FORENSIC SCIENCE COMMUNICATIONS

October 2009 - Volume 11 - Number 4

FSC Links

- Table of Contents
- Meetings and Conferences
- Editors
- Back Issues
- About FSC
- Instructions for Authors
- Search
- FBI Laboratory
- Current Issue

SWGDAM Recommendations to the FBI Director on the "Interim Plan for the Release of Information in the Event of a 'Partial Match' at NDIS"

Scientific Working Group on DNA Analysis Methods Ad Hoc Committee on Partial Matches

Introduction | Notes | Appendix | References

Introduction

The Scientific Working Group on DNA Analysis Methods (SWGDAM) Ad Hoc Committee on Partial Matches was asked to examine the population genetics and statistics behind partial DNA matches and to make recommendations about the disposition of Combined DNA Index System (CODIS) searches that result in oderate stringency matches. When such partial DNA profile matches ("partial matches") occur, although it is clear that the offender profile is not the source of the crime scene profile, the possibility does exist that a close biological relative ("relative") of the offender might be the source of the crime scene profile.

The Committee looked at several examples of partial DNA matches to illustrate typical casework situations that arise in performing CODIS searches. The Committee met twice and had one telephone conference call, as well as numerous individual e-mail and telephone communications. The following recommendations represent a consensus opinion of the committee members.[1]

The Committee heard presentations from Steven Myers and Ranajit Chakraborty in which analysis of simulation studies carefully examined the extent of false positives (i.e., two DNA profiles from unrelated people that match at moderate stringency) and false negatives (two DNA profiles from true siblings that are not partial matches).

Moderate stringency CODIS matches, in general, have very low efficiency in locating true relatives in offender databases. There is little useful probative value in the majority of partial matches using the current CODIS searching rules and algorithms. There are two main reasons for this: (1) true siblings will very rarely share alleles at all 13 CODIS loci; (2) as offender DNA databases get large, the number of unrelated people that do share at least one allele at all loci increases very rapidly. The original intent for allowing moderate stringency CODIS searches was the realization and acknowledgment that crime scene profiles often may be partially degraded and/or contain DNA from more than one contributor. Additionally, different primer sets may have been used between profiles. Allowing the detection of partial matches can help accommodate these two scenarios and allow the ultimate detection of full, high-stringency matches that might otherwise not have been found.

There are other documented criteria for detecting possible relatedness between two DNA profiles (number of shared alleles, number of identical loci, number of rare alleles shared). The Committee specifically examined the question of whether partial DNA matches found using moderate stringency search criteria have any practical value in identifying relatives of a person in a forensic sample. The Committee makes the following recommendations to guide a laboratory's decision-making process regarding whether to release the name of the offender whose relative may be the source of the DNA profile. The Committee feels that these recommendations should be used in toto and that offender names should not be released if one or more recommendations are not met.

Recommendation 1: The crime scene sample should be from a single source.

Rationale: When more than two alleles at a locus are used for searching, the number of partial matches from unrelated individuals increases very quickly. Although the Committee did not calculate the extent of this increase, it is clear that the added number of partial matches is expected to be high when using moderate stringency CODIS searching rules. Similarly, when only one allele (perhaps an obligate, nonvictim allele) is used in moderate stringency searches, there is also a large increase in the number of unrelated individuals in a database who would have partial matches to the crime scene evidence. To help limit the number of individuals in a database who would have partial matches to a crime scene DNA profile, the Committee recommends using only single-source forensic profiles in any evaluation of partial matches. For this same reason, mixtures should be deconvoluted into single-source profiles with as many heterozygous loci as possible before a database search is done.

Recommendation 2: Local DNA Index System (LDIS) and State DNA Index System (SDIS) searches should be performed before searching at the National DNA Index System (NDIS) level.

Rationale: The number of false-positive partial matches is strongly influenced by the size of the offender database searched. Large databases will have more false-positive partial matches. Another consideration is that LDIS and SDIS databases that are more geographically limited have a higher a priori chance of containing profiles actually related to the perpetrator. Both of these factors make a partial match more valuable (i.e., have a higher chance of resulting from the comparison of two related people). As an efficient search strategy, when no full or partial matches are found at the LDIS and SDIS levels, it would be logical to search geographically nearby states before searching the NDIS level. This strategy might be impractical, but it illustrates the general principle that the smaller the database searched, the lower the probability of finding false-positive partial matches and consequently the higher the chance of detecting a true relative.

