Letters to the Editor

Am. J. Hum. Genet. 62:000, 1998

Consanguinity and Relative-Pair Methods for Linkage Analysis

To the Editor:

The recent paper by Génin and Clerget-Darpoux (1996) makes an important and timely point: that linkage analyses using nonparametric statistics in an affected-sib-pair design can be made more powerful by assaying identity-by-descent sharing within inbred siblings. However, we believe some of the technical details of their mathematical model and derivations may be in error.

In Appendix A of their paper, Génin and Clerget-Darpoux derive "IBW-state probabilities" (or "condensed coefficients of identity" [Jacquard 1974]) for siblings from a population with a mean inbreeding coefficient α, which is assumed to be constant over time and to be equal to the mean kinship coefficient. These are perhaps unrealistic assumptions in terms of many inbred human populations, as mean kinship coefficients change over time because of changing demographic factors (Khoury et al. 1987), and mean inbreeding coefficients are often much higher than mean kinship coefficients (De Braekeleer et al. 1993, 1996). Even so, under the modeling assumptions made by Génin and Clerget-Darpoux, the probabilities that they present fail to satisfy the following two consistency checks:

1. Let Δ_i represent the probability of being in condensed identity state S_i, where the range of i is 1–9, and assume that the two sibs are numbered "3" and "4" and that the parents are numbered "1" and "2." Then, according to a formula presented by Jacquard (1974), the kinship coefficient between the siblings, Φ_{34}, should be

$$\Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8 \ . \qquad (1)$$

We can derive what the kinship coefficient should be by classical recursion methods:

$$\begin{aligned}\Phi_{34} &= \frac{1}{2}(\Phi_{31} + \Phi_{32}) \\ &= \frac{1}{4}(\Phi_{11} + \Phi_{21} + \Phi_{12} + \Phi_{22}) \\ &= \frac{1}{4}\left[\frac{1}{2}(1+\alpha) + \alpha + \alpha + \frac{1}{2}(1+\alpha)\right] \\ &= \frac{1}{4}(1+3\alpha) \ , \qquad (2)\end{aligned}$$

because, under Génin and Clerget-Darpoux's assumptions, the inbreeding coefficient of each parent is α and the kinship between the parents is also α. However, when we apply equation (1) to Génin and Clerget-Darpoux's Δ's, we get $\Phi_{34} = \frac{1}{4} + \frac{\alpha}{2} + \frac{\alpha^3}{4}$, which is clearly incorrect (except when $\alpha = 0$ or when $\alpha = 1$).

2. Karigl (1981, eq. [7]) presents a matrix that permits one to derive a vector of several different kinship coefficients from the vector of Δ's. However, when this matrix is applied to Génin and Clerget-Darpoux's Δ's, the correct kinship coefficients are not recovered.

In passing, it is important to point out that the derivation, in Génin and Clerget-Darpoux's Appendix B, of condensed identity coefficients for two sibs from a first-cousin marriage is also incorrect. For example, they appear to compute Φ_{224} (sampling twice from one individual—i.e., person 2—and once from an unrelated person) as α^2. However, kinship sampling is done with replacement, so that the chance that the same gene is sampled twice from person 2 is $\frac{1}{2}$, and then it is identical by descent (IBD) with the gene from person 4 with probability α. Likewise, the chance that different genes are sampled from person 2 is $\frac{1}{2}$, and then, since the three different genes are IBD with probability α^2, $\Phi_{224} = \frac{1}{2}\alpha + \frac{1}{2}\alpha^2$. There appear to be similar mistakes throughout the derivation (e.g., for three unrelated people—2, 4, and 6—Φ_{2426} is not α^3). Génin and Clerget-Darpoux's Δ's in their Appendix B do pass the first consistency check above, but they fail the second consistency check.

