Transcript of Dr. Mark Perlin's talk on "Preserving DNA Information" delivered on 2 October 2010 in Cleveland, OH at the National Association of Medical Examiners 2010 Annual Meeting.

*Dr. Perlin*: I would like to thank the organizers for inviting us to give this DNA talk. We will be talking about preserving DNA information with human and computer review with the computer system that the Allegheny lab is bringing on board this year.

This is a Blairsville dentist. Blairsville is about an hour east of Pittsburgh. This is John Yelenic and his bride Michelle.

(Next Slide)

This is John four years ago as a murder victim where he was slashed to death in his home and exsanguinated.

(Next Slide)

This is Pennsylvania State trooper Kevin Foley, who was arrested for the crime. He was the boyfriend of the estranged wife, Michelle, with whom he was living at the time.

(Next Slide)

This is the fingernail DNA evidence from the victim, John Yelenic, and this evidence was comprised of two people. It was a mixture of DNA. 93% came from the victim and about 6.7% came as a trace amount from an unknown second contributor.

(Next Slide)

There were three DNA match scores in this case. The first was a score of 13,000 from the FBI lab and a second of 23 million from Dr. Robin Cotton, former director of the Cellmark laboratory. There was a score of 189 billion by the TrueAllele computer system. So the question is why these scores are different.

(Next Slide)

The answer is that they arise from different interpretation methods. They vary based on the nature of the data used. The FBI lab uses a standard inclusion method, which does not use the victim profile, as seen on the left. Inclusion does not preserve the quantitative data, as I will show in a minute. Dr. Robin Cotton's method does use the victim profile as information to infer a genotype, but again it does not use all of the quantitative data. The computer system makes full use of all of the information to the extent that statistical modeling and computation will

 

permit. It did use the victim profile in this case, and it does preserve all of the quantitative information present in the data.

(Next Slide)

Here we see one of the 13 CODIS loci in the case. On the x-axis is the length of the DNA fragment, and on the y-axis is the quantity of DNA at each of these alleles. The two tall peaks correspond to the victim, who is over 90% of the two-person mixture. Then the question becomes, where in all this data might there be the genotype of this unknown second contributor? There might even be more than one possibility.

(Next Slide)

The quantitative approach is to use all of the data. The way this works, in this case, is that the victim genotype is assumed to be present because the sample was taken from the victim and is seen in all of the data. The computer will try out every possible allele pair, here a 10 and 13, in every possible ratio of different heights. So, it does a massive search on the unknown genotype, on the amount of DNA, on the relative heights of the two, and on dozens of other variables in the process. For each hypothesis of what the allele pair would be, it generates a pattern. That pattern is compared against the data. In this case, we see that the hypothesis of a [10,13] predicts the data very well with the two victim peaks and

 

two unknown peaks. For this to work, a computer has to try out every possibility and assign what its likelihood is based on the data.

(Next Slide)

The FBI lab method, which is the standard method in the US, discards the quantitative data. The way that is done is that a laboratory-wide threshold is applied to the data so that it is cut off. Instead of having peak heights, they have equal allele events of anything over threshold. Anything underneath would be discarded. Then, once they have this reduced data, anything, any pair of alleles, such as [8,8], [8,10], [8,12] and so on (there are 10 in this case), would be considered to be equally likely. The result of having 10 possibilities of equal likelihood, instead of only one as with the quantitative information method, is a considerable loss of information. How much? Well, we will see.

(Next Slide)

Let us look at all 13 CODIS loci in this case. If we look at the likelihood ratio (LR), which is a measure of the specificity of the DNA match, at each of 13 CODIS loci, this is the FBI's result. This is the LR amount of 10, 20, 30, and so on, moving up the y-axis. A product of those numbers at all loci comes out to 13,000.

(Next Slide)

Dr. Robin Cotton's method, which uses the victim information, obtains more information, particularly at those data loci where there were four alleles. As in the case we saw, at locus D7 we could clearly separate out the two tall peaks and the two very small peaks. In that situation, the extra information that was obtained by a more informative interpretation approach at these two loci increased the match strength three orders of magnitude to 23 million ($10^7$).

(Next Slide)

The computer was able to replicate what Dr. Cotton did at these two loci, but also, at other loci, it was able to extract more information by accounting for other quantitative effects that are in the data, adding yet another four orders of magnitude up to $10^{11}$ or 189 billion.

(Next Slide)

So, one might ask, how is this result used in court? There was a Frye hearing and a full day of testimony presenting articles and evidence on these topics, and at the end of the day, both Dr. Cotton's method and our computer method were accepted. This was the first time a computer interpretation of DNA evidence by a statistical program was accepted in court that we know of.

(Next Slide)

According to the prosecutor, the state trooper was convicted of first-degree murder of John Yelenic based largely on the DNA.