Recommendation 3: All available CODIS core loci should be used for searching.

Rationale: Even though omitting a locus from a search will increase the chance of finding a true relative that partially matches, the number of false positives increases. The result is that relaxing the partial search by omitting a locus or two is counterproductive and buries the true match to a relative in an even larger number of partial matches to nonrelatives. A minimum of 10 of the CODIS core loci is required for searching forensic DNA profiles at the National DNA Index System level.

Recommendation 4: Whenever possible, partial DNA matches that result from searching databases should have additional loci typed.

Rationale: Y-chromosome short tandem repeat (Y-STR) and mitochondrial DNA (mtDNA) can eliminate unrelated individuals, and although offenders in DNA databases are currently not typed for Y-STR and mtDNA, it is feasible to do this on a small number of candidate partial matches. This, of course, is useful only when enough forensic sample is available. Y-STR and mtDNA are lineage markers and are not highly discriminating on their own, but they eliminate pairs of unrelated people. Because most perpetrators are male, and in this discussion of partial matches we are typically looking for father-son or full-sibling pairs, Y-STR information can winnow relatives from false-positive partial DNA matches. In some cases, additional autosomal genetic loci can add probative information to a partial match.

Recommendation 5: An Expected Match Ratio (EMR) and an Expected Kinship Ratio (EKR) should be calculated for a partial match.

Rationale: The size of the database searched has a dominating effect on the probative value of a partial match. The larger the database, the greater the number of false-positive partial matches. It is possible to calculate a ratio of the probability of observing a partial match in true relatives to the probability of observing that same partial match when a database of size *N* is searched. The EMR and EKR can provide somewhat objective measures of the partial match value as a lead to an actual relative. Details of these calculations are given in an Appendix[2] to this report.

Recommendation 6: Four individual EMRs and EKRs should be calculated on the assumption that the database searched is made of (1) African Americans, (2) Caucasians, (3) S.E. (Southeastern) Hispanics, and (4) S.W. (Southwestern) Hispanics. The partial match is considered useful only if either the EMR or the EKR satisfies the following thresholds: at least one of the four database values is greater than or equal to 1.0 and all of the others are greater than or equal to 0.1.

Rationale: Because we do not know the actual ethnic composition of offender databases, we can do a pragmatic set of calculations under an assumption that the database is 100 percent of each of the four major ethnic groups in the FBI allele-frequency databases. Although the thresholds that we agreed upon (i.e., at least one EMR greater than or equal to 1.0 and all three of the other EMRs greater than or equal to 0.1) are somewhat arbitrary, they do set useful thresholds for the partial match to identify a true relative, if one exists in the database.

Recommendation 7: In order to implement these recommendations, it is important that CODIS Administrators have training in the evaluation of partial matches and in reporting the potential value of these matches.

Rationale: Partial matches come in many varieties, and the probative value of one, if any, can be determined only by further calculations and possibly by additional analytical tests. These calculations require the use of a spreadsheet or other software that currently is not in use but should be created and distributed. Report-wording suggestions need to be developed to stress the “limited” nature of the partial match and to state explicitly the possible family relationships.

Recommendation 8: All CODIS laboratories using these recommendations will report the profiles and associated EMRs and EKRs to the FBI, which will monitor the effectiveness of this approach.

Rationale: This is an emerging issue, and we have had little actual data to evaluate. With this or any other novel approach, an assessment seems necessary. It also would be useful if laboratories using alternate methods of identifying database partial matches report that method and data to the FBI. The FBI should evaluate these data provided by the LDIS and SDIS laboratories with the intent of modifying these recommendations and/or refining the thresholds as more data are collected.

