Am. J. Hum. Genet. 69:660–663, 2001

Report

The Significance of Not Finding a Gene

Michael A. Province

Division of Biostatistics, Washington University School of Medicine, St. Louis

As more investigators conduct extensive whole-genome linkage scans for complex traits, interest is growing in meta-analysis as a way of integrating the weak or conflicting evidence from multiple studies. However, there is a bias in the most commonly used meta-analysis linkage technique (i.e., Fisher's [1925] method of combining of *P* values) when it is applied to many nonparametric (i.e., model free) linkage results. The bias arises in those methods (e.g., variance components, affected sib pair, extremely discordant sib pairs, etc.) that truncate all "negative evidence against linkage" into the single value of LOD = 0. If incorrectly handled, this bias can artificially inflate or deflate the combined meta-analysis linkage results for any given locus. This is an especially troublesome problem in the context of a genome scan, since LOD = 0 is expected to occur over half the unlinked genome. The bias can be overcome (nearly) completely by simply interpreting LOD = 0 as a *P* value of $\frac{1}{2\ln(2)} \approx .72$ in Fisher's formula.

Fisher's (1925, p. 99) method of the "combination of probabilities from tests of significance" is becoming an increasingly used tool for the synthesis of linkage evidence across studies (Turecki et al. 1997; Allison and Heo 1998; Guerra et al. 1999; Merriman et al. 2001). One of the simplest yet most elegant techniques, Fisher's method was invented some 50 years before the term "meta-analysis" was coined by Glass in 1976 (Glass 1976), and yet it is still remarkably general and useful. It allows combination of evidence from multiple tests, which need not use the same statistic to produce their *P* values and which may even operate on different sampling units. All that is required for the validity of the combined test is that the individual *P* values be from tests of the same hypothesis (e.g., tests of linkage to the same locus) and be independent of one another. Thus, we can use Fisher's method to combine (*a*) parametric (i.e., model based) with nonparametric (i.e., model free) linkage tests (including Haseman-Elston, variance components, etc.); (*b*) dichotomized with continuous phenotype definitions (e.g., hypertension status with measured blood pressure, presence/absence of obesity with measured body-mass index, etc.); (*c*) samples of affected sib pairs, extremely discordant sib pairs, entire sibships, and/or extended pedigrees; and (*d*) two-point with multipoint linkage analyses. This flexibility is so attractive for the combining of evidence across different study designs that, even if we are given access to all of the original raw data for a combined analysis, we may well prefer to use such a meta-analysis technique instead, since that will allow us to optimize each study-specific analysis to best fit its own particular data, instead of deciding on a compromise, "one size fits all" common analysis strategy for the sake of pooling (M. A. Province, S. L. R. Kardia, K. Ranade, D. C. Rao, B. A. Thei, R. S. Cooper, N. Risch, S. T. Turner, D. R. Cox, S. C. Hunt, A. B. Weder, and E. Boerwinkle, unpublished data). This approach is especially attractive for the purposes of synthesizing the combined evidence from multiple genome scans into a single pooled, meta-analysis LOD surface.

Fisher's method is based on the observation that, if n independent tests are made of the same hypothesis, resulting in the P values P_1, P_2 ,..., P_n, then the quantity $\Sigma_{i=1}^{n}(-2\ln P_i)$ is distributed as a χ^2 with $2n$ df, which provides a combined P value for all n tests. In the case of linkage analysis, one can easily work on either the LOD score or the P-value scale, since there is a simple one-to-one correspondence between the two, given by the formula $P = 1 - \Phi[\text{sign(LOD)}\sqrt{2\ln(10)\,|\text{LOD}|}]$, where Φ is the gaussian (normal) distribution function (Ott 1991). However, Fisher's formulation was derived for the usual two-tailed situation, in which, under the null hypothesis, the P values are distributed uniformly

Received June 18, 2001; accepted for publication July 12, 2001; electronically published July 30, 2001.

Address for correspondence and reprints: Dr. Michael A. Province, Professor of Biostatistics, Division of Biostatistics, Box 8067, Washington University School of Medicine, 660 South Euclid, St. Louis, MO 63110. E-mail: mike@wubios.wustl.edu

0002-9297/2001/6903-0020$02.00

between 0 and 1 (with all values being equally probable). Unfortunately, many of the very popular nonparametric (i.e., model free) linkage methods (in particular, variance components, affected sib pair, and extremely discordant sib-pair linkage) produce one-tailed LOD scores that truncate all negative evidence against linkage at the single point LOD = 0. This can introduce a bias into the distribution of Fisher's original formula, if one is not careful.

