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

Likelihood Calculation Conditional on Observed Pedigree Structure

To the Editor:

The article by Vieland and Hodge (1995) has addressed the ascertainment problem from a theoretical point of view and has commendably explained the difference between the true and the observed likelihoods for a set of data. After pointing out the difference in the two likelihood expressions, they give an example, using a universe of sibships of size 3, to illustrate this difference. Their figure 1 shows a tree diagram for calculating $L_{\text{TRUE}}(\theta)$, and their figure 2 shows a tree diagram for calculating $L_{\text{OBS}}(\theta)$. We should like to clarify that the presence of figure 2 might mislead the reader, because it is irrelevant under the stated conditions.

The true likelihood is conditional on being ascertained and the true pedigree structure, whereas the observed likelihood is conditional on being ascertained and the observed pedigree structure. When the observed likelihood $L_{\text{OBS}}(\theta)$ is computed, the sampling rule for the relatives of the proband(s) is not taken into account and, hence, is irrelevant. We therefore find the tree diagram shown in Vieland and Hodge's figure 2, meant to aid the reader in the computation of $L_{\text{OBS}}(\theta)$, confusing. However, Vieland and Hodge correctly indicate, below the diagram, that the step from their level (ii) to their level (iii) in the tree diagram is not governed by the sampling rule and that the location of the probands within the sibships is irrelevant. Since the investigator constructs the likelihood conditional on the observed pedigree structure (i.e., on $s = 2$ or $s = 3$, where s is the observed sibship size in the example) with no allowance given for the sampling rule, the second column of Vieland and Hodge's table 1, which gives the probabilities for $L_{\text{OBS}}(\theta)$, is more simply obtained as follows, without referring to their figure 2.

First, there are three possibilities for a given subject: (1) not affected (probability $= 1 - \theta$); (2) affected proband (probability $= \theta\pi$); and (3) affected nonproband (probability $= \theta(1 - \pi)$).

When two sibs are observed, the first two entries of the second column of Vieland and Hodge's table 1 are, where $D_2 = 1 - (1 - \theta\pi)^2$,

$$\begin{aligned} &P(1 \text{ affected} \mid \text{ascertained}) \\ &\quad = P(1 \text{ proband and } 1 \text{ unaffected})/P(\text{ascertained}) \\ &\quad = 2\theta\pi(1 - \theta)/D_2, \end{aligned}$$

and

$$\begin{aligned} &P(2 \text{ affected} \mid \text{ascertained}) \\ &\quad = P(2 \text{ probands, or } 1 \text{ proband and} \\ &\qquad 1 \text{ affected nonproband})/P(\text{ascertained}) \\ &\quad = [(\theta\pi)^2 + 2\theta\pi\theta(1 - \pi)]/D_2 \\ &\quad = \theta^2[1 - (1 - \pi)^2]/D_2. \end{aligned}$$

Similarly, when three sibs are observed, the last three entries of the second column of Vieland and Hodge's table 1 are, where $D_3 = 1 - (1 - \theta\pi)^3$,

$$\begin{aligned} &P(1 \text{ affected} \mid \text{ascertained}) \\ &\quad = P(1 \text{ proband and } 2 \text{ unaffected})/P(\text{ascertained}) \\ &\quad = 3\theta\pi(1 - \theta)^2/D_3 , \end{aligned}$$

$$\begin{aligned} &P(2 \text{ affected} \mid \text{ascertained}) \\ &\quad = P(1 \text{ proband, } 1 \text{ affected nonproband and } 1 \\ &\qquad \text{unaffected, or } 2 \text{ probands and } 1 \text{ unaffected})/ \\ &\qquad P(\text{ascertained}) \\ &\quad = \{6\theta\pi[\theta(1 - \pi)](1 - \theta) + 3(\theta\pi)^2(1 - \theta)\}/D_3 \\ &\quad = 3\theta^2(1 - \theta)[1 - (1 - \pi)^2]/D_3 , \end{aligned}$$

and

$$\begin{aligned} &P(3 \text{ affected} \mid \text{ascertained}) \\ &\quad = P(1 \text{ proband and } 2 \text{ affected nonprobands,} \\ &\qquad 2 \text{ probands and } 1 \text{ affected nonproband,} \\ &\qquad \text{or } 3 \text{ probands})/P(\text{ascertained}) \\ &\quad = \{3\theta\pi[\theta(1 - \pi)]^2 + 3(\theta\pi)^2\theta(1 - \pi) + (\theta\pi)^3\}/D_3 \\ &\quad = \theta^3[1 - (1 - \pi)^2]/D_3 . \end{aligned}$$

We stress that the purpose of this letter is merely to make the example presented by Vieland and Hodge (1995) easier to understand and that it in no way detracts from the main point of their paper—namely, that the observed likelihood is correct for ascertained data only in very special situations. As genetic epidemiology moves—whether by segregation or linkage analysis or a combination of both—away from the mere detection of trait genes, toward the precise estimation of their effects (e.g., in terms of relative risks or attributable risks), it will be necessary, in order to obtain good parameter estimates, to design studies for which the observed likelihood (which is what we can calculate) is as close as possible to the intractable true likelihood. By under-

standing how these two likelihoods may differ when we ascertain large pedigrees, we can be guided toward sampling and analytical techniques that minimize their difference (Elston 1995). Furthermore, current attempts to locate, by model-based linkage analysis, genes underlying complex traits can be made more powerful by the use of a realistic model for the genetic mechanism underlying the trait phenotype. Clerget-Darpoux et al. (1986) showed how the LOD-score profile may be affected by misspecification of various genetic parameters, leading to biased estimates of the recombination fraction. In order to estimate the most appropriate trait models for linkage analysis, it is necessary both to allow for familial correlations due to causes other than the locus to be linked (Demenais and Lathrop 1993) and to take proper account of ascertainment considerations. Linkage analysis of multigenerational data allowing for residual correlations is implemented in the program package S.A.G.E. (1997), but only careful planning at the time when data are collected (Elston 1995; Zhao et al. 1997) will ensure that relevant likelihoods can be formulated.

Acknowledgments

We thank Drs. Veronica Vieland and Susan Hodge for clarifying their paper for us. This study was supported in part by U.S. Public Health Service research grant GM 28356 from the National Institute of General Medical Sciences, by National Heart, Lung and Blood Institute training grant HL 07567, and by National Center for Research Resources resource grant 1 P41 RR 03655.

P. Mahinda Karunaratne and Robert C. Elston
Department of Epidemiology and Biostatistics
Rammelkamp Center for Education and Research
MetroHealth Campus
Case Western Reserve University
Cleveland

References

Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J (1986) Effects of misspecifying genetic parameters in lod score analysis. Biometrics 42:393–399

Demenais F, Lathrop M (1993) Use of the regressive models in linkage analysis of quantitative traits. Genet Epidemiol 10:587–592

Elston RC (1995) Twixt cup and lip: how intractable is the ascertainment problem? Am J Hum Genet 56:15–17

S.A.G.E. (1997) Statistical analysis for genetic epidemiology, release 3.0. Computer program package available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland

Vieland VJ, Hodge SE (1995) Inherent intractability of the ascertainment problem for pedigree data: a general likelihood framework. Am J Hum Genet 56:33–43

Zhao LP, Hsu L, Davidov O, Potter J, Elston RC, Prentice RL (1997) Population-based family study designs: an interdisciplinary research framework for genetic epidemiology. Genet Epidemiol 14:365–388

Address for correspondence and reprints: Dr. Robert C. Elston, Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland, OH 44109. E-mail: rce@darwin.cwru.edu

0002-9297/98/6203-0036$02.00