Am. J. Hum. Genet. 53:314–323, 1993

# Violations of the Ceiling Principle: Exact Conditions and Statistical Evidence

Jennifer R. Slimowitz*,[1] and Joel E. Cohen[†]

*Department of Mathematics, Duke University, Durham, NC; and [†]Rockefeller University, New York

## Summary

The National Research Council recommended the use of the ceiling principle in forensic applications of DNA testing on the grounds that the ceiling principle was believed to be "conservative," giving estimates greater than or equal to the actual genotype frequencies in the appropriate reference population. We show here that the ceiling principle can fail to be conservative in a population with two subpopulations and two loci, each with two alleles at Hardy-Weinberg equilibrium, if there is some linkage disequilibrium between loci. We also show that the ceiling principle can fail in a population with two subpopulations and a single locus with two alleles if Hardy-Weinberg equilibrium does not hold. We give explicit analytical formulas to describe when the ceiling principle fails. By showing that the ceiling principle is not always mathematically reliable, this analysis gives users of the ceiling principle the responsibility of demonstrating that it is conservative for the particular data with which it is used. Our reanalysis of VNTR data bases of the FBI provides compelling evidence of two-locus associations within three major ethnic groups (Caucasian, black, and Hispanic) in the United States, even though the loci tested are located on different chromosomes. Before the ceiling principle is implemented, more research should be done to determine whether it may be violated in practice.

## Introduction

Forensic scientists analyze portions of DNA to determine whether a body fluid or hair sample left at a crime scene matches that of a suspect (Ballantyne et al. 1989; Office of Technology Assessment 1990; National Research Council Committee on DNA Technology in Forensic Science [hereafter NRC] 1992). Ideally, three to five loci from each of the two DNA samples are tested. If the two samples appear to differ, the suspect is excluded. If the two samples appear to match at the loci tested, it becomes necessary to estimate the frequency of the genotype they represent. If that genotype is common, a match need not indicate strongly that the suspect is the source of the specimen found at the crime scene, while if the genotype is rare, a match provides strong evidence that the suspect is the source of the specimen.

It is generally accepted that the estimated genotype frequency should be conservative, that is, greater than or equal to the actual frequency in the appropriate reference population. A method of estimating genotype frequencies called "the ceiling principle" has recently been strongly recommended on the grounds that it is believed to be conservative (NRC 1992). While the ceiling principle does give an upper bound if the product rule applies within and across loci in all genetically differentiated subpopulations, a recent counterexample shows that the ceiling principle need not always be conservative (Cohen 1992). However, the counterexample is highly artificial.

Here we show that the ceiling principle can fail under simple and natural conditions, in populations with two subpopulations and two alleles per locus. Since the ceiling principle is conservative when the product rule is

Received October 7, 1992; revision received January 14, 1993.

Address for correspondence and reprints: Joel E. Cohen, Rockefeller University, 1230 York Avenue, New York, NY 10021-6399.

1. Present address: Rockefeller University, New York.



0002-9297/93/5302-0003$02.00

valid (Cohen 1992), there are only two possible sources of failure for the ceiling principle: linkage disequilibrium between loci and Hardy-Weinberg disequilibrium within a locus. In the theoretical portion of this work, we give explicit analytical formulas that show when the ceiling principle fails in a population with two loci in linkage disequilibrium, if each locus has two alleles in Hardy-Weinberg equilibrium. We also give explicit formulas that show when the ceiling principle fails in a population with a single locus with two alleles in Hardy-Weinberg disequilibrium.

In the empirical portion of this work, we reexamine a statistical analysis of data bases of the Federal Bureau of Investigation (FBI) that contain VNTR DNA profiles (Weir 1992). The data provide compelling evidence of two-locus associations within three major ethnic groups (Caucasian, black, and Hispanic) in the United States, though the loci tested are located on different chromosomes. The publicly available data do not suffice to determine whether the ceiling principle would in fact be violated, but our secondary data analysis shows that it could potentially be violated. We reach a similar conclusion from the evidence of Hardy-Weinberg disequilibrium in data bases of VNTR DNA profiles, including the FBI's. This empirical evidence suggests that, before the ceiling principle is implemented, more research should be done to determine (*a*) the extent and nature of the deviations from Hardy-Weinberg equilibrium and linkage equilibrium and (*b*) whether the deviations would cause the ceiling principle to be nonconservative in practice.

## Methods of Estimating Genotype Frequencies

Three methods of estimating a genotype frequency have been widely considered in forensic applications of DNA testing. In historical order, they are the product rule, the counting method, and the ceiling principle (NRC 1992). (A fourth method, based on the Laplacian law of succession, proposed by Morton [1992, his equation 2], shares with the ceiling principle the drawback that the estimated genotype frequencies may sum to a number greater than 1.) All methods rely on the existence of a data base of DNA patterns obtained from an appropriate reference population.

The product rule multiplies the frequency of each allele at each of the loci being tested, to arrive at the frequency for the entire pattern (Jeffreys et al. 1985). The product rule assumes statistical independence between alleles, both within loci (Hardy-Weinberg equilibrium) and between loci (linkage equilibrium). Contrary to this assumption, statistical dependence among different alleles may arise from many sources, including the existence of subpopulations with differing allele frequencies, and may cause the product rule not to be conservative (Cohen 1990).

The counting method (Lander 1989) estimates the genotype frequency as the ratio of the number of matching patterns divided by the total number of patterns in the data base. This method assumes that matching patterns from different people have not been expunged from the data base, an assumption that may not be satisfied by the FBI data base (Sullivan et al. 1992).