Under the null hypothesis, the *P* values for these tests are actually distributed as a 50:50 mixture of (*a*) a uniform distribution between 0 and $\frac{1}{2}$ (for the positive linkage evidence) and (*b*) a point mass corresponding to LOD = 0 (Self and Liang 1987). This makes intuitive sense, since we would expect that, if there really is no linkage, then approximately half of all studies should show positive evidence while the remaining half should show negative evidence. If we apply Fisher's formula, then how close we get to the desired χ^2 distribution depends upon how we weight the point-mass evidence at LOD = 0. In fact, in the model-free–linkage case, it is not clear how to meaningfully translate LOD = 0 into a *P* value. For traditional parametric LOD scores (which can be negative as well as positive) LOD = 0 clearly corresponds to $P = .5$ (with $P < .5$ and $P > .5$ corresponding to positive and negative LOD scores, respectively). However, general maximum-likelihood theory would suggest that, for this truncated nonparametric case, LOD = 0 should be treated as $P = 1.0$, since this provides the same likelihood under both the alternative and the null hypotheses. In fact, one could formally define a functional mapping from LOD = 0 to any fixed *P* value of *a,* where $\frac{1}{2} \leq a \leq 1$, and achieve a valid test for a single hypothesis. In normal circumstances, the particular value chosen for *a* in a given study would not matter much, since investigators are interested in small *P* values, which tend to refute the null hypothesis; however, in the case of meta-analysis, exactly how LOD = 0 is mapped to the *P*-value scale is critical to Fisher's formula.

To demonstrate the impact, a 10,000-replication Monte Carlo simulation combining nine studies was conducted with different values for *a*—that is, different interpretations of the *P* value associated with LOD = 0 in Fisher's formula (shown in fig. 1). Under the null hypothesis, LOD scores from each of the nine studies will be a 50:50-mixture distribution of a point mass at 0, corresponding to $P = a$, and LOD > 0, corresponding to a uniform distribution of *P* values in the interval $(0, \frac{1}{2})$. For each replication, we generate nine LOD scores from this null-hypothesis–mixture distribution, translate each of them into *P* values by use of Ott's formula, and apply Fisher's method to combine all nine into a single meta-analysis *P* value. We then compare the empirical cumulative distribution of all 10,000 simulated meta-analysis *P* values to their nominal values, to examine the bias. We consider the impact of three possible values of *a* when used in Fisher's formula—that is, we examine three different mappings of the point mass at LOD = 0 into single *P* values. The simulation demonstrates that, if the point mass at LOD = 0 is interpreted as $P = \frac{1}{2}$, then Fisher's method will produce final combined *P* values that are too small (i.e., too liberal), resulting in the upper curve shown in figure 1. Here we have too many small *P* values (almost 10% are <.05, and 70% are <.50). This makes sense, because we are interpreting all negative LOD scores "optimistically" (i.e., pro linkage), as though they were really all at the upper limit of LOD = 0. On the other hand, if all negative LOD scores are taken to correspond to the single $P = 1.0$, then Fisher's method produces *P* values that are too large (i.e., too conservative), which corresponds to the lower curve shown in figure 1. Here, we have too few small *P* values (<2% are <.05), because we are now making the opposite mistake—that is, interpreting all negative linkage evidence as corresponding to the most "pessimistic" negative value (i.e., a LOD = $-\infty$). The exact modification

