John M. Hartmann,[1] *B.S., M.B.A.; Bruce T. Houlihan,*[1] *B.S.; Robert S. Keister,*[1] *B.S.; and Edward L. Buse,*[1] *B.S., M.A.*

The Effect of Ethnic and Racial Population Substructuring on the Estimation of Multi-Locus Fixed-Bin VNTR RFLP Genotype Probabilities

REFERENCE: Hartmann JM, Houlihan BT, Keister RS, Buse EL. The effect of ethnic and racial population substructuring on the estimation of multi-locus fixed-bin VNTR RFLP genotype probabilities. J Forensic Sci 1997;42(2):232–240.

ABSTRACT: Four East Asian ethnic and four racial VNTR RFLP Southern California databases were used to determine the impact of population substructure on fixed-bin genotype probability estimates. Two calculations were used for population-level probabilities: Stratified sampling, which takes substructuring into account, and pooling, which ignores it. Using 1000 four-locus genotypes, the relative difference between probabilities calculated with the stratified and the pooled methods did not exceed one order of magnitude out of about 11 orders of magnitude for East-Asian racial genotypes. Pooled estimates differed from cognate ethnic values by less than one order of magnitude out of about six. These findings suggest substructuring of races by major ethnic groups does not lead to large errors. Racial genotype probability variances were on average about twice the ethnic variances. Multi-racial total population probabilities calculated by the pooled and stratified methods differed by less than one order of magnitude out of five.

KEYWORDS: forensic science, DNA typing, restriction fragment length polymorphism, variable number of tandem reports, genotype, population genetics, population substructure

There has been an intense debate over a number of issues concerning forensic DNA profiling. These have been thoroughly reviewed by Weir (1), so they will not be repeated here. Among these issues, the application of VNTR RFLP analyses to forensic casework has been the subject of criticism that population substructure compromises the accuracy of genotype probability estimates. A number of analyses, both highly theoretical and empirical have been published, along with about as many proposals for calculating the rarity of VNTR RFLP profiles. This activity has resulted in the successive impaneling of two National Research Council (NRC) committees to examine the issues.

The NRC's first Committee on forensic DNA typing (2, NRC I) used a simple two-allele system to explain the effect of substructuring on genotype probability estimates and proposed what it termed a "ceiling" calculation to compensate for any substructure derived error (2). Cohen (3) and Slimowitz and Cohen (4) used a three-locus, three-allele system to demonstrate that even this ceiling calculation might underestimate a genotype probability. However, such contrived examples are almost useless for illuminating the nature of the substructuring problem and do nothing to determine its actual significance for real populations and analytical systems used by forensic laboratories (5).

Morton (6) suggested an upper bound on the expected error due to subpopulation derived disequilibria obtained from protein- and antigen-based kinship coefficients. However, such bounds may be excessive for the fixed-bin method of Budowle and Monson (7), which does not use actual allele frequencies. Their method calls for performing a two-step smoothing of the allele distributions, which tends to diminish observed differences between subpopulation distributions, and reduces the impact of substructuring. Weir (8) refers to fixed-bin genotypes as binotypes, to emphasize that bins are not the same as alleles, and each is likely to contain several different VNTR fragment lengths in addition to unknown sequence variants.

Crow and Denniston (9), Balding and Nichols (10), and Roeder (11) took approaches similar to Morton's in their use of some estimate of population inbreeding or homogeneity, such as Wright's F_{ST}, to compensate for possible substructure. However, there are very few VNTR RFLP derived values for these parameters. The second NRC committee to issue a report on forensic DNA typing (12) adopted Balding and Nichol's method for discrete allele systems but preferred a different approach for VNTR RFLP systems. The NRC II's analysis indicated that the compensation could be accomplished by substituting $2p$ for p^2 when calculating homozygote frequencies.

On a more pragmatic level, analyses of multi-racial and/or multi-ethnic databases have been published, which compared genotype probabilities calculated using cognate and non cognate databases. Weir (13) addressed the substructure issue by comparing multilocus genotype probabilities obtained using different ethnic and racial samples from the FBI's VNTR RFLP database together with Budowle and Monson's fixed-bin method (7). This approach has been amplified in the Federal Bureau of Investigation's (14) massive worldwide study of VNTR RFLP population data and similar but more limited studies (15–19). Such comparisons demonstrate that use of the wrong race or ethnic group database does not yield substantially different genotype probability estimates. However, the typical forensic case involves estimating an ethnic or racial probability from a possibly ethnically-substructured racial database, and not from a different race altogether. Hence, these studies

[1]Senior forensic scientists and forensic scientist III, respectively, Orange County Sheriff-Coroner Department, Santa Ana, CA.

Forensic Science Services, Orange County Sheriff-Coroner Department, P.O. Box 449, Santa Ana, CA 92703.

Received 15 June 1993; and in revised form 15 Nov. 1993, 21 Aug. 1995, and 8 Aug. 1996; accepted 9 Aug. 1996.

do not directly address the central issues in the debate over population substructuring. These concern the degree of error generated by substitution of a racial database for that of one, or a combination, of its constituent ethnic subpopulations, or by use of a population database where subpopulations exists. In this study, we directly compare genotype probability estimates derived from ethnic and racial subpopulations with those which would be obtained from population estimates.

Elsewhere, we have described a VNTR RFLP database consisting of well defined ethnic subpopulations (20). In that study, four racial samples were obtained: black, hispanic, white, and East Asian. The last sample was further stratified into Chinese, Japanese, Korean, and Vietnamese ethnic groups. East Asians were chosen because they provide a clear case of substructuring of a race by major ethnic groups in the United States (21,22). The Southern California East Asian racial population consists largely of recent immigrants and there has been little of the extensive inter-ethnic and even inter-racial marrying found among the other races.

We examined three commonly used methods of fragment-size distribution comparison using multivariate techniques to show that, for East Asians, within-race variation of fragment sizes was not greater than between-race variation. This led us to hypothesize that (1) within-race genotype probability variance could be conservatively estimated using between-race variance, and (2) the effect of ignoring ethnic substructuring by substituting pooled fixed-bin frequencies for stratified subpopulation frequencies at both racial and total population levels is small.