The NRC (1992) endorses a method of calculation called the "ceiling principle." It states (pp. 3-10–3-11): "(1) For each allele at each locus, determine a *ceiling frequency* 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." To estimate allele frequencies (p. 3-11), "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." Estimates given by this method are "believed to be conservative, given the available data, even if there are correlations among alleles because of population substructure" (p. 3-11). Further (NRC 1992, p. 3-13), "the calculation is fair to suspects, because the estimated probabilities are likely to be conservative in their incriminating power."

Pending the completion of the proposed studies of 15–20 populations, the NRC committee proposed an interim form of the ceiling principle (pp. 3-20–3-22), "provided that population studies have been carried out in at least three major 'races' . . . and that statistical evaluation of Hardy-Weinberg equilibrium and linkage disequilibrium has been carried out . . . and no significant deviations were seen." For the interim ceiling principle, the NRC committee (1992, p. 3-21) recommended that "the calculation should be carried out as follows. For each allele, a modified ceiling frequency should be determined by (1) calculating the 95% upper confidence limit for the allele frequency in each of the existing population samples and (2) using the largest of these values or 10%, whichever is larger. . . . A 10% lower bound is recommended while awaiting the results of the population studies of ethnic groups,

whereas a 5% lower bound will likely be appropriate afterwards. . . . Once the ceiling for each allele is determined, the multiplication rule should be applied. The race of the suspect should be ignored in performing these calculations."

The laudable aim of the ceiling principle, in both its ultimate and its interim forms, is to provide a practical procedure that is assured of estimating genotype frequencies that are at least as large as the unknown real genotype frequencies. To assess whether the ceiling principle meets that aim, we shall follow a common procedure in mathematical analyses of complex problems, by stripping away inessential variations of the method in order to concentrate more clearly on its essential core. First, we shall ignore the proposed ultimate 5% lower bound and the interim 10% lower bound; equivalently, we shall show (by numerical example) that the ceiling principle can fail to be conservative even when all allele frequencies are .1 or larger. Second, as is common in population genetics (e.g., see Morton 1992), we shall assume such large data bases that sampling variability in the estimates of allele frequencies can be ignored; equivalently, we assume that the upper 95% confidence intervals differ negligibly from the actual frequencies. Third, instead of requiring 15–20 populations, as in the final form, or at least three major races, as in the interim form, we shall examine the operation of the ceiling principle with just two subpopulations. This approach to examining the conservativeness of the ceiling principle under simplified conditions parallels testing the thrust of a rocket motor strapped to a launching pad or test bed before it is used under more realistic conditions. Just as rocket technicians are interested in diagnosing the conditions under which the engine may be expected to fail, our analysis aims to determine when, under these simplified conditions, the ceiling principle will not achieve its aim of giving conservative estimates. This analysis leaves open the performance of the ceiling principle under more complex conditions but cautions against assuming that the principle achieves its aims.

## Model with Linkage Disequilibrium

A previous example in which the ceiling principle gives nonconservative estimates considers an unrealistic theoretical population consisting of three subpopulations in which individuals are tested at three loci, each having three alleles, with perfect linkage between loci (Cohen 1992).

Here we show that the ceiling principle can be violated in a simple model with only two subpopulations, two loci with two alleles each, and less than perfect correlation between loci. We give exact conditions for the ceiling principle to give a nonconservative estimate for each genotype.

Consider a population with two subpopulations, labeled "1" and "2." Let $P$ represent the fraction of the population consisting of subpopulation 1, and let $Q$ represent the fraction consisting of subpopulation 2; hence, $P + Q = 1$. Suppose locus 1 has two alleles, A and B, and locus 2 has two alleles, X and Y. In subpopulation 1, let $a_1$ be the frequency of allele A and $b_1$ be the frequency of allele B; similarly, let $x_1$ be the frequency of allele X and $y_1$ the frequency of allele Y. Define $a_2$, $b_2$, $x_2$, and $y_2$ similarly for subpopulation 2. For $i = 1, 2$, $a_i + b_i = 1$, and $x_i + y_i = 1$.

If there were linkage equilibrium between the two loci, the conditional probability $P(\mathrm{A}\,|\,\mathrm{X})$ of a haplotype exhibiting the A allele given that it displayed the X allele would simply be the frequency of A in that population, i.e., $a_i$. However, since we are not assuming linkage equilibrium, let $P(\mathrm{A}\,|\,\mathrm{X})_i = r_i$ in subpopulation $i = 1, 2$, and similarly let $P(\mathrm{A}\,|\,\mathrm{Y})_i = s_i$, where $0 \leqslant r_i, s_i \leqslant 1$. (Thus when, for example, $r_i > a_i$, alleles A and X are positively associated in subpopulation $i$, while when the reverse inequality holds, A and X are negatively associated in subpopulation $i$.) Then

$$\begin{aligned} P(\mathrm{AX})_i &= P(\mathrm{A}\,|\,\mathrm{X})_i P(\mathrm{X})_i = r_i x_i \\ P(\mathrm{AY})_i &= P(\mathrm{A}\,|\,\mathrm{Y})_i P(\mathrm{Y})_i = s_i y_i \\ P(\mathrm{BX})_i &= P(\mathrm{B}\,|\,\mathrm{X})_i P(\mathrm{X})_i = (1 - r_i) x_i \\ P(\mathrm{BY})_i &= P(\mathrm{B}\,|\,\mathrm{Y})_i P(\mathrm{Y})_i = (1 - s_i) y_i \end{aligned} \tag{1}$$

are the haplotype frequencies for subpopulation $i$. Because $P(\mathrm{AX})_i + P(\mathrm{AY})_i = P(\mathrm{A})_i = a_i$, we know that $r_i x_i + s_i y_i = r_i x_i + s_i(1 - x_i) = a_i$ and similarly that $(1 - r_i)x_i + (1 - s_i)(1 - x_i) = b_i$. For each subpopulation there are three independent variables—$r_i$, $s_i$, and $x_i$. The three other variables—$a_i$, $b_i$, and $y_i$—are determined from $a_i = r_i x_i + s_i(1 - x_i)$, $b_i = 1 - a_i$, and $y_i = 1 - x_i$.

