

**UCLA**

**Department of Statistics Papers**

**Title**

Exact Inference for Several 2x2 and 2xK Tables with an Ordinal Stratification Factor

**Permalink**

<https://escholarship.org/uc/item/5f96h4q9>

**Author**

Karim F. Hirji

**Publication Date**

2011-10-24



## Interface Foundation of America

---

Exact Tests for Interaction in Several  $2 \times 2$  Tables

Author(s): Karim F. Hirji, Stein E. Vollset, Isildinha M. Reis, Abdelmonem A. Afifi

Source: *Journal of Computational and Graphical Statistics*, Vol. 5, No. 3 (Sep., 1996), pp. 209-224

Published by: [American Statistical Association](#), [Institute of Mathematical Statistics](#), and [Interface Foundation of America](#)

Stable URL: <http://www.jstor.org/stable/1390800>

Accessed: 09/08/2011 17:47

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of America are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Computational and Graphical Statistics*.

<http://www.jstor.org>

# Exact Tests for Interaction in Several $2 \times 2$ Tables

Karim F. HIRJI, Stein E. VOLLSET,  
Isildinha M. REIS, and Abdelmonem A. AFIFI

The investigation of interaction in a series of  $2 \times 2$  tables is warranted in a variety of research endeavors. Though many large-sample approaches for such investigations are available, the exact analysis of the problem has been formulated for the probability statistic only. We present several alternative statistics applicable in this context. We also give an efficient polynomial multiplication algorithm to compute exact distributions and tail areas for the family of stratum-additive statistics. Besides the probability statistic, these include the score, likelihood ratio, and other statistics. In addition to comparing, in empirical terms, the diverse computational strategies for exact interaction analysis, we also explore the theoretical linkages between them. Data from published papers are used for illustration.

**Key Words:** Exact inference; Fast Fourier transform; Network algorithm; Polynomial multiplication algorithm; Test of homogeneity.

## 1. INTRODUCTION

Epidemiologic, clinical, psychological, or sociological studies may produce data depicted as several  $2 \times 2$  contingency tables. As an example, consider a multicenter two-arm binary response clinical trial. Let  $\tau$  be the treatment indicator, and  $Y$  the response variable. Suppose that for  $k = 1, \dots, K$ , the results from the  $k$ th clinical center are

		<i>Response</i>		
		$Y = 1$	$Y = 0$	
<i>Treatment</i>				<i>Total</i>
$\tau = 1$	$a_k$	$n_k - a_k$	$n_k$	
$\tau = 0$	$b_k$	$m_k - b_k$	$m_k$	

Note that for each  $k$ , the row sums  $m_k$  and  $n_k$  are fixed by design. Let  $p_k(\tau)$  denote the probability that a patient at center  $k$  responds to treatment  $\tau$ . A stratified logit model

---

Karim F. Hirji is Associate Professor, Department of Biomathematics, UCLA School of Medicine, Los Angeles, CA 90095-1766; e-mail khirji@biomath.medsch.ucla.edu. Stein E. Vollset is Professor, Section for Medical Informatics and Statistics, University of Bergen, 5021 Bergen, Norway. Isildinha M. Reis is Graduate Student, and Abdelmonem A. Afifi is Professor, Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095.

©1996 American Statistical Association, Institute of Mathematical Statistics,  
and Interface Foundation of North America  
*Journal of Computational and Graphical Statistics*, Volume 5, Number 3, Pages 209-224

formulation of the dependence of response on treatment is

$$\log \left\{ \frac{p_k(\tau)}{1-p_k(\tau)} \right\} = \alpha_k + \beta_k \tau,$$

where  $\alpha_k$  reflects the stratum effect, and  $\theta_k = \exp(\beta_k)$  is the  $k$ th stratum treatment-response odds ratio. The common odds ratio model, which assumes  $\beta_k = \beta$  for all  $k$ , is widely applied to such data (Breslow and Day 1980; Gart 1971). Prior to fitting the model, however, one needs to assess the constancy of odds ratios across the strata. Such an assessment, for instance, constitutes a critical aspect of combining evidence from multiple studies (Chalmers 1991; Mosteller and Chalmers 1992). In the statistical literature, a test of the invariance of odds ratios across strata is referred to as a test of homogeneity or heterogeneity of odds ratios, or as a test for interaction.

A number of large sample tests for interaction are available (Breslow and Day 1980; Jones, O'Gorman, Lemke, and Woolson 1989; Paul and Donner 1989, 1992; Tarone 1985; Yusuf, Peto, Lewis, Collins, and Sleight 1985). Liang and Self (1985) gave two asymptotic tests suitable for sparse data settings. When the data are sparse, or the strata small, tests that do not invoke asymptotic approximations are desirable. In this article, we deal with such so-called exact tests for the homogeneity of odds ratios in several  $2 \times 2$  tables.

Zelen (1971) constructed an exact test for homogeneity that employs the ordering principle used in the Fisher-Irwin exact test for a single  $2 \times 2$  table. For computing an exact significance level, all collections of  $K$   $2 \times 2$  tables with fixed row, column, and strata margins are ranked by their conditional probabilities. Thomas (1975) gave a computer algorithm that enumerates all  $2 \times 2$  tables with given margins to implement this test. The algorithm of Pagano and Trichtler (1983) computes the Zelen exact  $p$  value by separately evaluating a numerator and a denominator. A partial enumeration recursion is used to obtain the former, and the fast Fourier transform (FFT) is applied to compute the latter. Thomas and Gart (1992) gave a program that incorporates an improved direct enumeration method. The software package StatXact-Turbo (Cytel Software Corporation 1992) also performs the Zelen exact test for the constancy of odds ratios in stratified data. According to the StatXact manual, the method used is an adaptation of the network algorithm of Mehta, Patel, and Wei (1988). However, no specific details are provided.

All the described algorithms share one feature—they provide exact significance levels for the Zelen (1971) exact test only. To date, neither an empirical study of this test nor an evaluation of the numerical algorithms proposed for it have been reported. In this article, we present an efficient algorithm that enables the exact analysis of interaction with a broad class of statistics. It can also be used to evaluate the disparate exact tests for homogeneity and their asymptotic counterparts. Before describing the algorithm, we first present several statistics pertinent to the problem at hand.

