Note

On the “NPD Ratio” as a Test for Crossover Interference

Franklin W. Stahl¹

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403

Manuscript received January 10, 2008

Accepted for publication February 27, 2008

ABSTRACT

The “NPD ratio,” widely used by yeast geneticists, is of limited applicability and is prone to falsely indicate significant crossover interference in a chi-square test. A simple, better chi-square test for interference in two-factor crosses is described.

FOR organisms, such as yeast, in which the four products of individual meioses can be examined, crossover interference can be assessed in two-factor crosses. In 1952, Haig Papazian noted that, for the null hypothesis of no interference, the frequency of nonparental ditype (NPD) tetrads expected (fN_{exp}) can be written as a function of the frequency of tetratype (TT) tetrads observed (fT_{obs}):

$$fN_{\text{exp}} = \frac{1}{2} \left[1 - fT_{\text{obs}} - \left(1 - \frac{3}{2} fT_{\text{obs}} \right)^{\frac{2}{3}} \right]. \quad (1)$$

Equation 1, which is equivalent to PAPAZIAN's (1952) Equation 5, assumes the absence of both chromatid and chromosome (chiasma) interference and is a rewritten form of equations by HALDANE (1931). Interference is indicated when the observed frequency of NPDs (fN_{obs}) differs significantly from fN_{exp} . The “NPD ratio” ($fN_{\text{obs}}/fN_{\text{exp}}$) is an indication of the intensity of interference, with smaller values indicating stronger positive interference (SNOW 1979a,b).

Papazian's approximate method: In 1952, in the absence of today's electronic calculators, Papazian provided an easily calculable approximation to Equation 1:

$$fN_{\text{exp}} \cong \frac{1}{8} fT_{\text{obs}}^2 \left(1 + \frac{2}{3} fT_{\text{obs}} \right). \quad (2)$$

Although PAPAZIAN (1952) advertised the approximation as generally valid, it is useful only for short intervals. For longer intervals, fN_{exp} calculated from this approximation falls ever farther below the values calculated from the exact equation. At the limit, when $fT_{\text{obs}} = \frac{2}{3}$, fN_{exp}

calculated from the approximate equation is only 0.080, which is less than half of its exact expectation of $\frac{1}{6}$. Subsequently, STRICKLAND (1958) proposed an approximation that is more useful over a wider range of fT_{obs} values and that, he says, changes Papazian's (allegedly incorrect) $\frac{2}{3}$ to $\frac{3}{2}$.

Some authors, noting correctly that Papazian's exact equation gives meaningless roots when $fT_{\text{obs}} > \frac{2}{3}$, have used Equation 2 to calculate fN_{exp} when confronted with data in which $fT_{\text{obs}} > \frac{2}{3}$. Apparently, these authors took PAPAZIAN (1952, p. 179) seriously when he wrote that his approximation “. . . is valid for long intervals and can be used where other methods of measuring interference cannot.” However, not only is the approximation a poor one for large fT_{obs} , its use when $fT_{\text{obs}} > \frac{2}{3}$ is as invalid as is that of the exact expression— fT_{obs} cannot (significantly) exceed $\frac{2}{3}$ in the absence of interference, as noted by PERKINS (1955). Furthermore, since Equation 1 describes relationships in the presumed absence of interference, the only value for fN_{exp} that can reasonably be based solely on an fT_{obs} that (insignificantly) exceeds $\frac{2}{3}$ is $\frac{1}{6}$, the value for fN_{exp} as map length approaches infinity. Some other authors have simply ignored the upper limit on fT for tetrads in which $fT_{\text{obs}} > \frac{2}{3}$ and extracted a meaningless “ fN_{exp} ” from the exact equation.

Tests for significance: Several users of Papazian's equation test for significance by chi square, but they fail to justify use of this test or to publish particulars. The most obvious, though not necessarily legitimate (see below), application of chi square would seem to be a test of the observed numbers of NPDs and of parental ditypes (PDs) *vs.* the numbers expected (d.f. = 1), where the frequency of PDs expected is $fP_{\text{exp}} = 1 - fN_{\text{exp}} - fT_{\text{obs}}$. Some investigators appear to ignore the PDs and calculate chi square as $(N_{\text{obs}} - N_{\text{exp}})^2/N_{\text{exp}}$ (d.f. = 1),

¹Address for correspondence: Institute of Molecular Biology, 1229, University of Oregon, Eugene, OR 97403-1229. E-mail: fstahl@uoregon.edu

where N_{exp} is $fN_{\text{exp}} \times \text{total tetrads}$, rounded to the nearest integer. (The rounding seems to imply the inappropriate use of a contingency chi-square test rather than the appropriate “goodness-of-fit” chi-square test.)

