appeal. Here, the stage 2 LOD score has no real meaning but just serves to act as a benchmark that must be attained. Values of 3.0, 3.3, and 3.6 are used, but no justification is given for choosing them; nor could there be one. It would be perfectly reasonable to find the power of a procedure to attain a certain LOD score or to attain a statistic having a certain *P* value (type I–error probability). What makes no sense is to aim to attain a certain LOD score in the second stage but then to achieve a specified type I–error probability overall by manipulating the threshold applicable for the first stage. The more natural approach, which I am sure would yield completely different results with regard to power and efficiency, would be to fix the threshold for the first stage (probably at .5–1.0) and, in the second stage, to aim either for some predetermined LOD score or for a LOD producing a certain overall *P* value.

The effects of Holmans and Craddock's approach are clear to see. The stage 1 criterion has to be made high enough so that only a small number of unlinked regions will achieve it and hence go on to produce false-positive results in stage 2. The more subjects and markers that are typed in the first stage, the more likely it is that high LOD scores will be thrown up by chance, and hence the higher the stage 1 criterion must be set. The higher this criterion is, the harder it may be for a truly linked locus to achieve it, and so such loci may be more frequently discarded. Thus, doing more genotyping in the first stage generally leads to a reduction in power, despite involving an increase in the total amount of genotyping required.

The first scenario that Holmans and Craddock present illustrates this clearly. A wide, 20-cM grid is used for stage 1, narrowing to 10 cM in stage 2, and the LOD score to be taken to indicate linkage, after stage 2, is chosen to be 3. When only 100 of the 200 sib pairs are typed, the threshold to move from stage 1 to stage 2 is set to a modest and sensible .89, and the overall power is .62. However, when all 200 pairs are typed, the stage 1 threshold has to be raised to 2.14, so many true linkages are missed, and the power falls to .57. Using 100 pairs together with their parents needs a threshold of 1.57 and yields a power of .52. Finally, initially using all 200 pairs and their parents apparently demands a stage 1 threshold of 3.1 and has a power of only .54. This would mean that, if one got a LOD of 3.05 with the initial 20-cM grid scan, one would not follow up this finding, even though it would count as a positive result if it were to be found in stage 2.

Given that genotyping is becoming ever cheaper and easier, given that linkage can easily be missed in sib-pair samples, and given that performing a genome scan but missing a disease locus is highly undesirable, my own personal view is that the initial scan should probably be fairly thorough, using all available subjects and a relatively narrow marker grid.

DAVID CURTIS
Department of Psychological Medicine
St. Bartholomew's and Royal London School of Medicine and Dentistry
London

Reference

Holmans P, Craddock N (1997) Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. Am J Hum Genet 60:657–666

Address for correspondence and reprints: Dr. David Curtis, Department of Psychological Medicine, St. Bartholomew's and Royal London School of Medicine and Dentistry, 3d Floor Alexandra Wing, Turner Street, London E1 2AD. E-mail: dcurtis@hgmp.mrc.ac.uk; URL: www.gene.ucl.ac.uk/~dcurtis

0002-9297/98/6201-0032$02.00

EDITOR'S NOTE.—This letter is a truncated version of the letter submitted by Dr. Curtis.

Am. J. Hum. Genet. 62:205–207, 1998

Reply to Curtis

To the Editor:

Curtis (1998 [in this issue]) has raised some criticisms regarding our paper on efficient strategies for genome screening for linkage (Holmans and Craddock 1997). We reply to them as follows:

Curtis has said that our decision to fix the stage 2 criterion "lacks any intrinsic appeal" and "makes no sense." However, we would like to point out that judgements regarding the significance of a linkage study are generally based on the final LOD score obtained. Therefore, to facilitate comparison between the various strategies, it is desirable that a given stage 2 LOD score should correspond to the same significance level in all the strategies, as far as possible. This can most easily be done by fixing the stage 2 criterion and varying the stage 1 criterion, to obtain the desired type I error probability. In practice, one would not regard such criteria as benchmarks of "significant" versus "nonsignificant" linkage—their purpose is to ensure a fair comparison of the power of the various strategies and as a guide to which LOD scores correspond at *P* value of $\leqslant .05$.

We chose 3.6 as one of our criteria since this was recommended by Lander and Kruglyak (1995) as corresponding to a genomewide *P* value of .05 and is in widespread use. The criterion of 3.0 was chosen as the traditional criterion for significant linkage. The criterion 3.3 was adopted when it became clear that 3.6 was too stringent for the strategies to give a *P* value of .05. It is clear from our results that higher criteria would make

it impossible for a type I error *P* value of .05 to be obtained, whereas lower criteria would reduce efficiency by requiring an excessively high stage 1 criterion.

As noted by Curtis, it is also possible to fix the stage 1 criterion and to vary the stage 2 criterion, although we do not see why this approach should be regarded as "more natural.". What is more important, we disagree with Curtis's comment (made without any justification) that such an approach "would yield completely different results", provided that the stage 2 criteria were chosen to produce the same overall *P* value. Our reasoning is as follows: Strategies using narrow grids in stage 1 will produce more false-positive results to be followed up in stage 2, because of the greater number of loci being tested. Therefore, a higher stage 2 criterion will be necessary, thereby reducing power. In addition, the degree of dependence between the stage 1 test and the stage 2 test is increased for strategies in which a large proportion of the total sample is typed in stage 1, because of the similarity in the data analyzed in the two stages. This means that loci giving false-positive results in stage 1 are more likely also to give high LOD scores in stage 2. Again, a higher stage 2 criterion will be needed. It is therefore quite possible that strategies involving a large amount of genotyping in stage 1 may not increase power, as we found in our simulations.