(Next Slide)

Now one can ask, how general is this phenomenon? Is this limited to only to this case? Is this well known? A study was done five years ago by the federal government at NIST (the National Institute of Standards and Technology) where samples were sent out to over 50 laboratories that do DNA testing in crime labs around the country. What they discovered was that there was a 10 order of magnitude difference between labs that used inclusion methods, without quantitative data, getting around $10^4$ (in the tens of thousands), and the few labs that were using all of the quantitative information by eye calculated of up to $10^{14}$ (around a hundred trillion). So, it was well known, at that point, that there was a 10 billion-fold discrepancy of interpretation of DNA information on the exact same data.

(Next Slide)

Collaborating with NIST, we undertook a study where they synthesized mixtures of known composition: five different mixture proportions in different dilutions. We

 

see those 40 experiments as dots on the scatter plot here in blue. The x-axis is the logarithm of the amount of unknown DNA, and the y-axis, again, is information measured in log(LR) units ($10^3$, $10^6$, $10^9$, and so on). What we found is that with computer interpretation that used all of the information down to about 100 pg of DNA, we would get all of the information, and then there would be a linear decrease down to about 15 pg of DNA to a million to one match score. Human review was shifted one entire order of magnitude to the right, and below 150 pg was not getting much information, which is perhaps why the current guidelines discourage crime labs from looking manually much below 100 pg. Interestingly, this experiment data was used in court, particularly at the admissibility hearing, to show that based on the amount of DNA under the fingernails, we would predict the match score that was found in the range of 100 billion, as opposed to what alternative methods would have found (which would have been very little). This work was published in December in PLoS One, which is an online peer reviewed journal.

(Next Slide)

We then did another study. This is with the New York State Police and is under review at the Journal of Forensic Sciences. Here we look at eight cases of two-person mixtures. Each mixture has two contributors, and in orange, we are seeing the information content from eight cases that the New York State police have reviewed using inclusion. In blue, we can see what the computer result is

 

on the identical data not assuming the victim but just using the information given. On average, the human review produced results of about 10 million ($10^7$), and the computer review was 10 trillion ($10^{13}$). We see about a million-fold ($10^6$) information gap on cases that the New York State police had solved.

(Next Slide)

The New York State Police and Cybergenetics then took yet another look at the data. Here, we looked at instead of how much information was obtained by computer given that a person was able to solve a case and put a number to it, let us just look at all of the cases. Here, we see, on the x-axis, a listing of all 85 cases. The y-axis again shows information in log units, 5, 10, 15, and so on. If we focus just on the large blue background, this is what the computer did. The height is the information, the logarithm of the match score, at each case, and it is shown in descending order. The median value is $10^{15}$ (or a quadrillion to one), and we get a meaningful result for every item. Maybe there are a few that are under a million to one at the very end on the right. By comparison, this is what human review was able to achieve. The different colors correspond to different methods, based on how difficult the case was. The four gray bars are assuming the mixture came from one person because there was a clear major contributor. The green bars assumed that we could use the victim, and the lower orange bars was human review where the victim was not used at all. There are two notable features of this particular slide. The first is, if we look at the cases where a

 

computer match score was produced, for how many did a human review produce a match score? The answer is less than 30%. This means that over 70% of the DNA cases' DNA items of evidence there was usable information that could identify a suspect, but nothing was reported. This was used to show that the productivity is such that 3.5 times the amount of work is being done redundantly if human review is used instead of a more informative computer review. The other interesting concept from here is, if we look at the total area of blue, then this is the total amount of information over the 85 items that was obtained. If we look at the area of the colored bars of human review and look at the ratio, it is only 20%. Overall, 80% of the information is discarded by human review that computer review actually preserves.

(Next Slide)

Let us draw some conclusions from these described and ongoing studies. The first is, as is well known and published almost monthly in the forensic science literature, DNA interpretation of challenging samples, mixtures, low-level, property crime, handguns, and so on varies greatly across laboratories, jurisdictions, countries, and so on typically by 10 orders of magnitude, depending on the methods used. Human interpretation discards data by imposing limits, saying what data can be used, above what level, what can be thrown out, and what can be kept. Computers can preserve far more of the information that is present for identification in the data with extensive computer modeling of

 

the quantitative data and modern statistics. The average loss between the standard threshold inclusion methods and quantitative methods is at least a million to one in match score. This loss may not affect the case for where the human result is a million to one and the computer result is a trillion to one, but when the computer result goes down to 1 million to one, humans have almost nothing to say about the results at all. The message for Medical Examiners is that, when doing excellent scientific work of collecting the evidence and sending it off for DNA interpretation, when the answer comes back “inconclusive”, that might be entirely an artifact of the human review method that was applied to it and not representative of the information the data contained. Thank you.