J Forensic Sci, Nov. 2004, Vol. 49, No. 6
Paper ID JFS2004253
Published 6 Oct. 2004
Available online at: www.astm.org

Authors' Response

Sir:

Krane et al. (1) contest the conclusion of our study of the "CODIS STR Loci Data from 41 Sample Populations" (2), that "there was little evidence for departures from Hardy-Weinberg expectations (HWE) in any of the populations" on the grounds that: (a) our application of the Bonferroni adjustment for multiple testing to our data on 12 to 13 loci, studied in each of the 41 populations, is inappropriate; (b) we disregarded the "significant" clustering of departures from HWE in two populations (Salishan and Navajo); and (c) we failed to pay attention to the distinctiveness of the Native American populations to explain the clusters of deviant test results in these populations. Further, Krane et al. (1) contend that our data "provides significant evidence that at least three loci in Navajos (FGA, D7S820, and TH01) as well as Salishans (D3S1358, FGA, and D7S820) do *not* adhere to HWE" and hence, "the product rule should not be used to estimate the rarity of genotypes involving those loci in those populations unless corrective factors are involved."

We demonstrate that their contentions are based on flawed statistical as well as population genetics logic. Moreover, even if their specifically chosen observations on our reported test results are taken in isolation (of the rest of the study results) as they did (1), their comments are of no consequence. The practices employed by the United States forensic community for the last decade generally do not use the strict product and, instead, allow for departures from HWE to estimate the rarity of genotypes in forensic computations (3).

The first statistical flaw of Krane et al.'s arguments is in regard to the Bonferroni adjustment. It is true that the Bonferroni adjustment for multiple tests applies when independent sets of data are used to test *the same hypothesis*. In the context of DNA Forensics, however, the issue is: "Do the genotype frequencies at the CODIS STR loci conform to HWE in population samples"? Our study (2) involves the examination of this specific hypothesis. What would differ from one locus-population combination to another in the dataset we analyzed is the magnitude of departure from HWE, should the null hypothesis be *not* true. In the terminology of meta-analysis, this is called the variation of size effects (4), and of the various alternative forms of meta-analytic multiple test correction of p-values (see Table 15.2 of Ref 5), the Bonferroni adjustment pays less attention to the varying size effects of multiple sets of data, though it is the simplest and most widely used (6). Notably, applications of Bonferroni-type of adjustment of significance levels are also common in the current genome-scan studies of mapping complex disease traits by multipoint linkage and association analyses, where the markers vary from one test to another (7,8). Thus, its application is not unique to forensic database analyses.

Krane et al. (1) committed another statistical error by implying that the Bonferroni adjustment assumes "the loci are equivalent with regard to HWE and discriminating power." This simply is not true, because the Bonferroni adjustment is based on the null distribution of the minimum p-values of all of the locus-population combination of tests (see Table 15.2 of Ref 5), each of which is, in turn, dependent on the product of effect size and sample size of the respective dataset (4). Thus, even though the order statistic (i.e., the minimum of all observed p-values), against which the revised significance level of the Bonferroni adjustment is compared, does not utilize the full spectrum of distribution of the p-values, it is incorrect to say that the variations of size effects and discriminatory power in the individual locus-population combinations are ignored in applications of the Bonferroni adjustment.

Nonetheless, even if the Bonferroni adjustment is regarded as crude, the approach taken by Krane et al. (1) to take clusters of deviant test results in isolation of the others, is statistically flawed. By concluding that "3 of the 31 noted departures from HWE would randomly be found in one of the 41 populations . . . is itself unlikely" they made two fundamental errors. First, their conclusion of unlikeliness is based on a p-value (the test procedure of which is not mentioned by Krane et al.) of 0.078, not supportive of the level of significance generally employed in data analysis. Second, and more importantly, in order to test the true randomness, they should have considered the entire array of our test results (including making distinctions of possible size effects of departures in different population-locus combinations), not merely the most clustered occurrence of significant results. Their approach is effectively a "vote-counting" method of meta-analysis (9), which has been known to be inappropriate for synthesis of multiple test results when the test statistic values (individual p-values of the exact test in our case) are reported for each study (10).

