1998) and for a quantitative phenotype (van den Oord 2000).

Furthermore, some of the assessments made by Labuda et al. miss the mark. They assume (see their fig. 1) that under scenario A, where the offspring genotype is the one that "counts," the parents of affected children will resemble control parents with respect to the gene under study. This ignores the fact that the genotypes of parents and their children are correlated. Just as the parents of offspring with Huntington disease will differ from population controls in their prevalence of the allele for Huntington disease, parents of offspring who have a complex disease will tend to differ from population controls. Thus, the case-control analyses reported in table 1 of Labuda et al. (2002) are not specific to parentally mediated genetic effects.

There are other reasons, biologic and technical, to doubt the interpretation offered by Labuda et al., who suggest that their data support a parent-mediated effect of *CYP2E1*5* on risk of childhood acute lymphoblastic leukemia. First, the mechanisms by which the maternal and paternal genotypes would influence offspring phenotype are very different (i.e., in utero environment vs. DNA replication errors that produce genetically abnormal sperm). It thus seems unlikely that the etiology of a given condition would be related to both maternal and paternal effects of a single gene. Rather, similar "effects" of the maternal and paternal genotypes, on the basis of case-control parental data, seem more likely due to the selection of a biased control group or to offspring-mediated effects that have confounded the comparison of the (correlated) parental genotypes. Thus, the data offered by Labuda et al., which show very similar odds ratios for the mother and for the father, may be seen more plausibly as reflecting either a systematic bias in the control group or a chance finding.

The final issue is analytic. The odds ratio parameter estimated by the case-control analysis is not the same as that estimated by transmissions. Labuda et al. evidently used a standard method for paired data, calculating the ratio of counts for discordant transmission pairs based on heterozygous parents. This approach estimates the relative penetrance for carriers of a single copy of the variant allele under a gene dose model in which the relative penetrance for two copies is the square of that for one copy. By contrast, in their case-control analysis, Labuda et al. use carrier status, which presumes a dominant model. The paired estimator based on transmissions can be shown to be biased toward 1.0 under such a model. Even if the two analyses were estimating the same parameter, there is considerable overlap in the CIs for the two estimates. For these reasons, the results presented by Labuda et al. (2002) should be seen as providing only very weak evidence for a parent-mediated effect of *CYP2E1*5*.

CLARICE R. WEINBERG[1] AND LAURA MITCHELL[2]
[1]*National Institute of Environmental Health Sciences, Research Triangle Park, NC; and* [2]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia*

References

Labuda D, Krajinovic M, Sabbagh A, Infante-Rivard C, Sinnett D (2002) Parental genotypes in the risk of a complex disease. Am J Hum Genet 71:193–197

Lande R, Price T (1989) Genetic correlations and maternal effect coefficients obtained from offspring-parent regression. Genetics 122:915–922

Mitchell L (1997). Differentiating between fetal and maternal genotypic effects, using the transmission test for linkage disequilibrium. Am J Hum Genet 60:1006–1007

van den Oord EJCG (2000) The use of mixture models to perform quantitative tests for linkage disequilibrium, maternal effects, and parent-of-origin effects with incomplete subject-parent triads. Behav Genet 30:335–343

Weinberg C, Wilcox A, Lie R (1998) A log-linear approach to case-parent triad data: assessing effects of disease genes that act directly or through maternal effects, and may be subject to parental imprinting. Am J Hum Genet 62:969–978

Wilcox AJ, Weinberg CR, Lie RT (1998) Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads." Am J Epidemiol 148:893–901

Address for correspondence and reprints: Dr. C. R. Weinberg, MD A3-03, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709. E-mail: Weinberg@niehs.nih.gov

0002-9297/2002/7105-0029$15.00

Am. J. Hum. Genet. 71:1240–1242, 2002

Reply to Comments by Kraft and Wilson and by Weinberg and Mitchell on "Parental Genotypes in the Risk of a Complex Disease"

To the Editor:

Kraft and Wilson (2002 [in this issue]) point out that there are other analytical options to a joint application of case-control and TDT analysis in our study of the effect of parental genetics in the risk of a complex disease. They propose a "pseudo-sibling controls" design as an alternative to the approach proposed earlier by Weinberg and colleagues (1998) to study parental effects in case-parent trios. However, these tests are directed to evaluate the effect within a presumed model and are not designed to estimate joint effects of both parents' genotypes, which appeared to be the case with our data. Our study, inspired by original experimental observations, led us to understand the underlying genetic effects

that did not follow established paradigms. We concluded that a number of complementary strategies will need to be used simultaneously to dissect genetic predisposition to complex disorders (Labuda et al. 2002). In this regard, we are in agreement with Kraft and Wilson (2002 [in this issue]) that additional collecting of control-parent trios would extend possibilities of testing the observed effects under a greater variety of genetic and statistical models.

