

Meta-Analysis of Linkage Data under Worst-Case Conditions: A Demonstration Using the Human *OB* Region

David B. Allison and Moonseong Heo

Obesity Research Center, St. Luke's/Roosevelt Hospital Center, Columbia University College
of Physicians & Surgeons, New York, New York 10025

Manuscript received February 28, 1997

Accepted for publication October 23, 1997

ABSTRACT

To date, few methods have been developed explicitly for meta-analysis of linkage analyses. Moreover, the methods that have been developed or suggested generally depend on certain ideal situations and have not been widely applied. In this article, we apply standard statistical theory and meta-analytic techniques in novel ways to five published papers discussing the evidence of linkage of body mass index (BMI) to the region of the human genome containing the *OB* gene. These methods are "inference based," meaning that they allow one to make statements about the statistical significance of the entire body of evidence. As currently developed, they do not allow specific statements to be made about the amount of variance explained by any putative locus or allow precise confidence intervals to be placed around the putative location of a linked locus. By applying these techniques to the literature on linkage in the human *OB* gene region, we are able to show that the evidence for linkage somewhere in the region is extremely strong ($P = 1.5 \times 10^{-5}$).

POOLING data to increase the precision of one's estimates and conclusions dates back at least to the early 1900s (Pearson 1904, 1933a,b; Tippett 1931; Fisher 1954) and was termed *meta-analysis* by Glass (1976). Since then, the use of meta-analysis has increased dramatically. However, there has been relatively little use of meta-analysis in assessing evidence of linkage between human diseases and genetic markers.

Recently, several authors have suggested the possibility of developing methods for meta-analysis of linkage studies. Lander and Kruglyak (1995, p. 245) state "careful meta-analysis of *all* studies may be useful to assess whether the overall evidence for linkage is convincing. . . . To combine results among studies it is always best to pool the raw data and re-analyze the entire data set. Lod scores can be added across studies, but only when they are computed with the same methods, with the same set of markers, and at the same map position." Lander and Kruglyak (1995) then state that other meta-analytic methods are available and cite Cox and Hinkley (1974). The meta-analytic technique mentioned by Cox and Hinkley (1974, p. 80) involves combining *P* values from several independent samples as outlined by Fisher (1954, p. 99). Recently, Li and Rao (1996) presented a meta-analytic method in which each study used the Haseman-Elston procedure (Haseman and Elston 1972) and the same markers.

We agree with Lander and Kruglyak (1995) that the preferable way to meta-analyze data is by obtaining the raw data from each study and conducting a pooled analysis (Jeng *et al.* 1995). In the context of pooling raw data, multipoint procedures (*e.g.*, Kearsley and Hyne 1994; Hyne and Kearsley 1995; Fulker *et al.* 1995; Xu and Atchley 1995) will have great utility. However, in many situations this is not feasible and the meta-analyst must work with the data available in published reports. Moreover, Li and Rao's (1996) method is limited in that not every study uses the Haseman-Elston procedure or the same markers. Herein, we present meta-analytic techniques that can be used under worst-case conditions, including:

1. Different studies use different genetic markers.
2. Different studies use different statistical techniques to test for linkage.
3. Some studies include multiple hypothesis tests by using multiple markers, multiple statistical techniques, or multiple phenotypic cutoff points. This creates issues of nonindependent multiple testing that must be managed.
4. Not all studies report all of the information required for easy extraction of the data. The techniques we will present are applicable even under these difficult circumstances.

GENERAL PRINCIPLES

Assume that there are *m* independent studies assessing linkage of a disease or trait to markers within a region of the genome. Suppose that a *P* value can be ob-

Corresponding author: David B. Allison, Obesity Research Center, St. Luke's/Roosevelt Hospital, 1090 Amsterdam Ave., 14th floor, New York, NY 10025. E-mail: dba8@columbia.edu

tained on each of the m data sets where the P value indicates the probability of obtaining data as extreme or more extreme than the data observed under the null hypothesis of no linkage in the region. Fisher (1954, p. 99) showed that m independent P values can be combined into a single test of significance. Specifically,

$$\chi_{2m}^2 = -2 \sum_{i=1}^m \ln(P_i). \quad (1)$$

This quantity is distributed as χ^2 with $2m$ degrees of freedom. One then can test the significance of the entire body of data by evaluating the probability of obtaining a χ^2 greater than or equal to the observed χ^2 under the null hypothesis and by using the significance level of their choice. (One could imagine situations in which the null hypothesis is rather different. However, in this article we confine our attention to the simple null hypothesis of no linkage in the region.)