Table 1 shows the relevant data (directly extracted from Ref 2) for a true test of randomness considering the entire array of our tests results. With this tabulation, the randomness of occurrences of significant (at 5% level) or non-significant departures from HWE is tested by $2 \times c$ contingency table analysis for each of the five major groups of populations, as well as for the pooled data, taking into account each individual locus-population test result. The significance test for randomness is done by following the permutation algorithm suggested by Roff and Bentzen (11), which does not assume any large sample property of the (Chi-square) test statistic. The last two columns of Table 1 clearly show that the occurrences of significant deviation from HWE in individual locus-specific tests are random between population samples within each of the five major

TABLE 1—*Data pertinent to test of randomness of observed departures from HWE.*

Population Group	Number of Population Samples	Total Number of Significant Departures*	χ^2	p-Value†
African Americans	11	7	7.41	0.888
US Caucasians	9	8	10.36	0.267
Hispanics	8	4	4.08	≈1.0
Asians	6	2	4.11	≈1.0
Native Americans	7	10	6.09	0.426
Pooled	41	31	41.79	0.376

* Extracted from exact tests of departure from the data presented in Ref 2, the significant departures from HWE at 5% level are: in African Americans: 0 of 13 (0/13) of the locus-specific tests for the FBI sample, 2/13 for Bahama, 1/13 for Jamaica, 1/13 for Trinidad, 1/13 for California, 0/13 for Alabama, 1/13 for Florida, 0/13 for Virginia, 0/13 for New York, 1/13 for Illinois, and 0/12 for Minnesota; in US Caucasians: 1/13 for FBI, 0/13 for California, 2/13 for Alabama, 1/13 for Florida, 0/13 for Virginia, 0/12 for New York, 3/13 for Michigan, 1/12 for Minnesota, and 0/12 for Canada; for Hispanics: 0/13 for FBI, 0/13 for California, 1/13 for Florida, 1/13 for New York, 1/13 for Michigan, 0/12 for Minnesota, 0/13 for Arizona, and 1/13 for Mexico; in Asians: 0/13 for Chinese, 0/13 for Japanese-1, 0/13 for Japanese-2, 0/13 for Koreans, 1/13 for Vietnamese, and 1/13 for General Asians; and in Native Americans: 0/13 for Michigan, 1/13 for Minnesota, 1/13 for Apache, 3/13 for Navajo, 1/12 for Northern Ontario, 3/12 for Salishan, and 1/12 for Saskatchewan samples.

† The significance levels (p-values) were obtained by permutation tests of the respective $2 \times c$ contingency table chi-square (with algorithm of ref. 11) with 10,000 replications of permutations.

population groups, as well as in the pooled sample of 41 populations. This randomness of deviations from HWE is in direct contradiction with the result obtained using the inappropriate method by Krane et al. (1).

To further illustrate that our observed 31 deviant test results out of the 524 tests performed are in accordance with the expectation of multiple testing, one can also use Fisher's (12) original theory; namely, under the null hypothesis that HWE applies for all of the 524 tests performed, the distribution of the *p*-values (of exact tests) should follow a uniform distribution. Thus, the array of individual locus-population specific *p*-values can be tabulated for each group of populations to check for conformity with the uniform distribution. In the total data, the 524 *p*-values do not deviate from a uniform distribution by the Kolmogorov-Smirnoff non-parametric test ($Z = 1.154$, with 2-sided *p*-value of 0.134 with 10,000 replications of simulation). Further, the observed distributions of *p*-values across the five groups of populations (utilizing the 142 test results in African Americans, 114 in U.S. Caucasians, 103 in Hispanics, 78 in Asians, and 87 in the Native Americans) are seen to be homogenous (by a 5×20 contingency table Chi-square test, obtained by grouping the *p*-values into 20 classes of equal interval length), since this test resulted in a Chi-square value of 57.99 whose empirical level of significance is 0.948 (with 10,000 replications of permutations). Thus, even if our original Bonferroni adjustment were to be questioned, when the data represented in (2) are re-analyzed by at least two other meta-analytic methods, each of which recognizes locus-population specific differences of discriminating power and size effects of possible departure from HWE, we still find no over-all significance of deviations from HWE (i.e., the hypothesis that at least one of the locus-population combination of genotype frequencies deviate from HWE is rejected) when the array of all 524 test results are considered simultaneously. In addition, homogeneity of distribution of *p*-values as well as randomness of occurrence of *p*-values below the nominal level of significance of 5% (in individual tests) suggest that there is no significant clustering of deviant test results in our data reported in (2).

