The Ceiling Principle Is Not Always Conservative in Assigning Genotype Frequencies for Forensic DNA Testing

To the Editor:

In forensic DNA typing for individual identification, when a suspect’s DNA pattern matches that from a crime scene specimen, a crucial step is the assignment of a probability that the specimen genotype would match that of a person randomly selected from the population of potential perpetrators. On the presumption that a suspect is innocent until proved guilty, a method of assigning a probability to a suspect’s genotype, given the same genotype from a crime specimen, should be conservative in the sense that the assigned probability should be greater than or equal to the true
probability. A “ceiling principle” has been recommended as a conservative method of providing an upper bound on the true match probability, assuming no laboratory mix-ups (Lander 1991; National Research Council Committee on DNA Technology in Forensic Science [hereafter NRC] 1992, p. 3-13). It is shown here that the ceiling principle does give an upper bound if the product rule applies within and across loci in all genetically differentiated subpopulations. However, a counterexample with correlations across loci is given here in which the ceiling principle gives an estimated “upper bound” that in fact is strictly smaller than the true match probability for every observed genotype. In this case, the ceiling principle exaggerates the power of the evidence to inculpate the suspect and is not conservative. Apparently, examples in which the ceiling principle fails to be conservative have not been described previously. The ceiling principle may be inappropriate for general use in forensic DNA typing unless additional information is available that justifies the use of the product rule within each subpopulation of a genetically heterogeneous population. Alternative methods of estimating a match probability should be explored.

The ceiling principle is presented (NRC, pp. 3-10–3-11) as “a practical and sound approach for accounting for possible population substructure . . . applying the ceiling principle involves two steps: (1) For each allele at each locus, determine a ceiling frequency [emphasis in original] that is an upper bound for the allele frequency that is independent of the ethnic background of a subject; and (2) To calculate a genotype frequency, apply the multiplication rule, using the ceiling frequencies for the allele frequencies. [¶] How should ceiling frequencies be determined? . . . The [NRC] committee strongly recommends the following approach: Random samples of 100 persons should be drawn from each of 15–20 populations, each representing a group relatively homogeneous genetically; the largest frequency in any of these populations or 5%, whichever is larger should be taken as the ceiling frequency.”

After giving a numerical example of the procedure, the NRC observes (NRC 1992, p. 3-11): “Because the calculation uses an upper bound for each allele frequency, it is believed to be conservative given the available data, even if there are correlations among alleles because of population substructure and even for persons of mixed or unknown ancestry.” Further (NRC 1992, p. 3-13), “The calculation is fair to sus-

pects, because the estimated probabilities are likely to be conservative in their incriminating power.”

The ceiling principle is indeed conservative if a population contains subpopulations in each of which there is no linkage disequilibrium between loci and in each of which Hardy-Weinberg equilibrium holds within every locus. In this case, the product rule is valid within each subpopulation. Since a product of non-negative numbers is a monotonically increasing function of each factor, the result given by the ceiling principle necessarily is an upper bound on the probability of any genotype in the population and is therefore conservative.

Without the assumption of independence between alleles, within and between loci, the ceiling principle need not be conservative. For example, consider a population with three subpopulations, S1, S2, and S3. For purposes of intuition, these subpopulations may be thought of as major ethnic groups, but the example is not intended to be realistic. Assume that each subpopulation is ⅓ of the whole population. Suppose a DNA test is performed at three loci—A, B, and C—each of which has exactly three alleles. The three alleles of the A locus are A1, A2, and A3; of the B locus, B1, B2, and B3; and, of the C locus, C1, C2, and C3. Suppose that the alleles at the three loci are so strongly associated that only three haplotypes are found in the population: A1B1C1, A2B2C2, and A3B3C3. These three haplotypes may be called “H1”, “H2”, and “H3”, respectively. (A haplotype such as A1B2C1 is assumed not to occur at all.) Suppose the haplotype frequencies are as shown in table 1 and that any two haplotypes combine at random within each subpopulation to form a genotype. Thus each locus is assumed to be in Hardy-Weinberg equilibrium within each subpopulation. For example, since H1 and H2 occur with

### Table 1

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency for Subpopulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1C1</td>
<td>S1</td>
</tr>
<tr>
<td>A2B2C2</td>
<td>.5</td>
</tr>
<tr>
<td>A3B3C3</td>
<td>0</td>
</tr>
<tr>
<td>All others</td>
<td>0</td>
</tr>
</tbody>
</table>
frequency .5 in S1, the frequency of the triple heterozygote H1H2 = A1A2B1B2C1C2 genotype in S1 is 
2 \times .5 \times .5 = .5. The factor of 2 allows for the fact 
that the H1 haplotype may come from the father or 
the mother. Likewise the frequency of the triple homozygote H1H1 = A1A1B1B1C1C1 genotype in S1 is 
.5 \times .5 = .25.