Let  $\beta_1 = \beta$ , and for  $k \geq 2$ ,  $\beta_k = \beta + \gamma_k$ . Then  $H_0 : \beta_1 = \dots = \beta_K = \beta$  is equivalent to  $H_0 : \gamma_2 = \dots = \gamma_K = 0$ . Furthermore, let  $r_k = a_k + b_k$ ,  $l_k = \max(0, r_k - m_k)$ ,  $t_k = a_k - l_k$ , and  $u_k = \min(r_k, n_k) - l_k$ . With  $T_k$  denoting the random variate whose realization is  $t_k$ ,  $0 \leq t_k \leq u_k$ , we consider a family of stratum-additive statistics of the form

$$W_K = \sum_{k=1}^K s_k(T_k).$$

Table 1. Statistics for the Exact Analysis of Interaction in a Series of 2 x 2 Tables

Statistic	$s_k(t_k)$
Zelen	$\ln \left\{ \binom{n_k}{t_k + l_k} \binom{m_k}{r_k - t_k - l_k} \right\}$
Score	$\frac{(t_k - E[T_k   R_k = r_k; \hat{\beta}])^2}{\text{var}[T_k   R_k = r_k; \hat{\beta}]}$
Likelihood ratio	$(t_k + l_k) \ln \{(t_k + l_k) / \hat{a}_k\}$ $+ (n_k - t_k - l_k) \ln \{(n_k - t_k - l_k) / (n_k - \hat{a}_k)\}$ $+ (r_k - t_k - l_k) \ln \{(r_k - t_k - l_k) / \hat{b}_k\}$ $+ (m_k - r_k + t_k + l_k) \ln \{(m_k - r_k + t_k + l_k) / (m_k - \hat{b}_k)\}$
Chi-square	$\frac{n_k(t_k + l_k - \hat{a}_k)^2}{\hat{a}_k(n_k - \hat{a}_k)} + \frac{m_k(r_k - t_k - l_k - \hat{b}_k)^2}{\hat{b}_k(m_k - \hat{b}_k)}$
Mixture Model	$(t_k - E[T_k   R_k = r_k; \hat{\beta}])^2$
$X^2$	$(t_k + l_k - E_k)^2 / V_k$

The expressions for  $s_k(t_k)$  for the Zelen statistic and five other statistics appear in Table 1. We constructed the five statistics from their published asymptotic counterparts (Jones et. al. 1989; Liang and Self 1985; Norton 1945; Paul and Donner 1989, 1992; Yusuf et. al. 1985). When we contemplate the exact distribution of a statistic for interaction, the scale and shift factors invariant upon fixing all the margins of the  $2 \times 2 \times K$  layout can be omitted. Consequently, we see from Table 1 that the exact versions of the  $X^2$  and mixture model statistics are markedly simple in contrast to their asymptotic formulations in the article cited previously.

Some explanatory remarks concerning these statistics are in order. Consider the score statistic. Here,  $E[.]$  and  $\text{var}[.]$  denote the mean and variance, respectively, of the  $k$ th conditional hypergeometric distribution when  $\beta_k = \hat{\beta}$ . Paul and Donner (1992) considered seven versions of the score statistic. In our work, we deploy two of these, namely the statistics labeled  $T_3$  and  $T_5$ . For the first one, which is an unconditional score statistic,  $\hat{\beta}$  is the unconditional maximum likelihood estimator (mle) of the common log-odds ratio, but for the second, the conditional score statistic, it is the conditional mle. We also note that for the purpose of exact analysis, the conditional score statistic is equivalent to the modified sparse data statistic of Liang and Self (1985).

The exact version of the likelihood ratio statistic shown in Table 1 derives from that named LRT by Jones et. al. (1989), and the chi-square statistic is the one labeled  $P$  in the article. In both of these,  $\hat{a}_k$  and  $\hat{b}_k$  correspond to mle's for the cell counts  $a_k$  and  $b_k$  computed under the common odds ratio model by the iterative proportional fitting algorithm. Liang and Self (1985) also presented a sparse data statistic for a mixture model. Computing the exact distribution for it in the context of the stratified logit model is equivalent to setting  $s_k(t_k) = (t_k - E[T_k | R_k = r_k; \hat{\beta}])^2$ , where  $\hat{\beta}$  is the conditional mle of  $\beta$ . Another exact test is obtained from the heterogeneity statistic of Yusuf et. al. (1985). In Table 1, its exact version is denoted the  $X^2$  statistic. Here,  $E_k$  and  $V_k$  are the central hypergeometric mean and variance, respectively.

A common avenue for exact analysis employs a distribution of the relevant statistic that is devoid of nuisance parameters (Agresti 1992). Noting that  $\Sigma T_k$  is sufficient for  $\beta$ , and using arguments akin to those of Zelen (1971), we can show that the appropriate reference distribution for testing  $H_0$  is

$$f_K(w, t) = \Pr[W_K = w | \Sigma T_k = t, R_k = r_k, k = 1, \dots, K],$$

where  $t = \sum(a_k - l_k)$ . Now, let  $w_0$  be the observed value of  $W_K$ . Then, for the Zelen statistic, the exact significance level is computed as

$$\sum_{w \leq w_0} f_K(w, t; H_0),$$

but for the other statistics, it is obtained as

$$\sum_{w \geq w_0} f_K(w, t; H_0).$$

An advantage of tests based on  $p$  values computed this way is the guaranteed protection of the specified type I error rate. Tests based on asymptotic considerations may, on the other hand, exhibit erratic features in this regard in finite samples.

In this article, we describe an efficient algorithm that computes exact distributions and significance levels for the family of stratum-additive statistics. This family includes all the previously mentioned statistics. For the Zelen statistic, we show that our algorithm is more versatile and efficient compared to those published earlier. The interconnections between diverse algorithmic approaches are also pointed out. We now present our algorithm for exact homogeneity testing in stratified  $2 \times 2$  tables.

## 2. RECURSIVE POLYNOMIAL MULTIPLICATION (RPM) ALGORITHM

We show that determination of the null distribution of  $W_K$  can be depicted as evaluating the product of several polynomials. The problem of computing a tail area of the distribution then becomes one of multiplying selected terms from these polynomials. To establish these assertions, we first derive, in Lemma 1, the null conditional probability generating function of  $W_K$ . Lemma 2 shows that the generating function is obtainable from a partial product of a series of bivariate polynomials. Lemma 3 then indicates how the null distribution of  $W_K$  is extracted from the generating function. From here on we let  $T_k^+ = T_1 + \dots + T_k$ .