Methods

Sampling

Blood samples were obtained from volunteers of four Southern California racial groups: Afro-American (black), European Caucasian (white), Southwestern Hispanic (hispanic), and East Asian. The last group was further stratified into Chinese, Japanese, Korean, and Vietnamese. The above races and ethnic groups comprise the large majorities of the region's total and East Asian populations. Details are described elsewhere (20). The sample sizes for each ethnic/racial group and locus (Table 1) were not optimal in the sense of the sample sizes not being proportional to their ethnic/racial fractions of the Southern California population (Table 2), which were derived from the 1990 U.S. Census (23), by normalizing after excluding relatively minor subpopulations.

RFLP Analysis

The samples were subjected to RFLP analysis according to the method of Budowle and Baechtel (24), with minor modifications (25), and the fragments sized as described by Monson and Budowle (26).

TABLE 1—*Number of fragments.**

Race or Ethnic Group	Locus			
	D1S7	D2S44	D4S139	D10S28
Chinese	218	216	206	240
Japanese	274	252	250	274
Korean	200	198	186	200
Vietnamese	426	426	430	386
Black	404	426	420	444
White	424	430	434	430
SW Hispanic	514	496	486	512

*Single-band patterns counted as two fragments.

TABLE 2—*Normalized East Asian and racial population percentages (Southern California).*

Black	8%
Hispanic	33%
White	53%
East Asian	6%
Chinese	37%
Japanese	21%
Korean	23%
Vietnamese	19%

Fixed-Bin Distributions

Following Budowle and Monson (7), raw band sizes in base pairs (bp) were aggregated into 31 bins. The 31 bins resulting from this first stage were subjected to a further multi-step aggregation to eliminate classes (rebins) having fewer than five observations. Hence, the product of their rebinning algorithm is directly dependent upon sample size.

Balanced Fixed-Bin Distributions

Because of the size dependence noted above, rebin distributions derived from the same bin frequency distribution but from samples of different sizes can yield different rebin (final bin) frequencies. The larger database will produce, on average, lower rebin frequencies and genotype probabilities. When comparing genotype probabilities derived from rebinned databases of different sizes, compromise is unavoidable and several methods for accomplishing this have been proposed, some more amenable to pairwise comparisons than to multi-subpopulation analyses.

One approach is to delete at random observations from the larger databases until the sample sizes are equal prior to rebinning. Another method is to determine rebin boundaries using the smallest database and apply those boundaries to the larger database as well. These and similar methods suffer reduced detail in the allele frequency distributions either directly by discarding alleles or indirectly from the increased rebinning of the larger, more informative, database. Although perhaps being amenable to pairwise comparisons, these methods are ill suited to studies such as this one, in which not only do the subpopulation databases differ in size, but their sizes are not in census proportions.

Another method, and the one used here, is to first balance the bin counts of all the databases to be compared by multiplying all of their bin frequencies by the same total count, T_b, based upon the least represented subpopulation. Balanced bin counts thusly obtained were then subjected to rebinning, and these rebin frequencies were in turn used to calculate genotype probabilities.

Because the subpopulation sample sizes in this study did not reflect their census proportions, the various subpopulations were over or under represented to varying degrees. The Chinese and white subpopulations were the least represented ethnic and racial groups, respectively. A T_b of 600 was obtained for the East Asians after dividing the average number of alleles across loci for each ethnic group by its fraction of the population. The Chinese subpopulation comprised about one-third of the normalized East-Asian population, but its database averaged only about 200 fragments. This yielded an effective East-Asian sample size of about 600

fragments. The remaining three ethnic groups yielded larger, and so less conservative, effective sample sizes. By the same method, a conservative total population sample size of 800 for the T_b was obtained using the white racial group, the sample sizes of which averaged only about 400 alleles but comprised over 50% of the population.

This method also enabled the use of the FBI furnished binning-rebinning program, BINDATA written by K. Monson, which does not provide for binning separate from rebinning. With the particular databases used here, only bins containing zero observations were merged. Bins containing even a single observation, following balancing, yielded at least 6 ethnic or 8 racial balanced counts. Because the Budowle-Monson rebinning algorithm merges only bins with fewer than 5 counts, all non empty original bins in this study remained unmerged.

Pooled Fixed-Bin Distributions

The East Asian ethnic bin frequencies were combined to simulate the pooled Asian racial database obtained if representative rather than stratified sampling had been used. Similarly, the racial bin frequencies were used to simulate those of a total Southern California population. For a given locus, the pooled bin frequency, $\bar{p}_i$, is calculated by weighting the frequency for each subpopulation, p_{ik}, by its proportion in the population, w_k.

$$\bar{p}_i = \sum_{k=1}^{4} w_k p_{ik}$$

These pooled bin frequencies were balanced as described above and then rebinned to yield the final rebin frequencies used in calculating the pooled genotype probabilities. Genotypes estimated from these are equivalent to those that would have been obtained had population substructuring been ignored (but representative sampling been used).

Test Genotypes

Two sets of test genotypes were prepared, both of which used sampling with replacement. In the first method (population proportion) one thousand one-, two-, three-, and four- locus ethnic and racial genotypes were generated by resampling with replacement fragments from each subpopulation in its proportion of the East Asian race or total population, as appropriate. For some analyses, a second method was used, which used equal proportions (25%) for each of the ethnic groups or each of the racial groups.

Independent resampling to generate genotypes results in a lower frequency of single band phenotypes than in the original databases. Thus, it was necessary to model coalescence so as to maintain a similar distribution of single band phenotypes. For each locus, resampled fragment pairs were tested for probable coalescence as described elsewhere (25), and both replaced by their mean if coalescence was indicated.