Our parameters are easily related to two measures, $D_i$ and $Z_i$, of linkage disequilibrium, described by Crow and Kimura (1970, p. 197): $D_i \equiv P(\mathrm{AX})_i P(\mathrm{BY})_i - P(\mathrm{AY})_i P(\mathrm{BX})_i = x_i y_i (r_i - s_i)$, and $Z_i \equiv P(\mathrm{AX})_i P(\mathrm{BY})_i / [P(\mathrm{AY})_i P(\mathrm{BX})_i] = [r_i(1 - s_i)]/[s_i(1 - r_i)]$.

To calculate the frequency of each genotype, we assume Hardy-Weinberg equilibrium within each subpopulation and multiply haplotype frequencies, multiplying by two for each heterozygous locus. The frequencies

**Table 1**

**Exact Formulas for Genotype Frequencies and Estimated Frequencies Given by the Ceiling Principle in the Model of Two Loci with Hardy-Weinberg Equilibrium and Linkage Disequilibrium, Assuming $x_1 \geq x_2$**

| Genotype | Actual Frequency | Ceiling Estimate: Case 1 $a_1 > a_2$ | Ceiling Estimate: Case 2 $a_2 > a_1$ |
|---|---|---|---|
| AAXX ..... | $Pr_1^2x_1^2 + Qr_2^2x_2^2$ | $a_1^2x_1^2$ | $a_2^2x_1^2$ |
| AAXY ..... | $2Pr_1s_1x_1y_1 + 2Qr_2s_2x_2y_2$ | $2a_1^2x_1y_2$ | $2a_2^2x_1y_2$ |
| AAYY ..... | $Ps_1^2y_1^2 + Qs_2^2y_2^2$ | $a_1^2y_2^2$ | $a_2^2y_2^2$ |
| ABXX ..... | $2Pr_1(1 - r_1)x_1^2 + 2Qr_2(1 - r_2)x_2^2$ | $2a_1b_2x_1^2$ | $2a_2b_1x_1^2$ |
| ABXY ..... | $2Pr_1(1 - s_1)x_1y_1 + 2P(1 - r_1)s_1x_1y_1 + 2Qr_2(1 - s_2)x_2y_2 + 2Q(1 - r_2)s_2x_2y_2$ | $4a_1b_2x_1y_2$ | $4a_2b_1x_1y_2$ |
| ABYY ..... | $2Ps_1(1 - s_1)y_1^2 + 2Qs_2(1 - s_2)y_2^2$ | $2a_1b_2y_2^2$ | $2a_2b_1y_2^2$ |
| BBXX ..... | $P(1 - r_1)^2x_1^2 + Q(1 - r_2)^2x_2^2$ | $b_2^2x_1^2$ | $b_1^2x_1^2$ |
| BBXY ..... | $2P(1 - r_1)(1 - s_1)x_1y_1 + 2Q(1 - r_2)(1 - s_2)x_2y_2$ | $2b_2^2x_1y_2$ | $2b_1^2x_1y_2$ |
| BBYY ..... | $P(1 - s_1)^2y_1^2 + Q(1 - s_2)^2y_2^2$ | $b_2^2y_2^2$ | $b_1^2y_2^2$ |

for the entire population are listed in the second column of table 1.

To calculate the frequencies given by the ceiling principle (apart from the recommended lower bounds of 5% or 10%), we multiply the higher frequency for each allele from the two subpopulations, with a factor of two for single heterozygotes and a factor of four for double heterozygotes. Without loss of generality, we label the two subpopulations so that $x_1 \geq x_2$, and (since $x_i + y_i = 1$) we have $y_2 \geq y_1$. For ease in calculation, we consider two cases. In case 1, $a_1 > a_2$ and thus $b_2 > b_1$. In case 2, $a_2 > a_1$ and $b_1 > b_2$. The ceiling frequencies for each genotype for each case are listed in columns 3 and 4 of table 1.

For the ceiling principle to be nonconservative, the ceiling estimate must be strictly lower than the actual frequency. The exact conditions for each genotype may be read off from the corresponding line in table 1. Thus, for genotype AAXX, the ceiling estimate is nonconservative if $Pr_1^2x_1^2 + Qr_2^2x_2^2 > a_1^2x_1^2$ when $a_1 > a_2$ (case 1). As R. C. Lewontin (personal communication) pointed out, this is not an onerous requirement. For example, if $r_1 = r_2$ and $x_1 = x_2$, then since $P + Q = 1$, the ceiling principle will be violated for genotype AAXX if $r_1 > a_1 > a_2$ or, in other words, if the frequency of allele A in subpopulation 1 exceeds its frequency in subpopulation 2 and also, within subpopulation 1, allele A is associated with allele X more than it is with allele Y. The other inequalities derived from table 1 have equally simple interpretations in special cases. These interpretations show that the ceiling principle can easily fail to be conservative for an individual genotype.