The population genetic errors committed by Krane et al. (1) are more fundamental than the relatively obvious statistical errors. For example, in response to our comment that the majority of the initially found departures from HWE are due to genotypes consisting of rare alleles (2), they assert that in the Navajo population, "not even one of the homozygotes observed" at the deviant loci (FGA, D7S820, and TH01) "were homozygous for rare alleles." However, they failed to recognize that for this population all of the rare (below five counts) alleles (FGA-28, D7S820-7, and TH01-9) occur in heterozygote forms. This should have been obvious to Krane et al (1). Moreover, the frequencies of the heterozygotes with rare alleles are in excess of their respective HWE expectations, contributing to the overall significant departures from HWE at these loci. Incidentally, 11 of the 31 initial deviations from HWE, noted in our data analysis, are due to excess overall heterozygosity at the respective loci, and hence our statement with regard to rare alleles contributing to deviations from HWE cannot be equated to excess homozygosity. Of particular note, of the clusters of deviations from HWE in the Navajo and Salishan samples (three per population), erroneously claimed by Krane et al. (1), three (D7S820 in Navajo, and D3S1358 and FGA in Salishan) show heterozygosity excess (see Table 5 of Ref 2); thus population substructure within these populations, or genetic drift are not likely to account for these departures (because both of these factors would have caused heterozygote deficiency and not excess as observed).

The second population genetic error committed by Krane et al. (1) relates to their statement that the clusters of HWE deviations in the Navajo and Salishan samples is consistent with the relative large F_{ST} values for the Native Americans (average F_{ST} of 0.0282) that we reported. This is clearly wrong; F_{ST}, computed in our work, reflects the standardized allele frequency variation across the seven Native American populations sampled. This has no direct relationship with deviations from HWE within each population, which would have been reflected if they had estimated F_{IS} (13,14). Of course, as noted in our prior work (15), loci/populations with smaller within-population gene diversity generally exhibit larger F_{ST} in substructured populations. Nevertheless, as noted above, since at least three of the six HWE deviations in the Navajo and Salishan samples cannot be ascribed to hidden population substructure within these populations, there is no foundation to suggest a consistency of the observed deviations from HWE (within each sample) with their group-level F_{ST}.

By claiming that "the Navajos have significantly higher rates of allele sharing than any other population, which suggests a greater degree of substructure within the Navajo population," Krane et al. (1) made an even more egregious population genetic error. The data presented in their Table 1 are not at all surprising, and not novel either, as our group already has shown that with respect to VNTR as well as STR markers, individuals of smaller populations generally share more alleles, compared with those of larger populations (16,17). However, Krane et al. (1) apparently are not aware that the larger rate of allele sharing is a consequence of a lower (reduced) level of genetic diversity in smaller sized populations, and not necessarily due to their inherently higher hidden substructuring (17). This is so, because in each of these populations, one can compute the distribution of allele sharing based on allelic independence within and between loci (18), and as shown in (17), in spite of the larger rates of allele sharing between individuals of populations of reduced genetic diversity, the distribution of allele sharing between individuals can be in accordance with their expectations based on the assumption of mutual independence of alleles within and across loci. Data presented in Table 2 on the 23 population samples in which the initial 31 deviations from HWE had been observed in our work (2) exhibit this phenomenon. Clearly this shows that in each of the 23 populations where at least one initial deviation from HWE was observed (2), the observed rates of allele sharing are in accordance with their respective expectations under the hypothesis of mutual independence of alleles within and across loci (i.e., HWE for each locus, and pairwise as well as higher-order linkage equilibria (LE) between loci). This is true, in spite of the fact that the populations with lower average heterozygosity exhibit a higher rate of allele sharing. Thus, the contention that a higher degree of allele sharing in a population is reflective of its greater degree of substructuring is clearly incorrect. In fact, the analysis shown in Table 2 addresses another issue raised by Krane et al. (1), which is that LE tests should be carefully evaluated. The distribution of allele sharing is consistent with the assumption of LE (and more strongly, that of mutual independence of alleles within and across loci), even in the samples where some initial departures of HWE in locus-specific tests were observed in our original investigation (2).