Genotype Probability Estimation

Forensic scientists may be called upon to make several different kinds of estimates in situations in which information regarding subpopulations or databases for them do not exist. The total human population of an area can be divided into races, which may be further subdivided into ethnic subpopulations. Most often, concern has focused on the effect of ethnic (subpopulation) admixture on the accuracy of racial (population) genotype probability estimates. In this study, we examine the problem at this level and also explore the impact of racial substructuring on total population estimates.

Three methods were used for estimating population genotype probabilities from subpopulation data. For these, we assumed genetic equilibrium of the subpopulations, both within and between loci, because it was desired to limit variance to a single inter-level effect. All three methods calculated single-locus genotype probabilities the same way. Heterozygote probabilities were determined by $2pq$. Homozygote probabilities were calculated using $2p(1 - p)$, which should be noted is a different treatment from the $2p$ method of Budowle and Monson (7). (Because all rebin frequencies are small, this method yields results only slightly different from theirs and the results below apply to both homozygote treatments.)

In the first method, the *cognate* ethnic or racial frequencies pertaining to each subpopulation alone were used. In the second method, referred to here as *stratified*, one first obtains the probabilities for each subpopulation by the first method and then after weighting these by their normalized census proportions, sums them to obtain the overall probability. This method explicitly takes substructuring into account. Given two alleles, i and j, at locus, l, and k population weights, w_k:

$$P_s = \sum_{k=1}^{4} w_k \prod_{l=1}^{4} p_{kl} \qquad \text{where} \begin{cases} p_{kl} = 2p_{kli}p_{klj} & \text{if } i \neq j \\ p_{kl} = 2p_{kli}(1 - p_{kli}) & \text{if } i = j \end{cases}$$

Quasi-random convenience sampling of racial groups is perhaps the method in most widespread forensic use. The third method modeled this method of sampling the East Asian racial population and also the total Southern California population and ignored substructuring by using *pooled* (average) bin frequencies calculated as above.

$$\bar{P}_p = \prod_{l=1}^{4} \bar{p}_l \qquad \text{where} \begin{cases} \bar{p}_l = 2\bar{p}_{li}\bar{q}_{lj} & \text{if } i \neq j \\ \bar{p}_l = 2\bar{p}_{li}(1 - \bar{p}_{li}) & \text{if } i = j \end{cases}$$

For our analyses, we transformed final genotype probabilities by $\log_{10}(1/P)$.

Ethnic and Racial Genotype Probability Variance Comparison

In earlier work, we showed that racial fragment-size distributions varied more than ethnic ones (20), a finding which also suggested that the variance of racial genotype probability estimates would exceed that of ethnic estimates. One thousand racial genotypes comprised of equal proportions of each subpopulation were generated by resampling with replacement. Logs of the four single-locus probabilities were determined for each genotype using the ethnic or racial rebin frequencies as appropriate. The grand medians of the racial and ethnic genotype (log-transformed) probabilities were compared using the Wilcoxon Signed Rank test (WSR) (27). The grand median of the variances of the four ethnic and four racial estimates were compared by the WSR test. Also, the ratio of the racial to ethnic variances (F-Statistic) was calculated for each test genotype. Because each of these variances was obtained from only four points, the F-Statistics are only descriptive.

Correlation and Regression

The three above methods of genotype probability estimation, ethnic/racial, stratified, and pooled, were correlated using scatter

plots and log-log linear regression. All used 1000 subpopulation test genotypes in proportion to their fraction of the population and balanced rebin frequencies. Four loci were used here because this is a commonly used number of loci in forensic casework. We also compared the stratified and pooled calculation methods with one-, two-, three-, and four-locus Asian ethnic genotypes. We used D2S44 for the single locus comparison; D2S44 and D1S7 for the two-locus comparison; D2S44, D1S7, and D4S139 for the three-locus comparison; and finally D2S44, D1S7, D4S139, and D10S28 for the four-locus comparison. Naturally, small differences in results can occur if other combinations of loci are used.

Estimation Errors

Occasionally, it is necessary to make an inference about a subpopulation genotype probability, P_p, from a population-level database. Now, it is the difference between the subpopulation estimate and the true subpopulation value that is of consequence. To make such differences comparable across differing databases and number of loci, it was useful to employ a function that maps to a convenient scale. The first function used as the reference the presumably more accurate of the two population level estimates to be compared, namely the stratified probability estimate, P_s.

$$D_{p/s} = \frac{\log(1/P_p) - \log(1/P_s)}{\log(1/P_s)}$$

The second method compared the stratified population level estimate with that of the ethnic or racial subpopulation for the reference value.

$$D_{s/sub} = \frac{\log(1/P_s) - \log(1/P_{sub})}{\log(1/P_{sub})}$$

The last method compared the pooled population level estimate with that of the ethnic or racial subpopulation for the reference value.

$$D_{p/sub} = \frac{\log(1/P_p) - \log(1/P_{sub})}{\log(1/P_{sub})}$$

With all three methods, positive differences indicate underestimation of the reference estimate and so non conservative error. With these functions, for example, a relative difference of 0.1 means that two estimates differ by one order out of ten orders of magnitude, or less than one part per billion.

Estimation Intervals

Much of the debate concerning forensic DNA testing has focused on the use of racial (population level) databases in the absence of ethnic (subpopulation level) databases to estimate ethnic genotype probabilities of individuals, who may be members of just one or a few of the component ethnic groups. We investigated if either of two simple measures of genotype probability variation would provide adequate coverage of ethnic subpopulation probabilities using a racial population database.