In this article, we illustrate the application of this technique to published evidence for linkage in the human *OB* gene region to body mass index (BMI; kg/m²). Most of the exposition involves the application of standard statistical methods and meta-analytic “tricks of the trade” to derive one P value from each study. Following the example, a reiteration and discussion of the general approach are included.

EXAMPLE

To our knowledge, there are five published studies concerning linkage of BMI with markers in the human *OB* gene region. A sixth study (Stirling *et al.* 1995) examined linkage of diabetes mellitus to markers in the *OB* region but only examined associations (not linkage) with obesity and, therefore, is not included. We review each study below. The use of only published data is addressed in the discussion section.

Clement *et al.* (1996): Clement *et al.* (1996) evaluated linkage to BMI dichotomized as “greater than 35”

or “less than or equal to 35” with markers ranging from D7S651 to D7S509. The main results are displayed in Table 1. The data consist of sib-pair analyses testing whether the mean proportion of alleles shared identical by descent (IBD) among sib pairs differs from the expectation of 1/2 under the null hypothesis. The extraction of a single P value from this study is not easy because (1) in some cases the information offered is imprecise; (2) the data really consist of two different studies, one of obese-lean pairs and one of obese-obese pairs; and (3) multiple markers are used.

However, the first problem of incomplete data can be solved rather easily in this case. Because the exact t values are provided along with n (the sample size) it is easy to obtain the exact P values by integrating the t distribution with $n - 1$ degrees of freedom (d.f.). In this case, the second problem of having two separate studies can also be solved. Although the two samples involved contain overlapping individuals, it has been shown that the IBD status of different sibling pairs from within the same sibship are pairwise independent (*e.g.*, Hodge 1984; Amos *et al.* 1989). Therefore, test statistics obtained on the two samples will be independent, and this one paper (Clement *et al.* 1996) can then be considered to simply contain two statistically independent studies, each of which yields one P value. The final problem, multiple markers, is more challenging. Had a multipoint procedure been used to yield a single P value for the point in the interval that yielded the maximum evidence of linkage, the single P value could be corrected and used. However, because a multipoint procedure was not used, it is necessary to combine the eight P values observed into a single P value. Because the markers are linked to each other, test statistics obtained at each marker, and therefore the P values of each marker, are not statistically independent. Any method of combining them must take this nonindependence into account.

To accomplish this, we begin by converting each P value to a corresponding (standard normal) Z-score by

TABLE 1

Proportion of alleles shared IBD in *OB* markers for discordant (obese-lean) and concordant (obese-obese) sib pairs (Clement *et al.* 1996)

Marker	Extreme obesity (BMI > 35 kg/m ²)									
	Full sample		Obese-lean pairs				Obese-obese pairs			
	n	$\bar{\pi}$	n	$\bar{\pi}$	t	P	n	$\bar{\pi}$	t	P
D7S651	145	0.56	60	0.56	-1.42	NS	66	0.57	1.98	0.03
D7S692	136	0.53	57	0.54	-1.29	NS	59	0.52	0.68	NS
D7S677	94	0.47	37	0.48	0.76	NS	46	0.49	-0.29	NS
D7S680	123	0.52	54	0.48	0.42	NS	57	0.59	2.47	0.008
D7S514	111	0.52	45	0.45	1.02	NS	53	0.59	2.44	0.009
D7S530	144	0.51	58	0.43	2.04	0.02	65	0.59	2.96	0.002
D7S640	126	0.50	51	0.43	1.87	0.03	57	0.55	0.99	NS
D7S509	134	0.52	57	0.50	0.14	NS	56	0.54	1.01	NS

means of the inverse standard normal distribution function Φ^{-1} , that is, $Z = \Phi^{-1}(1 - P)$. Under the null hypothesis of no linkage, the eight Z-scores have a multivariate normal distribution with zero means, unit variances, and a correlation matrix R . According to Carey and Williamson (1993), the correlation of IBD status (and therefore Z-scores based on IBD status) between two markers is equal to $(1 - 2\theta)^2$, where θ denotes the recombination fraction between the markers. Note that one centimorgan (cM) is equal to θ of 0.01, equivalent, on average, to 1 million base pairs (bp) (Department of Energy 1992).