Finally, we note that, irrespective of the results of tests of independence of alleles within and across loci in database analyses, the current forensic calculations of multilocus genotype profiles do not use the strict product rule and always incorporate conservative features (3). These include: (a) application of population substructure adjustment for homozygotes, with levels of θ generally larger than the F_{ST} found in empirical data; (b) invoking a minimum threshold frequency for rare alleles; and (c) using the upper confidence limit on the point estimate of the multilocus profile frequency. Some laboratories even include computing conditional probabilities that

TABLE 2—*Allele sharing statistics in the population samples with at least one apparent deviation from HWE.*

Population Group/Sample	Av. Heterozygosity (in %)		Sample Size†	Number of Pairs of Subjects	Mean (SD) Allele Sharing	
	Obs.	Exp.*			Obs.	Exp.‡
African Americans:						
Bahama	78.82	79.49	153	11,628	8.15 (2.13)	8.24 (2.17)
Jamaica	79.35	78.40	157	12,246	8.61 (2.25)	8.54 (2.18)
Trinidad	79.40	80.25	76	2,850	7.92 (2.18)	7.97 (2.15)
California	78.27	78.89	200	19,900	8.35 (2.20)	8.42 (2.18)
Florida	79.36	79.44	94	4,371	8.16 (2.11)	8.20 (2.16)
Illinois	79.19	79.37	150	11,175	8.23 (2.16)	8.29 (2.17)
U.S. Caucasians:						
Alabama	79.51	79.02	150	11,175	8.67 (2.13)	8.67 (2.18)
FBI	78.52	78.12	194	18,721	8.67 (2.18)	8.64 (2.17)
Florida	76.25	77.93	201	20,100	8.57 (2.18)	8.74 (2.19)
Michigan	78.83	78.36	146	10,585	8.62 (2.22)	8.58 (2.18)
Minnesota§	79.22	78.48	150	11,175	7.89 (2.11)	7.85 (2.09)
Hispanics:						
Florida	77.79	78.56	191	18,145	8.48 (2.22)	8.52 (2.18)
Mexico	78.91	77.69	143	10,153	8.81 (2.21)	8.76 (2.19)
Michigan	79.13	78.71	150	11,175	8.50 (2.17)	8.46 (2.17)
New York	79.36	79.55	150	11,175	8.20 (2.19)	8.23 (2.16)
Asians:						
General Asians	77.79	78.64	196	19,110	8.45 (2.18)	8.52 (2.17)
Vietnam	76.79	77.27	200	19,900	8.84 (2.24)	8.89 (2.19)
Native Americans:						
Apache	70.20	70.81	198	19,503	10.47 (2.41)	10.53 (2.24)
Minnesota§	74.50	76.37	200	19,900	8.28 (2.20)	8.46 (2.12)
Navajo	68.93	70.26	182	16,471	10.46 (2.26)	10.59 (2.23)
N. Ontario§	68.87	69.47	125	7,750	9.93 (2.27)	10.01 (2.15)
Salishan§	75.72	73.96	93	4,278	9.05 (2.23)	8.91 (2.12)
Saskatchewan§	73.31	73.58	79	3,081	9.04 (2.08)	9.07 (2.14)

NOTE: The observed rates of allele sharing in the FBI Caucasian, Jamaican, Bahamaian, Trnidadian, and Navajo populations reported in this table are slightly different from the computations shown by Krane et al. (1), the reason for which is unclear. Nonetheless, even if the genotype records used by Krane et al. (1) were different from ours, the major conclusion, namely, the observed distribution of allele sharing is in accordance with the expectations of mutual independence of alleles (i.e., HWE and LE) holds for all populations, irrespective of lower genetic diversity (and consequently, larger extent of allele sharing) within populations.