In one method examined here, one thousand ethnic genotypes comprised of two-hundred and fifty genotypes generated from each of the four East Asian ethnic databases and their four sets of ethnic (unbalanced) rebin frequencies were used to calculate the ethnic probability. Balanced rebin frequencies were not used because these are not normally employed by forensic laboratories. Also, the pooled population level estimate was used as the reference value because in the absence of subpopulation samples, only a pooled estimate would be available as a reference value. The difference between the ethnic probability, P_e, and East Asian pooled probability, P_p, was divided by the latter probability to yield R:

$$R = \frac{\log(1/P_e) - \log(1/P_p)}{\log(1/P_p)}$$

The second method divided the same difference as above by the standard deviation of the four racial probabilities, using for the East Asian racial probability, the pooled estimate.

$$Z = \frac{\log(1/P_e) - \log(1/P_p)}{\hat{s}_{\log}}$$

Results

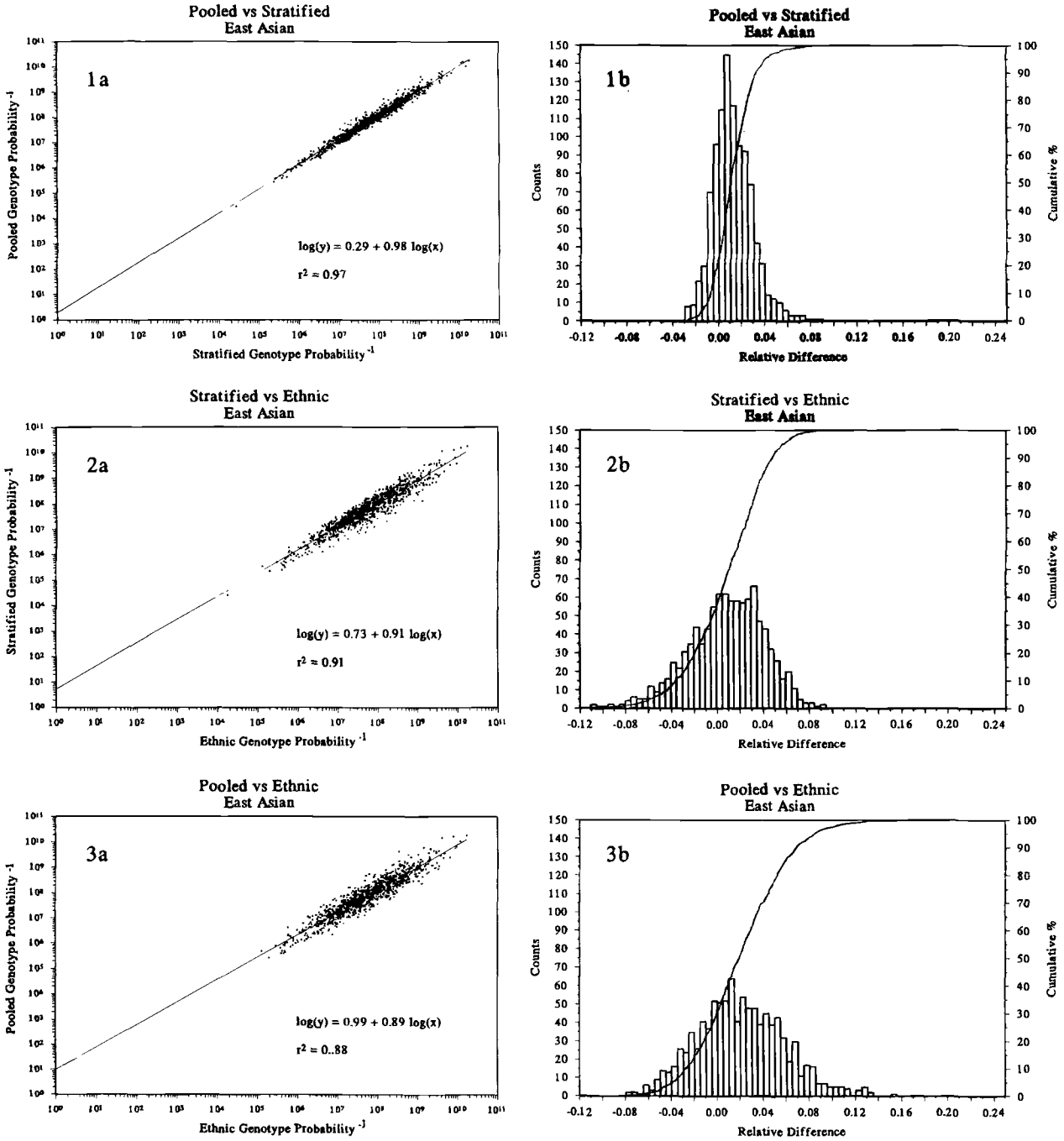
Ethnic and Racial Genotype Probability Variance Comparison

The mean (-log) four-locus racial genotype probability, 8.406 ± 0.969 was not significantly different from the ethnic mean, 8.388 ± 0.814 (WSR, P(S) = .5943). The mean (-log) four-locus racial genotype probability variance, 0.209 ± 0.181, was significantly greater than the mean ethnic variance, 0.100 ± 0.199 (WSR, P(S) = .0001). Approximately 70% of the individual genotype F-statistics exceeded 1, but only 11% of the ratios exceeded a normality based .05% critical value of 9.28. Note that testing for individual F-statistics should not be interpreted as a test of significance.

This result had been anticipated, because we had shown previously that between-race bin frequency distribution variation was greater than that between ethnic groups (25). These results also suggest that, using East Asians as a model for other races, ethnic genotype probability variance can be conservatively estimated using racial variance. The pooled and stratified racial population and the ethnic subpopulation four-locus genotype probabilities differed significantly from each other. The pooled and stratified total population and the racial subpopulation probabilities also differed significantly from each other. Hence, there are subpopulation effects detectable even in multilocus estimates.

Correlation and Regression

Scatter plots of the three methods of genotype probability estimation at the Asian ethnic subpopulation level are contained in Fig(s). 1a–3a, along with the linear regression line and correlation coefficient. Scatter plots using one to four loci comparing the stratified and pooled methods with ethnic groups are found in Fig(s). 4a–4d. All demonstrate excellent correlation. The apparent paucity of data points, especially high probability ones, in Fig. 4a is due to there being fewer than 200 possible single-locus fixed-bin genotypes, and the 10 most common genotypes account for over one-half of the thousand types. Scatter plots comparing the various calculation methods at the racial subpopulation level are presented in Fig(s). 5a–7a. Correlations are somewhat poorer, but the stratified and pooled methods remain highly correlated.