Notes

1. These recommendations were approved by the SWGDAM membership on July 17, 2008, and are the recommendations proposed by the Ad Hoc Committee on Partial Matches, with minor revisions. The Ad Hoc Committee on Partial Matches comprised George Carmody (Chair), Ranajit Chakraborty, W. David Coffman, Kenneth Kidd, Steven P. Myers, Taylor Scott, and Ted Staples.

2. The Appendix was approved by the SWGDAM membership on July 17, 2008, and is the Appendix proposed by the Ad Hoc Committee on Partial Matches, with minor revisions.

Appendix to the SWGDAM Recommendations to the FBI Director on the “Interim Plan for the Release of Information in the Event of a ‘Partial Match’ at NDIS”

When DNA profiles are compared, there are well-understood statistical patterns of expected allele sharing that can be calculated for unrelated as well as closely related individuals. These patterns of genetic sharing have been used in medical genetics as well as to help in the identification of disaster victims and missing persons. All of the genetic sharing between two DNA profiles can be distilled into a likelihood ratio that can then allow a statistical evaluation of how much more likely the match resulted from related versus unrelated individuals.

However, when partial DNA matches occur as the result of a moderate stringency search of an offender database of size *N*, attempts to evaluate the significance of that partial match have two complications. The first complication is that when many comparisons have been done, there will be partial matches between unrelated profiles, and the number of these increases with the size of the database, *N*. Even though on average we expect profiles from close relatives to have more genetic sharing, a mere partial match is far from being a guarantee that we have found profiles from two related people when large databases are searched for only 13 STR (short tandem repeat) loci.

The second complication is a bias that is introduced when a large number of comparisons (*N*) have been made and only the partial matches are examined. If a second calculation were now done on only the comparison that matched at moderate stringency (for example, a kinship index [KI]), the statistical properties of this number are not the same as if we had done the same calculation on all *N* comparisons and ranked them.

Because some unknown fraction of moderate stringency partial matches will come from people who are closely related, it is vital to find those matches that have a higher prospect of coming from related pairs of people. To this end, the Committee suggests two approaches to address the complications that arise in searching databases at moderate stringency. The first is a statistical calculation devised by a member of the Committee, Steven Myers. The Committee named this calculation the Expected Match Ratio (EMR). The second is a modification of the standard kinship index. The Committee named this calculation the Estimated Kinship Ratio (EKR).

In recommendations 5 and 6, we suggest that when a moderate stringency match is found, the EMR and EKR are calculated. What follows in this Appendix to the Ad Hoc Committee Recommendations are the genetic background, mathematical derivation, and statistical logic of these calculations, as well as examples of their application.

Definition of Moderate Stringency Partial Matches Between Single-Source Profiles

Using the CODIS moderate stringency match criteria, a candidate match between two single-source profiles is declared at a locus when the alleles entered for one profile are completely detected within the alleles entered for the second profile. This includes both high-stringency matches where the alleles entered are identical, as well as moderate stringency matches where a profile with only one allele entered is concordant with either one of two alleles entered for the second profile.

Table 1: Example of a Moderate Stringency Partial Match

Locus	Forensic Unknown	Candidate Offender	Match Stringency
A	7	7	High
B	7	7, 8	Moderate
C	13,14	13,14	High
D	13,14	13	Moderate
E	13,14	14	Moderate

What Do the Expected Match Ratio and Estimated Kinship Ratio Measure?

The EMR assesses what is more likely to occur, a partial match between the perpetrator and one of his or her relatives or a partial match between the perpetrator and one or more unrelated people in a database of size *N*. It does this by comparing two probabilities.

- The numerator is the probability of a moderate stringency partial match ("msMatch") between the donor of the forensic unknown profile and one relative. This is a kinship-testing comparison that includes estimations of alleles' being identical by descent and/or identical only by state (Thompson 1991).
    - *P*(msMatch | Forensic Unknown & 1 Relative)
- The denominator is the probability of a moderate stringency partial match between the donor of the forensic unknown profile and one unrelated person, multiplied by the number of profiles searched (*N*). This product estimates the number of moderate stringency matches we expect to see in comparing the forensic profile to *N* unrelated individuals.
    - *P*(msMatch | Forensic Unknown & 1 Non-Relative) * *N*

The final EMR calculation is in the form of a likelihood ratio:

- *P*(msMatch | Forensic Unknown & 1 Relative)
  P (msMatch | Forensic Unknown & 1 Non-Relative) * *N*

In contrast, the EKR focuses on the specific pair of profiles observed in the partial match. It asks whether these profiles are more likely to be seen in one pair of related people or one pair of unrelated people found by searching a database of size *N*. It does this by comparing two probabilities:

- The numerator is the probability of observing the profiles from the moderate stringency partial match if the donor of the forensic unknown profile and the candidate offender are related. This is a kinship-

testing comparison that includes estimations of alleles being identical by descent and/or identical only by state (Thompson 1991).

- *P*(msMatch | Forensic Unknown & Candidate Offender are related)

- The denominator is the probability of observing the profiles from the moderate stringency partial match if the donor of the forensic unknown profile and the candidate offender are unrelated, multiplied by the number of profiles searched (*N*).
  - *P*(msMatch | Forensic Unknown & Candidate Offender are unrelated) **N*

The final EKR calculation is in the form of a likelihood ratio:

- *P*(msMatch | Forensic Unknown & Candidate Offender are related)
  P(msMatch | Forensic Unknown & Candidate Offender are unrelated) * *N*

When the EMR or EKR is greater than 1.0, the expectation of a partial match between the donor of the forensic unknown profile and one relative is greater than the expectation of a coincidental partial match (or matches) in a search of *N* unrelated individuals in the database, supporting the hypothesis that the match is between relatives.

What Are the Formulae for the Probabilities Used in the EMR and EKR?

At each locus, the numerator and denominator are the probabilities describing the ways two profiles could match under the moderate stringency rules.

- For loci where the forensic unknown has a single allele, this is the probability of having any profile that matches in the manner of locus A and locus B in Table 1.
- For loci where the forensic unknown profile has two alleles, this is the probability of having any profile that matches in the manner of locus C, locus D, and locus E in Table 1.
- The calculations incorporate a subpopulation correction (Balding 2005). In accordance with the recommendations of the National Research Council (1996), the Committee recommends a theta (θ) of 0.01.
- The kinship coefficients are per Thompson (1991):

Relationship	k_2	k_1	k_0
Full Siblings	0.25	0.5	0.25
Parent-Child	0	1	0

Loci where the forensic unknown profile has a single allele:

- Formula 1a—The candidate offender profile might have a single allele (see locus A in Table 1).
  - *P*(msMatch | Forensic Unknown & 1 Relative)

$$k_2 + \frac{k_1[2\theta + (1-\theta)p]}{(1+\theta)} + \frac{k_0[2\theta + (1-\theta)p][3\theta + (1-\theta)p]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $k_2 + k_1 p + k_0 p^2$.

  - P(msMatch | Forensic Unknown & 1 Non-Relative)

$$\frac{[2\theta + (1-\theta)p][3\theta + (1-\theta)p]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to p^2.

- Formula 1b—The candidate offender profile might have two alleles (see locus B in Table 1).
  - P(msMatch | Forensic Unknown & 1 Relative)

$$\frac{k_1(1-\theta)q}{(1+\theta)} + \frac{k_0 2[2\theta + (1-\theta)p](1-\theta)q}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $k_1 q + k_0 2pq$.

  - *P*(msMatch | Forensic Unknown & 1 Non-Relative)

$$\frac{2[2\theta + (1-\theta)p](1-\theta)q}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $2pq$.

  - For the EMR calculation, *Q* could be any allele that is not *P*. Therefore, the frequency *q* would equal (1 – *p*).
  - For the EKR calculation, *q* would be the frequency of the candidate offender's nonmatching allele.

- The EMR would use the sum of 1a and 1b.

- The EKR would use 1a or 1b depending upon the profile of the candidate offender.

Loci where the forensic unknown profile has two alleles:

- Formula 2a—The candidate offender profile might have two alleles (see locus C in Table 1).
  - P(msMatch | Forensic Unknown & 1 Relative)

$$\frac{k_2 + k_1[\theta + (1-\theta)p]}{2(1+\theta)} + \frac{k_1[\theta + (1-\theta)q]}{2(1+\theta)} + \frac{k_0 2[\theta + (1-\theta)p][\theta + (1-\theta)q]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $k_2 + k_1 p/2 + k_1 q/2 + k_0 2pq$.