Table 1 makes it easy to calculate exactly by how much the ceiling principle fails. For example, the ratio of the ceiling bound to the actual probability for the genotype AAXX in case 1 ($a_1 > a_2$) equals $(a_1^2x_1^2)/(Pr_1^2x_1^2 + Qr_2^2x_2^2)$. In the special case considered in the preceding paragraph, this ratio simplifies to $(a_1/r_1)^2$. As might be expected, the ceiling principle would fail most (i.e., the ratio would be smallest) when the A allele is rare but the conditional frequency $r_1$ of A, given the X allele, is large in subpopulation 1.

## Hypothetical Numerical Examples

To illustrate a theoretical situation where most individuals have genotypes that are more frequent than the ceiling principle would estimate, suppose that

$$\begin{aligned} &a_1 = x_1 = .90, \quad a_2 = x_2 = .82, \quad r_1 = .99, \quad s_1 = .09, \\ &b_1 = y_1 = .10, \quad b_2 = y_2 = .18, \quad r_2 = .99, \\ &s_2 = .045\overline{5} = .045555\ldots, \\ &P = Q = \tfrac{1}{2}. \end{aligned} \tag{2}$$

Since here $a_1 > a_2$, this example illustrates case 1. No allelic frequency in this hypothetical example is less

**Table 2**

**Hypothetical Genotype Frequencies Compared with Estimates from the Ceiling Principle in the Model of Two Loci with Hardy-Weinberg Equilibrium and Linkage Disequilibrium, Assuming Parameter Values Given in (2)**

| Genotype | Actual Frequency | Ceiling Estimate |
|---|---|---|
| AAXX ..... | .72645012 | .6561[a] |
| AAXY ..... | .01467576 | .26244 |
| AAYY ..... | .00007412 | .026244 |
| ABXX ..... | .01467576 | .26244 |
| ABXY ..... | .22069648 | .104976[a] |
| ABYY ..... | .00222776 | .0104976 |
| BBXX ..... | .00007412 | .026244 |
| BBXY ..... | .00222776 | .0104976 |
| BBYY ..... | .01889812 | .00104976[a] |

[a] Ceiling estimate violation.

than .1, so no lower bound adjustment would be required. Then

$$P(\mathrm{AX})_1 = P(\mathrm{A}\,|\,\mathrm{X})_1 P(\mathrm{X})_1 = r_1 x_1 = (.99)(.90) = .891$$

$$P(\mathrm{AY})_1 = P(\mathrm{A}\,|\,\mathrm{Y})_1 P(\mathrm{Y})_1 = s_1 y_1 = (.09)(.10) = .009$$

$$P(\mathrm{BX})_1 = P(\mathrm{B}\,|\,\mathrm{X})_1 P(\mathrm{X})_1 = (1 - r_1)x_1 = (.01)(.90) = .009$$

$$P(\mathrm{BY})_1 = P(\mathrm{B}\,|\,\mathrm{Y})_1 P(\mathrm{Y})_1 = (1 - s_1)y_1 = (.91)(.10) = .091 \qquad (3)$$

and similarly, $P(\mathrm{AX})_2 = .8118$, $P(\mathrm{AY})_2 = .0082$, $P(\mathrm{BX})_2 = .0082$, and $P(\mathrm{BY})_2 = .1718$.

Table 2 lists the actual, as well as ceiling, frequencies for this scenario. The ceiling principle is violated for three genotypes—AAXX, ABXY, and BBYY—with a combined frequency greater than 96%. Whether the ceiling principle would give nonconservative estimates so frequently in practice remains to be determined.

Numerical simulations show that it is very easy to find hypothetical examples that violate the ceiling principle. We simulated numerically case 1 with $P = Q = \frac{1}{2}$ on a computer, specifically to investigate whether the ceiling principle could fail to be conservative for all possible genotypes, whether failure for one genotype implies failure for another genotype, and whether validity of the ceiling principle for one genotype implies validity for another genotype. For each trial, random numbers uniformly distributed between 0 and 1 were chosen for $r_1$, $r_2$, $s_1$, $s_2$, $x_1$, and $x_2$. When $x_2 > x_1$, the two numbers were switched so that $x_1 > x_2$. Then $a_1$ and $a_2$ were calculated from $a_i = r_i x_i + s_i(1 - x_i)$; whenever $a_2 > a_1$, new random numbers were chosen for $r_i$, $s_i$, and $x_i$ until $a_1 > a_2$. The remaining variables were determined by $y_1 = 1 - x_1$; $y_2 = 1 - x_2$; $b_1 = 1 - a_1$; and $b_2 = 1 - a_2$. Then the actual and the ceiling frequencies were calculated for each genotype for each trial. This numerical simulation is not intended necessarily to reflect the distribution of allelic frequencies and associations in real populations.

Based on 10,000 trials, the ceiling principle was found to be nonconservative for at least one genotype in 0.2633 (±0.0044 SD) of the trials. The ceiling principle failed to be conservative for each genotype at least once. Violation of the ceiling principle for any one genotype did not invariably imply violation for any other genotype. Adherence to the ceiling principle for any one genotype did not invariably imply adherence for any other.

