study included the following tests:

- **10 mixtures each with 1, 2, 3 and 4 contributors were incorrectly interpreted as having 2, 3, 4 and 5 contributors, respectively; and**
- **3 mixtures each with 2 and 3 contributors, and 4 mixtures each with 4 and 5 contributors were incorrectly interpreted as having 1, 2, 3 and 4 contributors, respectively**

Each mixture was then evaluated against each of the known contributors and against 134 known non-contributors.

Overestimation of the number of contributors correctly produced likelihood ratios greater than 1 for the known contributors. It produced a few likelihood ratios greater than 1 for known non-contributors, but their order of magnitude is much lower than the likelihood ratios produced for the known contributors.[6]

Underestimation of the number of contributors did not have any influence on the likelihood ratios for the known major and minor contributors. It correctly produced lower likelihood ratios for the known trace contributors.

The FBI internal validation study included similar tests on an additional 30 mixtures which produced the same expected trends as the DFS internal validation results [6].

> (2) *How does the method perform as a function of the number of alleles shared among individuals in the mixture? Relatedly, how does it perform when the mixtures include related individuals?* (PCAST, page 79)

The DFS internal validation study performed sensitivity and specificity studies on mixtures with different amounts of alleles shared among the contributors across the loci. These tests correctly and reliably produced the expected results. Given that continuous probabilistic genotyping models take allele sharing into account in their peak height models, this method can handle the entire range of possible allele sharing among the DNA's contributors.

[5] This includes partial profiles.

[6] Note that DFS has defined likelihood ratios between 1 and 100 as being "uninformative" based on the results of their internal validation study.

With regard to related individuals, the FBI internal validation study tested the method on mixtures with 3 contributors that consisted of 2 parents and 1 child. This type of mixture entails a risk of an underestimation of the number of contributors if only the number of peaks is counted and peak height information is disregarded. An underestimation of the number of contributors has no impact on the likelihood ratios of the known major and minor contributors, yet lowers the likelihood ratio for the known trace contributor.

(3) *How well does the method perform—and how does accuracy degrade—as a function of the absolute and relative amounts of DNA from the various contributors? For example, it can be difficult to determine whether a small peak in the mixture profile represents a true allele from a minor contributor or a stutter peak from a nearby allele from a different contributor. (Notably, this issue underlies a current case that has received considerable attention.)* (PCAST, page 79)

The DFS internal validation study included sensitivity and specificity studies on DNA mixtures of varying amounts of DNA. These ranged from an average peak height of about 20 rfu to >25,000 rfu (saturation). The mixture ratios ranged from 25:1 to 1:1 for two person mixtures, with the full range in between for three, four and five person mixtures. As expected for all methods, this method correctly and reliably produced uninformative results for contributors with very low template. For contributors with higher template, this method correctly and reliably produced high likelihood ratios greater than 1 for known contributors, and low likelihood ratios less than 1 for known non-contributors, which clearly separated the results of the known contributors from the results of the known non-contributors. On the high-template end, the method correctly interprets the profile qualitatively for saturated profiles.

Probabilistic genotyping does not determine whether a small peak in the mixture profile represents a true allele from a minor contributor or a stutter peak from a nearby allele from a different contributor. It takes all reasonable possibilities into account, and assigns probabilities to the observations given each of the possibilities. In other words, it assigns weights to the different possibilities, and must therefore not choose between the category of a true allele and the category of a stutter peak.

(4) Under what circumstances—and why—does the method produce results (random inclusion probabilities) that differ substantially from those produced by other methods? (PCAST, page 80)

The method used by DFS uses a fully continuous probabilistic genotyping model to produce likelihood ratios which express the relative support the DNA typing results provide for one proposition with regard to an alternative proposition. A likelihood ratio is a different statistical quantity from a random match probability or a combined probability of inclusion, and will therefore produce different numerical results than either of the latter quantities. In addition, a fully continuous model can produce likelihood ratios that are different from likelihood ratios obtained from a binary model or a semi-continuous model: the reason for these differences is that a fully continuous model takes into account all of the available peak height information above the analytical threshold in the electropherogram, whereas binary and semi-continuous models only take a very limited amount of this information into account (e.g., comparing observed peak heights to a stochastic threshold), if at all. Hence a fully continuous model will produce results different from those produced by binary and semi-continuous models in circumstances where the electropherogram contains peak height information that is taken into account by the fully

continuous model and not taken into account by the binary and semi-continuous models. Taking into account more information makes this method produce higher likelihood ratios in support of the DNA contribution of known contributors and lower likelihood ratios (or exclusions) in support of no DNA contribution of known non-contributors (e.g., [8-11]). This is the expected performance for all likelihood ratio methods.