In this case, since the tests on individual markers are one-sided, we defined more extreme data in terms of the sum of the Z-scores

$$S_k = \sum_{i=1}^k Z_i,$$

that is, $P = P(S_k > s_k)$, where s_k represents the observed sum of the Z-scores. Specifically, because the variance of the sum of variates is the sum of the variances plus twice the sum of the covariances,

$$S_k / \sqrt{k + 2 \sum_{i < j} r_{ij}} \quad (2)$$

will be distributed as the standard normal, where r_{ij} is the correlation of the i and j th Z; hence, the covariance. Using this approach, the single P values were obtained as 0.294 and 0.031 for the obese-lean pairs and the obese-obese pairs, respectively.

Duggirala *et al.* (1996): Although a multiple-marker linkage analysis was conducted by Duggirala *et al.* (1996), in contrast to the previous case of Clement *et al.* (1996), all the information contained in their multiple markers is already combined into a single P value for the interval investigated by their use of a multipoint procedure. For BMI, the P value was 0.003. However, one needs to correct this P value by taking into account that it was calculated from a multipoint procedure including markers spanning a 211-cM interval (D7S531 to D7S483). This can be accomplished by using the formula provided by Lander and Kruglyak (1995, p. 244). Let us denote a corrected P value by P^* ; then, according to their formula, we have $P^* = 1 - \exp(-\mu(T))$, where

$$\mu(T) = [C + 2\rho GT^2]\alpha(T), \quad (3)$$

where C is the number of chromosomes, G is the genome length measured in Morgans, ρ is the crossing-over rate between the genotypes being compared, and T is the threshold level that yields the significance level $\alpha(T)$. In Duggirala *et al.* (1996), these values are as follows: $C = 1$ (one chromosome used for the study), $G = 2.11$ (equivalent to 211 cM), $\rho = 2$ (for sib-pair tests; Lander and Kruglyak 1995), $\alpha(T) = 0.003$ (observed P value), and $T = \Phi^{-1}(1 - 0.003)$, which is the threshold level of the observed P value. This yields $\mu(T) = 0.073$, and hence $P^* = 0.070$. Therefore, after correct-

ing the P value of 0.003 for the fact that it was obtained by a multipoint procedure with a 211-cM interval, the corrected single P value is 0.070.

Borecki *et al.* (1994): Borecki *et al.* (1994) used only one marker in the area of the human *OB* gene. This one marker was KELL, located at 7q33. Four hundred two sibling pairs were included and the Haseman-Elston procedure was used. Thus, this study already yields a single P value requiring no correction. However, the P value is reported as 0.000. This implies that the actual P value was less than 0.001. From the published text there is no way of knowing how much less. Although it might have been reasonable to use the value 0.001 in the analysis, we contacted the original authors who provided us with an exact t of -4.4383 with 400 d.f., which translates to a single P value of 4.8×10^{-6} .

Norman *et al.* (1996): Norman *et al.* (1996) also used the Haseman-Elston procedure with 488 sibling pairs from the Pima Indian community. Only two markers were used, D7S530 and HCPA2. These markers are estimated to be 1 cM away from each other (Duggirala *et al.* 1996). Z-scores obtained for the two markers therefore will be highly correlated. The exact correlation between these two statistics is estimated to be 0.96 using again the formula from Carey and Williamson (1993). As was done with the data from Clement *et al.* (1996), the P values observed in Norman *et al.* (1996) were converted to Z statistics. The probability of obtaining two Z's as great or greater than the two Z's observed was again calculated in terms of the sum of the Z-scores, which yielded a single P value of 0.544 for this study.