* Based on HWE.

† The sample sizes (n) refer to the number of individuals with complete multi-locus profile available, so that the number of comparisons for computing the allele sharing statistics becomes n(n – 1)/2, shown in the next column.

‡ Based on mutual independence of alleles within and across loci (see Ref 17).

§ Data on these population samples consist of 12 loci (data of D16S539 missing), and hence, the allele sharing statistics are based on 12-locus genotype profile comparisons.

allow for substructure adjustments at the individual locus level for homozygotes as well as heterozygotes. With these protocols in place, corrective actions are always imposed in forensic computations, whether or not the population data show any sporadic departure from HWE and LE. In conclusion, the entire commentary by Krane et al. (1) is scientifically incorrect, both in the use of statistics and population genetics. There is no basis for their assertions that question the forensic practices for estimating DNA profile frequencies. Their commentary also is a useless exercise since it does not address the current practices employed by forensic laboratories. Krane, et al. (1) state the loci that depart from HWE in Navajos and Salishans "should not be used when the product rule is employed to compute the frequency of multi-locus genotypes. . . unless corrective actions are taken. . . " Perhaps, they are unaware of the (stated above) practices employed for many years for estimating the rarity of a DNA profile.

References

1. Krane DE, Doom TE, Mueller L, Raymer ML, Shields WM, Thompson WC. Commentary on: Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J Forensic Sci 2001;46:453–89. [PubMed]
2. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J Forensic Sci 2001;46:453–89. [PubMed]
3. National Research Council. The evaluation of forensic DNA evidence. Washington DC: National Academy Press, 1996.
4. Rosenthal R. Parametric measures of effect size. In: The handbook of research synthesis. Cooper H, Hedges LV, editors. New York: Russell Sage Foundation, 1994;231–44.
5. Becker BJ. Combining significance levels. In: The handbook of research synthesis. Cooper H, Hedges LV, editors. New York; Russell Sage Foundation, 1994:215–30.
6. Weir BS. Genetic data analysis – II. Sunderland, MA: Sinauer Press, 1996.
7. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 1995;11: 241–7. [PubMed]
8. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7. [PubMed]
9. Hedges LV, Olkin I. Vote-counting methods in research synthesis. Psychological Bulletin 1980;88:359–69.
10. Bushman BJ. Vote-counting procedure in meta-analysis. In: The handbook of research synthesis. Cooper H, Hedges LV, editors. New York: Russell Sage Foundation, 1994;193–213.
11. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small populations. Mol Biol Evol 1989; 6:539–45. [PubMed]
12. Fisher RA. Statistical methods for research workers. 2nd ed. London: Oliver and Boyd, 1928.

13. Wright S. The interpretation of population substructure by F-statistics with special regards to systems of mating. Evolution 1965;19:395–420.
14. Chakraborty R, Danker-Hopfe H. Analysis of population structure: A comparative study of different estimators of Wright's fixation indices. In: Handbook of statistics, volume 8: statistical methods for biological and medical sciences. C. R. Rao and R. Chakraborty, editors. New York: North-Holland, 1991;203–54.
15. Chakraborty R, Jin L. Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. Hum Genet 1992;88:267–72. [PubMed]
16. Chakraborty R. Impact of reduced genetic variation in small human populations: Data and inference from repeat loci. Proceedings of XVIIIth Intl Congress of Genetics, Kunming Satellite Conference, Kunming. 1998 Aug 16–18.
17. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of STR loci beyond human identification: Implications for development of new DNA typing systems. Electrophoresis 1999; 20:1682–96. [PubMed]
18. Chakraborty R, Jin L. Determination of relatedness between individuals by DNA fingerprinting. Hum Biol 1993;65:875–95. [PubMed]

Ranajit Chakraborty, Ph.D.
Center for Genome Information
Department of Environmental Health
University of Cincinnati College of Medicine
Cincinnati, OH 45267-0056

Hee S. Lee, M.D.
Center for Genome Information
Department of Environmental Health
University of Cincinnati College of Medicine
Cincinnati, OH 45267-0056

Bruce Budowle, Ph.D.
Federal Bureau of Investigation Laboratory
Quantico, VA 22135