## Model with Hardy-Weinberg Disequilibrium

The ceiling principle can also be violated in a population with two subpopulations and one locus with two alleles if Hardy-Weinberg equilibrium does not hold. The analysis is essentially identical to the previous analysis after a simple reinterpretation of the symbols. Now interpret A and B as the maternally inherited alleles at a single locus; interpret X as the paternally inherited A allele at the same locus and Y as the paternally inherited B allele at the same locus. The symbols previously interpreted as haplotypes at two loci are now interpreted as genotypes at a single locus; thus AX is now interpreted as the AA homozygote at this locus, BY as the BB homozygote, and both AY and BX as the heterozygous genotypes AB at the given locus. From (1), we immediately have the genotype frequencies for subpopulation $i$

$$P(\mathrm{AA})_i = r_i x_i$$

$$P(\mathrm{AB})_i = s_i y_i + (1 - r_i)x_i \qquad (4)$$

$$P(\mathrm{BB})_i = (1 - s_i)y_i\,.$$

Table 3 lists the formulas for the actual genotype frequencies and the ceiling estimates for this model. As before, the exact conditions for the ceiling principle to be nonconservative may be read off from the line corresponding to each genotype in table 3. Thus, for the homozygote AA, the ceiling estimate is nonconservative if $Pr_1 x_1 + Qr_2 x_2 > a_1 x_1$. As before, this is not an onerous requirement. For example, if $r_1 = r_2$ and $x_1 = x_2$, then, since $P + Q = 1$, the ceiling principle will be violated for genotype AA if $r_1 > a_1 > a_2$ or, in other words, if the frequency of allele A in subpopulation 1

**Table 3**

**Exact Formulas for Genotype Frequencies and Estimated Frequencies, Given by the Ceiling Principle, in the Model of One Locus in Hardy-Weinberg Disequilibrium, Assuming $x_1 \geq x_2$**

| | | Ceiling Estimate | |
|---|---|---|---|
| Genotype | Actual Frequency | Case 1 $a_1 > a_2$ | Case 2 $a_2 > a_1$ |
| AA ..... | $Pr_1x_1 + Qr_2x_2$ | $a_1x_1$ | $a_2x_1$ |
| AB ..... | $Ps_1y_1 + Qs_2y_2 + P(1 - r_1)x_1 + Q(1 - r_2)x_2$ | $a_1y_2 + b_2x_1$ | $a_2y_2 + b_1x_1$ |
| BB ..... | $P(1 - s_1)y_1 + Q(1 - s_2)y_2$ | $b_2y_2$ | $b_1y_2$ |

exceeds its frequency in subpopulation 2 and also, within subpopulation 1, the maternal allele A is associated with the paternal allele A more than the maternal allele A is associated with the paternal allele B. The other inequalities derived from table 1 have equally simple interpretations. (In a special case, with different notation and slightly different assumptions, Bruce Weir [personal communication] independently showed that Hardy-Weinberg disequilibrium could lead to violations of the ceiling principle for heterozygotes.)

Using the hypothetical parameter values in (2) in the formulas in (4) gives, according to (3),

$$P(\mathrm{AA})_1 = .891$$
$$P(\mathrm{AB})_1 = .018 \qquad (5)$$
$$P(\mathrm{BB})_1 = .091$$

and, similarly, $P(\mathrm{AA})_2 = .8118$, $P(\mathrm{AB})_2 = .0164$, and $P(\mathrm{BB})_2 = .1718$. Table 4 lists the actual and ceiling frequencies for this hypothetical scenario. Here the ceiling principle is nonconservative for both homozygous genotypes, which constitute more than 98% of the population. Once again, whether the ceiling principle would give nonconservative estimates so frequently in practice remains to be determined.

**Table 4**

**Hypothetical Genotype Frequencies Compared with Estimates from the Ceiling Principle in the Model of One Locus in Hardy-Weinberg Disequilibrium, Assuming Parameter Values Given in (2)**

| Genotype | Actual Frequency | Ceiling Estimate |
|---|---|---|
| AA ..... | .8514 | .81[a] |
| AB ..... | .0172 | .324 |
| BB ..... | .1314 | .0324[a] |

[a] Ceiling estimate violation.

## Evidence of Two-Locus Associations within U.S. Ethnic Groups

The ceiling principle could be nonconservative in forensic practice only if there are associations within or among loci. We next review briefly some evidence and arguments concerning the existence and extent of possible linkage disequilibrium and Hardy-Weinberg disequilibrium in the United States.

Weir (1992) carefully investigated three FBI data bases (B5 for blacks, C4 for Caucasians, and H4 for Hispanics) for evidence of associations within and among loci. The data bases contain VNTR profiles at six loci (numbered 1, 2, 4, 10, 14, and 17; see table 5 for key to loci). No two loci lie on the same chromosome. According to Weir (1992, p. 881), the FBI uses four of the six loci in forensic work (loci 1, 2, 4, and 17).

Using fixed bins, as the FBI does in forensic practice, and the bootstrap resampling method to simulate the distribution of a likelihood-ratio statistic, Weir (1992, p. 884, his table 10) estimated empirical significance levels for all possible two-locus associations. "Note that the test is not a test for linkage disequilibrium—significant results would obtain if any subset of the four bins in a two-locus genotype had dependent frequencies" (Weir 1992, p. 883). If there were no two-locus associations, the empirical significance levels should be uniformly distributed between 0 and 1. The observed distribution of empirical significance levels appears to differ from the expected uniform distribution. For example, among the 144 reported empirical significance levels, 15 were equal to .01 (whereas only $1.4 = 144 \times .01$ would be expected to be .01 or smaller), and 47 fell in the range .02–.05 (whereas only $5.8 = 144 \times .04$

would be expected in the range .02–.05). The largest reported empirical significance level was .33.

Assigning statistical significance to these apparent deviations from a uniform distribution is difficult because the empirical significance levels are not all pairwise independent. For example, the test for association among all Caucasians at loci 1,2 is not independent of the test among Florida Caucasians or the test among Texas Caucasians, because both states are included in "all" Caucasians. In spite of this dependence, the marginal distribution of empirical significance levels should be uniform if there are no two-locus associations.