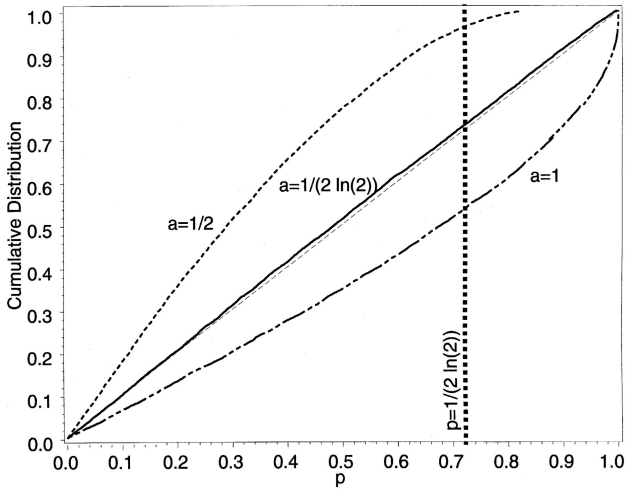

Figure 1 Cumulative distribution of *P* values for the modified Fisher method defining LOD = 0 as $P = a$ (for $a = \frac{1}{2}$, $a = \frac{1}{2\ln(2)} \approx .72$, and 1): a 10,000-replication Monte Carlo simulation experiment demonstrating the correction to the bias in Fisher's combining-of-*P*-values method when applied to nonparametric linkage methods that do not produce negative LOD scores. The upper curve is the empirical cumulative distribution of meta-analysis *P* values when nine studies are combined under the null hypothesis (no linkage), in which we interpret the all-negative evidence (LOD = 0) to be *P* values of $a = \frac{1}{2}$, resulting in too many significant meta-analysis *P* values; in the lower curve, the same nine studies are combined, but each LOD = 0 is interpreted to be a *P* values of $a = 1$, resulting in too few significant meta-analysis *P* values; finally, in the middle curve, the same nine studies are combined, but all negative linkage evidence at LOD = 0 is interpreted as a *P* value of $a = \frac{1}{2\ln(2)} \approx .72$, which nearly completely removes the bias.

of Fisher's method that should be used for a completely unbiased result is a somewhat complicated formula (below), which depends on *n,* the number of studies combined. But we can eliminate almost all of the bias if we simply interpret LOD = 0 as corresponding to a *P* value of $a = \frac{1}{2\ln(2)}$, which nearly "splits the difference" between $\frac{1}{2}$ and 1, at ~.72. This produces the middle curve shown in figure 1 (which closely follows the fainter dashed "identity" line), indicating that nominal *P* values are close to their empirical cumulative-distribution values and that we obtain approximately the expected number of significant *P* values for any alpha level of significance. This simple modification of Fisher's method is completely unbiased (except for the discrete part of the distribution) for combining exactly two studies and is nearly unbiased for a small number (i.e., <100) of combined studies.

To derive this bias-correction formula analytically, for each study, $i = 1, \ldots, n$, under the null hypothesis of no linkage at a given locus, we define a family of probability spaces under which the point mass at LOD = 0 is mapped to each *P* value, *a,* in the interval $\frac{1}{2} < a < 1$, obtaining a consistent family of probability-density functions (PDFs). We solve for that *a* value that makes the PDF for the sum of the Fisher-transformed *P* values nearly a χ^2 (as it is in the parametric case). For each such *a,* let $f_a(P_i)$ be the PDF that maps LOD = 0 to the *P* value *a*. Then $f_a(P_i)$ is a 50:50 mixture of a uniform $(0, \frac{1}{2})$ random variable (RV) (corresponding to the positive LOD scores) and a point mass at *a* (corresponding to LOD = 0, where all negative linkage evidence is concentrated). Applying Fisher's *P* transform to the *P* values, $y_i = -2\ln(P_i)$, we obtain the transformed PDFs, $g_a(y_i)$, as 50:50 mixtures of a truncated χ^2 distribution with 2 df and a point mass at $-2\ln(a)$.

$$g_{Y_i a}(y_i) = \frac{1}{2} e^{-y/2} I_{[2\ln(2),\infty)}(y_i) + I_{[-2\ln(a)]}(y_i)$$

for each study, *i,* under the null hypothesis. To obtain the combined meta-analysis *P* value, we sum these *n* transformed *P* values across all *n* studies at each locus $Y = \Sigma_{i=1}^{n} Y_i = \Sigma_{i=1}^{n} (-2\ln P_i)$. We use the basic convolution theorem to find the PDF of the sum of independent RVs, applying induction on *n,* the number of studies. In particular, for $n = 2$,

$$g_{(Y_1+Y_2)a}(y) = \frac{1}{4} e^{-y/2} \left\{ y + \left[\frac{2}{a} - 4\ln(2) \right] \right\} I_{[4\ln(2),\infty)}(y) + \frac{1}{4} e^{-y/2} \left(\frac{2}{a} \right) I_{[2\ln(2)-2\ln(a),4\ln(2)]}(y) + I_{[-4\ln(a)]} .$$

