Comments on the PCAST REPORT TO THE PRESIDENT Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods – John Buckleton

Thursday, September 01, 2016

The comments in this statement are my own and do not necessarily represent the views of my organisation.

It was disappointing to see the dismissal by the committee of much of the work published by the STRmix developers. They seem to be only aware of some of the published body of work [1-33] We feel that this dismissal is a veiled accusation of bias, which challenges both the our scientific integrity and the scientific journal review process. The committee probably have no way of knowing that the STRmix developers all get no material benefit direct or indirect from STRmix. We are all salaried employees of our respective states and we certainly try to do honest reporting.

In addition to the considerable published work, internal validation studies have been undertaken by every laboratory that is now live with STRmix in casework. These now number 11 in the US, 6 in Australia, one each in Canada, England, Scotland, Ireland and New Zealand, as generally required for accreditation. In many cases this has been done more than once for different multiplexes or different versions of the software. Most of these meet or exceed the SWGDAM guidelines [34]. The committee might argue that they are not all independent since we are associated in a support role with many of these labs but I feel the committee needs to realise that the laboratories adopting STRmix are professionals working within a quality assurance framework and such are able to judge whether we are acting in good faith. In addition to these live labs there are many labs part way through validations. These are not published and it is unlikely that a journal would publish them. The committee could argue that they had no way of knowing but they could have asked or inferred that the live labs must have done an internal validation. I feel that before such sweeping comments are made more study could have been warranted.

PCAST suggest that analyses should be done without knowledge of the number of contributors. We have deliberately used the wrong number in several publications to test the effect [1, 15, 18]. Certainly the title of one of these should have attracted attention: The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. In addition it is a SWGDAM requirement to try alternate propositions, usually one too many and one too few contributors and this has been done by the internal validations.

PCAST simply have the wrong number of mixtures reported in Bright et al [35]. They state 31 and it is 93, 2 to 4 person mixtures with total template from 10-500pg. There are 264 true donor comparisons and 17,000 false donor tests.

This would be a tiny fraction of the test mixtures run in STRmix by ourselves and others.

Other published ground truth known tests include, but are not limited to:

1. Identifiler 94 \( H_p \) true, 3,077,734 \( H_d \) true [15]
2. Identifiler™ in silico 54 \( H_p \) true 54,000 \( H_d \) true [36]
3. Varying 21 \( H_p \) true 123,230,000 \( H_d \) true [2].
4. Identifiler 50 \( H_p \) true [20]
5. Identifiler 240 \( H_p \) true [37]
A sample of unpublished internal validations are listed in appendix 1.

Independent studies have been attempted but were affected by politicking of the parties other than STRmix. In some cases pressure has been applied to these parties.

It is again disappointing to see a committee of this standing relying on the tabloid newspaper article about the miscode that affected Queensland. Again they could have asked for further information to better understand the issue. Perhaps they could question why it was Queensland that was mainly affected when most labs in Australasia were already live. The answer is that it related to an unusual way that Queensland health were using STRmix. Queensland health regularly sent us profiles up to and including the first in the affected series. They have declined to send any more profiles and this has hampered any investigation into the cause.

In summary I suggest that insufficient research was undertaken by the committee. The simplest course was to ask us or Dr Butler. The conclusions of the committee are incorrect and need to be revisited.

References
A Summary of some Internal validations

Forensic Science South Australia (additional to published material)
Profiler Plus
1 person 10 x 1 Hp true
2 person 70 x 2 Hp true
3 person 90 x 3 Hp true

Northern Territory Forensic Science Service
During verification of STRmix we put through 36 mixtures that had either 2p, 3p or 4p contributing.

Erie County validation
Identifiler
2 person 24 Hp true, Hd true 200 per mix
3 person 23 Hp true, Hd true 200 per mix
4 person 16 Hp true, Hd true 200 per mix
Fusion
2 person 17 Hp true, Hd true unknown
3 person 19 Hp true, Hd true unknown
4 person 14 Hp true, Hd true unknown
66 adjudicated ID Plus cases and 28 adjudicated Fusion cases were tested.

San Diego Police Department
Twenty four single source samples as well as two, three, four, and five-person mixtures (a total of 186 mixtures) were created as part of the GlobalFiler validation. The mixtures samples were designed for STRmix that had a range of contributor compositions – from balanced mixtures to mixtures where there are one or two contributors that are the source of most of the DNA in the mixture. There are also mixtures in every set that have at least one contributor dropping out.