Reed *et al.* (1996): The data for Reed *et al.* (1996) present a unique challenge. The data consist of 213 concordant obese sibling pairs. Reed *et al.* (1996) used markers contained in and surrounding the interval D7S504 through D7S1875. They combined the marker information into haplotypes and conducted their analysis by looking at sharing of haplotypes rather than alle-

TABLE 2

Mean proportion of the *OB* gene haplotypes identical by descent for obese-obese sibling pairs (Reed *et al.* 1996)

Obese cutoff	Pairs (n)	Proportion identical by descent	t value	Nominal P value	Corrected P value
≥30	213	0.51 ± 0.33	0.24	0.4038	NS
≥35	135	0.50 ± 0.34	0.03	0.4333	NS
≥40	59	0.60 ± 0.33 ^a	2.28	0.0132*	0.0396**

Data are means ± SD. Obesity is defined at ascending values of BMI (30–40 kg/m²). The levels of statistical significance were corrected for three tests using a Bonferroni correction: * $P \leq 0.01$; ** $P \leq 0.05$.

^a This combination does not produce the corresponding t value. However, the Z-score was extracted from the P value, which was confirmed by D. R. Reed (personal communication).

TABLE 3

Transmission disequilibrium of a haplotype flanking the OB locus (D7S504 and D7S1875) (Reed *et al.* 1996)

BMI of sibling	1-5 transmitted/ not transmitted	% transmitted	χ^2 (1 d.f.)	Nominal <i>P</i> value	Corrected <i>P</i> value
≥ 30	71/50	58.7	3.64	0.056	NS
≥ 35	60/39	60.6	4.45	0.035*	NS
≥ 40	46/24	65.7	6.91	0.009**	0.027*

The size of allele 1 from D7S504 is 145 bp; the size of allele 5 from D7S1875 is 214 bp. The levels of statistical significance were corrected for three tests using a Bonferroni correction: * $P < 0.05$; ** $P < 0.01$.

les. This aspect of their analysis simplifies the extraction of a single P value since significance is assessed only for IBD sharing at the single haplotype rather than at each individual locus. The challenge of the data is twofold. First, Reed *et al.* (1996) used two different statistical approaches to analyze their data. They used the sibling-pair approach as described, and they used transmission disequilibrium testing (TDT; Spielman *et al.* 1993) on a subset of offspring for whom heterozygote parents were available. Second, within each statistical analysis, Reed *et al.* (1996) analyzed the data three times, each at a different cutpoint of BMI to define obesity. The results from Reed *et al.* (1996) are reproduced here in Tables 2 and 3.

We begin by considering the data for the sibling-pair sets in Table 2, in which for each BMI cutoff, the corresponding nominal P value was provided. Hence, by applying Φ^{-1} to those P values, the corresponding Z -scores are obtained. In this situation, the 3×3 correlation matrix \mathbf{R} for the Z -scores can be calculated by assuming that the IBD status for the sibling pairs is independently and identically distributed (i.i.d.). Under this assumption, the correlation between the Z -score for a sample and a subset within the sample is the square root of the proportion of subjects in the subsample from the larger sample. For example, the correlation between the Z -scores for sibling pairs above 35 and sibling pairs above 30 is

$$\sqrt{135/213} = 0.796.$$

As above, once the three Z 's and their correlation matrix are obtained, a single P value, in terms of the sum, can be obtained. In this case, it was observed to be 0.159 for the sib-pair test. (Note that the sib-pair tests in Table 2 are one-sided.)

Turning to the TDT results in Table 3, a single P value can be obtained in a similar fashion. The chi-squares in Table 3 are converted to Z -scores by taking their square root. Assuming that the data are i.i.d., the correlation among the Z 's can again be estimated as the square root of the proportion of subjects in a subset divided by the number of subjects in the larger set. For example, the estimated correlation of the Z -score in subjects with a BMI ≥ 40 and the Z -score for subjects with a BMI > 30 is

$$\sqrt{70/121} = 0.761.$$

With the Z -scores and the 3×3 correlation matrix among them derived, a single P value for the TDT in Reed *et al.* (1996) can be calculated. However, the TDT is two-sided, unlike the previous sib-pair tests, which must be considered. Denoting the vector of Z -scores by \underline{Z} , we defined the P value in terms of a quadratic form

$$Q = \underline{Z}'\mathbf{R}^{-1}\underline{Z}, \quad (4)$$

that is, $P = P(Q > q)$, where q represents the observed value of Q . In general, it is well known that Q is chi-square-distributed with d.f. equal to the number of Z -scores. However, in this particular example, since the correlation matrix \mathbf{R} has one restriction (if two correlations are given, then the remaining third one can be decided from the two), the d.f. is the number of Z -scores minus one, that is two. More specifically, if there are m Z -scores and the correlations are determined by means of the proportion of sample sizes, the d.f. would be $m - 1$. From this consideration, the single P value was obtained as 0.026.

