have selected a single model a priori, often on the basis of very weak epidemiological data, and have confined linkage analyses exclusively to evaluations assuming this model. The suggested advantage of this approach lies in avoiding concerns about inflation of the LOD score because of "multiple comparisons."

We acknowledge, as discussed in our paper, that adjustment of *P*-values associated with LOD scores obtained after testing multiple analytical models is no simple matter. However, despite this disadvantage, we very strongly advocate the testing of a wide range of alternative models varying both disease classification definitions and disease locus parameters. We view such analyses as exploratory in nature and accept that definitive evidence of a gene responsible for a highly complex and heterogeneous disorder will therefore almost certainly require replication using independent data. The independent data could consist either of additional information obtained from the original collection of pedigrees by using additional markers located in the candidate region to increase net marker informativeness close to 1.0 or of replication using another collection of families. We favor the exploratory data analysis strategy, because we fear the negative consequences of failing to detect a true linkage more than we do the chance of temporarily pursuing a false-positive LOD score generated as an artifact of multiple analyses. Our exploratory analysis strategy is exemplified in our recently reported linkage studies of schizophrenia (Su et al. 1993) and early-onset periodontitis (Hart et al., in press). In both studies we searched for linkage by using 12 different models while also allowing for the possibility of locus heterogeneity.

With respect to the other point raised by Hodge et al., in which they question our assertion that phenocopies are quite rare in high-density pedigrees (MacLean et al. 1993*a*, p. 355), we both agree and disagree. It seems that the answer depends on precisely what kind of "high-density" pedigree is considered. We do not question the validity of the interesting data presented by Durner et al. (1992) which clearly indicate that, under some sampling schemes, intrafamilial disease locus heterogeneity may often occur. However, their sampling strategy resulted in very high levels of intrafamilial heterogeneity only if more than five affected individuals were required from each family, which consequently led to affected individuals being distributed as first-, second-, and third-degree relatives. In contrast, we have explored levels of intrafamilial heterogeneity for nuclear families consisting of six siblings, with three affected (MacLean et al. 1993*b*). For such samples, two disease loci (both dominant and with equal allele frequencies and with penetrance of 80%) virtually never segregated in the same family if they act independently. Intrafamilial disease locus heterogeneity occurred at high levels only for models with very strong epistasis, as would be expected.

SCOTT R. DIEHL AND CHARLES J. MACLEAN
Departments of Psychiatry and Human Genetics
Medical College of Virginia
Virginia Commonwealth University
Richmond

References

Durner M, Greenberg DA, Hodge SE (1992) Inter- and intrafamilial heterogeneity: effective sampling strategies and comparison of analysis methods. Am J Hum Genet 51: 859–870

Hart TC, Marazita ML, Schenkein HA, McCanna KS, Diehl SR. Exclusion of the candidate region on chromosome 4q for early onset periodontitis. Hum Genet (in press)

MacLean CJ, Bishop DT, Sherman SL, Diehl SR (1993*a*) Distribution of lod scores under uncertain mode of inheritance. Am J Hum Genet 52:354–361

MacLean CJ, Sham PC, Kendler KS (1993*b*) Joint linkage of multiple loci for a complex disorder. Am J Hum Genet 53:353–366

Su Y, Burke J, O'Neill A, Murphy B, Nie L, Kipps B, Bray J, et al (1993) Exclusion of linkage between schizophrenia and the D2 dopamine receptor gene region of chromosome 11q in 112 Irish multiplex families. Arch Gen Psychiatry 50: 205–211

0002-9297/93/5303-0023$02.00

Am. J. Hum. Genet. 53:777–779, 1993

Is MoM Bashing Justified?

To the Editor:

Recently, Bishop et al. (1993) suggested that it is inappropriate to use multiples of the median (MoM) for interpreting maternal serum alpha-fetoprotein (MSAFP) measurements. This prompted us to reexamine this issue once again.