A more sensitive and exact way to investigate possible two-locus associations consists of combining independent tests by using Fisher's (1970, pp. 99–100) method. If $p_j$ is the significance level of test $j$, $j = 1, \ldots, J$, and if all $J$ tests are statistically independent, then

$$\chi^2 = -2 \sum_{j=1}^{J} \log(p_j)$$

has the distribution of $\chi^2$ with $2J$ df.

Since six loci were tested, three independent tests of two-locus association can be combined for each different combination of ethnic group (e.g., Caucasian) and specific state (e.g., Florida). For example, the test for association of loci 1 and 2 is practically independent of the test for association of loci 4 and 10, and both are practically independent of the test for association of loci 14 and 17. (There may be a correlation between disequilibrium at two distinct pairs of loci due to statistical sampling [Hill and Weir 1988], but "the effect is small" [Bruce Weir, personal communication].) In table 5, the combined test for association in these three pairs of loci is denoted by 1,2+4,10+14,17. For this trio of pairwise tests, the $\chi^2$ statistic with 6 df for Caucasians in Florida is shown in table 5 as 16.670. According to the critical values given at the bottom of table 5, this combined test gives evidence of two-locus association at a significance level between .01 and .02.

Table 5 shows the $\chi^2$ statistic for all 15 possible trios of two-locus tests of association. Among Florida Caucasians, all 15 statistics are significant at the .05 level, while among California Caucasians, none is significant. Texas Caucasians are intermediate. Among Florida and Texas blacks, all 30 statistics are significant at the .01 level, and a majority of the Texas black statistics are significant at the .001 level. Among California blacks, only one statistic is significant at the .05 level. Among Florida Hispanics, 11 of 15 statistics are significant at the .05 level, while among Texas Hispanics, all statistics are significant at the .05 level, including 8 that are significant at the .01 level. In summary, California Caucasians and blacks show little suggestion of two-locus associations, Texas Caucasians show weak or variable evidence of two-locus associations, and the remaining ethnic × state groups show consistent and strong evidence of two-locus associations.

Bruce Weir (personal communication) suggested that the significant values in his tests for two-locus associations (Weir 1992, his table 10) may largely reflect significant single-locus associations reported in his table 6 (Weir 1992, p. 880). To exclude any associations due to the uncertainty about whether single-band patterns represent homozygotes or heterozygotes, Weir (1992, pp. 883–884) tested each ethnic data base (i.e., Caucasian, black, and Hispanic, not separated by state) for associations between loci, by using only double heterozygotes for each pair of loci. He found no empirical significance level smaller than .05. Weir (personal communication) provided us with the 45 empirical significance levels based on double heterozygotes only (3 ethnic data bases ×15 pairwise comparisons for the six loci). We carried out an analysis parallel to the analysis used to produce table 5. Among the 45 $\chi^2$ statistics for all possible trios of two-locus tests of association, only one was significant at the .05 level, and none was significant at the .02 level. These results confirm Weir's conclusion that the data on double heterozygotes only provide no evidence of pairwise associations among the loci analyzed.

Tests of association between a single-band pattern at one locus and a heterozygous genotype at another locus have not yet been reported, so it is not yet entirely clear whether the two-locus associations in the FBI data bases are entirely due to the single-locus Hardy-Weinberg disequilibrium or whether they may be partially due to linkage disequilibrium between loci.

The NRC (1992) did not propose that the ceiling principle be applied to data on double heterozygotes only, though it did endorse the procedure of doubling the single-band frequencies (NRC 1992, p. 3-5). (Morton [1992, p. 2557] appears not to favor such a doubling.) The apparent absence of linkage among double heterozygotes does not justify the use of the product rule or the ceiling principle with data bases that include single-band patterns. Nor (in response to a referee's question) does the apparent absence of linkage for double heterozygotes necessarily imply the validity of the product rule for triple or quadruple heterozygotes, since, in theory, the combined alleles at two loci could easily affect the conditional probability of alleles at a

**Table 5**

**Tests for Two-Locus Associations in FBI Data Bases, Combining Triples of Independent Two-Locus Tests**

| | Caucasian | | | Black | | | Hispanic | |
|---|---|---|---|---|---|---|---|---|
| Loci[a] | Florida | Texas | California | Florida | Texas | California | Florida | Texas |
| 1,2+4,10+14,17 ..... | 16.670 | 12.542 | 4.605 | 20.464 | 25.434 | 5.991 | 14.450 | 17.691 |
| 1,4+2,10+14,17 ..... | 18.502 | 12.154 | 4.605 | 23.026 | 24.048 | 5.991 | 14.004 | 17.691 |
| 1,10+2,4+14,17 ..... | 18.056 | 12.332 | 4.605 | 22.661 | 23.237 | 15.202 | 13.665 | 16.184 |
| 1,14+2,4+10,17 ..... | 15.048 | 11.540 | ... | 20.464 | 22.215 | 9.210 | 14.375 | 15.373 |
| 1,17+2,4+10,14 ..... | 15.413 | 11.540 | ... | 19.889 | 23.237 | 9.210 | 12.413 | 17.806 |
| 1,2+4,14+10,17 ..... | 16.937 | 12.520 | ... | 18.918 | 24.412 | 6.438 | 15.334 | 14.683 |
| 1,2+4,17+10,14 ..... | 15.944 | 14.375 | ... | 21.275 | 24.858 | 4.605 | 11.349 | 16.995 |
| 1,4+2,14+10,17 ..... | 16.145 | 11.557 | 4.605 | 19.442 | 21.640 | 7.824 | 14.122 | 14.473 |
| 1,4+2,17+10,14 ..... | 16.479 | 12.943 | 2.939 | 20.829 | 20.829 | 5.627 | 14.523 | 13.086 |
| 1,10+2,14+4,17 ..... | 16.230 | 14.067 | 4.605 | 22.661 | 23.472 | 12.429 | 11.046 | 15.162 |
| 1,10+2,17+4,14 ..... | 18.943 | 13.758 | 2.939 | 21.115 | 25.434 | 12.065 | 13.471 | 15.859 |
| 1,14+2,10+4,17 ..... | 15.579 | 13.872 | ... | 24.048 | 24.858 | 4.605 | 11.638 | 17.570 |
| 1,14+2,17+4,10 ..... | 15.668 | 12.989 | 2.939 | 20.464 | 25.434 | 5.627 | 13.297 | 18.056 |
| 1,17+2,10+4,14 ..... | 16.937 | 12.018 | ... | 21.115 | 25.434 | 6.438 | 13.662 | 17.691 |
| 1,17+2,14+4,10 ..... | 14.312 | 11.442 | 4.605 | 19.078 | 24.048 | 7.824 | 12.896 | 17.481 |