So why do Génin and Clerget-Darpoux's results fail to satisfy these checks? They derive their Δ's for the sibs by first deriving the parental Δ's and then multiplying these by a transition matrix. Since the transition matrix

Table 1

Probabilities of Ordered Pairs, Kinship Coefficients, and Condensed Identity Coefficients for Two People Drawn at Random from a Population Consisting of 25% Sibs from the Same First-Cousin Marriage (C) and 75% Noninbred Unrelated Individuals (U)

			Condensed Identity Coefficient								
Pair	Probability	Kinship	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9
(C,U)	$\frac{3}{16}$	0	0	0	0	$\frac{1}{16}$	0	0	0	0	$\frac{15}{16}$
(U,C)	$\frac{3}{16}$	0	0	0	0	0	0	$\frac{1}{16}$	0	0	$\frac{15}{16}$
(C,C)	$\frac{1}{16}$	$\frac{9}{32}$	$\frac{1}{64}$	0	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{15}{64}$	$\frac{15}{32}$	$\frac{3}{16}$
(U,U)	$\frac{9}{16}$	0	0	0	0	0	0	0	0	0	1

appears to be correct to us, we believe that the problem lies in the specification of the parental Δ's. This can be illustrated by the following scenario: Suppose that we have a population in which $\frac{1}{4}$ of the individuals, C, are offspring (siblings) of the same first-cousin marriage and in which the remaining $\frac{3}{4}$ of the individuals, U, are noninbred unrelated individuals. This population has a mean inbreeding coefficient, α, of $\frac{1}{64}$. Then, if we randomly sample two individuals to form the parents of our sib pair, we see, from the data in table 1, that we have a *zero* chance of getting a parental pair in state S_2, whereas Génin and Clerget-Darpoux's Appendix A indicates that the parents should be in state S_2 with a *nonzero* probability $\alpha^2(1 - \alpha)$. If we judge on the basis of this slightly artificial example, it seems that one cannot specify correctly the condensed identity coefficients for the parents in terms of α alone; rather, one must take the *specific type and frequency* of consanguineous matings into account. (Similarly, one cannot recover condensed identity coefficients for a pair of individuals on the basis of knowledge of their kinship coefficient alone, since a parent-offspring pair has the same kinship coefficient as does a pair of siblings.) At least, one should use any available information about population sizes over time: Jacquard (1974, pp. 167–171) discusses how to adjust kinship coefficients properly for the background level of inbreeding due to finite population size (and, in fact, derives eq. [2] under slightly different modeling assumptions).

Finally, at the end of Génin and Clerget-Darpoux's paper, they state that the affected-pedigree-member (APM) method of linkage analysis (Weeks and Lange 1988) fails to take "full advantage of IBW states, since they only use a part of the information that concerns IBD between individuals" (Génin and Clerget-Darpoux 1996, p. 1158). Although this is true, we would like to point out that one of us (D.E.W.) has explored, in his dissertation, assaying for increased marker similarity *within* inbred individuals in the context of the APM method (Weeks 1988). The Appendix presented here contains a relevant (and slightly edited) extract regarding the theoretical development of this procedure. Note that to take full advantage of this extension of the APM method requires that the relationships of the affected individuals be known and specified. As our critique of Génin and Clerget-Darpoux's paper suggests, it may be difficult to properly analyze pedigrees from an inbred population, unless one devotes much effort to determining, as best as possible, the precise structure of each family.

Acknowledgments

This work was supported by the Wellcome Trust Centre for Human Genetics at the University of Oxford, the Association Française Contre Les Myopathies, NIH grant HG00719, and the University of Pittsburgh.

Daniel E. Weeks[1,2] and Janet S. Sinsheimer[1,3]
[1]*The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford;* [2]*Department of Human Genetics, University of Pittsburgh, Pittsburgh; and* [3]*Departments of Biomathematics and Biostatistics, University of California, Los Angeles, Los Angeles*

Appendix

Usually "a sample of unrelated individuals can give no information about linkage" (Smith 1953, p. 159), but this principle does not hold in the case of inbred individuals. For example, consider an affected offspring of a first-cousin marriage. If the disease is very rare and recessive, then the affected person is almost certainly homozygous by descent at the disease locus and so would also show increased homozygosity at any marker loci closely linked to the disease locus. If this increased homozygosity can be observed, then, as Smith (1953) first observed, a sample of inbred individuals may contain information about linkage. Lander and Botstein (1987) later elaborated this approach, naming it "homozygosity mapping."