  - P(msMatch | Forensic Unknown & 1 Non-Relative)

$$\frac{2[\theta + (1-\theta)p][\theta + (1-\theta)q]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $2pq$.

- Formula 2b—The candidate offender profile might have a single allele matching the first allele of the forensic unknown (see locus D in Table 1).
  - *P*(msMatch | Forensic Unknown & 1 Relative)

$$\frac{k_1[\theta + (1-\theta)p]}{2(1+\theta)} + \frac{k_0[\theta + (1-\theta)p][2\theta + (1-\theta)p]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $k_1 p/2 + k_0 p^2$.

  - P(msMatch | Forensic Unknown & 1 Non-Relative)

$$\frac{[\theta + (1-\theta)p][2\theta + (1-\theta)p]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to p^2.

- Formula 2c—The candidate offender profile might have a single allele matching the second allele of the forensic unknown (see locus E in Table 1).
  - *P*(msMatch | Forensic Unknown & 1 Relative)

$$\frac{k_1[\theta + (1-\theta)q]}{2(1+\theta)} + \frac{k_0[\theta + (1-\theta)q][2\theta + (1-\theta)q]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to $k_1 q/2 + k_0 q^2$.

  - *P*(msMatch | Forensic Unknown & 1 Non-Relative)

$$\frac{[\theta + (1-\theta)q][2\theta + (1-\theta)q]}{(1+\theta)(1+2\theta)}$$

    At θ = 0, this reduces to q^2.

- The EMR would use the sum of 2a, 2b, and 2c.
- The EKR would use 2a, 2b, or 2c depending upon the profile of the candidate offender.

Note: The probability of selecting someone with the forensic unknown profile is identical in both the numerator and denominator and cancels out in the final equations. Therefore, those probabilities were not included in the equations listed above.

What Value of N Should Be Used?

The intent of this report is to address partial matches between single-source forensic unknown profiles and candidate offenders, because it is the partial matches to candidate offenders that will generally be investigated further. For this purpose, the value of N used would be the total number of offender profiles searched.

Combining additional genetic testing with the EMR or EKR:

When additional autosomal loci are tested after a moderate stringency search has been done, these new genetic loci, which were not used in the original CODIS database search, can be evaluated separately as a standard KI. A KI from loci inherited independently of the searched loci can then be multiplied by the EMR or EKR. If the resultant values from either equation are above the thresholds in Recommendation 6 (i.e., at least one of the four values is greater than or equal to 1.0 and all three of the others are greater than or equal to 0.1), the intent of the AD Hoc Committee is satisfied.

When adding information from Y-STRs (Y-chromosome short tandem repeats), Y-SNPs (Y-chromosome single nucleotide polymorphism), or mtDNA (mitochondrial DNA), the same multiplication can be done by expressing the upper confidence level of the match as a likelihood ratio. In practice, should further genetic testing of any of these lineage markers disclose a match, the two profiles have a very high probability of coming from related individuals.

Calculating the EMR and EKR

Example 1

A hypothetical partial match:

Locus	Forensic Unknown	Candidate Offender	Match Stringency
D8S1179	12	12, 14	Moderate
D21S11	30, 31	30, 31	High

Using U.S. Caucasian allele frequencies (Budowle et al. 2001) for the loci D8S1179 and D21S11, an example of the EMR calculation (at θ = 0.01) for full siblings is given below:

The forensic unknown D8S1179 profile has only a 12 allele (p = 0.1454):

