

Address for correspondence and reprints: Dr. Rubin D. L. Hundscheid, Department of Human Genetics, University Hospital Nijmegen, P. O. Box 9101, Nijmegen, 6500 HB, The Netherlands. E-mail: R.Hundscheid@obgyn.azn.nl

© 2000 by The American Society of Human Genetics. All rights reserved.
0002-9297/2000/6701-0033\$02.00

Am. J. Hum. Genet. 67:258–259, 2000

Inflated False-Positive Rates in Hardy-Weinberg and Linkage-Equilibrium Tests Are Due to Sampling on the Basis of Rare Familial Phenotypes in Finite Populations

To the Editor:

If it is assumed that genotypes of some locus (G_D) are in Hardy-Weinberg equilibrium (HWE) in a population and that these genotypes are correlated with some phenotype (Ph), then, among “cases” in the tail of the distribution of Ph (equivalently, affected with rare disease), the G_D will show Hardy-Weinberg disequilibrium (HWD) (Nielsen et al. 1999; Deng et al. 2000; Göring and Terwilliger 2000). However, this does not imply that “generally, in individuals at either end of the quantitative-trait distribution, HWD exists if and only if there exists a whole-population LD [i.e., “linkage disequilibrium”]” (Deng et al. 2000, p. 1030). The “only if” part of this sentence is not correct. Even Deng et al. (2000, p. 1044) point out that “an absence of HWD does not imply that a marker locus and a QTL are not in LD” and that, for completely random marker loci, there will be inflated false-positive rates in tests for HWD (and LD as well), because “cases” of familial disease tend to be more related than “controls,” for the following reasons.

Assume that Ph is correlated in families, without specifying whether this is due to genetic or shared environmental factors. Let the prevalence, $\phi = P(\text{individual B is a case})$, and the familial relative risk $\lambda = P(\text{individual B is a case} | \text{relative A is a case}) / \phi$ (Weiss et al. 1982; Risch 1990). Then, $P(A \text{ and } B \text{ are affected} | A \text{ and } B \text{ are relatives}) = \lambda\phi^2$, and $P(A \text{ and } B \text{ are affected}) = \phi^2$ if they are randomly ascertained. This implies that $P(A \text{ and } B \text{ are relatives} | A \text{ and } B \text{ are affected}) = \lambda\phi^2 P(A \text{ and } B \text{ are relatives}) / \phi^2 = \lambda P(A \text{ and } B \text{ are relatives})$. If $\lambda > 1$, then ascertainment of “cases” ascertains relatives with greater probability than does random ascertainment of “controls,” leading to increased false-positive evidence of HWD and LD throughout the genome. This effect will be largest when λ is large, ϕ is small, and the population is small and/or structured (such that $P[A \text{ and } B \text{ are relatives}]$ is nontrivial). In a sense, this is related to the problem of population stratification when the phenotype

being studied correlates with a familial stratum, regardless of whether the trait is “genetic” (see Chase 1977).

If the “case” phenotype is a good predictor of G_D (a prerequisite for mapping to be powerful), then a large portion of the “case” sample will share some risk allele IBD from some common ancestor. The coalescent path connecting these chromosomes historically defines the most distant possible relationship among the “cases” carrying this allele, defining an upper bound on how “unrelated” they could possibly be. Again, this implies that ascertainment of affected individuals increases the probability of ascertainment of relatives. And the less frequent the shared risk allele is, the more closely related the “case” individuals will be (see Terwilliger, in press), leading to potential deviations from HWE and LE in unrelated parts of the genome as well.

The more closely related two people are, the larger the proportion of their genomes that they will share, as measured by their kinship coefficient (also see Terwilliger et al. 1997). If cases are “more related” than controls, then they will, with higher probability than will be seen in controls, share alleles IBD at random places in the genome, leading to increased false-positive rates in HWD and LD tests. This anticonservative behavior may be minor in studies of a single marker locus, but, when one considers the effects of testing hundreds of thousands of markers jointly in a genome scan, often making inferences based on the most significant values of the test statistic over the genome, the inflation of the type I error can have significant import. Furthermore, because the effect of small deviations, from HWE and/or LE, that are induced by such sampling is to shift the distribution slightly upward, the anticonservative bias will increase as we look farther out into the tail of the pointwise distribution (data not shown—but similar in shape to what appears in fig. 4 of Göring and Terwilliger 2000), leading to potentially gross inflation of genome-wide false-positive rates. To test for such problems, one can do a Monte Carlo randomization, as was done, in a case-control study of a small genetically homogeneous population isolate, by Hovatta et al. (1999), who kept the genotypes (for the whole genome scan) of all individuals constant and randomized their phenotypes (“case” and “control”). The simulation showed that their sample had approximately twice as many positives as would be expected from the randomization test, consistent with what is expected for reasons described in this note. When the fundamental assumption that “cases” and “controls” are independent and identically distributed with respect to random marker-locus genotype frequencies throughout the genome appears to have been rejected, it is essential to maintain skepticism in the interpretation of the results of such an analysis.