Problems with chi square: PAPAIZIAN’s (1952) equation uses only a fraction of the data available for testing interference. The resulting inefficiency, which is minimal at small values of fN_{obs} , increases with increasing fN_{obs} . Because the Papazian-based chi-square tests ascribe all of the deviation to two of the three classes, their chi-square values, based on the sum of the squares of those deviations, are enlarged. Accordingly, such Papazian-based methods will be prone to falsely claiming interference.

Below, we offer a simple, better way to test two-factor tetrad data for interference and compare it with the above-mentioned Papazian-based methods.

The better way: In the presumed absence of interference, map length (X , in morgans) may be calculated from the observed recombinant frequency (R) using HALDANE’s (1919) equation

$$X = -\frac{1}{2} \ln(1 - 2R), \quad (3)$$

where R , by definition, is

$$R = fN_{\text{obs}} + fT_{\text{obs}}/2. \quad (4)$$

The expected frequencies of the three classes of tetrads are written below in a form that is user friendly for calculation. They appear in different form in HALDANE (1931), PAPAIZIAN (1952), and SNOW (1979a),

$$fT_{\text{exp}} = \frac{2}{3}(1 - e^{-3X}) \quad (5)$$

$$fN_{\text{exp}} = \frac{1}{2}[(1 - e^{-2X}) - fT_{\text{exp}}] \quad (6)$$

$$fP_{\text{exp}} = 1 - fN_{\text{exp}} - fT_{\text{exp}}, \quad (7)$$

with all equations written for the presumed absence of interference. Multiplying these expected frequencies by the total number of tetrads in the analysis gives a corresponding set of expected numbers, which we test by chi-square goodness-of-fit (d.f. = 1) against the observed numbers (see below). For further convenience, the calculations are automated at Stahl Lab Online Tools at <http://molbio.uoregon.edu/~fstahl/>.

With respect to false positives, the performance of the better way (BW) can be compared with those of the Papazian-based methods and the maximum-likelihood estimation (MLE) (FISHER 1925), using the appropriate likelihood equation of estimation (Equation 7 of SNOW

FIGURE 1.—Frequencies of type I errors (false positives) as calculated by the Papazian-based methods, the better way (BW), and maximum-likelihood estimation (MLE) for simulated samples of populations of tetrads lacking interference. For each of 14 map lengths, a virtual population of tetrads was constructed in which the frequencies of each of the three types of tetrads were as expected for no interference (Equations 5–7). Fifty thousand samples of 1000 tetrads were simulated from each of these populations, and the fraction of times in which interference was indicated at the 5% significance level (a type I error) was plotted against the map length of the interval. Triangles, Papazian-based chi square (d.f. = 1) calculated as $(P_{\text{obs}} - P_{\text{exp}})^2/P_{\text{exp}} + (N_{\text{obs}} - N_{\text{exp}})^2/N_{\text{exp}}$; diamonds, Papazian-based chi square (d.f. = 1) calculated as $(N_{\text{obs}} - N_{\text{exp}})^2/N_{\text{exp}}$, with N_{exp} rounded to the nearest whole number; squares, BW; star, MLE.

1979a). When a universe of tetrads lacking interference is sampled, the BW behaves as well as the MLE over the range tested, giving “significance” 5% of the time (Figure 1). The false positives of the Papazian-based methods, on the other hand, increase with increasing map length. For instance, a 2003 yeast article in *Molecular and Cellular Biology* uses Papazian-based chi-square tests and reports interference significant at the 0.02 level for an interval that has 188 PDs, 16 NPDs, and 235 TTs. When tested by the better way, the P -value for that interval is 0.07. A 2003 article in *GENETICS* reports negative interference ($P = 0.03$) on the basis of 542 PDs, 80 NPDs, and 617 TTs. When tested by the BW, the P -value is 0.12.

In Table 1, we compare the properties of the three methods for some of the *Schizosaccharomyces pombe* data of MUNZ (1994) and for fictitious data that vary with respect to map length, interference, and total sample size. Table 1 shows that chi-square values for the BW and the MLE are practically identical. The Papazian-based method, on the other hand, often gives appreciably larger chi-square values. The entries in Table 1 suggest that, in the presence of positive interference, MLEs of X are slightly larger than BW estimates. This suggestion is supported by extensive comparisons (APPENDIX) made (unintentionally) by SNOW (1979a,b).