The original APM statistic involves comparisons of marker identity-by-state (IBS) status *between* affected individuals (Weeks and Lange 1988). In order to apply

this statistic in the context of homozygosity mapping, it must be modified to include comparisons *within* individuals. The APM statistic was constructed in terms of a random variable, Z_{ij}, which measures the IBS marker similarity between two affected individuals, i and j. Let i have maternal marker allele G_{ix} and paternal marker allele G_{iy}. Likewise, let j have maternal marker allele G_{jx} and paternal marker allele G_{jy}. The original definition of Z_{ij}, for $i \neq j$, was

$$Z_{ij} = \frac{1}{4}\delta(G_{ix}, G_{jx})f(p_{G_{ix}}) + \frac{1}{4}\delta(G_{ix}, G_{jy})f(p_{G_{ix}}) + \frac{1}{4}\delta(G_{iy}, G_{jx})f(p_{G_{iy}}) + \frac{1}{4}\delta(G_{iy}, G_{jy})f(p_{G_{iy}}) ,$$

where the Kronecker delta is defined as

$$\delta(G, G') = \begin{cases} 1 & G \text{ and } G' \text{ match in state} \\ 0 & G \text{ and } G' \text{ do not match in state} \end{cases} .$$

Each match is weighted by the function $f(p)$ of the frequency of the allele involved. Z_{ij} can be interpreted as a conditional expectation. In the usual notation for conditional expectations,

$$Z_{ij} = \mathrm{E}[\delta(G_i, G_j)f(p_{G_i})|\text{obs. marker genotypes of } i \text{ and } j] ,$$

where G_i and G_j are randomly selected marker genes from i and j, respectively. If we permit i to equal j, then this definition gives

$$Z_{ii} = \frac{1}{4}f(p_{G_{ix}}) + \frac{1}{2}\delta(G_{ix}, G_{iy})f(p_{G_{ix}}) + \frac{1}{4}f(p_{G_{iy}}) .$$

To construct an analytically computable statistic, we need to find formulas for the mean and variance of Z_{ii}. The mean of Z_{ii} is easy to calculate by exploiting the logic of Weeks and Lange (1988):

$$\mathrm{E}(Z_{ii}) = \Phi_{ii}\sum_{k=1}^{n} p_k f(p_k) + (1 - \Phi_{ii})\sum_{k=1}^{n} p_k^2 f(p_k) .$$

In order to simplify this expression, let Φ_{FM} be the probability that the paternal and maternal genes of person i are IBD. Recall that $\Phi_{ii} = \frac{1}{2}(1 + \Phi_{\mathrm{FM}})$. Then,

$$\mathrm{E}(Z_{ii}) = \frac{1}{2}\sum_{k=1}^{n} p_k f(p_k) + \frac{1}{2}\sum_{k=1}^{n} [\Phi_{\mathrm{FM}} p_k f(p_k) + (1 - \Phi_{\mathrm{FM}})p_k^2 f(p_k)] .$$

From the intimidating formula (6) presented by Weeks and Lange (1988) for $\mathrm{E}[Z_{ij}Z_{kl}]$, we find that, when $i = j = k = l$,

$$\mathrm{E}(Z_{ii}Z_{ii}) = \left[\sum p_m f(p_m)^2\right]\Phi\{(G_i, G_i, G_i, G_i)\} + \left[\sum p_m f(p_m)\right]^2 \Phi\{(G_i, G_i)(G_i, G_i)\} + \left[\sum p_m^2 f(p_m)^2\right]\left\{\begin{array}{l} 2\Phi\{(G_i,G_i)(G_i,G_i)\} \\ +4\Phi\{(G_i,G_i,G_i)(G_i)\}\end{array}\right\} ,$$

where the $\Phi()$'s are generalized kinship coefficients as described by Weeks and Lange (1988). Now,

$$\begin{aligned}\Phi\{(G_i, G_i, G_i, G_i)\} &= P[(G_i, G_i, G_i, G_i) \mid \text{picked the same allele four times}] \\ &\quad \times P_{[\text{picked the same allele four times}]} \\ &\quad + P[(G_i, G_i, G_i, G_i)|\text{picked both alleles}] \\ &\quad \times P_{[\text{picked both alleles}]} = (1)\left(\frac{1}{8}\right) + \Phi_{\mathrm{FM}}\left(\frac{7}{8}\right) .\end{aligned}$$