**Lemma 1.** For  $|\psi| < 1$  and  $k \leq K$ , the conditional probability generating function of  $W_k$  under  $H_0$ ,  $E[\psi^{W_k} | T_k^+ = m, R_j = r_j, j = 1, \dots, k; H_0]$ , denoted as  $G_k(m, \psi)$ , is

$$\{ \Delta_k \delta_k(m) \}^{-1} \sum_{\Sigma t_j = m} \{ \prod_{j=1}^k \pi_j(t_j) \psi^{S_j(t_j)} \},$$

where

$$\delta_k(m) = \Pr[T_k^+ = m | R_j = r_j, j = 1, \dots, k; H_0],$$

$$\pi_j(t_j) = \binom{n_j}{t_j + l_j} \binom{m_j}{r_j - t_j - l_j},$$

and

$$\Delta_k = \prod_{j=1}^k \binom{n_j + m_j}{r_j}.$$

**Lemma 2.** Define the bivariate polynomial  $H_k(\theta, \psi)$  as

$$\prod_{j=1}^k \{ \sum_{0 \leq x \leq u_j} \theta^x \psi^{S_j(x)} \pi_j(x) \}.$$

Also let  $F_k(m, \psi)$  comprise all terms of  $H_k(\theta, \psi)$  in which the power of  $\theta$  is  $m$ . Then  $G_k(m, \psi) = F_k(m, \psi) / \{ \Delta_k \delta_k(m) \}$ . In particular,  $G_K(t, \psi) = F_K(t, \psi) / \{ \Delta_K \delta_K(t) \}$ .

**Lemma 3.** Suppose  $\Omega$  is the conditional sample space of  $W_K$ , and let  $F_K(t, \psi) = \sum_{w \in \Omega} c(w) \psi^w$ . Then  $\sum_{w \in \Omega} c(w) = \Delta_K \delta_K(t)$ , and  $f_K(w, t; H_0) = c(w) / \{ \sum_{w \in \Omega} c(w) \}$ .

Lemma 1 is readily proved for  $k = 2$ , and the result extended to  $k > 2$  by induction. The main portion of the proof in both cases is almost the same as the derivation of the convolution of two independent discrete random variables. Lemma 2 is a restatement of a basic result in polynomial algebra regarding the coefficient of a term in the convolution of several polynomials. Lemma 3 follows from it by means of straightforward algebra. We note that lemma 3 indicates that if the complete distribution of  $W_K$  is extracted from the polynomial  $F_K(t, \psi)$ , then a separate computation of the denominator  $\delta_K(t)$  is not necessary. The main question then is how to determine  $F_K(t, \psi)$  in an efficient manner. In Theorem 1, proved in Appendix A, we give a recursive algorithm to compute this polynomial.

**Theorem 1.** For  $k = 1, \dots, K$ , let  $N_{1k} = \max(0, t + u_k^+ - u_K^+)$ ,  $N_{2k} = \min(t, u_k^+)$ , and for  $N_{1k} \leq n \leq N_{2k}$ ,  $m_1(n) = \max(0, n - N_{2, k-1})$  and  $m_2(n) = \min(u_k, n - N_{1, k-1})$ . Also let  $F_1(n, \psi) = \pi_1(n) \psi^{S_1(n)}$ . Then  $F_K(t, \psi)$  obtains from implementing, for  $k = 2, \dots, K$ , the recursion

$$F_k(n, \psi) = \sum_{x=m_1(n)}^{m_2(n)} \pi_k(x) \psi^{S_k(x)} F_{k-1}(n - x, \psi). \tag{2.1}$$

Even with such a recursion, however, computing the complete distribution of  $W_K$  is generally an onerous task, especially if there are many strata. For an exact  $p$  value, we need the area in a tail of the distribution. Theorem 2 shows how this is obtained in an efficient manner by imposing two conditions on recursion (2.1).

**Theorem 2.** Let  $w_0$  be the observed value of  $W_K$ , and for  $k \leq K$ , and  $n \leq t$ , let

$$\mu_k(n) = \max \sum_{j=k}^K s_j(t_j), \nu_k(n) = \min \sum_{j=k}^K s_j(t_j),$$

and

$$\lambda_k(n) = \sum \left\{ \prod_{j=k}^K \pi_j(t_j) \right\},$$

where the maximization, minimization, and summation, respectively, are over all  $(t_k, \dots, t_K)$  satisfying  $t_k + \dots + t_K = n$ ,  $0 \leq t_j \leq u_j$ ,  $t_j$  integer,  $j = k, k+1, \dots, K$ . Also let  $\mu_{K+1}(n) = \nu_{K+1}(n) = 0$  and  $\lambda_{K+1}(n) = 1$  for all  $n$ .

Assume we are at the  $k$ th stage of the recursion (2.1) and we consider a term with  $n = t_k^+$  and power of  $\psi$  equal to  $w_k$ . We check the potential for this term to eventually converge into the right tail by the following two criteria:

1. Trimming criterion. Suppose  $w_k + \mu_{k+1}(t - t_k^+) < w_0$ , then we omit this term in the  $k$ th stage polynomial  $F_k(t_k^+, \psi)$  from further consideration.
2. Augmentation criterion. Suppose  $w_k + \nu_{k+1}(t - t_k^+) \geq w_0$ , then we add the contribution of all subsequent terms emanating from this term to the right tail as

$$\text{Contribution} = \text{Contribution} + d_k(w_k)\lambda_{k+1}(t - t_k^+),$$

where  $d_k(w_k)$  is the coefficient of the term in  $F_k(t_k^+, \psi)$  for which the exponent of  $\psi$  is  $w_k$ . Then we also delete this term from the recursion.

After trimming and augmenting at stage  $K$  of the recursion, we compute the right tail probability as  $\text{Contribution}/\lambda_1(t)$ , where  $\lambda_1(t) = \Delta_K \delta_K(t)$ .

A proof of this theorem appears in Appendix A. To calculate the left tail area, as for the Zelen statistic, we just reverse roles of the quantities used in the trimming and augmentation criteria. In Appendix B we give a backward induction scheme to compute  $\mu_k(\cdot)$ ,  $\nu_k(\cdot)$ , and  $\lambda_k(\cdot)$  for all  $k$ . Note that backward induction also yields the denominator  $\lambda_1(t) = \Delta_K \delta_K(t)$ .