NOTE.—The original empirical significance levels of tests for individual two-locus associations are given by Weir (1992, p. 884). Under the null hypothesis of no two-locus associations, the entries in the table should have the distribution of $\chi^2$ with 6 df, for which the critical values are 12.592 for $P = .05$, 15.033 for $P = .02$, 16.812 for $P = .01$, and 22.475 for $P = .001$. Thus, for Caucasians in Florida, the entry of 16.670 shows that the associations of loci 1 and 2, loci 4 and 10, and loci 14 and 17, in combination, would occur by chance alone with a probability between .01 and .02.

[a] Locus 1 = D1S7; locus 2 = D2S44; locus 4 = D4S139; locus 10 = D10S28; locus 14 = D14S13; and locus 17 = D17S79.

third or fourth locus; the question must be settled empirically.

## Evidence of Hardy-Weinberg Disequilibrium within U.S. Ethnic Groups

Odelberg et al. (1989) report significant deviations from Hardy-Weinberg equilibrium at three of the eight VNTR loci examined in samples of 78–151 Utah Caucasians. The exact significance values they quote should not be taken literally because they appear to make no allowance for multiple tests; nevertheless, three significant results at the nominal .05 significance level is substantially larger than the number ($8 \times .05 = .4$) of significant tests at the .05 level expected by chance alone, assuming Hardy-Weinberg equilibrium. For the three loci (D2S44, D14S13, and D1S74) with reportedly significant deviations from Hardy-Weinberg equilibrium, homozygotes are two to three times more frequent than expected from Hardy-Weinberg equilibrium. Devlin et al. (1990), Weir (1992), and Chakraborty et al. (1992) also report substantial, and often statistically significant, deviations of raw counts of banding phenotypes from Hardy-Weinberg equilibrium in FBI data.

It is not possible to estimate the parameters of the above model for two subpopulations with Hardy-Weinberg disequilibrium on the basis of the information reported by Odelberg et al. (1989), Devlin et al. (1990), Weir (1992), and Chakraborty et al. (1992), so it is not clear whether the reported disequilibrium would suffice to cause a violation of the ceiling principle. Were access to the FBI data unrestricted (Anderson 1992), the appropriateness of the ceiling principle for these data could readily be tested.

Whether the observed excess of homozygotes and observed deficiency of heterozygotes (relative to Hardy-Weinberg equilibrium) reflect population substructure or artifacts of biochemical technique is a controverted question. Devlin et al. (1990, p. 1416) argue that the apparent excess of homozygotes "is not necessarily real because many heterozygotes with similar allele sizes are misclassified as homozygotes." Cohen et al. (1991) raise extensive questions concerning the analysis and conclusions of Devlin et al. (1990).

Chakraborty et al. (1992) propose that the observed excess of homozygotes "is caused by the inability to detect extremely small-sized alleles (called 'non-detectable' alleles). . . . If 'non-detectable' alleles are the pre-

dominant source of observed heterozygote deficiency, then gene-count estimates of all detectable alleles provide enough cushion to prescribe such upper bounds" on the true frequency of each observable genotype, provided that the frequency of each single-band phenotype is estimated by twice the observed frequency of the corresponding allele (see Chakraborty et al. 1992, p. 53). Devlin and Risch (1992) reach a similar conclusion that allowance for null alleles obviates deviations from Hardy-Weinberg equilibria, on the basis of a comparison of FBI and Lifecodes data bases.

Unfortunately, Chakraborty et al. (1992) omit an easy but crucial empirical test of this elegant and ingenious proposal. With the data at their disposal, they could compare the observed frequency of every "homozygous" (i.e., single-banded) and "heterozygous" (i.e., double-banded) phenotype with the corresponding estimated upper bound (in their notation) $2\hat{p}_i$ and $2\hat{p}_i\hat{p}_j$ derived from their model. If "non-detectable" alleles suffice to explain the observed heterozygote deficiency, then (apart from occasional sampling error due to small sample sizes) the observed frequencies should be smaller than the corresponding estimated upper bounds. The proposed multilocus extension of their approach using the product rule (Chakraborty et al. 1992, p. 54) should be tested similarly by comparing observed multilocus phenotype frequencies with their estimated upper bounds (which differ from those of the ceiling principle). Until such empirical analysis is reported in detail, their proposed analysis remains conjectural. As Chakraborty et al. (1992, p. 55) candidly concede, "It is true that in principle the possibility of heterogeneity within a population (population substructuring) cannot be distinguished from the scenario presented here." The test just proposed could either exclude or be consistent with the scenario presented by Chakraborty et al. (1992).