Following Curtis's suggestion, we investigated the power of the various strategies when, as suggested, the stage 1 criterion was fixed and the stage 2 criterion was varied to give a type I error *P* value of .05. A stage 1 criterion of .9 was used—this is similar to the stage 1 criteria of the best-performing tests in our original paper and also is within the range suggested by Curtis. The results are displayed in table 1. Fixing the stage 1 criterion certainly improves the performance of the strategies in which the whole sample (200 pairs + parents) is typed in stage 1, compared with the results displayed in our original paper. However, it can still be seen that typing the whole sample in stage 1 gives, at best, a minimal power increase over that of strategies in which only the affected pairs are typed, while requiring considerably more genotyping. Use of a tight (10-cM) grid in stage 1 gives no more power than the use of a wide (20-cM) grid, and it requires a large increase in genotyping. The main conclusions of our original paper—that is, that both sample-splitting and grid-tightening increase efficiency and that a relatively wide initial grid is preferable to a narrow one—therefore stand.

Of course, if the stage 2 criterion is "predetermined," as suggested by Curtis, then strategies involving large amounts of genotyping in stage 1 will give the highest "power"—that is, the highest number of LOD scores exceeding the criterion. However, this would be meaningless, since such strategies would also have the highest number of false positives.

Table 1

Performance of Two-Stage Strategies

Grid and Stage 1 Sample [a]	Stage 2 LOD Criterion [b]	Power [c]	No. (SD) of Genotypings
20 cM/10 cM:			
100 Pairs	2.96	.622	47,359 (400)
200 Pairs	3.15	.709	78,166 (333)
100 Pairs + parents	3.10	.665	78,493 (350)
200 Pairs + parents	3.15	.722	140,134 (265)
20 cM/5 cM:			
100 Pairs	3.10	.652	59,180 (705)
200 Pairs	3.25	.782	90,829 (619)
100 Pairs + parents	3.19	.717	91,209 (654)
200 Pairs + parents	3.41	.778	154,671 (562)
10 cM/5 cM:			
100 Pairs	3.19	.733	94,174 (627)
200 Pairs	3.41	.783	156,274 (527)
100 Pairs + parents	3.19	.785	155,800 (564)
200 Pairs + parents	3.41	.786	278,908 (423)

[a] Grid-tightening strategies are denoted as the intermarker interval in stage 1 (to the left of the slash) followed by that in stage 2 (to the right of the slash). The stage 1 LOD-score criterion is .9; $\lambda_s = 2$.

[b] Fixed to make genomewide type I error probability ~.05.

[c] SD $\leqslant$.01.

By means of the reasoning mentioned above, the explanation for the seemingly counterintuitive results noted by Curtis becomes apparent. Strategies in which a large number of loci are typed in stage 1 require a higher stage 1 test criterion, to restrict the number of false positives being tested in stage 2 to the correct level. Similarly, strategies using a high proportion of the sample in stage 1 require an increased stage 1 test criterion, to offset the increased dependence of the stage 1 and stage 2 tests. We can therefore see that strategies that utilize a "more thorough search in the first stage" require a higher stage 1 criterion, for a given type I error probability. This is why the power of such strategies may be reduced, despite the increase in the amount of genotyping, as Curtis himself notes. The numerical example that Curtis gives illustrates the point perfectly.

As Curtis notes, in the extreme case, in which the entire sample is genotyped in stage 1, the required stage 1 criterion may actually exceed the stage 2 criterion. However, we are not recommending such strategies; we are merely quoting the stage 1 criteria that make these strategies as efficient as possible, given the stage 2 criterion. These strategies perform very poorly, requiring a large amount of genotyping and giving only low power. In fact, the implausibly high stage 1 criteria should act as a warning not to use these strategies—in the Discussion section of our original paper (Holmans and Crad-

dock 1997), we note that the best-performing strategies have stage 1 criteria of ~1.

It is true that we have not considered multipoint analysis, and this would be an interesting area for further work. Given that our conclusions hold up under two-locus analysis (Holmans and Craddock 1997), we would be hopeful that they would also be true under multilocus analysis.

In conclusion, although there is certainly scope for further work—particularly that involving multilocus analysis—we disagree strongly with Curtis's statements that our work is "fundamentally flawed" and that our results "do not provide useful information." The results presented here show that the alternative approach advocated by Curtis would result in conclusions similar to ours, despite his assertions to the contrary.

PETER HOLMANS AND NICK CRADDOCK*
Division of Psychological Medicine
University of Wales College of Medicine
Heath Park
Cardiff

References

Curtis D (1998) Efficient strategies for genome scanning with affected sib pairs. Am J Hum Genet 62:000–000 (in this issue)

Holmans P, Craddock N (1997) Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. Am J Hum Genet 60:657–666

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Address for correspondence and reprints: Dr. Peter Holmans, Division of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, Wales. E-mail: wpcpah@cardiff.ac.uk

* Present affiliation: Department of Psychiatry, University of Birmingham, Birmingham, United Kingdom.

0002-9297/98/6201-0033$02.00