We have implemented the polynomial multiplication algorithm with trimming and augmentation in FORTRAN-77. The source code was compiled in an IBM-PC compatible environment with the Lahey F77L-EM/32 compiler (Lahey Computer Systems 1993). Several practical issues need to be mentioned. The efficiency of the algorithm depends on available RAM memory, degree of rounding used for the statistic, and the method used for storing the recursion vectors  $(k, t_k^+, w_k)$  and the coefficients  $d_k(w_k)$ . The choice of the rounding factor will be discussed subsequently. We used hashing with sequential search to store intermediate stage vectors (Sedgewick 1983). Another key issue is whether vectors generated at each stage are tested for trimming and augmentation as they are generated, or are stored first and tested at the end. The first strategy reduces storage and retrieval time but increases the number of tests done. On the other hand, when there are many intermediate stage vectors, the second strategy can be quite inefficient. Our implementation uses both of these techniques, switching from one to the other as the number of vectors varies. A further improvement derives from processing identical strata in a single recursion step—this involves a slight modification of the recursion given in Theorem 1. Changing the order by which the strata are processed also impacts computational efficiency. This will be illustrated in the following, when we compare our algorithm with the existing methods.



Table 2. BD6.DAT

<i>Cancer</i>		<i>Controls</i>	
<i>80 + g</i>	<i>0 - 79g</i>	<i>80 + g</i>	<i>0 - 79g</i>
1	0	9	106
4	5	26	164
25	21	29	138
42	34	27	139
19	36	18	88
5	8	0	31

NOTE: Column headings represent level of alcohol consumption. Source: Breslow and Day (1980, p. 137).

### 3. COMPARISONS WITH OTHER ALGORITHMS

Because all previous algorithms for the exact analysis of interaction in a series of  $2 \times 2$  tables invoke the Zelen statistic only, our comparisons are confined to it. All computations reported in this article were done on an IBM compatible PC with a Pentium microprocessor running at 60 MHz and with 8 Mb of random access memory.

We first compare our algorithm with those of Pagano and Tritchler (1983) and Thomas and Gart (1992), referred to as P&T and T&G, respectively. P&T was implemented as in Pagano and Tritchler (1983): the numerator was computed using their partial enumeration method, and the denominator with a fast Fourier transform (FFT). Double-precision versions of the FFT subroutines given in Press, Flannery, Teukolsky, and Vetterling (1992) were used. Specifically, the subroutine TWOFFT was used to compute the forward transform of two strata at a time, and the subroutine REALFT used for computing the inverse transform of the convolution. We also obtained the program IC2X2.EXE (version 2.04) from Donald Thomas which implements the algorithms in Thomas and Gart (1992) under a 32-bit DOS environment.

The following data sets highlight the relative features of these algorithms. (1) BD6.DAT: These data, stratified by age and shown in Table 2, are reproduced from Breslow and Day (1980), and are from a study of alcohol consumption and esophageal cancer. (2) P&T91.DAT: These data, with nine strata, are from Pagano and Tritchler (1983). They

Table 3. P&amp;T91.DAT

<i>Y = 1</i>		<i>Y = 0</i>	
<i><math>\tau = 1</math></i>	<i><math>\tau = 0</math></i>	<i><math>\tau = 1</math></i>	<i><math>\tau = 0</math></i>
7	6	1	2
1	3	5	3
4	5	6	7
3	6	7	5
6	6	7	5
5	14	2	2
7	2	5	3
1	6	7	3
8	5	7	3

NOTE: Strata order different from the original. Source: Pagano and Tritchler (1983).

Table 4. Exact Zelen Test Computing Times of Four Algorithms

Data set	p value	CPU Time (seconds)				
		Sort	P&T (1983)	T&G (1993)	RPM	SXT
BD6.DAT	.09924	No	5	<1	2	4
		Asc	4	<1	2	4
		Des	7	<1	1	4
P&T91.DAT	.56745	No	60	7	2	4
		Asc	55	7	2	4
		Des	76	7	1	4
P&T92.DAT	.07688	No	6908	880	22	28
		Asc	6905	880	29	36
		Des	8811	880	20	42
P&T181.DAT	.69962	No	Inf	Inf	53	23
		Asc	Inf	Inf	32	18
		Des	Inf	Inf	22	16

NOTE: Sorting by range of the stratum hypergeometric distribution (Asc = ascending sort, Des = descending sort); Inf = Infeasible, which means more than 24 hours; SXT = StatXact-Turbo version 2.11 run at four significant digit accuracy; rounding factor used for RPM =  $10^4$ .

appear in Table 3. (3) P&T92.DAT: These data were generated from P&T91.DAT by doubling each cell count. (4) P&T181.DAT: Here, a data set with 18 strata was formed from P&T91.DAT by attaching to it a copy of itself.

Table 4 shows the comparative results for these data. In all cases, P&T is the least efficient of the three methods. The computing times of P&T and T&G rise rapidly as the number of strata, or the stratum size, increases. In the example with  $K = 18$ , both these methods did not converge in 24 hours. The computing time of RPM, on the other hand, increased only moderately. The efficiency of T&G is unaffected by sorting as the program always sorts the strata prior to computation. P&T performs best when the strata are sorted in an increasing order by stratum range, and RPM performs best when a decreasing sort is used.

We now compare the RPM method with the network algorithm for the homogeneity testing problem. The latter has not been published, but a version of it is programmed in StatXact-Turbo (Cytel Software Corporation 1992). Like the other algorithms, it too uses the Zelen statistic only. Our comparison of these two algorithms rests on the following theoretical result.

**Theorem 3.** *The RPM algorithm for testing interaction in several  $2 \times 2$  tables is also a network algorithm.*

The proof of this appears in Appendix A. There we show a one-to-one correspondence between all the steps of the RPM algorithm and a network algorithm for the problem. Given this equivalence, we were not surprised to find that the computing times of the two methods were of the same order of magnitude. Differences in both directions were observed; these may be due to the specifics of implementation or, for larger problems, they may flow from the possible use in StatXact of some Monte Carlo method. Publication of the details of the StatXact algorithm is needed to resolve these questions.