Apparently independently of Devlin et al. (1990) and Chakraborty et al. (1992), Morton (1992) analyzes the combination of population substructuring (or endogamy within subpopulations), binning, and nondetectable, or null, alleles. He shows that they all contribute to an "apparent inbreeding coefficient" that can be estimated from the observed frequency distribution of genotypes or banding phenotypes. Separating the components of the apparent inbreeding coefficient into the proportions due to population substructuring, binning, and null alleles would appear to require that the distributions of banding phenotypes be measured separately for relatively endogamous subpopulations, as called for by the NRC (1992, p. 3-20), and that the biochemical technique be refined or replaced so that merged and null alleles can be measured or eliminated altogether.

## Discussion

We have shown that the ceiling principle recommended by the NRC (1992) can fail to be conservative in a simplified hypothetical population with two subpopulations and two alleles per locus. Exact conditions for failure of the ceiling principle are given for a single locus in Hardy-Weinberg disequilibrium and for a pair of loci, each in Hardy-Weinberg equilibrium, with linkage disequilibrium between them. Hardy-Weinberg disequilibrium and linkage disequilibrium in combination could amplify the failures of the ceiling principle. The conditions for the ceiling principle to fail need not be onerous, and hypothetical numerical examples of failures are plentiful.

A necessary condition for failure of the ceiling principle is the existence of associations within or between loci. The NRC report mentions that the interim version of the ceiling principle should only be put into effect after possible linkage disequilibrium and Hardy-Weinberg disequilibrium have been checked. This is an important requirement. FBI data publicly available without restriction do not presently make it possible to estimate whether the associations among loci are sufficient to cause the ceiling principle to fail in practice. However, our analysis of Weir's (1992) empirical significance levels based on FBI data gives strong evidence of two-locus associations for certain ethnic groups in certain states (table 5). It is not yet clear whether the two-locus associations are due to deviations from Hardy-Weinberg equilibrium, when single-banded patterns are regarded as homozygotes. Nevertheless, the existence of two-locus associations suggests that, in those ethnic groups and states, the validity of the ceiling principle should be established before it is implemented in either its interim or its final form.

Current FBI VNTR data on double heterozygotes only, as analyzed by Weir (1992, and personal communication), give no evidence for two-locus associations. Sample sizes for such VNTR profiles have not been made public, so this conclusion could be a result of low power, and larger samples could change it. However, at the moment, the product rule appears to apply to the limited case of double heterozygotes only.

Our results emphasize the need for empirical study of allelic associations within and between loci in defined subpopulations, as called for by the NRC (1992, p. 3-20). Our results also emphasize the need for theoreti-

cal study of methods of estimating upper bounds on genotype frequencies, which are robust to allelic associations within and across loci. Miron L. Straf's suggestion to use the Bonferroni inequalities (see Cohen 1992, p. 1167) deserves further exploration.

## Acknowledgments

Richard C. Lewontin suggested finding explicit formulas for the conditions under which the ceiling principle is nonconservative. Bruce S. Weir suggested analyzing failures of the ceiling principle at a single locus in Hardy-Weinberg disequilibrium and also generously made available to us the results of his unpublished analysis of the FBI's double heterozygote data. Lewontin, Weir, and Danny J. Boggs and two anonymous referees commented helpfully on previous drafts. This collaboration was supported in part by the SURF Program of Rockefeller University and in part by U.S. National Science Foundation grant BSR 87-05047. J.E.C. is grateful for the hospitality of Mr. and Mrs. William T. Golden.

## References

Anderson C (1992) FBI attaches strings to its DNA database. Nature 357:618

Ballantyne J, Sensabaugh G, Witkowski JA (1989) DNA technology and forensic science. Banbury report 32. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 56:45–57

Cohen JE (1990) DNA fingerprinting for forensic identification: potential effects on data interpretation of subpopulation heterogeneity and band number variability. Am J Hum Genet 46:358–368

——— (1992) The ceiling principle is not always conservative in assigning genotype frequencies for forensic DNA testing. Am J Hum Genet 51:1165–1168

Cohen JE, Lynch M, Taylor CE, Green P, Lander ES, Devlin B, Risch N, et al (1991) Forensic DNA tests and Hardy-Weinberg equilibrium. Science 253:1037–1041

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York

Devlin B, Risch N (1992) A note on Hardy-Weinberg equilibrium of VNTR data by using the Federal Bureau of Investigation's fixed-bin method. Am J Hum Genet 51:549–553

Devlin B, Risch N, Roeder K (1990) No excess of homozygosity at loci used for DNA fingerprinting. Science 249:1416–1420

Fisher RA (1970) Statistical methods for research workers, 14th ed. Oliver & Boyd, Edinburgh

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific "fingerprints" of human DNA. Nature 316:76–79

Lander ES (1989) DNA fingerprinting on trial. Nature 339:501–505

Morton NE (1992) Genetic structure of forensic populations. Proc Natl Acad Sci USA 89:2556–2560

National Research Council Committee on DNA Technology in Forensic Science (1992) DNA technology in forensic science. National Academy Press, Washington DC

Odelberg SJ, Plaetke R, Eldridge JR, Ballard L, O'Connell P, Nakamura Y, Leppert M, et al (1989) Characterization of eight VNTR loci by agarose gel electrophoresis. Genomics 5:915–924

Office of Technology Assessment (1990) Genetic witness: forensic uses of DNA tests. Rep no OTA-BA-438. U.S. Government Printing Office, Washington, DC

Sullivan PJ, Risch N, Devlin B, Budowle B (1992) DNA fingerprint matches. Science 256:1743–1746

Weir BS (1992) Independence of VNTR alleles defined as fixed bins. Genetics 130:873–887