In the mixture study, different ranges of template DNA were targeted. The high level samples were prepared such that the average RFU for the highest percentage contributor was between 3K and 10K RFU. The mid-range samples were prepared such that the average RFU for the highest percentage contributor was between 1K and 3K RFU. The low level samples were prepared such that average RFU for the lowest percentage contributor had between 200 and 500 RFU.

The validation samples were used to assess the MCMC process used by STRmix to deconvolute the mixtures. The validation samples were evaluated for the percent contribution
of each contributor which was compared to the results from the STRmix MCMC. The evaluation of the STRmix MCMC was also done by determining whether the correct genotypes included in the genotype probability distribution, whether correct combination was in the top 99%, and whether the STRmix genotype possibilities were intuitive.

**Sensitivity and Specificity**
A subset of the 2-, 3-, and 4-person mixtures created as part of the GlobalFiler validation were used for STRmix sensitivity (the ability to correctly identify a true contributor) and specificity (the ability to exclude known non-contributors) testing. A database file containing DNA profiles used in the validation was created based on the STRmix file requirements. The file contained 76 known DNA profiles. This database file was used to test the subset of mixture samples for sensitivity and specificity.

**2-person mixtures**
Ten 2-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. Each 2-person mixture resulted in likelihood ratios favoring inclusion for the individuals known to comprise the mixtures. All other non-contributors in the database had likelihood ratios of zero (i.e., excluded).

**3-person mixtures**
Seventeen 3-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. In eight of the mixtures, including two of the low level mixtures, all non-contributors had likelihood ratios of zero (i.e., exclusion). In the remaining nine mixtures, the non-contributors all had negative log likelihood ratios favoring exclusion.

**4-person mixtures**
Sixteen 4-person mixtures were deconvoluted using STRmix. These were mixtures that included high, mid, and low level mixtures with a range of contributor proportions. Seven of the mixtures had contributors with dropout. After running the mixtures through STRmix, the deconvolution results were compared to the database file of 76 individuals. All sixteen of the 4-person mixtures resulted in the inclusions of the individuals known to comprise the mixtures. In addition to the correct inclusions, one mixture (mixture ID: 4-63) also had a single non-contributing profile from the database that also resulted in a likelihood ratio that favored inclusion. When the result was conditioned on the contributor with the highest contribution, that non-contributor was excluded as a possible contributor to the mixture.

**Adjudicated Cases**
Six adjudicated cases with a sexual component (sex crimes and child abuse) were selected because these cases contained samples known to have mixtures of DNA, are representative of the types of cases encountered, had a high probative value, and represented a range of previously validated DNA typing kits (Profiler Plus, COFiler, Identifiler, and Identifiler Plus) for comparison to the GlobalFiler DNA typing kit. One additional case with a sexual component did not have prior DNA typing of the evidence or the victim’s reference, but had multiple samples for testing with a range of mixture proportions.

**US Army Criminal Investigation Laboratory validation**
Identifiler Plus/3130xl
42 ground truth mixtures were prepared covering 2- (x18), 3- (x12), and 4- (x12) contributors.

These were all amplified in replicate (total of 84 samples) at various templates and contributor proportions to include trace contributor(s) with dropout.

Studies with these mixtures were to assess the reliability and limitations of the software and to better predict the LR outputs for interpretation guidelines. These studies evaluated the mixture deconvolution tool, proposition settings (i.e. number of contributors, Type 1 and Type II errors), reproducibility, and sensitivity.

In addition, 204 single source samples were evaluated through Model Maker for peak variance and in determining LSAE. Of these, 64 profiles were from non-probative casework data (i.e. partial, degraded, and preferentially amplified) to better model the variance observed in actual casework.

**FBI Laboratory validation**

Identifiler Plus, 3130x1, AT=50 RFU, drop-in=0 (27 cycles). With the exception of the N-1 Contributor study as noted below, typing data were generated from laboratory-prepared single-source and mixed DNA samples. Adjudicated case data were also evaluated as specified. Many amplifications/analyses were replicated (not indicated).