Table 5. Pregnancy Data Sets

<i>PREG1.DAT</i> ( $K = 7$ )				<i>PREG2.DAT</i> ( $K = 8$ )			
$Y = 1$		$Y = 0$		$Y = 1$		$Y = 0$	
$\tau = 1$	$\tau = 0$	$\tau = 1$	$\tau = 0$	$\tau = 1$	$\tau = 0$	$\tau = 1$	$\tau = 0$
4	74	2	74	24	54	18	58
3	77	3	85	2	48	9	40
8	31	3	37	6	74	8	80
4	11	7	7	5	34	14	26
0	18	7	18	3	12	3	11
1	96	1	97	35	28	33	30
12	48	13	40	0	18	9	16
				1	59	1	52

Source: Goldstein, Sacks, and Chalmers (1987).

Now consider the matter of accuracy. For the RPM algorithm, the results in Table 4 were obtained with the rounding factor  $10^4$  applied to  $s_k(t_k)$ . The resulting  $p$  values agreed with those from StatXact, P&T (with one exception noted in the following), and T&G to four significant digits. This issue is considered further in Section 5.

For some data, we noted that P&T produced highly inaccurate results. The source of the errors was traced to the use of the FFT to compute the denominator,  $\delta_K(t)$ , for data sets in which it is very small. We illustrate this point using BD6.DAT. For these data, the correct  $\delta_K(t)$ , as determined by backward induction and verified by direct computation of the conditional distribution for the common odds ratio model, is  $.5236 \times 10^{-18}$ . The FFT method, however, gave  $\delta_K(t) = .1186 \times 10^{-15}$ . Thus, while the correct Zelen statistic exact  $p$  value is .09924, P&T yields .0004381. The impact of the FFT-induced error on testing the null hypothesis in the common odds ratio model is inconsequential (Vollset, Hirji, and Elashoff 1991); however, for testing the heterogeneity of odds ratios, that is seen not to be the case. For BD6.DAT, the unconditional and conditional score asymptotic  $p$  values are .1074 and .1079, respectively. With such data, the algorithm of Pagano and Trichtler (1983) may thus yield results substantially less precise compared to that provided by an approximate test.

#### 4. APPLICATIONS

Using the Zelen, score, and  $X^2$  statistics, we present analyses of interaction in three real data sets. The first, BD6.DAT, is given in Table 2. The other two, labeled PREG1.DAT and PREG2.DAT in Table 5, are from a review of the use of hormone administration for maintenance of pregnancy (Goldstein, Sacks, and Chalmers 1987). PREG1.DAT is from table 38.9 in that report, and shows the effect of progestogens on miscarriage, stillbirth, or neonatal death. PREG2.DAT is extracted from table 38.10 of the report; we have deleted two uninformative strata with no events. It shows the effect of progestogens on prematurity as observed in ten studies.

The results of our analyses are in Table 6. Consider first BD6.DAT. The three exact  $p$  values here are similar, ranging from .09 to .10. The two score asymptotic  $p$  values are slightly larger. But the  $X^2$  asymptotic  $p$  value is much smaller. This statistic, advocated by Yusuf et al. (1985), is identical to the asymptotic statistic for homogeneity

Table 6. Interaction Test Results for BD6.DAT, PREG1.DAT, and PREG2.DAT

<i>Data: Statistic</i>	<i>CPU Time</i>		<i>p value</i>	
	<i>Exact</i>	<i>Exact</i>	<i>Exact</i>	<i>Asymptotic</i>
<i>BD6.DAT</i>				
Zelen	1	.09924		
Conditional Score	1	.09168		.10789
Unconditional Score	1	.09151		.10739
$X^2$	1	.08563		.00682
<i>PREG1.DAT</i>				
Zelen	2	.05935		
Conditional Score	2	.07921		.08758
Unconditional Score	2	.07919		.08750
$X^2$	.08090	.09211		
<i>PREG2.DAT</i>				
Zelen	16	.00761		
Conditional Score	17	.01132		.01542
Unconditional Score	17	.01132		.01538
$X^2$	17	.01203		.01809

NOTE: Data sorted in a descending order by range of hypergeometric distribution; rounding factor used for RPM is  $10^4$ .

of Zelen (1971). Halperin et al. (1977) critiqued the validity of the asymptotic distribution designated for it by Zelen (1971). However, for BD6.DAT, as in the two other examples to follow, the exact  $X^2$   $p$  value is similar to the two exact score statistic  $p$  values. The problem may therefore lie not with the use of the statistic per se as with the validity of the approximations used for it.

Now consider the results for PREG1.DAT and PREG2.DAT. For both these data sets we note that, in terms of exact analysis, the Zelen statistic provides stronger evidence for heterogeneity than do the other three. Moreover, the exact  $p$  values for these three statistics are similar. The Zelen statistic is the only one in current use. This discordance between it and the other statistics warrants further study.

It is pertinent to mention here that in the three examples described earlier the total sample size is not small, and even the counts in the strata are not that low. That even for such data exact and asymptotic  $p$  values differ to the extent they do, both between and within themselves indicates the need for an appropriate evaluation of the ways of computing exact and asymptotic significance levels. Such an evaluation should examine issues like conservativeness, bias, and power over a comprehensive configuration of model parameters and data (Agresti 1992).

Finally, we comment on the mixture model statistic. Liang and Self (1985) derived it from a model distinct from the logit model assumed by us. However, we use this statistic to address the following question: If the logit model is the true model for the data, does the use of the mixture model statistic yield significance levels similar to those from statistics based on the correct model? Ordinarily, answers to such questions are confounded by inaccuracies in the asymptotics employed. The algorithm we give permits us to circumvent that problem. From the comparisons done with several data sets, we found that the mixture model statistic tends to give exact  $p$  values markedly different from those derived from other statistics. For example, while the exact  $p$  values

for BD6.DAT for the four statistics shown in Table 6 are between .09 and .10, the exact mixture model statistic  $p$  value is .2095. A more detailed study of this issue is also warranted.

### 5. ALGORITHMS FOR EXACT ANALYSIS

We now explore the linkages between the algorithms for exact inference on the interaction in a series of  $2 \times 2$  tables. Let  $\mathbf{t} = (t_1, \dots, t_K)$ , and consider the set  $\mathbf{T} = \{\mathbf{t} : t_1 + \dots + t_K = t, 0 \leq t_k \leq u_k, k = 1, \dots, K\}$ . Each element of  $\mathbf{T}$  corresponds to a  $2 \times 2 \times K$  contingency table with the margins fixed.