Bishop et al. conclude that SDs of MSAFP measurements differ by gestational age at 15–19 wk. This does not agree with data from our laboratory. Table 1 shows a similar analysis for 9,126 initial serum samples from

Table 1

Population Parameters for MSAFP Measurements, by Completed Week of Gestation, Before and After Outliers Are Trimmed

Gestational Age (completed wk)	Median MSAFP (IU/ml)	No. of Pregnancies	SD (log_{10})	No. of Outliers[a]	Trimmed SD (log_{10})[b]
15	26.9	903	.202	4	.194
16	29.4	4,265	.185	6	.181
17	34.0	2,526	.177	2	.175
18	38.0	898	.180	1	.175
19	43.0	371	.189	2	.176
20	49.2	163	.220	2	.181

[a] Seventeen observations were trimmed (6 values >250 IU/ml and 11 values <5 IU/ml).
[b] By Levene's test for equal variances ($F=1.66$; df=5, 9103; $P=.14$).

women with unaffected singleton pregnancies screened at the Foundation for Blood Research and published elsewhere (Haddow et al. 1992). Prior to trimming, the SDs are significantly higher at both 15 and 20 wk gestation than at 16–19 wk. After trimming 17 of the most extreme values (<0.2%), only slight variability remains in the SDs, there is no apparent trend, and the differences are not statistically significant. Extreme values such as those trimmed are known to be associated with fetal death, multiple gestations, and badly misdated pregnancies. Bishop et al. may not have trimmed their data. Inclusion of outliers could distort the SDs, especially when there are relatively few observations.

Bishop et al. show a probability plot of MSAFP levels at 16 wk gestation. They conclude that the data fit a log-Gaussian distribution only between the 3d and 98th centiles. Figure 1 displays the 4,265 MSAFP measurements at 16 wk gestation in the present data set. Overall, our fit is better, extending between the 1st and 99th centiles. The deviation from the log-Gaussian distribution in extreme centiles for both data sets can be dealt with by the current standard practice of using truncation limits for the MoM values (Palomaki and Haddow 1987; Wald et al. 1992).

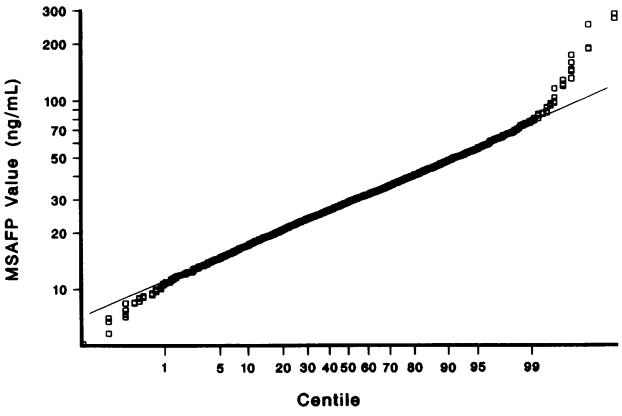

Figure 1 Expected vs. actual distribution of MSAFP measurements in 4,265 singleton pregnancies at 16 wk gestation, under a log-Gaussian distribution.

The authors conclude that "mixing" gestational week-specific distributions of MoM levels (with varying SDs) causes a "difference" between the observed (3.9%) and expected (4.2%) percentage of women assigned an MoM of ≤0.5. They have not tested the statistical significance of this small difference. The 95% confidence interval (CI) (3.5%–4.4%) of the observed percentage includes the expected value. Thus, the "difference" is not statistically significant. Similarly, the observed 1.4% (95% CI 1.1%–1.7%) of MSAFP levels at or above 2.5 MoM is not different from the expected 1.2%. The expected and observed percentages at each gestational week also do not differ beyond what can be explained by chance.

Bishop et al. do not refer to much of the literature that relates directly to the use of MoM in MSAFP screening. For example, they suggest that pooling the data in order to estimate MSAFP variance in pregnancies with Down syndrome was done without considering whether the variances were equal at each gestational week. This issue was addressed in 1984 (Cuckle et al. 1984) and again in 1987 (Wald et al. 1987). In the latter analysis, 240 cases of Down syndrome from four published and three unpublished data sets were stratified by completed week of gestation. There were no statistically significant differences in the variances. As a second example, Bishop et al. suggest that the MoM may not be useful, because the medians (and therefore the MoM levels) are influenced by such factors as maternal

weight, race, and insulin-dependent diabetes. In fact, it is currently standard practice to take these and other factors into account in both unaffected and affected populations (Wald et al. 1984, 1987; Cuckle et al. 1990). Any other interpretative unit (such as centiles, Z scores, or risk) will also need to take these variables into account.