**Sensitivity & Specificity:**

- 106 x 2-person (neither, one or both contributors’ DNA in the 2-person mixtures were UV-degraded) (0.075 to 1 ng) (20:1 to 1:1): 212 $H_p$ true, 22,504 $H_d$ true
- 66 x 3-person (0.375 to 3 ng) (16:1:1 to 1:1:1): 187 $H_p$ true, 13,625 $H_d$ true
- 84 x 4-person (0.95 to 4 ng) (19:1:1:1 to 1:1:1:1): 336 $H_p$ true, 17,808 $H_d$ true
- 19 x 5-person (0.25 to 2 ng) (10:1:1:2:2 to 1:1:1:1:1): assuming one contributor (major, if present) in both $H_p$ and $H_d$, 95 $H_p$ true, 4161 $H_d$ true. *Incorrect number of contributors:*

**N+1 trials:**

- 5 x 1-person, 10 $H_p$ true, 1000 $H_d$ true
- 9 x 2-person, 18 $H_p$ true, 1800 $H_d$ true
- 9 x 3-person, 27 $H_p$ true, 1800 $H_d$ true

**N-1 trials:**

- 3-person profiles constructed using 1 x 2-person, with a third “contributor” (“child,” sharing alleles, so as to appear to underrepresent the true number of contributors) added artificially at 50, 100 & 200 RFU avg, such that:
  3 x 3-person, 18 $H_p$ true, 400 $H_d$ true, includes testing as both 2-person and 3-person.

**Allele sharing among mixture contributors:**

- 2 x 2-person (1 parent-child, 1 sib-sib), each: (1:10, 1:5, 1:2, 1:1) and (2 to 0.03 ng)
- 5 x 3-person (2 parent-parent-child, 1 sib-sib, 1 parent-child-unrelated, 3 unrelated), each: (1:1:1, 1:2:2, 1:2:4, 1:3:6)
- 7 to 24 propositions were tested for each mixture: $H_p$ including contributors, related non-contributors and unrelated non-contributors, as well as simultaneously hypothesized contributors and non-contributors.
- Both correct and incorrect numbers of contributors were evaluated.

**Adjudicated cases:**
Nineteen previously examined 2-, 3- and 4-person mixtures were examined using STRmix™ and the results were compared to the reported statistics and conclusions. Each of four known POIs was hypothesized individually as a contributor, regardless of whether the known individual had been previously excluded as a possible contributor. For all STRmix™ analyses, the contributor number was set to the minimum number of contributors previously reported for the mixture (e.g., 3 for a reported “3 or more”). In two instances (a 2-person and a 3-person mixture), the STRmix™ results indicated that the reported minimum number may be incorrect, and the contributor number was increased by 1 for additional STRmix™ analyses. The STRmix™ results were considered with respect to the qualitative verbal equivalent scale.

District of Columbia Department of Forensic Sciences (DC DFS) STRmix™ Validation Summary

The Forensic Biology Unit at the DC DFS validated STRmix™ v2.3.06 using AmpFlSTR® Identifiler® Plus on the Applied Biosystems 3130xl Genetic Analyzer according to the applicable validation sections of the FBI’s Quality Assurance Standards and the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. This validation was completed on January 7, 2016.

Parameters for STRmix™ were established prior to starting validation studies. These parameters included a new set of color-specific analytical thresholds determined according to the most current published methods. Forty samples from various validation projects and performance checks were used to calculate the thresholds which were then verified using another set of 265 samples. Allele-specific stutter ratios were determined using 102 single source samples previously run for quality assurance. The saturation parameter was set at 7000 relative fluorescence units (rfu) based on 60 samples which were amplified from 0.03125ng to 4ng. Allele variance, stutter variance and locus specific amplification efficiency parameters were determined by the Model Maker using 10 samples amplified 10 times from 0.1ng to 1.0ng (100 samples total).

Validation studies were then conducted using the established parameters to evaluate the performance of STRmix™. Two single source samples amplified at low level target quantities (0.025ng to 0.4ng) were interpreted to show that the most supported genotypes received the highest weights. Additionally, the likelihood ratio for one of the samples was calculated by hand and compared to the value calculated by STRmix™ to confirm concordance. Seventy-two (72) two-person, thirty-two (32) three-person and seventy-two (72) four-person mixtures amplified at various ratios and targets were interpreted under different conditions including different hypotheses, assuming a true contributor, assigning one additional/less contributor. Comparisons were made to true contributors and a database of approximately 300 non-contributors. All data produced the following expected results:

1. As the complexity of a mixture increased due to additional contributors or low peak heights, the weightings applied by the software decreased.
2. The software was able to reliably resolve mixtures and correctly determine contributors and non-contributors. False inclusions and exclusions were only observed in samples with very low peak heights and/or a high number of contributors.
3. The assumption of a true contributor improved the performance of the software by increasing the likelihood ratio for true inclusions and decreasing the likelihood ratio for true exclusions.
4. Assigning one additional/less contributor did not affect the results of the major or minor contributors in the mixtures; however some false inclusions and exclusions were obtained for trace level contributors.