#### 5.1 CELL-WISE VERSUS STAGE-WISE RECURSION

We define a *cell-wise* recursive process for computing the exact distribution of  $W_K$  as one founded upon generation of the individual members of  $\mathbf{T}$ . In terms of polynomial multiplication, this involves selection of a term at a time from each polynomial

$$\sum_{0 \leq x \leq u_j} \pi_j(x) \theta^x \psi^{S_j(x)},$$

provided the  $K$ -fold combination of terms is constrained so that the powers of  $\theta$  sum to  $t$ . Repetitive multiplication, over a maximum  $\nu_K = (1 + u_1) \dots (1 + u_K)$  combinations of such terms, generates the required distribution. In the network context, cell-wise recursion enumerates all the paths in the network. The effort entailed in cell-wise recursion rises rapidly as  $K$  or  $\max(u_K)$ , and hence  $\nu_K$ , increase. The computation of the numerator in the algorithm of Pagano and Tritchler (1983) uses cell-wise recursion. Judging from its empirical performance and the partial description given in the article, the algorithm of Thomas and Gart (1992) also appears to employ cell-wise recursion. When we need a tail area, cell-wise recursion is improved by trimming and augmentation. Pagano and Tritchler (1983) gave a technique to trim at the last two steps of generating an element of  $\mathbf{T}$ . The backward-induction-derived criteria are also applicable in the cell-wise context, provided the statistic under consideration is stratum additive.

Another avenue to compute the exact distribution of  $W_K$  is the *stage-wise* recursive approach. This is what we implemented. Let  $\mathbf{t}_k = (t_1, \dots, t_k)$ , and  $\mathbf{T}_k = \{\mathbf{t}_k : \max(0, t - (u_{k+1} + \dots + u_K)) \leq t_k^+ \leq \min(u_k^+, t), 0 \leq t_j \leq u_j, j = 1, \dots, k\}$ . The  $\mathbf{T} = \mathbf{T}_K$ . The stage-wise approach is based on generating the classes of pairs  $\{(t_k^+, w_k) : \mathbf{t}_k \in \mathbf{T}_k\}, k = 1, \dots, K$  in a recursive manner. All computations at a stage are completed before embarking on the next one; unlike in cell-wise recursion, there is no return to an earlier stage. In polynomial terms, this implies performing recursion (2.1) for all powers of  $\theta$  and  $\psi$  before proceeding further, while under network theory it requires processing all records at all the nodes of a particular stage before going on.

Stage-wise recursion derives its efficiency from the aggregation of multiple subsequences in  $\mathbf{T}_k$  into the same pair  $(t_k^+, w_k)$ . This obviates separate computation for each subsequence, and enables joint performance of the trimming and augmentation tests. We implemented the polynomial approach using both the stage-wise and the cell-wise approach. The former was distinctly more efficient. For example, for the data set

P&T92.DAT with  $K = 9$ , the cell-wise method with descending sort, and trimming and augmentation using backward induction bounds, gave the Zelen test exact  $p$  value in 229 seconds; the stage-wise approach took only 21 seconds.

We note that when  $K$  is small, or with moderately large  $K$  but sparse strata, the Thomas and Gart (1992) cell-wise algorithm is very efficient, outperforming all the other methods. However, the differences in computing time between various methods in such problems do not exceed a few seconds. This cell-wise approach, and more so that of Pagano and Tritchler (1983), can witness a dramatic decrease in efficiency as the range of the stratum hypergeometric distributions, or the number of strata increase. Thus, for P&T92.DAT, the cell-wise polynomial approach at 229 seconds is about 30 times faster than the Pagano and Tritchler (1983) approach, and four times faster than the Thomas and Gart (1992) algorithm (see Tab. 4). This is due to the use of backward-induction-derived trimming and augmentation criteria.

Computer implementation of the cell-wise approach requires less memory than the stage-wise approach does. The higher memory requirements of the latter are necessitated by the need to store, for the current and previous stage, all triplets  $(t_k^+, w_k, d_k(w_k))$ . In order to keep the number of records at a manageable level,  $w_k$  can be rounded. This tends to introduce inaccuracies in the final result. If a low rounding factor is used, a result generally less accurate than that from the cell-wise approach may ensue. The question of rounding in stage-wise algorithms arises for other problems as well. For example, for stratified linear rank statistics, the StatXact-Turbo manual recommends the default rounding factor  $10^3$ . For the problem of exact interaction analysis, we experimented with factors ranging from  $10^2$  to  $10^6$  in a variety of data, and for different statistics. Our general conclusion is that the rounding factor  $10^4$  suffices to rapidly give  $p$  values almost always accurate to four significant digits. That holds for data with very small or very large  $p$  values, and also for data for which the cell-wise approach is either too slow or infeasible. The four significant digit concordance of the  $p$  values obtained by the different methods for the data in Table 4 illustrate this claim.

An additional point of note is that the stage-wise approach is applicable to conditionally stratum-additive statistics only. All the statistics described earlier are of this type. A commonly used large sample statistic violating this property is the Breslow and Day (1980) statistic for homogeneity. For performing exact analysis with this statistic we may set

$$s_k(t_k) = \frac{(t - e_k[\hat{\beta}_m])^2}{v_k[\hat{\beta}_m]},$$

where  $\hat{\beta}_m$  is the Mantel-Haenszel estimate of the common log-odds ratio, and  $e_k[\hat{\beta}_m]$  and  $v_k[\hat{\beta}_m]$  are asymptotic estimates of  $E[T_k|R_k; \hat{\beta}_m]$  and  $\text{var}[T_k|R_k; \hat{\beta}_m]$ , respectively. The fact that  $\hat{\beta}_m$  depends on the interior cell values makes the statistic nonstratum-additive, precluding the use of the stage-wise approach for it. Furthermore, we note that for this statistic, even if a cell-wise recursive approach was used, we cannot apply the backward-induction-derived trimming and augmentation criteria, since even the use of these is confined to stratum-additive statistics. Another statistic not conditionally stratum-additive is that of Cochran (1954). The development of efficient algorithms for exact analysis using such statistics is an unexplored research area.