The final challenge with the data from Reed *et al.* (1996) is that one now has two P values, one from TDT and one from sibling-pair testing. If the correlation among these two tests could be determined, then one could combine these into a single P value. However, it is not immediately apparent how to estimate this correlation. There are several alternatives. First, one could, on some *a priori* grounds of preference, choose one test over another. For example, one might argue that because all of the other studies are using a sibling-pair approach rather than TDT it would be more appropriate to combine sibling-pair data rather than the TDT data and be consistent with the others. Second, one could multiply the lowest P value by two as a form of Bonferroni correction. However, this is overly conservative because it does not take the correlation between the two tests into account. Third, one could estimate the correlation via simulation. Fourth, one could conduct the overall meta-analysis with the results of the sibling-pair tests and then conduct the analysis again using the results of the TDT tests as a form of sensitivity analysis (Greenhouse and Iyengar 1994). This is the strategy that we adopt.

TABLE 4
***P* values obtained for each study in overall meta-analysis and sensitivity analysis**

Reference		<i>P</i> value	χ^2 (<i>P</i> value) ^a	χ^2 (<i>P</i> value) ^b
Clement <i>et al.</i> (1996)	obese-lean	0.294	41.66 (0.9×10^{-5})	45.28 (0.2×10^{-5})
	obese-obese	0.031	37.16 (5.3×10^{-5})	40.78 (1.2×10^{-5})
Duggirala <i>et al.</i> (1996)		0.070	38.78 (2.8×10^{-5})	42.41 (0.6×10^{-5})
Borecki <i>et al.</i> (1994)		4.8×10^{-6}	19.61 (0.0332)	23.23 (0.0099)
Norman <i>et al.</i> (1996)		0.544	42.89 (0.5×10^{-5})	46.51 (0.1×10^{-5})
Reed <i>et al.</i> (1996)	sib pairs	0.159	40.43 (1.4×10^{-5})	
	TDT	0.026		40.43 (1.4×10^{-5})
	Overall		44.10 (1.5×10^{-5})	47.73 (3.5×10^{-6})

^a Chi-squares with 10 d.f., using Reed *et al.* (1996) sib-pair test result, when the corresponding reference is deleted from the analysis; the last row is for the overall chi-square with 12 d.f. and its *P* value.

^b Chi-squares with 10 d.f., using Reed *et al.* (1996) TDT result, when the corresponding reference is deleted from the analysis; the last row is for the overall chi-square with 12 d.f. and its *P* value.

Overall meta-analysis: The (single) *P* values for the five papers calculated as described above are presented in Table 4. In the primary overall meta-analysis the *P* values were pooled using Fisher's method described in the Introduction. For Reed *et al.* (1996) the *P* value used was that from the sib-pair study. Based on all these *P* values, the χ^2 for the overall analysis was 44.10 with 12 d.f. ($P = 1.5 \times 10^{-5}$). These results apparently provide strong evidence of linkage somewhere in the *OB* region. When the *P* value for Reed *et al.* (1996) was replaced with that for TDT the results were even more significant, the overall χ^2 being 47.73 ($P = 3.5 \times 10^{-6}$). Thus, these results suggest that there is clear evidence for linkage of BMI to something in the *OB* region regardless of whether one uses the TDT or sib-pair analyses from Reed *et al.* (1996).

As a sensitivity analysis, each study result was removed from the analysis, and the chi-square statistic with 10 d.f. (from the remaining study results) was computed. The corresponding *P* values are given in Table 4. This table shows that while the Borecki *et al.* (1994) study provides the extreme significance observed, even excluding this study, the remaining results still provide a significant value, again regardless of whether one uses the TDT or sib-pair analyses from Reed *et al.* (1996). It follows that the strong overall statistical evidence is not necessarily due to a particular study with extreme significance but due to the accumulation of many study results.