Unfortunately, the conditions in which “cases” are most likely to be relatives (e.g., small populations, rare

diseases, large familial correlations) are the same cases in which LD and HWD tests are likely to be useful (see Zöllner and von Haeseler 2000; Terwilliger, in press). In a study of more-common phenotypes and larger, more diverse populations, it is highly unlikely that marginal effects of single-risk alleles of a given locus are going to be etiologically important—in which case, LD and HWD tests will have little or no power (see Terwilliger and Weiss 1998; Terwilliger and Göring 2000; Weiss and Terwilliger, in press). And small populations with unusual histories are also more likely to have some population-level deviation from HWE in general, and, if one does not ascertain population controls, then there is no way to validate this critical assumption of the model. Although the paranoia about population stratification that leads people to mistrust case-control samples may be exaggerated, the absence of a sample of controls poses even greater danger.

Acknowledgments

Support from a Hitchings-Elion Fellowship from the Burroughs-Wellcome Fund is greatly appreciated. Discussions with Drs. Iris Hovatta, Harald H. H. Göring, John Blangero, Patrik Magnusson, and Kenneth M. Weiss are gratefully acknowledged.

JOSEPH D. TERWILLIGER

*Department of Psychiatry, Columbia University,
Columbia Genome Center, and Division of
Neuroscience, New York State Psychiatric Institute,
New York*

References

- Chase GA (1977) Genetic linkage, gene-locus assignment, and the association of alleles with diseases. *Transplant Proc* 9: 167–171
- Deng HW, Chen WM, Recker RR (2000) QTL fine mapping by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. *Am J Hum Genet* 66:1027–1045
- Göring HHH, Terwilliger JD (2000) Linkage analysis in the presence of errors. IV. Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 66:1310–1327
- Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajärvi R, Juvonen H, et al (1999) A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. *Am J Hum Genet* 65:1114–1124
- Nielsen DM, Ehm MG, Weir BS (1999) Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 63:1531–1540
- Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46:222–228
- Terwilliger JD. On the resolution and feasibility of genome scanning approaches to unraveling the genetic components of multifactorial phenotypes. In: Rao DC, Province MA (eds) *Genetic dissection of complex phenotypes: challenges for the new millennium*. Academic Press, New York (in press)
- Terwilliger JD, Göring HHH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 72:63–132
- Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE (1997) True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Am J Hum Genet* 61:430–438
- Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 9:578–594
- Weiss KM, Chakraborty R, Majumder PP, Smouse PE (1982) Problems in the assessment of relative risk of chronic disease among biological relatives of affected individuals. *J Chronic Dis* 35:539–551
- Weiss KM, Terwilliger JD. How many diseases do you have to study to map one gene with SNPs? *Nat Genet* (in press)
- Zöllner S, von Haeseler A (2000) A coalescent approach to study linkage disequilibrium between single-nucleotide polymorphisms. *Am J Hum Genet* 66:615–628

Address for correspondence and reprints: Dr. Joseph D. Terwilliger, Columbia University, Unit 109, 1150 St. Nicholas Avenue, Room 520C, New York, NY 10032. E-mail: jdt3@columbia.edu

© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6701-0034\$02.00

Am. J. Hum. Genet. 67:259–261, 2000

QTL Fine Mapping, in Extreme Samples of Finite Populations, for Complex Traits with Familial Correlation Due to Polygenes

To the Editor:

Recently, Deng et al. (2000) developed a QTL fine-mapping approach on the basis of Hardy-Weinberg (HW) or linkage-disequilibrium (LD) patterns in extreme samples of large and random-mating populations in which HW equilibrium holds. This approach is based on robust linkage results that have already localized a quantitative-trait locus (QTL) to a large genomic region (e.g., ~30 cM). The purpose is to fine map the QTL to a small region of ~1 cM, through examination of the patterns of deviation from HW and linkage equilibrium at a series of closely linked marker loci in extreme samples of populations. The deviation can be measured by a number of indices (including some test statistics—e.g., the test statistics for HW equilibrium) (Deng et al. 2000). Our approach is an extension of those of Feder et al. (1996) and Nielsen et al. (1998) for fine mapping of disease-