In the context of simple Mendelian disorders, fig. 1 could be troubling, but our paper was intended to divert the reader from this paradigm. Indeed, in a highly penetrant autosomal recessive condition, such as cystic fibrosis, in which two defective gene copies mean disease, collecting patients obviously identifies heterozygous parents, who otherwise would be difficult to find in such numbers in a control population. The example of Huntington chorea used by Weinberg and Mitchell (2002 [in this issue]) is rather unfortunate, since, at time of diagnosis, the case carrier-parent would already succumb to this disease. In a complex, multifactorial, and multilocus disease, in which the effect of a given allele is likely due to gene-gene (i.e., the presence of another variant at a different locus) and/or gene-environment interaction (e.g., exposure in only a fraction of carriers), one does not necessarily expect the enrichment in the at-risk alleles among patients' parents, expected in turn to transmit these alleles "preferentially" to their case-offspring (fig. 1*A*). In other words, we believe that figure 1 provides a good illustration of the experimental situation we faced.

We apologize for not giving satisfactory credit to earlier developments, which was pointed out by Weinberg and Mitchell (2002 [in this issue]). The fact that reference to Lande and Price (1989) is also absent among articles cited by Weinberg et al. (1998) is not an excuse. Rather, it reflects the fact that these excellent methodological contributions were reported in the absence of experimental data, in contrast, for example, to a recent paper by Infante-Rivard et al. (2002*b*) where both the sampling scenarios include control-parent trios as well as testing for maternally mediated effects.

Obviously, the mechanisms through which the maternal and paternal genotypes could influence child phenotype might be very different, but their net effect relevant to cancer risk, such as an increased mutation burden, need not be. In respect to this, different pathways controlling the metabolism of carcinogens, the level of oxidative stress, or the efficiency of DNA repair may have their unique contributions to the increase in the level of DNA lesions and, consequently, cancer risk. The observed effect with *CYP2E1*5* is, therefore, not at all unlikely. It is, however, possible that, for the effect to occur, the *CYP2E1*5* carriers would have to undergo a particular environmental exposure (Infante-Rivard et al. 2002*a*). However, as with all such results, this is the first report that will have to be confirmed by other studies that include different populations.

Here, our population of case and control subjects, both of French-Canadian origin, seems to be excellent for association studies because of the common genetics and lifestyle. Moreover, as presented in our report, we independently tested this population for a possibility of stratification that, in light of the recent results of Ardlie et al. (2002), appears to be less of a problem in appropriately designed studies. Kraft and Wilson (2002 [in this issue]) evaluated an underestimation of 11% in the odds ratio, related to the use of "surrogate" parental controls. This 11% arises from the elevated disease probability in the chosen numerical example and actually corresponds only to 1 SD in the variant frequency (0.250 ± 0.023) estimated in a sample of 350 chromosomes. Such extent of variation is expected under experimental conditions.

Weinberg and Mitchell (2002 [in this issue]) in their comments were also concerned by the effect of the *CYP2E1*5/*5* homozygotes. Because of the rarity (see table 1 in Labuda et al. 2002) of the variant in question, we did not need to consider the effect of its homozygotes. The dominant effect is, therefore, the one to be assumed to be consistent with presumed phenotypic outcome of this allele, leading to higher inducibility and, therefore, to higher activity of the enzyme (see references in Labuda et al. 2002).

For the reasons discussed above, we believe that our study provides solid evidence for the parental effects. It provides also an experimental illustration of genetic effects that, although escaping a simple Mendelian paradigm, were anticipated in earlier studies such as that of Weinberg and her colleagues (Weinberg et al. 1998). There is, therefore, no reason to believe that these effects should not be expected in other complex diseases.

Damian Labuda,[1,2] Maja Krajinovic,[1,2] Audrey Sabbagh,[1,*] Claire Infante-Rivard,[1,3] and Daniel Sinnett[1,2]

[1]*Centre de recherche, Hôpital Sainte Justine,* [2]*Département de pédiatrie, Université de Montréal, and* [3]*Department of Epidemiology and Biostatistics and Occupational Health, McGill University, Montreal*

References

Ardlie KG, Lunetta KL, Seielstad M (2002) Testing for population subdivision and association in four case-control studies. Am J Hum Genet 71:304–311

Infante-Rivard C, Krajinovic M, Labuda D, Sinnett D (2002*a*) Childhood acute lymphoblastic leukemia associated with parental alcohol consumption and polymorphisms of carcinogen-metabolizing genes. Epidemiology 13:277–281