EMR

- The comparison profile might have only a 12 allele—Formula 1a.
  - P(msMatch | Forensic Unknown & 1 Full Sibling) = 0.3381
  - P(msMatch | Forensic Unknown & 1 Non-Relative) = 0.0277
- The comparison profile might have a 12 allele and some other non-12 allele (q = 1 – 0.1454 = 0.8546)—Formula 1b.
  - P(msMatch | Forensic Unknown & 1 Full Sibling) = 0.4862
  - P(msMatch | Forensic Unknown & 1 Non-Relative) = 0.2693
- The EMR ratio for D8S1179:

$$\frac{(0.3381 + 0.4862)}{(0.0277 + 0.2693)} = \frac{0.8243}{0.2970}$$

EKR

- The candidate offender profile has the 12 and 14 (q = 0.2015) alleles—Formula 1b.
- P(msMatch | Forensic Unknown & Candidate Offender are full siblings) = 0.1146
- P(msMatch | Forensic Unknown & Candidate Offender are unrelated) = 0.0635
- The EKR ratio for D8S1179:

$$\frac{0.1146}{0.0635}$$

The forensic unknown D21S11 profile has the 30 and 31 alleles (p = 0.2321, q = 0.0714):

EMR

- The comparison profile might have the 30 and 31 alleles—Formula 2a.
  - P(msMatch | Forensic Unknown & 1 Full Sibling) = 0.3387
  - P(msMatch | Forensic Unknown & 1 Non-Relative) = 0.0376
- The comparison profile might have only a 30 allele—Formula 2b.
  - P(msMatch | Forensic Unknown & 1 Full Sibling) = 0.0739
  - P(msMatch | Forensic Unknown & 1 Non-Relative) = 0.0581
- The comparison profile might have only a 31 allele—Formula 2.
  - P(msMatch | Forensic Unknown & 1 Full Sibling) = 0.0217
  - P(msMatch | Forensic Unknown & 1 Non-Relative) = 0.0071
- The EMR ratio for D21S11:

$$\frac{(0.3387 + 0.0739 + 0.0217)}{(0.0376 + 0.0581 + 0.0071)} = \frac{0.4343}{0.1028}$$

EKR

- The candidate offender profile has the 30 and 31 alleles—Formula 2a.
- P(msMatch | Forensic Unknown & Candidate Offender are full siblings) = 0.3387
- P(msMatch | Forensic Unknown & Candidate Offender are unrelated) = 0.0376

The EKR ratio for D21S11:

$$\frac{0.3387}{0.0376}$$

The EMR for these two loci is the product of the individual ratios, adjusted for the size of the database (*N*):

$$\frac{0.8243 * 0.4343}{0.2970 * 0.1028 * N}$$

If *N* equals 10, the EMR = 1.2.

If *N* equals 100, the EMR = 0.12.

The EKR for these two loci is the product of the individual ratios, adjusted for the size of the database (*N*):

$$\frac{0.1146 * 0.3387}{0.0635 * 0.376 * N}$$

If *N* equals 10, the EMR = 1.6.

If *N* equals 100, the EMR = 0.16.

These results demonstrate that a moderate stringency partial match found when searching a smaller database is more likely to be from a relative than when searching a larger database.

The EMR for these two loci using U.S. African American, Caucasian, and S.E. (Southeastern) and S.W. (Southwestern) Hispanic allele frequencies (Budowle et al. 2001):

N	African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics	Satisfies Recommen-dations?
10	1.7	12	0.93	1.4	Yes
100	0.17	0.12	0.093	0.14	No

The EKR for these two loci using U.S. African American, Caucasian, and S.E. and S.W. Hispanic allele frequencies (Budowle et al. 2001):

N	African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics	Satisfies Recommen-dations?
10	2.0	1.6	1.5	1.9	Yes
100	0.20	0.16	0.15	0.19	No

Calculating the EMR and EKR:

Example 2

A hypothetical partial match:

Locus	Forensic Unknown	Candidate Offender	Match Stringency
D8S1179	13	13, 14	Moderate
D21S11	28, 31.2	28, 31.2	High
D7S820	12	10, 12	Moderate
CSF1PO	10, 12	10	Moderate
D3S1358	15, 17	15, 17	High
TH01	8	7, 8	Moderate
D13S317	9, 12	9	Moderate
D16S539	11, 12	12	Moderate
VWA	17	15, 17	Moderate
TPOX	8, 11	8	Moderate
D18S51	24	16, 24	Moderate
D5S818	9, 12	12	Moderate
FGA	24, 25	24, 25	High