FIGS. 1–3—*1000 simulated East Asian four-locus fixed-bin genotype probabilities using balanced rebin frequencies: a) log-log correlations of pooled, stratified, and cognate probabilities; b) histogram and cumulative distributions of errors from Figs. 1a–3a, relative to the abscissa:* (log(ord) − log(abs))/(log(abs)). *Positive differences indicate non conservative errors. Regardless of calculation method, errors rarely, if ever, exceeded one order of magnitude out of ten, or in other words, one part per billion. The proportions of the ethnic groups are as found in Table 2.*

There was excellent agreement among all of the methods, with the pooled method yielding slightly lower probabilities. The high correlations are the result of three factors. First, each subpopulation database is contained in the larger population-level database. Second, human populations do not exhibit extreme divergence, and the binning method affords some smoothing of the allele distribution. And third, the numerous independent variables afford substantial cancellation of errors. Although all these errors may coincide in direction, it is unlikely that many actually would do so. Finally, it should be noted that these three factors are not artificial, as the same factors are operative in actual casework.

Estimation Errors

Beside each scatter plot is the histogram of the relative difference between the (-log) probabilities obtained by test and reference estimation methods (Figs. 1b–3b, 5b–7b). Positive values indicate underestimation of the probability by the test method compared with the reference method. In general, the relative differences were not normally distributed. Of the three comparisons, only the stratified/cognate error distributions were negatively skewed, yielding four-locus over estimations as great as about one order of magnitude out of ten.

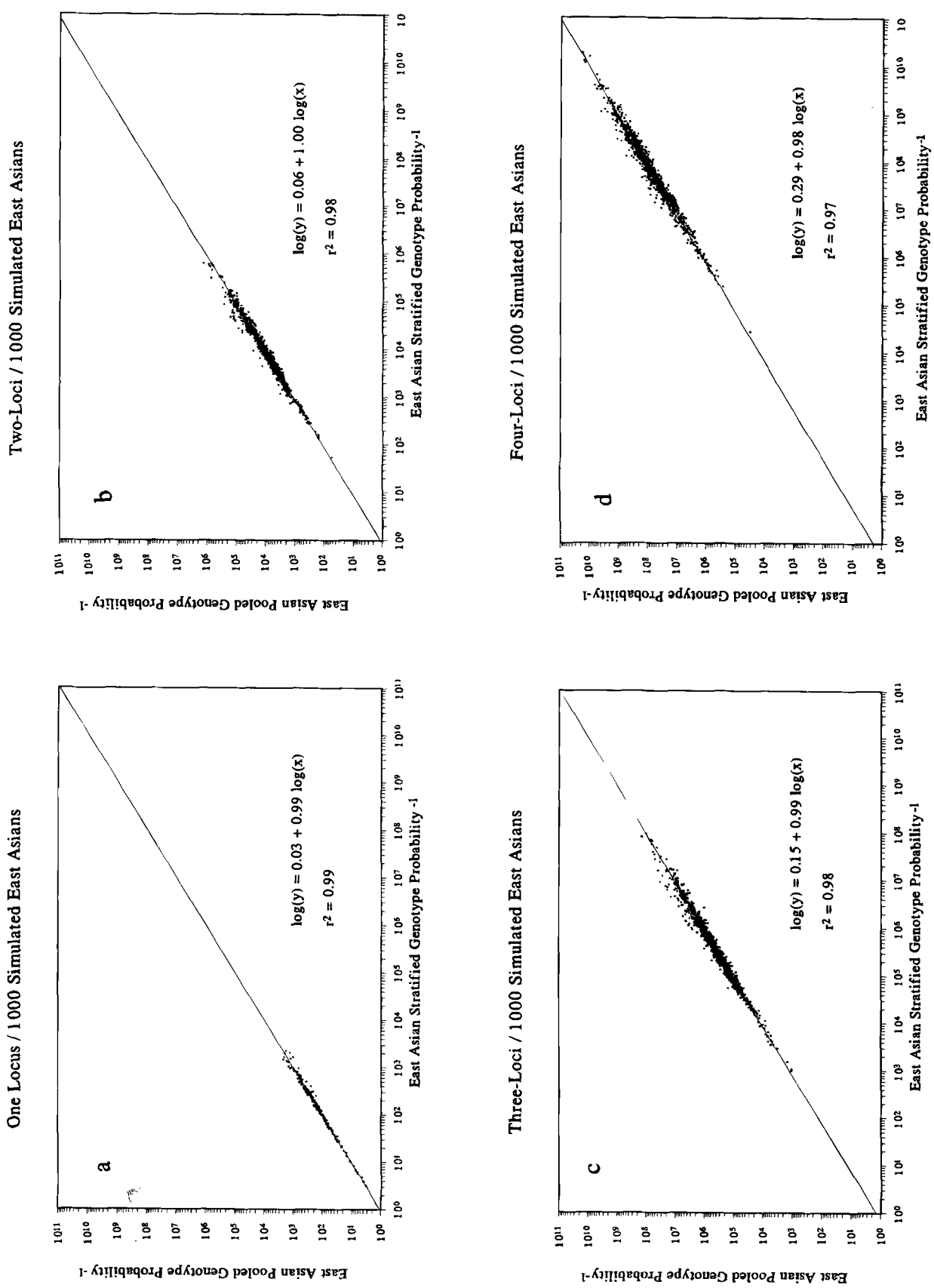
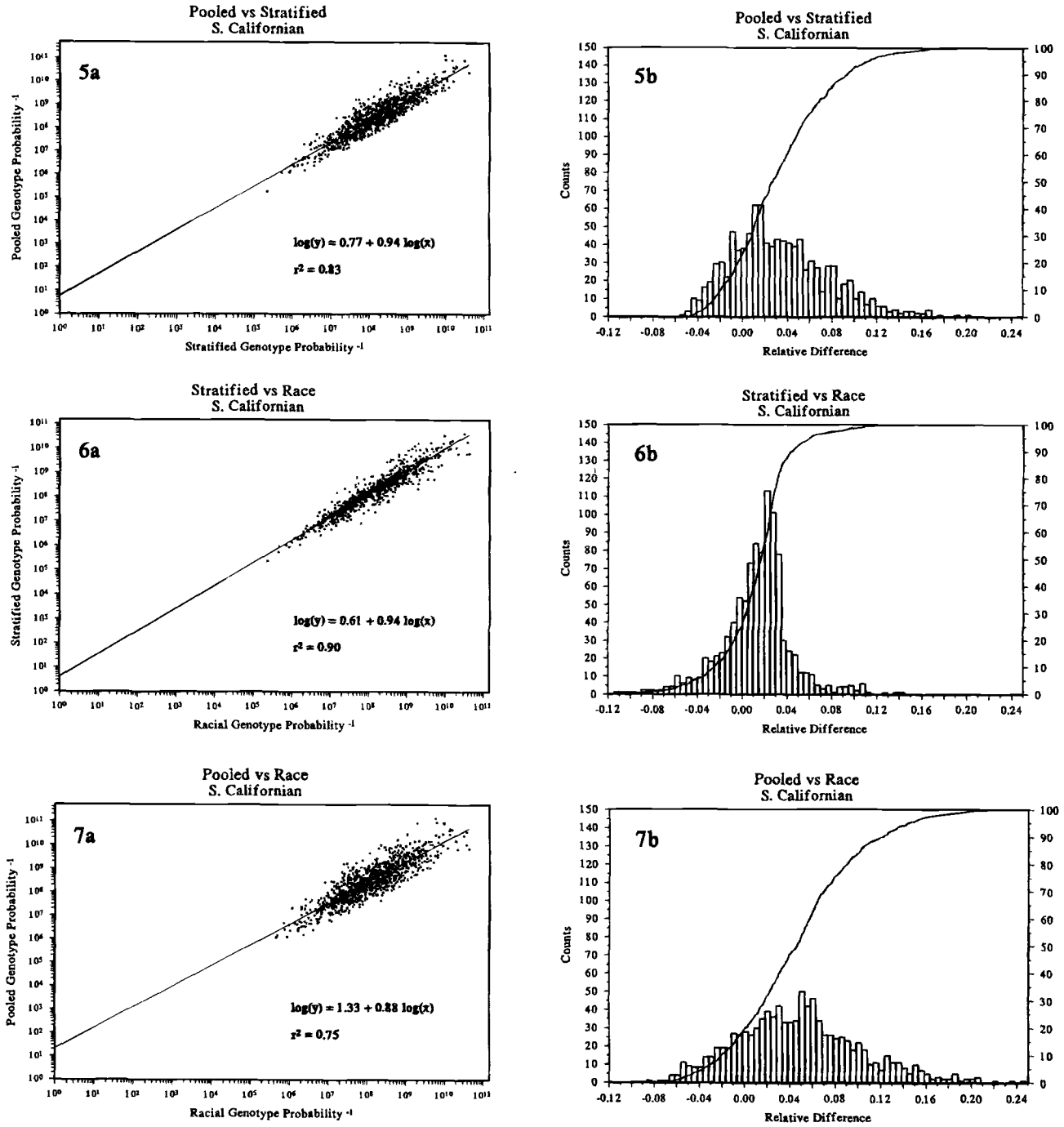