Now the genotype frequency obtained from the ceiling 
principle will be compared with the correct genotype frequency. The maximum frequency (over all 
three subpopulations) of each allele at each locus is .5. 
Hence the predicted frequency, by the ceiling principle, 
for each triple heterozygote H1H2, H1H3, and 
H2H3 is \((2 \times .5 \times .5)^3 = .125.\) The predicted 
frequency, by the ceiling principle, for each triple homozygote 
H1H1, H2H2, and H3H3 is \((.5 \times .5)^3 = .015625.\) These are the only genotypes that can possibly 
occur in a crime specimen from the hypothetical 
population constructed in table 1 and therefore are the 
only genotypes that need be considered.

The actual frequency for each triple heterozygote 
H1H2, H1H3, and H2H3 is \((1/3) \times .5 = 1/6,\) because 
H1H2 can occur only in S1, H1H3 only in S2, and 
H2H3 only in S3; the frequency of each triple heterozygote is .5 in the subpopulation in which it occurs, 
and each subpopulation is 1/3 of the whole population. 
The actual frequency for each triple homozygote 
H1H1, H2H2, and H3H3 is \((1/3) \times (.5 \times .5) \times 2 = 1/6,\) because each triple homozygote occurs (with 
frequency .5 \times .5) in two subpopulations, each of 
which is 1/3 of the whole population. Since there are 
three triple heterozygotes and three triple homozygotes, 
the actual frequencies of all possible genotypes 
add up to 1, as they must.

In this example, for every possible genotype that 
could be found in a crime specimen, the ceiling 
principle gives a frequency that is lower than the actual 
frequency. For triple heterozygotes, .125 < 1/6. For 
triple homozygotes, .015625 < 1/6. While the ceiling 
principle gives positive estimates for genotypes other 
than triple heterozygotes and triple homozygotes, 
these other genotypes would never be observed in a 
crime specimen from the hypothetical population and 
are therefore irrelevant to evaluating the method. In 
this example, the ceiling principle is uniformly more 
inocriminating than the evidence justifies.

It is not yet clear whether, in practice, the ceiling 
principle is likely to be conservative or nonconserva-
tive. Some evidence presented by Risch and Devlin 
(1992) apparently favors the applicability of the prod-
uct rule within very broad ethnic groups, but the popu-
lation-sampling procedure by which the data base is 
constructed is not specified and may not correspond 
in a meaningful way to a random sample of the popu-
lation of potential crime perpetrators. Further, by 
matching genotypes, the analysis of Risch and Devlin 
(1992) did not test the applicability of the product rule 
within each locus, whereas the ceiling principle applies 
a product rule within (as well as between) loci after 
taking ceiling frequencies. Moreover, in practice, the 
FBI used a binning procedure different than that used 
by Risch and Devlin, and the effect of the actual binning 
procedure on match probabilities was not studied. Other evidence, extensively reviewed elsewhere 
(e.g., NRC 1992, pp. 3-6-3-8, and references given 
there), argues against the applicability of the product 
rule, within very broad ethnic groups. This evidence 
on DNA and protein polymorphisms and genetic dis-
wases, which is also controversial, suggests that, within 
broad ethnic categories, there may exist genetically 
differentiated subgroups with differing allele frequen-
cies, resulting in an association of alleles at the level 
of the broad ethnic group or the whole population.

One alternative to the ceiling principle is the count-
ing method—dividing the frequency of the observed 
genotype in a reference data base by the number of 
individuals in the data base. Risch and Devlin (1992, 
p. 720) object "that such an approach is unnecessarily 
conservative." A second possible approach, suggested 
by Miron L. Straf (personal communication), is to 
apply Bonferroni's inequalities (e.g., see Feller 1968, 
pp. 110 and 142), which make no assumptions of 
independence within or across loci. The practical 
application of Bonferroni's inequalities remains to be 
investigated, as do the general conditions under which 
the ceiling principle fails to be conservative.

Acknowledgments

I thank Miron L. Straf for raising the question whether 
the ceiling principle is always conservative and for suggesting 
the use of the Bonferroni inequalities, and I thank Eric S. 
Lander and Richard C. Lewontin for helpful comments. 
This work was supported in part by U.S. National Science 
Foundation grant BSR87-05047 and the hospitality of Mr. 
and Mrs. William T. Golden.

Joel E. Cohen

Rockefeller University

New York
References


© 1992 by The American Society of Human Genetics. All rights reserved.
0002-9297/92/5105-0028$02.00