Using U.S. African American, Caucasian, and S.E. and S.W. Hispanic allele frequencies (Budowle et al. 2001) for the 13 CODIS core loci, the overall probabilities (at θ = 0.01) for the forensic unknown profile:

EMR

- *P*(msMatch | Forensic Unknown & 1 Full Sibling)

African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics
2.319 E-03	2.651 E-03	3.009 E-03	2.519 E-03

- *P*(msMatch | Forensic Unknown & 1 Non-Relative)

African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics
1.941 E-09	2.683 E-09	2.305 E-09	2.575 E-09

The final EMR calculation incorporating various SDIS (State DNA Index System) and NDIS (National DNA Index System) database sizes (Federal Bureau of Investigation [hereafter "FBI"] 2008) (*N*):

- $\frac{P(\text{msMatch} \mid \text{Forensic Unknown \& 1 Full Sibling})}{P(\text{msMatch} \mid \text{Forensic Unknown \& 1 Non-Relative}) * N}$

N	Index	African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics	Satisfies Recommen-dations?
197	WY	6100	5000	6600	5000	Yes
834	RI	1400	1200	1600	1200	Yes
7,127	ME	170	140	180	140	Yes
65,141	CO	18	15	20	15	Yes
216,083	NY	5.5	4.6	6	4.5	Yes
397,500	FL	3	2.5	3.3	2.5	Yes
893,147	CA	1.3	1.1	1.5	1.1	Yes
5,070,473	NDIS	0.24	0.19	0.26	0.19	No

EKR

- *P*(msMatch | Forensic Unknown & Candidate Offender are full siblings)

African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics
4.330 E-12	1.820 E-12	6.461 E-12	2.826 E-12

- *P*(msMatch | Forensic Unknown & Candidate Offender are unrelated)

African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics
3.533 E-17	5.293 E-17	1.753 E-16	1.153 E-16

The final EKR calculation incorporating various SDIS and NDIS database sizes (FBI 2008) (*N*):

- $\frac{P(\text{msMatch} \mid \text{Forensic Unknown \& Candidate Offender are full siblings})}{P(\text{msMatch} \mid \text{Forensic Unknown \& Candidate Offender are unrelated}) * N}$

N	Index	African Americans	Caucasians	S.W. Hispanics	S.E. Hispanics	Satisfies Recommen-dations?
197	WY	620	170	190	120	Yes
834	RI	150	41	44	29	Yes
7,127	ME	17	4.8	5.2	3.4	Yes
65,141	CO	1.9	0.53	0.57	0.38	Yes
216,083	NY	0.57	0.16	0.17	0.11	No
397,500	FL	0.31	0.086	0.093	0.062	No
893,147	CA	0.14	0.038	0.041	0.027	No
5,070,473	NDIS	0.02t	0.0068	0.0073	0.0048	No

Assessment of the Partial Match

Because the EMR values for the California SDIS search exceed the minimum levels, the recommendations of the Committee would be satisfied for all of these SDIS searches. The EMR calculated for a national search would likely rise above the recommended thresholds if an inclusion was determined after testing the forensic unknown and candidate offender with, for example, one of the commercially available Y-STR kits.

References

Balding, D. J. *Weight-of-evidence for Forensic DNA Profiles*. John Wiley & Sons, Hoboken, New Jersey, 2005, p. 64 (applying equation 5.6).

Budowle, B., Shea, B., Niezgoda, S., and Chakraborty, R. CODIS STR loci data from 41 sample populations, *Journal of Forensic Sciences* (2001) 46:453–489.

Federal Bureau of Investigation. *CODIS—NDIS Statistics*. Available: http://www.fbi.gov/hq/lab/codis/clickmap.htm. Accessed January 31, 2008, for NDIS Database sizes as of October 2007.

National Research Council. Committee on DNA Technology in Forensic Science. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington, D.C., 1996.

Thompson, E. A. Estimation of relationships from genetic data. In: *Handbook of Statistics 8: Statistical Methods in Biological and Medical Sciences*. C. R. Rao and R. Chakraborty, Eds. North-Holland–Elsevier Science, Amsterdam, 1991, pp. 255–269.