Apart from the overall significance of these results, one may question whether the variation of the strength of the evidence (in terms of the *P* value) seen from study to study is simply random variation or represents some statistically significant heterogeneity in the linkage. Although this is clearly a legitimate question, it is not possible to conduct such a heterogeneity test in the current situation when only the inferential information

available from *P* values is used, since the heterogeneity of the strength of statistical evidence across the studies is not identifiable from the heterogeneity of the linkage *per se*.

Summary of meta-analytic approach used: In this section we reiterate in general terms the approach taken in the meta-analysis. The first step is to extract a single *P* value from each study. The studies contained herein provide an illustration of several different techniques for extracting the single *P* value from each study. The techniques used can briefly be summarized as follows:

1. If a single *P* value was provided from a multipoint procedure, then the Lander-Kruglyak correction in Equation 3 was applied; Duggirala *et al.* (1996) case.
2. If a separate *P* value for each of several markers is used, then the Z-scores and the correlations among them are obtained by using the inverse standard normal distribution function Φ^{-1} and the formula in Carey and Williamson (1993), respectively. Since the *P* values are one-sided for linkage studies, a single *P* value was determined in terms of the sum of the Z-scores in Equation 2, which is normally distributed; cases of Clement *et al.* (1996) and Norman *et al.* (1996).
3. For the multiple cutoff study defining affected or unaffected sib pairs (*e.g.*, Reed *et al.* 1996), the correlations among Z-scores were estimated by means of the proportion of sample sizes. For the one-sided sib-pair tests, the sum in (2) was used for extracting the single *P* value. In contrast, for the two-sided TDT, the single *P* value was extracted by means of the quadratic form *Q* in (3), which is chi-square distributed.
4. For a single marker study, no correction needs to be applied; Borecki *et al.* (1994) case.

Once a single *P* value is derived for each study, the *P* values are pooled using Fisher's method in Equation 1.

DISCUSSION

The above example indicates that by judicious use of standard statistical theory and meta-analytic techniques, one can meta-analyze data from multiple linkage studies of the same phenotype even under a number of worst-case conditions. We believe such methods will become essential to understanding the overall significance of the research literature when raw data are not available. This is all the more important in the field of gene mapping for complex traits because the power to detect significance in any one study is often quite below what one would desire. This power can be enhanced by pooling the data in multiple studies.

The efficacy of these procedures is demonstrated by the example involving the assessment of linkage of BMI to the genome region containing the human *OB* gene. Taken individually, no one study is terribly convincing, except Borecki *et al.* (1994). Moreover, a cursory glance at the body of literature suggests that the results may be inconsistent across studies (*i.e.*, some studies obtain significant results and others do not), which might lead readers to believe that the overall body of evidence does not support linkage in this region. However, as the foregoing analyses demonstrate, such a conclusion would be erroneous. Thus, in this case at least, meta-analysis of the entire body of data is able to demonstrate clarity, whereas individual studies or informal review of the individual studies yields no such clarity.

We wish to stress again that we do not consider our approach outlined in the current paper to be the optimal approach. Clearly, a better approach is to obtain the raw data of all of the investigators, whether the data were published or not, and conduct a pooled analysis. However, for the near future, we suspect that there will be many situations in which meta-analysts have only one alternative available to them: the statistical integration of data from multiple published studies that have used different statistical methods and different markers and, in some cases, have presented incomplete information. Moreover, although we hope that our paper acts as a call for the presentation of more complete information, such calls have been issued by meta-analysts for over two decades (Jackson 1980; Feinstein 1995). Our subjective assessment is that such calls have resulted in an only modest increase in the completeness of information presented in published reports. Thus, we suspect that meta-analysts will continue to be faced with similar situations.