## 5.2 POLYNOMIAL GENERATION VERSUS POLYNOMIAL EVALUATION

Another conceptual distinction in algorithmic design is between polynomial generation and polynomial evaluation. Both the complete distribution of  $W_K$  and its tail are embodied in the polynomial  $F_K(t, \psi)$ . The cell-wise and stage-wise approaches are polynomial generation approaches because they deploy recursion (2.1) to either directly construct this polynomial, or to compute the sum of the coefficients of selected terms of the polynomial. A polynomial evaluation approach, on the other hand, uses recursion (2.1) to evaluate the polynomial at several values of  $\psi$ , and then uses a polynomial inversion algorithm to construct the required polynomial. If the evaluations are done at the complex roots of unity, the fast Fourier transform is used in the inversion step (Press et al. 1992). The use of trimming and augmentation has not been developed for this approach. Thus, even a significance level computation necessitates the generation of the entire distribution. Furthermore, rounding errors associated with the Fourier transform introduce inaccuracies in the final result (Duhamel and Vetterli 1991). The memory requirements of the evaluation approach are comparable to those of cell-wise recursion.

The algorithm of Pagano and Tritchler (1983) is an amalgam of polynomial generation and evaluation; the numerator is computed by cell-wise polynomial generation, and the denominator by an FFT-based evaluation approach. A comprehensive FFT approach for this problem would compute both the numerator and the denominator using the FFT. However, the potential for error noted earlier would be compounded by the introduction of errors in both the numerator and the denominator. Vollset, Hirji, and Elashoff (1991) showed that for the common odds ratio model the FFT method gives precise  $p$  values but very inaccurate confidence limits. Our results show that for testing interaction in several  $2 \times 2$  tables, the FFT-derived  $p$  values can be highly inaccurate. Moreover, this method is less efficient compared to the other alternatives. The tendency to yield substantially imprecise  $p$  values in the Pagano and Tritchler (1983) algorithm can be rectified by using backward induction instead of the FFT to compute the denominator. This, however, does not improve efficiency, since computation of the denominator ordinarily takes up less than .1% of the total time.

To summarize, we have shown that the algorithmic approaches for exact analysis of interaction in stratified  $2 \times 2$  tables are based, in one way or another, upon recursion (2.1). This recursion embodies all the elements of the isomarginal set  $\mathbf{T}$ . What distinguishes the algorithms is how the recursion is used: in a cell-wise or stage-wise manner, or for an evaluative or generative purpose. Connecting and contrasting computational algorithms through such a scheme has ramifications beyond the problem under study, and is applicable to many models for exact analysis of discrete and continuous data.

In conclusion, the recursive polynomial multiplication method presented in this article gives accurate exact significance levels for a variety of conditionally stratum-additive statistics in a reasonable time even for moderately large data sets. Using this algorithm, we have embarked on a comprehensive study of exact and asymptotic tests for interaction in several  $2 \times 2$  tables. In addition to the statistics already mentioned, we will implement our algorithm for the exact likelihood ratio and chi-square statistics. The results will be reported in a future article. A FORTRAN program implementing the RPM method is available from the authors free of charge.

## APPENDIXES

### A. PROOFS OF THEOREMS

We now give proofs of Theorems 1, 2, and 3.

**Proof of Theorem 1:** This theorem is proved by consideration of the product of relevant terms of  $H_{k-1}(\theta, \psi)$  with the terms of the  $k$ th stratum polynomial to get terms of  $H_k(\theta, \psi)$  in which the power of  $\theta$  is  $n$ .  $\square$

**Proof of Theorem 2:** It is clear that vectors identified by the trimming criterion cannot be in the required tail of the distribution, and dropping them does not affect the  $p$  value. Furthermore, the augmentation criterion identifies a group of vectors all of which will be in the required tail of the distribution. The total probability of this group of vectors is  $\Pr[T_k^+ = t_k^+, W_k = w_k, W_K \geq w_0 | T_K^+ = t, R_j = r_j, j = 1, \dots, k]$ , which can be written as  $\Pr[T_k^+ = t_k^+, W_k = w_k | T_K^+ = t, R_j = r_j, j = 1, \dots, k] \times \Pr[W_K \geq w_0 | T_k^+ = t_k^+, W_k = w_k, T_K^+ = t, R_j = r_j, j = 1, \dots, k]$ . The expression  $d_k(w_k)\lambda_{k+1}(t - t_k^+)$  is, within a normalization constant, just this probability.  $\square$

**Proof of Theorem 3:** We show that a network algorithm for this problem is actually a pictorial representation of recursion (2.1). Consider an acyclic network with  $(K + 1)$  stages. At stage 0, we have one node  $(0,0)$ ; at stage  $K$ , we also have a single node  $(K, t)$ ; and at stage  $k$ , for  $0 < k < K$ , we have nodes denoted by  $(k, t_k^+)$ , where  $t_k^+$  is the cumulative sum  $t_1 + \dots + t_k$ . For each  $k$ , let  $N_{1k} \leq t_k^+ \leq N_{2k}$ , with  $N_{1k}$  and  $N_{2k}$  as defined in Theorem 1. Then  $(k, N_{1k}), (k, N_{1k} + 1), \dots, (k, N_{2k})$  constitute the set of nodes for stage  $k$ .

Now suppose from each node  $(k - 1, t_{k-1}^+)$  at stage  $k - 1$ , there is an arc to a node in stage  $k$ ,  $(k, t_k^+)$ , provided the restrictions on the final value of  $t_1 + \dots + t_K$  and on the value of each  $t_k$  are not violated. For each such arc, let  $\pi_k(t_k^+ - t_{k-1}^+) = \pi_k(t_k)$  be the probability arc length, and  $s_k(t_k^+ - t_{k-1}^+) = s_k(t_k)$  be the rank length. In network theory parlance, the set of values of  $W_k$  with cumulative sum  $t_k^+$  constitutes the set of records at node  $(k, t_k^+)$ .

With these terminological conventions in mind, consider now the recursion (2.1). Each term of  $F_k(n, \psi)$  is then equivalent to a record at node  $(k, n)$ . A term in the summation in the right side of (2.1) represents the contribution of an arc from a node  $(k - 1, n - x)$  at stage  $k - 1$  to this node at stage  $k$ . The multiplication of coefficients in this summation is equivalent to updating the probability length along this arc, while adding the powers of  $\psi$  is identical to updating the cumulative rank length. Furthermore, the shortest and longest paths computed for this network are identical to the quantities  $\nu_k(n)$  and  $\mu_k(n)$  given in Theorem 2, and are used for an identical purpose.  $\square$

### B. BACKWARD INDUCTION SCHEME

For  $k \leq K$ , and  $q \leq t$ , let  $\mu_k(q) = \max \sum_{j=k}^K s_j(t_j)$ ,  $\nu_k(q) = \min \sum_{j=k}^K s_j(t_j)$ , and  $\lambda_k(q) = \sum \{ \prod_{j=k}^K \pi_j(t_j) \}$ , where the maximization, minimization, and summation, respectively, are over all  $(t_k, \dots, t_K)$  satisfying  $\sum_{j=k}^K t_j = q$ ,  $0 \leq t_j \leq u_j$ ,  $t_j$  integer,



$j = k, k + 1, \dots, K$ . For any of the  $\mu_k(q), \nu_k(q)$  or  $\lambda_k(q)$  to be defined, we must have that  $\sum_{k=1}^K u_k \geq q$ .