Similar reasoning leads to $\Phi\{(G_i,G_i)(G_i,G_i)\} = \frac{1}{8}(1-\Phi_{\mathrm{FM}})$ and $\Phi\{(G_i, G_i, G_i), (G_i)\} = \frac{1}{8}(1 - \Phi_{\mathrm{FM}})$. Thus,

$$\mathrm{E}(Z_{ii}Z_{ii}) = \left[\sum p_m f(p_m)^2\right]\left[\frac{1}{8} + \frac{7}{8}\Phi_{\mathrm{FM}}\right] + \left[\sum p_m f(p_m)\right]^2\left[\frac{1}{8}(1 - \Phi_{\mathrm{FM}})\right] + \left[\sum p_m^2 f(p_m)^2\right]\left[\frac{6}{8}(1 - \Phi_{\mathrm{FM}})\right] .$$

Note that, in order to find $\mathrm{E}(Z^2)$, we also need to calculate terms such as $\mathrm{E}(Z_{ii}Z_{ij})$. However, in the computer program, it is easier to use the general formula (6) presented by Weeks and Lange (1988) than to use specific expressions for each special case.

For a pedigree, from the various statistics Z_{ij}, we form the overall statistic $Z = \Sigma_{i \leqslant j} Z_{ij}$. The mean and variance of Z may be computed by the following equations:

$$\mathrm{E}(Z) = \sum_{i \leqslant j} \mathrm{E}(Z_{ij})$$

and

$$\mathrm{E}(Z^2) = \sum_{\substack{i \leqslant j \\ k \leqslant l}} \mathrm{E}(Z_{ij}Z_{kl}) .$$

When combining the Z statistics from different pedigrees into the test statistic T (see Weeks and Lange 1988, eq. [7]), we cannot use the weights w_m in equation (8) of Weeks and Lange (1988), since we may now have pedigrees with only one inbred affected individual. If r_m is the number of affected and typed individuals in the mth pedigree and if Z_m is the Z statistic for this pedigree, then we choose to use $w_m = \sqrt{r_m} / \sqrt{\mathrm{Var}(Z_m)}$. There is no rigorous justification for this choice of weights. However, intuitively it seems better than giving all pedigrees equal weight.

Weeks (1988) investigated some sample applications of this extended APM statistic, which is based on measurement of IBS marker similarity *within* affected individuals as well as *between* pairs of affected relatives. Note that the computation of this statistic requires that the pedigree structure linking the affected relatives be known.

References

De Braekeleer M, Giasson F, Mathieu J, Roy M, Bouchard JP, Morgan K (1993) Genetic epidemiology of autosomal recessive spastic ataxia of Charlevoix-Saguenay in northeastern Quebec. Genet Epidemiol 10:17–25

De Braekeleer M, Perusse L, Cantin L, Bouchard JM, Mathieu J (1996) A study of inbreeding and kinship in intracranial aneurysms in the Saguenay Lac-Saint-Jean region (Quebec, Canada). Ann Hum Genet 60:99–104

Génin M, Clerget-Darpoux F (1996) Consanguinity and the sib-pair method: an approach using identity by descent between and within individuals. Am J Hum Genet 59: 1149–1162

Jacquard A (1974) The genetic structure of populations. Springer-Verlag, New York

Karigl G (1981) A recursive algorithm for the calculation of identity coefficients. Ann Hum Genet 45:299–305

Khoury MJ, Cohen BH, Diamond EL, Chase GA, McKusick VA (1987) Inbreeding and prereproductive mortality in the old order Amish. I. Genealogic epidemiology of inbreeding. Am J Epidemiol 125:453–461

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570

Smith CAB (1953) The detection of linkage in human genetics. J R Stat Soc B 15:153–184

Weeks DE (1988) New mathematical methods for human gene mapping. PhD dissertation, Department of Biomathematics, University of California, Los Angeles

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

Address for correspondence and reprints: Dr. Daniel E. Weeks, University of Pittsburgh, Department of Human Genetics, 130 DeSoto Street, A310 Crabtree Hall, Pittsburgh, PA 15261. E-mail: dweeks@watson.hgen.pitt.edu

0002-9297/98/6203-0033$02.00