FIG. 4—*Log-log correlations and regressions calculated with pooled (average) and stratified (weighted) allele frequencies for 1000 simulated East Asian one- (D2S44) (a), two- (D2S44, D1S7) (b), three- (D2S44, D1S7, D4S139) (c), and four- (D2S44, D1S7, D4S139, D10S28) (d), locus fixed-bin genotype probabilities. The proportions of the ethnic groups are as found in Table 2. Correlations of 0.97 to 0.99 show that pooled racial databases can substitute for databases consisting of individual ethnic groups.*



FIGS. 5–7—*1000 simulated racial four-locus fixed-bin genotype probabilities using balanced frequencies: a) log-log correlations of pooled, stratified, and racial probabilities. b) histogram and cumulative distributions of errors from Figs. 5a–7a, relative to the abscissa:* (log(ord) − log(abs))/(log(abs)). *Positive differences indicate non conservative errors. The greater kurtosis of the Stratified versus Race histogram is due to the 53% weighting of the Caucasian allele frequencies. (The proportions of the racial groups are as found in Table 2.) A relative error of one order of magnitude out of four is no worse than one part per thousand.*

Racial Population Level

At the racial population level, the pooled method gave only very slightly worse correlation compared with the stratified method for estimating East Asian probabilities. This is made clear in the pooled versus stratified scatter plot (Fig. 1a). The greatest difference was about one order of magnitude out of eleven, and in 99% the differences were less than one out of fifteen (Fig. 1b). This result clearly demonstrated that for racial groups such as East Asians, which present arguably an extreme case of population substructure, the impact on racial genotype probability estimates is small. Furthermore, the data suggested that representative (racial) sampling and pooled rebin frequencies can adequately substitute for ethnic (stratified) sampling when estimating racial genotype probabilities.

Using a stratified East Asian estimate to substitute for the particular ethnic estimate resulted in only one of the thousand ethnic four-locus genotype probabilities being underestimated by more

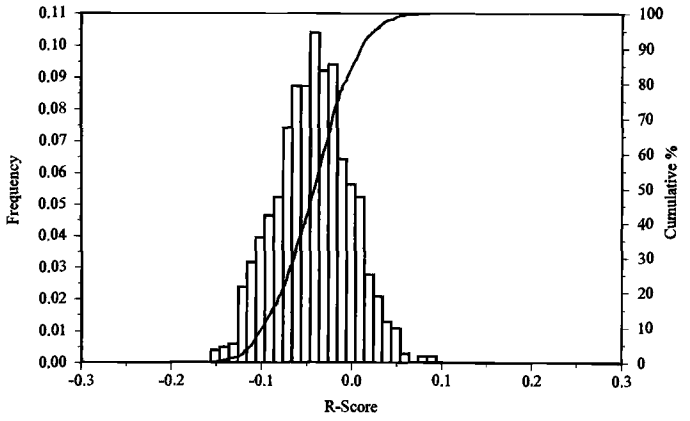


FIG. 8—*Distribution of 1000 East Asian four-locus R-scores:* R = $(\log(1/P_c) - \log(1/P_p))/(\log(1/P_p))$, *where* P_c *is the cognate ethnic probability, and* P_p *is probability obtained from pooled racial allele frequencies. Chinese, Japanese, Korean, and Vietnamese genotypes are equally represented. A negative score indicates that the ethnic probability is greater than the pooled racial probability, or in other words a non conservative error. The worst error among all 1000 four-locus genotypes could be negated by adding one order of magnitude to a typical* 10^{-6} *genotype probability.*