Converting MSAFP measurements to MoM levels has proved to be worthwhile during the past 15 years, both in practice and for research purposes. No other normalization methodology has been shown to be as useful for this area of prenatal screening. The statement by Bishop et al. (1993, p. 429) that "other statistical methods should be sought" carries an obligation to provide a recommended alternative for expert evaluation.

GLENN E. PALOMAKI, GEORGE J. KNIGHT, AND JAMES E. HADDOW
Foundation for Blood Research
Scarborough, ME

References

Bishop JC, Dunstan FDJ, Nix BJ, Reynolds TM, Swift A (1993) All MoMs are not equal: some statistical properties associated with reporting results in the form of multiples of the median. Am J Hum Genet 52:425–430

Cuckle HS, Wald NJ, Lindenbaum RH (1984) Maternal serum alpha-fetoprotein measurement: a screening test for Down's syndrome. Lancet 1:926–929

Cuckle HS, Wald NJ, Stevenson JD, May HM, Ferguson-Smith MA, Ward AM, Barbour HM, et al (1990) Maternal serum alpha-fetoprotein screening for open neural tube defects in twin pregnancies. Prenat Diagn 10:71–77

Haddow JE, Palomaki GE, Knight GJ, Williams J, Pulkkinen A, Canick JA, Saller DN, et al (1992) Prenatal screening for Down's syndrome with use of maternal serum markers. N Engl J Med 327:588–593

Palomaki GE, Haddow JE (1987) Maternal serum α-fetoprotein, age, and Down syndrome risk. Am J Obstet Gynecol 156:460–463

Wald NJ, Cuckle HS (1987) Recent advances in screening for neural tube defects and Down's syndrome. Bailliere's Clin Obstet Gynaecol 1:649–676

Wald NJ, Cuckle HS, Boreham J (1984) Alpha-fetoprotein screening for open spina bifida: effect of routine biparietal diameter measurement to estimate gestational age. Rev Epidemiol Sante Publique 32:62–68

Wald NJ, Cuckle HS, Densem JW, Kennard A, Smith D (1992) Maternal serum screening for Down's syndrome: the effect of routine ultrasound scan determination of gestational age and adjustment for maternal weight. Br J Obstet Gynaecol 99:144–149

0002-9297/93/5303-0024$02.00

Am. J. Hum. Genet. 53:779–781, 1993

Reply to Palomaki et al.

To the Editor:

Before giving a detailed reply, we would like to stress that, while our article appeared negative when investigating the statistical properties of MoMs, we are committed to trying to find a more robust approach, since it is our belief that advances in screening could well be limited by the type of analyses performed; we are not simply engaging in "MoM bashing." Our paper was primarily written to clarify the assumptions that are needed for MoM analyses to be justified and not to imply that such assumptions could never be met. For the record, we would agree with the authors that the MoM approach has helped advance a better understanding of screening for Down syndrome over the past 15 years, but we also believe that constant reassessment of current techniques, both biochemical and statistical, is necessary if the momentum is to be maintained.

The authors offer a reason as to why our SDs of log(MoM) might be different. Our data were derived by using ultrasound measurements based on BPD. The data were screened for fetal death and multiple gestations. Unexplained outliers were also removed, this accounting for 0.09% of the original data set. We think our data are as clean as possible, and consequently the derived estimates should be realistic summary measures. One of the reviewers of our paper also raised the point about testing the significance of the variation in our SD values. To answer the question, we performed a formal statistical test and found that the variation was highly significant. We argued that the result of this test should not be included in our paper, since it may have encouraged readers to believe that we were suggesting that a linear trend in the SDs was present in all data sets. We did not wish to give this impression, since, while the data set that we had used seemed to indicate this, we had seen other data sets with other patterns, including a smoothly curved profile as indicated by Palomaki et al.'s own data. We find the preponderance of these