In Figure 2, the BW and the MLE are compared with respect to the efficiency with which they reflect the universe sampled. Figure 2A shows that, even for small samples (100 tetrads each), the mean values of X obtained by the BW reflect the universe mean about as well

TABLE 1

Some comparisons of Papazian, better way (BW), and maximum-likelihood estimation (MLE) as tests for interference

	X^a	PD	NPD	TT	χ^{2b}	P	Interference ^c by BW
Obs.		330	70	600			Insignificant
Exp. Pap.		307.7	92.3	600	7.0	0.008	
Exp. BW	0.67	340.9	80.9	578.3	2.62	0.11	
Exp. MLE	0.68	338.3	81.7	580.0	2.57	0.11	
Obs.		200	100	700			Positive
Exp. Pap.		—	—	—	—	—	
Exp. BW	1.15	227.2	127.2	645.6	13.66	0.0002	
Exp. MLE	1.21	220.0	131.0	649.0	13.16	0.0003	
Obs.		400	100	500			Negative
Exp. Pap.		448.4	51.6	500	50.62	<0.0001	
Exp. BW	0.60	371.4	71.4	557.1	19.47	<0.0001	
Exp. MLE	0.59	377.1	69.8	553.1	19.55	<0.0001	
Obs.		175	165	660			Insignificant
Exp. Pap.		193.2	146.8	660	3.97	0.046	
Exp. BW	2.30	172	162	666	0.16	0.689	
Exp. MLE	2.26	172.5	161.6	665.9	0.16	0.689	
Obs.		40	10	50			Insignificant
Exp. Pap.		44.84	5.16	50	5.06	0.025	
Exp. BW	0.60	37.14	7.14	55.71	1.95	0.16	
Exp. MLE	0.59	37.71	6.98	55.31	1.96	0.16	
Obs.		42	4	54			Insignificant
Exp. Pap.		39.52	6.48	54	1.1	0.29	
Exp. BW	0.48	43.47	5.47	51.05	0.618	0.43	
Exp. MLE	0.49	43.10	5.57	51.34	0.610	0.43	
Obs.		20	20	20			Negative
Exp. Pap.		38.90	1.10	20	—	—	
Exp. BW	∞	10	10	40	30	<0.0001	
Exp. MLE	4.5	10.00	9.996	39.999	30	<0.0001	
Obs.		40	0	60			Positive
Exp. Pap.		30.77	9.228	60	12.0	0.0005	
Exp. BW	0.46	45.09	5.099	49.80	7.76	0.0053	
Exp. MLE	0.47	44.33	5.273	50.39	7.53	0.0061	
Obs.		55	15	30			Negative
Exp. Pap.		68.6	1.4	30	—	—	
Exp. BW	0.46	45.10	5.099	49.80	29.27	<0.0001	
Exp. MLE	0.43	47.00	4.684	48.315	—	—	
Obs. (MUNZ 1994) ^d		124	16	148			Insignificant
Exp. Pap.		123.9	16.1	148		1	
Exp. BW	0.49	124.0	16.0	147.9		0.99	
Exp. MLE (MUNZ 1994)	0.49	124.0	16.1	147.9		0.99	
Obs. (MUNZ 1994) ^d		239	27	232			Insignificant
Exp. Pap.		244.9	21.1	232		0.18	
Exp. BW	0.43	235.1	23.1	239.8		0.32	
Exp. MLE (MUNZ 1994)	0.42	235.9	22.9	239.2		0.33 ^e	
Obs. (MUNZ 1994) ^d		280	9	114			Insignificant
Exp. Pap.		283.9	5.1	114		0.08	
Exp. BW	0.20	276.7	5.7	120.5		0.13	
Exp. MLE (MUNZ 1994)	0.20	277.1	5.7	120.2		0.16 ^f	

Expectations for the Papazian method (Exp. Pap.) are from Equations 1 and 7. Expectations for the BW (Exp. BW) and the MLE (Exp. MLE) are from Equations 5–7, and see Stahl Lab Online Tools. The calculations for Exp. BW tetrad types were made prior to rounding the X values to two decimal places. “Observed” data (Obs.) are concocted except where indicated.

^a Assuming no interference. BW values were calculated with the aid of Stahl Lab Online Tools; MLE values, based on equations of SNOW (1979a) with successive approximation carried out to two places, were calculated with the aid of an inherited Hewlett-Packard 15C calculator (RPL) and were checked, but not double checked.

^b No correction for continuity.

^c Chromosome (chiasma) interference, assuming no chromatid interference.

^d Data from *Schizosaccharomyces pombe*, which appear to lack interference.

