American Journal of Epidemiology
Copyright © 2003 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

Vol. 158, No. 5
Printed in U.S.A.
DOI: 10.1093/aje/kwg152

Lee Responds to "Testing for Hardy-Weinberg Disequilibrium"

Wen-Chung Lee

From the Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan.

Received for publication April 28, 2003; accepted for publication May 8, 2003.

Abbreviation: HWT, Hardy-Weinberg disequilibrium test.

I appreciate Weinberg and Morris' thoughtful commentary (1) on my paper (2). In their article, they put my work under the perspective of gene mapping in the postgenomic era. I share the same view with them that the method proposed in my paper amounts to a tree-shaking approach to harvesting the high-hanging fruit (a low-cost approach to generating hypotheses aimed at localizing disease-susceptibility genes for complex human diseases). However, some issues raised by Weinberg and Morris (1) deserve scrutiny. These are 1) the power of the Hardy-Weinberg disequilibrium test (HWT) when a single-nucleotide polymorphism is a "marker" but is not a disease-susceptibility "gene" itself; 2) the utility of the proposed method as a gene-localization tool; and 3) the false alarm due to unmeasured ethnicity.

To address the first issue, consider a marker, *M*, which is in linkage disequilibrium with a disease-susceptibility gene, *A*. Jiang et al. (3) showed that, for the *M* marker, the Hardy-Weinberg disequilibrium coefficient in the affected population is (with the notations changed to be consistent with my paper (2)):

$$D = \left[\frac{q(1-f)}{R}\right]^2 \times (1-\theta)^{2t} \times (\Psi_2 - \Psi_1^2),$$

where f is the allele frequency of M in the source population, θ is the recombination fraction between M and A, t is the generation elapsed since the A gene was first introduced to the population, and q, R, Ψ_1, and Ψ_2 are defined the same as in my paper (2). The equation shows that the Hardy-Weinberg disequilibrium coefficient of the M marker decays according to the function, $(1-\theta)^{2t}$. However, the $\Psi_2 - \Psi_1^2$ term still appears in the equation, meaning that the effect of the mode of inheritance of the A gene is largely preserved even though we are looking at the M marker. Weinberg and Morris' assertion that "[s]uch a marker will display a gene-dose relation to risk, even if the linked risk-related gene for which it serves as a surrogate works according to a recessive or a dominant model" (1, p. 401), is therefore incorrect.

A second consequence of the above equation is that the Hardy-Weinberg disequilibrium coefficient, D, decays more quickly than the linkage disequilibrium coefficient, $\delta = q(1-f) \times (1-\theta)^t$, as the genomic distance between M and A increases (3). Thus, if a disease gene is not of too recent origin, a marker has to be closer to the gene to reach statistical significance using the HWT more than a marker has to be using the transmission/disequilibrium test. This implies that, in a Hardy-Weinberg population, a genome-wide HWT scan can fine map the putative disease-susceptibility gene(s), because in the very vicinity of the marker(s) with significant HWT, there may exist disease-susceptibility gene(s). This fine-mapping ability should be better for a HWT scan as compared with a transmission/disequilibrium test scan.

As for the problem of unmeasured ethnicity (hidden stratification), the "genomic control" method of Reich and Goldstein (4) can be used for a correction of the HWT. (Their method was proposed originally to correct the allelic chi-square statistic of a case-control design.) To be precise, a number of markers (e.g., 50 markers) are to be selected at random throughout the genome. It is unlikely that any such randomly selected marker will be tightly linked to a disease-susceptibility gene. Therefore, the mean square HWT (denoted as λ) of these "null markers" will be close to one if the population is a Hardy-Weinberg population. (A chi-square distribution with 1 df has the expectation of one.) On the other hand, λ will tend to be greater than one if the population is stratified. By the principle of multiplicative scaling of chi-square distribution (4), one refers the adjusted statistic, HWT^2/λ, to a 1-df chi-square distribution for each and every marker typed in the study. Such a correction procedure should reduce the number of false positive results.

Reprint requests to Dr. Wen-Chung Lee, Graduate Institute of Epidemiology, National Taiwan University, No. 1, Jen-Ai Road, Section 1, Taipei, Taiwan (e-mail: wenchung@ha.mc.ntu.edu.tw).

REFERENCES

1. Weinberg CR, Morris RW. Invited commentary: testing for Hardy-Weinberg disequilibirum using a genome single-nucleotide polymorphism scan based on cases only. Am J Epidemiol 2003;158:401–3.
2. Lee W-C. Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibirum in a gene bank of affected individuals. Am J Epidemiol 2003;158:397–400.
3. Jiang R, Dong J, Wang D, et al. Fine-scale mapping using Hardy-Weinberg disequilibrium. Ann Hum Genet 2001;65: 207–19.
4. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 2001;20:4–16.