> *Most importantly, current studies have adequately explored only a limited range of mixture types (with respect to number of contributors, ratio of minor contributors, and total amount of DNA). The two most widely used methods (STRMix and TrueAllele) appear to be reliable within a certain range, based on the available evidence and the inherent difficulty of the problem. Specifically, these methods appear to be reliable for three-person mixtures in which the minor contributor constitutes at least 20 percent of the intact DNA in the mixture and in which the DNA amount exceeds the minimum level required for the method.* (PCAST, page 80)

The DFS internal validation study has shown that STRmix™ version 2.4 is reliable for DNA mixtures with 2, 3, 4 and 5 contributors. Independently, the FBI internal validation study has shown that STRmix™ version 2.3.06 is reliable for DNA mixtures with 2, 3, 4, and 5 contributors [6]. The results of additional internal validation studies of STRmix™ conducted by other laboratories can be found at https://johnbuckleton.wordpress.com/strmix/strmix-validations/ (accessed October 24, 2017).

Again, we note that the PCAST notion of a lower limit percentage of the minor contributor as a criterion for deciding whether a DNA profile is interpretable or uninterpretable is scientifically unsound. The scientific criterion for making this decision is the quantity of information in the electropherogram(s) for a particular contributor (e.g. [12]).

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] G.S. Morrison, D.H. Kaye, D.J. Balding, D. Taylor, A.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.d. Zoete, R.B. Ostrum, A. Caliebe, A comment on the PCAST report: Skip the 'match'/'non-match' stage, Forensic Science International 272 (2017) e7-e9.
[3] B. Budowle, Response to the PCAST report, 2017.
[4] J. Buckleton, U.S. v. Benito Valdez, 2017.
[5] J.-A. Bright, R. Richards, M. Kruijver, H. Kelly, C. McGovern, A. Magee, A. McWhorter, A. Ciecko, B. Peck, C. Baumgartner, C. Buettner, S. McWilliams, C. McKenna, C. Gallacher, B. Mallinder, D. Wright, D. Johnson, D. Catella, E. Lien, C. O'Connor, G. Duncan, J. Bundy, J. Echard, J. Lowe, J. Stewart, K. Corrado, S. Gentile, M. Kaplan, M. Hassler, N. McDonald, P. Hulme, R.H. Oefelein, S. Montpetit, M. Strong, S. Noël, S. Malsom, S. Myers, S. Welti, T. Moretti, T. McMahon, T. Grill, T. Kalafut, M.M.

Greer-Ritzheimer, V. Beamer, D. Taylor, J. Buckleton, Internal validation of STRmix™ - A multi laboratory response to PCAST, Forensic Science International: Genetics, in press (2018).
[6] T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J. Buckleton, J.-A. Bright, D. Taylor, A.J. Onorato, Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics 29 (2017) 126-144.
[7] J.-A. Bright, D. Taylor, C. McGovern, S. Cooper, L. Russell, D. Abarno, J. Buckleton, Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Science International: Genetics 23 (2016) 226-239.
[8] T.W. Bille, S.M. Weitz, M.D. Coble, J. Buckleton, J.-A. Bright, Comparison of the performance of different models for the interpretation of low level mixed DNA profiles, Electrophoresis 35 (2014) 3125-3133.
[9] H. Kelly, J.-A. Bright, J. Buckleton, J.M. Curran, A comparison of statistical models for the analysis of complex forensic DNA profiles, Science & Justice 54 (2014) 66-70.
[10] J.-A. Bright, I.W. Evett, D. Taylor, J.M. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Science International: Genetics 14 (2015) 125-131.
[11] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Science International: Genetics 11 (2014) 144-153.
[12] J.-A. Bright, D. Taylor, S. Gittelson, J. Buckleton, The paradigm shift in DNA profile interpretation, Forensic Science International: Genetics 31 (2017) e24-e32.