Infante-Rivard C, Rivard GE, Yotov WV, Genin E, Guiguet M, Weinberg C, Gauthier R, Feoli-Fonseca JC (2002*b*) Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. N Engl J Med 347:19–25

Kraft P, Wilson M (2002) Family-Based Association Tests Incorporating Parental Genotypes. Am J Hum Genet 71: 1238–1239 (in this issue)

Labuda D, Krajinovic M, Sabbagh A, Infante-Rivard C, Sinnett D (2002) Parental genotypes in the risk of a complex disease. Am J Hum Genet 71:193–197

Lande R, Price T (1989) Genetic correlations and maternal effect coefficients obtained from offspring-parent regression. Genetics 122:915–922

Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 62: 969–978

Weinberg CR, Mitchell, L (2002) Regarding "Parental Genotypes in the Risk of a Complex Disease." Am J Hum Genet 71:1239–1240 (in this issue)

Address for correspondence and reprints: Dr. Damian Labuda, Centre de recherche, Hôpital Sainte-Justine, 3175 Côte-Sainte Catherine, Montréal, Québec H3T 1C5 Canada. E-mail: damian.labuda@umontreal.ca

* Present affiliation: CNRS-Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris, France.

0002-9297/2002/7105-0030$15.00

Am. J. Hum. Genet. 71:1242–1247, 2002

Partition-Ligation–Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms

To the Editor:

The mapping of SNPs in human genomes has generated a lot of interest from both the biomedical research community and industry. In conjunction with SNP mapping, researchers have shown that haplotypes possess considerably greater potential than the traditional single-SNP approach in disease-gene mapping and in our understanding of complex landscapes of linkage disequilibrium (LD) (Goldstein 2001). *In silico* methods for haplotype reconstruction have attracted much attention because of their cost-effectiveness and accuracy (Tishkoff et al. 2000) and have played an important role in the definition of human haplotype block structure and in candidate-gene studies of complex traits (Tabor et al. 2002). In a recent publication, Niu et al. (2002) proposed a partition-ligation (PL) strategy and implemented it together with Gibbs sampling, to estimate haplotype phases for a large number of SNPs. Although the resulting program, HAPLOTYPER, has been in high demand from many research groups, a significant portion of researchers are also strongly interested in using an expectation-maximization (EM)–based algorithm. In the present letter, we describe how to combine the PL strategy with the EM algorithm and how to handle the local-mode problem. We also present a fast and robust method of computing the variance of the estimated haplotype frequencies. Some related issues concern the handling of missing data and the multiple imputations of haplotype phases.

The EM algorithm is arguably the most popular statistical algorithm, because of its interpretability and stability. Compared to the Gibbs sampler, the EM approach is a deterministic procedure, requires less computing time, and is easier for convergence check. The output of the EM algorithm, if not trapped in a local mode, is the maximum-likelihood estimate (MLE), which possesses well-established statistical properties. However, the capability of most EM-based approaches is restricted to approximately one dozen loci, because of the memory constraint. A recently developed program, SNPHAP (see David Clayton's Web site [SNPHAP: A Program for Estimating Frequencies of Large Haplotypes of SNPs]), is an exception that, although different from the PL strategy, can handle many more linked loci by using a progressive-extension technique.

The essential steps of the PL strategy (Niu et al. 2002) are as follows: One first breaks down all of the marker loci into stretches of "atomistic" units and then uses either the EM algorithm or the Gibbs sampler to construct haplotypes for each unit and to rebuild the phase hierarchically, through a bottom-up approach. For example, an individual represented in the lipoprotein lipase (LPL) gene SNP data set (Nickerson et al. 1998) has the genotype (01200001000000000100010), where 0 stands for heterozygote and 1 and 2 stand for wild-type and mutant homozygotes, respectively. Since there are 18 heterozygous loci, the standard EM algorithm has to consider 2^{18} possible haplotypes, making it extremely costly for haplotype estimation. Using the PL strategy, we divide the linked loci into four "atomistic" units—(012000), (010000), (000001), and (00010)—and use the EM algorithm to estimate partial haplotypes within each unit. Afterward, two adjacent partial haplotypes are "ligated" by using the EM algorithm again, just like phasing two linked multiallelic markers. The ligation process is repeated until the complete phase is determined.

It is well known that the EM algorithm can be trapped in a local mode. This problem becomes a more serious issue for the PL-EM strategy, because every atomistic haplotype construction or ligation step involves a complete EM algorithm implementation. A naive implementation of the ligation step considers only the partial haplotypes that have nonzero estimated frequencies in the previous EM step. However, it appears that one phase configuration (and the corresponding haplotypes with nonzero es-