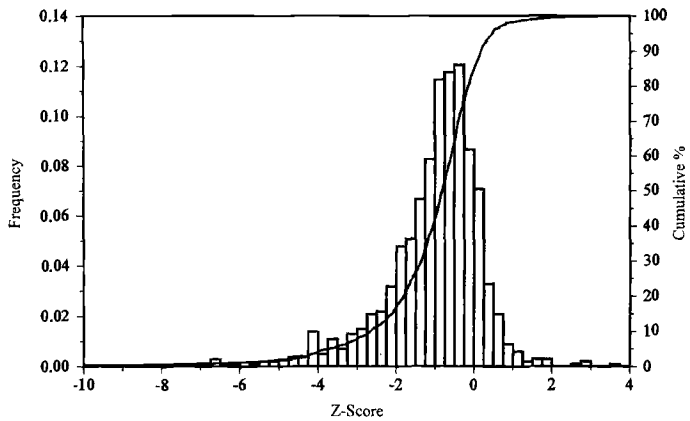


FIG. 9—*Distribution of 1000 East Asian four-locus Z-scores:* Z = $(\log(1/P_c) - \log(1/P_p))/\hat{s}_{\log}$, where P_c *is the cognate ethnic probability,* P_p *is the probability calculated using pooled racial allele frequencies, and* $\hat{s}$ *is the standard deviation of the logs of the four racial probability estimates: Black, East Asian, Hispanic, and White. Chinese, Japanese, Korean, and Vietnamese genotypes are equally represented. A negative score indicates the ethnic probability is greater than the pooled racial probability, or in other words a non conservative error. This score's large range limits its utility as an error predictor. The greater non conservative errors were due not to large discrepancies between cognate and pooled estimates, but to very small variances of some of the pooled racial estimates.*

than one order of magnitude out of eleven, and 99% by less than one out of thirteen (Fig. 2b). After substituting the pooled East Asian estimate for the ethnic one, all relative errors were less than one out of six, with 99% of these less than one out of eight (Fig. 3b). To put these last figures in perspective, they are equivalent to errors of less than one part in one-hundred thousand and one part in ten million, respectively.

Total Population Level

The relative differences at the racial level were significantly greater, with pooled and stratified total population estimates differing by almost one order of magnitude out of five, although 99% were less than one out of six (Fig. 5b). Using a stratified total population estimate for a racial one yielded relative errors of less than one out of seven (Fig. 6b). The marked kurtosis of the stratified-racial error distribution was due to the self-correlation of the whites, who comprised a large part (53%) of the total. In contrast, the East Asians were almost evenly divided among the four ethnic groups (Table 2). Pooled total population estimates differed from the racial ones by, at the maximum, nearly one order of magnitude out of four, although 99% were less than one out of five (Fig. 7b). These results also suggest that representative sampling of a racially substructured population might be substituted for stratified sampling were a total population genotype probability estimate sought. However, using a total population probability to substitute for racial probability incurs greater error. Furthermore, errors as large as those observed here are likely to occur even more often in practice, as the proportions of some racial subpopulations appearing as suspects and defendants in criminal cases are greater than their census proportions.

Errors Relative to Pooled Estimates

Using the pooled East Asian figures as the reference values for calculating the relative error of the ethnic frequency estimation did not yield errors greater than one order of magnitude out of seven and 99% less than one out of nine. The total population probabilities differed from the racial probabilities (relative to the total population probability) by less than one order of magnitude out of five, and 99% by less than one out of six. These ratios are slightly less than those above, as the pooled estimates, which appear in the denominator, tend to underestimate the correct value and so are larger.

Estimation Intervals

The distribution of 1000 ethnic *R* and *Z* scores are plotted in Fig(s). 8 and 9. The *R* scores suggest that to underestimate the probability of fewer than one ethnic four-locus genotype per thousand, it would be necessary to increase by about one order of magnitude a pooled racial probability of 10^{-6}. The *Z* statistic calculation yielded a large range and so is not sufficiently stable for determining ethnic genotype probability confidence limits from pooled racial probabilities. Inspection of some of the extreme *Z* scores revealed their being due not to large differences between East Asian ethnic and pooled racial probabilities, but due to very small variances of some of the four-race probability sets used to calculate $\hat{s}_{\log}$.

All of the errors here employ estimates using Budowle and Monson's fixed-(re) bin method. In unpublished work, we have found these fixed-bin estimates are conservative in comparison with those made using a match window of ±3.4% (i.e., a total integration window of 6.8%) of mean band size for about 94% of four-locus genotypes. For the remaining under-estimated genotypes, the maximum relative error was less than one order of magnitude out of thirteen.

Discussion

Our results show that, using East Asians as a model for extreme substructuring of American races by major ethnic groups, the impact of such substructuring on racial genotype probability estimates is negligible. It is difficult to comprehend how estimation errors due to substructuring effects on the order of one part per million could be material to forensic fact finding.

The East Asian model used here was chosen to be representative of extreme substructuring by major ethnic groups. Substructuring effects for other races, which have already undergone substantial ethnic and even racial intermarriage, can be expected to be less than those observed here. We are not suggesting, however, that greater errors cannot occur where minor subpopulations are concerned. Clearly, the results here do not apply to small, isolated, and inbred subpopulations such as the Karitiana Indians of the Amazon forest (28). At the same time, it is equally obvious that such subpopulations are irrelevant, as well as immaterial, to forensic fact finding when such groups are not directly involved.