A limitation of the meta-analysis conducted herein concerns the potential for publication bias. Publication bias occurs when the probability of a study being published is dependent upon the results of the study. Publication bias has been shown to exist in other fields (*e.g.*, Allison *et al.* 1996). Again, although calls have been issued to the research community to correct this problem (*e.g.*, Chalmers *et al.* 1990), the problem does not appear to be going away (*e.g.*, Dickersin and Min

1993). There are several ways to approach the problem of publication bias. Clearly, the best way is to obtain unpublished studies as well as published studies. We did not attempt to do that in the context of this paper because our primary goal was to present our approach to conducting the meta-analysis as an example. The second approach one might take is to calculate a quantity called the "fail-safe sample size," which is the number of unpublished studies needed to overturn the overall conclusion of statistical significance from the published studies. Procedures for calculating the fail-safe sample size are presented in Iyengar and Greenhouse (1988). We calculated the fail-safe sample sizes for the current example and obtained values of 15 and 18 depending on whether Reed *et al.*'s (1996) sib-pair test or TDT data were used, respectively. Because it seems very unlikely that there are 15-to-18 unpublished studies on this topic, the overall conclusion of statistical significance seems reasonably safe.¹

Finally, the methods presented herein only allow one to conduct inferential tests of whether or not there is significant linkage in a particular region. The appropriate conclusion to such an analysis, assuming a statistically significant result is obtained, is that there is statistically significant evidence for linkage somewhere in the region examined, in this case the region from D7S531 to D7S483, which is the same as Duggirala *et al.*'s (1996) interval. Unfortunately, one is unable, based solely on an analysis of the type presented herein, to estimate where in that interval the putative QTL lies or in any way to place a confidence interval around that estimated location. It is hoped that methods are developed to accomplish this goal in the future.

In conclusion, as more studies assessing the linkage between genetic markers and complex diseases or traits are possible, we suspect that there will be increasing need for meta-analysis to objectively and quantitatively pool the resulting information. We hope that the procedures outlined in this paper are a useful step in that direction.

This research was supported by National Institutes of Health grants R29DK47256, R01DK51716, and P30DK26687.

LITERATURE CITED

- Allison, D. B., M. S. Faith and B. S. Gorman. 1996 Publication bias in obesity treatment trials? *Int. J. Obes.* **20**: 931-937.
- Amos, C. I., R. C. Elston, A. F. Wilson and J. E. Bailey-Wilson. 1989 A more powerful robust sib-pair test of linkage for quantitative traits. *Genet. Epidemiol.* **6**: 435-449.
- Borecki, I. B., T. Rice, L. Pérusse, C. Bouchard and D. C. Rao. 1994 An exploratory investigation of genetic linkage with body composition and fatness phenotypes: the Québec family study. *Obes. Res.* **2**: 213-219.

¹ Since preparation of this article, we have become aware that there are at least three additional articles on this topic (Bray *et al.* 1996; Oksanen *et al.* 1997; Hasstedt *et al.* 1997). The interested reader is referred directly to those papers.