Solving to make the inner parenthetical term equal to 0 (which is a χ^2 PDF with 4 df), we find that, when $a = \frac{1}{2\ln(2)}$,

$$g_{(Y_1+Y_2)a}(y) = \frac{1}{4} e^{-y/2} y I_{[4\ln(2),\infty)}(y) + o(y) I_{[2\ln(2)-2\ln(a),4\ln(2)]}(y) + I_{[-4\ln(a)]}(y) .$$

Induction on *n,* by application of the convolution theorem, gives the result, for any $n > 2$, after simplification. Note that the χ^2 approximation covers more of the range of the random variable *y* as *n* increases but that it is also a more biased estimate as *n* increases. It is an exact expression for $n = 2$, in the range $y \geq 4\ln(2)$—that is, where $P_1 P_2 \leq e^{-[4\ln(2)/2]} = \frac{1}{4}$. A more exact correction could be obtained by simply taking higher-order Taylor-series–expansion terms, but these would have the defect of depending on *n,* the number of studies combined, which means that the *P*-value interpretation for LOD = 0 in the same study would differ, depending on the number of other studies with which it was combined. The correction given above has the virtue of simplicity and constancy over the number of studies.

This modification of Fisher's method has the added benefit that we can obtain meta-analysis *P* values that lie within the interval $(\frac{1}{2}, 1)$, which corresponds to evidence against linkage when we translate back from the combined *P* value to the combined LOD scale; that is, we can actually obtain *negative* LOD-score evidence by combining several nonparametric LOD scores, each of which is nonnegative! If the evidence were truly neutral with respect to linkage in a region (LOD = 0 exactly and not simply truncated negative), we would expect half of our studies to have LOD > 0 and half to have LOD = 0 (truncated). Thus, when we observe multiple studies having LOD = 0 (truncated) at a locus, this is actually combined evidence against linkage. The more studies with LOD = 0 (truncated) that we combine, the stronger the negative evidence.

By their very nature, genome linkage scans are classic examples of *hypothesis-generating* exercises—the in-vogue term is "data mining," but such exploratory data analyses have sometimes been more cynically denounced as a "fishing expedition" (Holmes 1915). Of course, for many important complex traits (e.g., heart disease, hypertension, hyperlipidemia, stroke, cancer, obesity, asthma, diabetes, etc.), we know, from many lines of scientific investigation (animal studies, twin studies, etc.), that the genes are there to be found somewhere in the genome. So, if a linkage scan is a fishing expedition, at least we know the fish are there for the catching. However, as more investigators fish the genetic waters, trolling the genome with linkage markers, they are finding

it increasingly difficult to catch the elusive complex-trait gene on their own. It is etymologically fitting, therefore, that the first and most famous statistical "Fisher" gave us an easy, elegant, and powerful way to combine our individual efforts into a pooled "catch." But we must be careful how we interpret it when an individual effort fails (i.e., when LOD = 0), so as not to bias our combined results. For Fisher's purposes in this case, the significance of not finding a gene is $\frac{1}{2\ln(2)}$.

Acknowledgment

This work was partially supported by National Institutes of Health grants GM28719 and HL54473.

References

Allison DB, Heo M (1998) Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. Genetics 148:859–866

Fisher RA (1925) Statistical methods for research workers, 13th ed. Oliver & Loyd, London

Glass GV (1976) Primary, secondary and meta-analysis of research. Educ Researcher 5:3–8

Guerra R, Etzel CJ, Goldstein DR, Sain SR (1999) Meta-analysis by combining p-values: simulated linkage studies. Genet Epidemiol 17 Suppl 1:S605–S609

Holmes OW (1915) Ellis v Interstate Commerce Commission, 237 US 434, no 712, US Supreme Court

Merriman TR, Cordel HJ, Eaves IA, Danoy PA, Coraddu F, Barber R, Cucca F, et al (2001) Suggestive evidence for association of human chromosome 18q12-q21 and its orthologue on rat and mouse chromosome 18 with several diseases. Diabetes 50:184–194

Ott J (1991) Analysis of human genetic linkage. John's Hopkins University Press, Baltimore

Self ST, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Turecki G, Rouleau GA, Joober R, Mari J, Morgan K (1997) Schizophrenia and chromosome 6p. Am J Med Genet 74: 195–198