^e We get $P = 0.32$ when we calculate chi square from Munz’s data and obtain P (d.f. = 1) from the calculator at VassarStats web site, which was used for all chi-square to P calculations in Table 1 except, probably, those credited to Munz (1994).

^f We get $P = 0.13$ (see footnote e).

FIGURE 2.—Tests of efficiency of BW and MLE. For each of 11 map lengths, a virtual population of tetrads was constructed in which the frequencies of each of the three types of tetrads were as expected for no interference (Equations 5–7). Four million samples of 100 tetrads were simulated from each of these populations. Squares, BW; stars, MLE. (A) Mean map lengths of samples. (B) Variances of samples.

as those obtained by the MLE over the range tested. Figure 2B shows that the variances of the sample distributions obtained by the BW are also practically indistinguishable from those obtained by the MLE. The test range was limited by the frequency with which sample R values $> \frac{1}{2}$ prevented the use of the BW.

The similarity in performance of the MLE and the BW justifies the use of chi square as an easily calculated test for significance when the estimate of map length (X) is based on Equation 3 rather than on the tedious (or computerized) MLE. The BW is valid for values of fT_{obs} that are $> \frac{2}{3}$ (as long as the observed R is not $> \frac{1}{2}$).

Elizabeth Housworth kindly conducted the Monte Carlo assessments of the various methods, prepared the graphs for Figures 1 and 2, and was indispensably helpful throughout. Jette Foss edited the text for clarity.

LITERATURE CITED

- BARRATT, R. W., D. NEUMEYER, D. PERKINS and L. GARNJOBST, 1954 Map construction in *Neurospora crassa*. *Adv. Genet.* **6**: 1–93.
- FISHER, R. A., 1925 *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.
- HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* **8**: 299–309.
- HALDANE, J. B. S., 1931 The cytological basis of genetical interference. *Cytologia* **3**: 55–65.
- MUNZ, P., 1994 An analysis of interference in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **137**: 701–707.
- PAPAZIAN, H. P., 1952 The analysis of tetrad data. *Genetics* **37**: 175–188.
- PERKINS, D. D., 1955 Tetrads and crossing over. *J. Cell. Comp. Physiol.* **45**(Suppl. 2): 119–149.
- SNOW, R., 1979a Maximum likelihood estimation of linkage and interference from tetrad data. *Genetics* **92**: 231–245.
- SNOW, R., 1979b Comment concerning maximum likelihood estimation of linkage and interference from tetrad data. *Genetics* **93**: 285.
- STRICKLAND, W. N., 1958 An analysis of interference in *Aspergillus nidulans*. *Proc. R. Soc. Lond. Ser. B* **140**: 82–101.

Communicating editor: G. R. SMITH

APPENDIX

SNOW (1979a) developed maximum-likelihood equations for estimating interference according to a model by BARRATT *et al.* (1954). The MLE equations contain parameters for the map length (X) and for the strength of interference (k). Snow calculated what he called $X(i)$ for “ X with interference.” He was later advised that what he had, in fact, calculated was X as it would be if there were no interference (SNOW 1979b).

SNOW (1979a, p. 237) writes: “Because of the nature of tetrad data and the conditions of the interference model, the estimate of x obtained by (9) and (10) is equal to $-(1/2)\ln p(0)$ (i.e., $p(0) = e^{-2x}$), where $p(0)$ is the proportion of tetrads with no exchanges, that is, $PD - NPD/N$. For example, a *Saccharomyces* cross involving the markers *mat1* and *his4* on chromosome 3 (see Table 2) produced 97 PD, 7 NPD, and 174 T. The maximum likelihood estimate of x with interference is 0.56390, which is $-(1/2)\ln(90/278)$.”

Note that $fP - fN = 1 - 2fN - fT = 1 - 2R$. Thus, Snow calculated the $X(i)$ values of his Table 2 from equations exactly equivalent to our Equations 3 and 4, which assume no interference (HALDANE 1919). Snow appeared at that time not to realize what he had done, judging from his remark that “the estimates for $x(i)$ are somewhat smaller than those for x , in keeping with expectations of the interference model” (p. 241). In fact, Snow’s estimates of $X(i)$ tend to be less than his MLEs of X in the no-interference model simply because the BW procedure for estimating X assuming the absence of interference is slightly biased. By comparing MLEs and BW estimates of X to five decimal places, SNOW (1979a) has given us examples of the extent of bias in the X values of the BW method. The degree of the bias can be appreciated by examining the many entries for X and $X(i)$ in SNOW’s (1979a) Table 2.