- Bray, M. S. E. Boerwinkle and C. L. Hanis, 1996 OB gene not linked to human obesity in Mexican American affected sib pairs from Starr County, Texas. *Hum. Genet.* **98**: 590–595
- Carey, G., and J. Williamson, 1993 Linkage analysis of quantitative traits: increased power by using selected samples. *Am. J. Hum. Genet.* **49**: 786–796.
- Chalmers, T. C., C. S. Frank and D. Reitman, 1990 Minimizing the three stages of publication bias. *J. Am. Med. Assoc.* **263**: 1392–1395.
- Clement, K., C. Garner, J. Hager, A. Philippi C. LeDuc *et al.*, 1996 Indication for linkage of the human OB gene region with extreme obesity. *Diabetes* **45**: 687–690.
- Cox D. R., and D. B. Hinkley, 1974 *Theoretical Statistics*. Chapman & Hall, London.
- Department of Energy Human Genome Program, 1992 *Primer on Molecular Genetics*. Oak Ridge National Laboratory, Oak Ridge.
- Dickersin, K., and Y. Min, 1993 Publication bias: the problem that won't go away. *Ann. NY Acad. Sci.* **703**: 135–148.
- Duggirala, R., M. P. Stern, B. D. Mitchell, L. J. Reinhart, P. A. Shipman *et al.*, 1996 Quantitative variation in obese-related traits and insulin precursors linked to the OB gene region on human chromosome 7. *Am. J. Hum. Genet.* **59**: 694–703.
- Feinstein, A. R., 1995 Meta-analysis: statistical alchemy for the 21st century. *J. Clin. Epidemiol.* **48**: 71–79.
- Fisher, R. A., 1954 *Statistical Methods for Research Workers*, 12th ed. Hafner Publishing Company Inc., New York.
- Fulker, D. W., S. S. Cherny and L. R. Cardon, 1995 Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am. J. Hum. Genet.* **56**: 1224–1233.
- Glass, G. V., 1976 Primary, secondary, and meta-analysis of research. *Educ. Res.* **5**: 3–8.
- Greenhouse, J. B., and S. Iyengar, 1994 Sensitivity analysis and diagnostics, pp. 383–409 in *The Handbook of Research Synthesis*, edited by H. Cooper and L. V. Hedges. Russell Sage Foundation, New York.
- Haseman, J. K., and R. C. Elston, 1972 The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**: 3–19.
- Hasstedt, S. J., M. Hoffman, M. F. Leppert and S. C. Elbein, 1997 Recessive inheritance of obesity in familial non-insulin-dependent diabetes mellitus, and lack of linkage to nine candidate genes. *Am. J. Hum. Genet.* **61**: 668–677.
- Hodge, S. E., 1984 The information contained in multiple sibling pairs. *Genet. Epidemiol.* **1**: 109–122.
- Hyne, V., and M. J. Kearsy, 1995 QTL analysis: further uses of 'marker regression.' *Theor. Appl. Genet.* **91**: 471–476.
- Iyengar, S., and J. B. Greenhouse, 1988 Selection models and the file drawer problem. *Stat. Science* **3**: 109–135.
- Jackson, G. B., 1980 Methods of integrative reviews. *Rev. Educ. Res.* **50**: 438–460.
- Jeng, G. T., J. R. Scott and L. F. Burmeister, 1995 A comparison of meta-analytic results using literature vs individual patient data. *J. Am. Med. Assoc.* **274**: 830–845.
- Kearsy, M. J., and V. Hyne, 1994 QTL analysis: a simple 'marker-regression' approach. *Theor. Appl. Genet.* **89**: 698–702.
- Lander, E. S., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**: 241–247.
- Li, Z., and D. C. Rao, 1996 Random effects model for meta-analysis of multiple quantitative sibpair linkage studies. *Genet. Epidemiol.* **13**: 377–384.
- Norman, R. A., R. L. Leibel, W. K. Chung, L. Power-Kehoe, S. C. Chua *et al.*, 1996 Absence of linkage of obesity and energy metabolism to markers flanking homologues of rodent obesity genes in Pima Indians. *Diabetes* **45**: 1229–1232.
- Oksanen, L., M. Öhman, M. Heiman, K. Kainulainen, J. Kapiro *et al.*, 1997 Markers for the gene *ob* and serum leptin levels in human morbid obesity. *Hum. Genet.* **99**: 559–564.
- Pearson, K., 1904 Report on certain enteric fever inoculation statistics. *Brit. Med. J.* **3**: 1243–1246.
- Pearson, K., 1933a Appendix to Dr. Elderton's paper on "The Larkshire milk experiment." *Ann. Eugen.* **5**: 337–338.
- Pearson, K., 1933b On a method of determining whether a sample size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika* **25**: 379–410.
- Reed, D. R., Y. Ding, W. Xu, C. Cather, E. D. Green *et al.*, 1996 Extreme obesity may be linked to markers flanking the human OB gene. *Diabetes* **45**: 691–694.
- Spielman, R. S., R. E. McGinnis and W. J. Ewens, 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *Am. J. Hum. Genet.* **52**: 506–516.
- Stirling, B., N. J. Cox, G. I. Bell, C. L. Hanis, R. S. Spielman *et al.*, 1995 Identification of microsatellite markers near the human OB gene and linkage studies in NIDDM-affected sib pairs. *Diabetes* **44**: 999–1001.
- Tippett, L. H. C., 1931 *The Methods of Statistics*. Williams & Norgate, London.
- Xu, S., and W. R. Atchley, 1995 A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**: 1189–1197.

Communicating editor: R. R. Hudson