The type of racial level subpopulation sampling most often in forensic use, namely convenience sampling, was modelled here using pooled frequencies constructed from stratified samples. The difficulty of conducting representative sampling should not be underestimated. The Southern California East Asian population is so highly substructured that, without deliberate effort, it is unlikely a representative East Asian sample would have been obtained had the sampling methods ordinarily used by forensic laboratories been employed. However, the results obtained here by substituting a multiracial pooled sample, and those of Weir (13) substituting the wrong racial sample entirely, suggest errors greater than one part per thousand are unlikely.

To date, the focus of criticisms of forensic DNA testing has been on the underestimation of genotype frequencies due to population substructuring by ethnic groups. In this report, we have shown that for real VNTR RFLP databases, such errors are negligible. Elsewhere, we have explored the impact of sampling and measurement error (25). The combined effect of these two sources of variation, both trivial, actually exceeds that of ethnic substructuring. Realistically, sample confusion and contamination, misinterpretation, and other such laboratory mistakes are the only feasible sources of error great enough to impact fact finding in criminal cases.

Acknowledgments

We thank Dr. Eva Steinberger for her many helpful comments on earlier versions of this manuscript.

References

1. Weir B. Population genetics in the forensic DNA debate. Proc Nat Acad Sci USA 1992;89:11654–9.
2. National Research Council Committee on DNA Technology in Forensic Science. DNA technology in forensic science. Washington (DC): National Academy Press 1992;74–96.
3. Cohen JE. The ceiling principle is not always conservative in assigning genotype frequencies for forensic DNA testing. Am J Hum Genet 1992;51:1165–8.
4. Slimowitz JR, Cohen JE. Violations of the ceiling principle: Exact conditions and statistical evidence. Am J Hum Genet 1993;53: 314–23.
5. Morton NE. Genetic structure of forensic populations. Am J Hum Genet 1994;55:587–8.
6. Morton NE. Genetic structure of forensic populations. Proc Nat Acad Sci USA 1992;89:2556–60.
7. Budowle B, Monson KL. A statistical approach for VNTR analysis. In: Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis; 1989; Quantico (VA). Washington (DC): Government Printing Office;121–6.
8. Weir B. Independence of VNTR alleles defined as floating bins. Am J Hum Genet 1992;51:992–7.
9. Crow JF, Denniston C. Population genetics as it relates to human identification. Proceedings of the Fourth Promega International Symposium on Human Identification; 1993 Phoenix (AZ); Madison (WI): Promega, 1993;31–6.
10. Balding DJ, Nichols RA. DNA profile match probability calculations: How to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64:125–40.
11. Roeder K. DNA fingerprinting: A review of the controversy. Stat Sci 1994;9:222–78.
12. National Research Council Committee on DNA Forensic Science. The evaluation of forensic DNA evidence. Washington (DC): National Academy Press, 1996;4,1–5,41.
13. Weir B. Independence of VNTR alleles defined as fixed bins. Genetics 1992;130:873–7.
14. Federal Bureau of Investigation. VNTR population data: A worldwide study. Washington (DC): Government Printing Office, 1993; Vol III.
15. Krane DE, Allen RW, Sawyer SA, Petrov DA, Hartl DL. Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations. Proc Nat Acad Sci USA 1992;89:10583–87.
16. Budowle B, Monson KL, Giusti AM. A reassessment of frequency estimates of Pvu II-generated VNTR profiles in a Finnish, an Italian, and a general U.S. Caucasian database: No evidence for ethnic subgroups affecting forensic estimates. Am J Hum Genet 1994;55:533–9.
17. Budowle B, Monson KL, Giusti AM, Brown BL. The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases. J Forensic Sci 1994;39:319–52.
18. Budowle B, Monson KL, Giusti AM, Brown BL. Evaluation of Hinf I-generated VNTR profile frequencies determined using various ethnic databases. J Forensic Sci 1994;39:94–112.
19. Duncan GT, Noppinger K, Carey J, Tracey M. Comparison of VNTR allele frequencies and inclusion probabilities over six populations. Genetica 1993;88:51–7.
20. Hartmann J, Keister R, Houlihan B, Thompson L, Baldwin R, Buse E, et al. An analysis of a multi-ethnic and multi-racial VNTR RFLP database for ethnic and racial diversities. Am J Hum Genet 1994; 55:1268–78.
21. Department of Health Services. Report on Los Angeles area Asian births for 1986–1988. Los Angeles (CA): Los Angeles County Data Collection and Analysis Unit; 1991.
22. Ong PM, Azores T. Asian immigrants in Los Angeles: Diversity and divisions. Los Angeles (CA): Univ. of Calif Asian American Studies Center; 1993, working paper.
23. Bureau of the Census. General population characteristics. Washington (DC): Government Printing Office; 1990; US Census: CP-1-6, Sec 1, 27–32.
24. Budowle B, Baechtel FS. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophoresis 1990;1:181–7.
25. Hartmann JM, Houlihan BT, Thompson LD, Chan C, Baldwin RA, Buse EL. The effect of sampling error and measurement error and its correlation on the estimation of multilocus fixed-bin VNTR RFLP genotype probabilities," J Forensic Sci 1997;42: in press.
26. Monson KL, Budowle B. A system for semi-automated analysis of DNA autoradiograms. In: Proceedings of an International Symposium on the Forensic Aspects of DNA Analysis. 1989; Quantico (VA). Washington (DC): Government Printing Office, 1989; 127–32.
27. Bradley, JV. Distribution-free statistical tests. 2nd ed. New York: John Wiley and Sons, 1968;252–6.
28. Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK. Studies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 1991;63:775–94.

Additional information and reprints requests:
John Hartmann
Forensic Science Services
Orange County Sheriff-Coroner Department
P.O. Box 449, 320 N. Flower Street
Santa Ana, CA 92703