Let  $N_{1k}^* = \max(0, t - u_k^+)$ ,  $N_{2k}^* = \min(t, u^+ - u_k^+)$ , and for  $N_{1k}^* \leq q \leq N_{2k}^*$ ,  $m_{1k}^*(q) = \max(0, q - N_{2k}^*)$  and  $m_{2k}^*(q) = \min(u_k, q - N_{1k}^*)$ . If, for  $k < K$ , and  $j = k, k + 1, \dots, K$ ,  $t_j$  units are allocated to  $j$ th stage, and  $y_k = t_k + \dots + t_K$ , then  $y_k$  and  $t_k$  satisfy

$$N_{1,k-1}^* \leq y_k \leq N_{2,k-1}^* \tag{B.1}$$

and, with  $y_k$  fixed,

$$m_{1k}^*(y_k) \leq t_k \leq m_{2k}^*(y_k). \tag{B.2}$$

Consider first  $k = K$ . Then for  $N_{1K}^* \leq q \leq N_{2K}^*$ , it is clear that  $\mu_K(q) = \nu_K(q) = s_K(q)$ , and  $\lambda_K(q) = \pi_K(q)$ . For  $k \leq K - 1$ , to compute  $\mu_k(y_k), \nu_k(y_k), \lambda_k(y_k)$ , we implement, for  $y_k$  allowable by (B.1), and  $t_k$  allowable by (B.2),

$$\mu_k(y_k) = \max_{t_k} \{ s_k(t_k) + \mu_{k+1}(y_k - t_k) \},$$

$$\nu_k(y_k) = \min_{t_k} \{ s_k(t_k) + \nu_{k+1}(y_k - t_k) \},$$

and

$$\lambda_k(y_k) = \sum_{t_k=m_{1k}^*(y_k)}^{m_{2k}^*(y_k)} \pi_k(t_k) \lambda_{k+1}(y_k - t_k).$$

Implementing these recursions in  $K - 1$  stages for  $k = K - 1, K - 2, \dots, 2, 1$ , yields, by the optimality principle, the required maximum, minimum, and sum (Hadley 1964).

### ACKNOWLEDGMENTS

This work was supported by USHHS Grant No. CA 16042 and a grant from the Norwegian Research Council.

[Received September 1994. Revised July 1995.]

### REFERENCES

Agresti, A. A. (1992), "A Survey of Exact Inference for Contingency Tables" (with discussion), *Statistical Science*, 7, 131-177.

Breslow, N. E., and Day, N. E. (1980), *Statistical Methods in Cancer Research: I. The Analysis of Case-Control Studies*, Lyon: International Agency for Research on Cancer.

Chalmers, T. C. (1991), "Problems Induced by Meta-Analysis," *Statistics in Medicine*, 10, 971-980.

Cochran, W. G. (1954), "The Combination of Estimates from Different Experiments," *Biometrics*, 10, 101-129.

Cytel Software Corporation (1992), *StatXact-Turbo: Statistical Package for Exact Nonparametric Inference*, (ver. 2.11), Boston, MA: Author.

- Duhamel, P., and Vetterli, M. (1991), "Fast Fourier Transforms: A Tutorial Review and a State of the Art," *Signal Processing*, 19, 259–299.
- Gart, J. J. (1971), "The Comparison of Proportions: A Review of Significance Tests, Confidence Intervals and Adjustments for Stratification," *Review of the International Statistical Institute*, 39, 148–169.
- Goldstein, P. A., Sacks, H. S., and Chalmers, T. C. (1987), "Hormone Administration for the Maintenance of Pregnancy," in *Effective Care in Pregnancy and Childbirth*, eds. M. W. Enkin, M. J. N. C. Keirse, and I. Chalmers, London: Oxford University Press, pp. 612–623.
- Hadley, G. (1964), *Nonlinear and Dynamic Programming*, Reading, MA: Addison-Wesley.
- Halperin, M., Ware, J. H., Byar, D. P., Mantel, N., Brown, C. C., Koziol, J., Gail, M., and Green, S. B. (1977), "Testing for Interaction in an  $I \times J \times K$  Contingency Table," *Biometrika*, 64, 271–275.
- Jones, M. P., O'Gorman T. W., Lemke, J. H., and Woolson, R. F. (1989), "A Monte-Carlo Investigation of Homogeneity Tests of the Odds Ratio under Various Sample Size Configurations," *Biometrics*, 45, 171–181.
- Lahey Computer Systems, Inc. (1993), *F77LEM-32 FORTRAN*, (ver. 5.2), Incline Village, NV: Author.
- Liang, K. Y., and Self, S. G. (1985), "Tests of Homogeneity of Odds Ratio when the Data are Sparse," *Biometrika*, 72, 353–358.
- Mehta, C. R., Patel, N. R., and Wei, L. J. (1988), "Computing Exact Significance Tests with Restricted Randomization Rules," *Biometrika*, 75, 295–302.
- Mosteller, F., and Chalmers, T. C. (1992), "Some Progress and Problems in Meta-Analysis of Clinical Trials," *Statistical Science*, 7, 227–236.
- Norton, H. W. (1945), "Calculation of Chi-square for Complex Contingency Tables," *Journal of the American Statistical Association*, 40, 251–258.
- Pagano, M., and Trichtler, D. (1983), "Algorithms for the Analysis of Several  $2 \times 2$  Contingency Tables," *SIAM Journal of Scientific and Statistical Computing*, 4, 302–309.
- Paul, S. R., and Donner, A. (1989), "A Comparison of Tests of Homogeneity of Odds Ratios in  $K \times 2 \times 2$  Tables," *Statistics in Medicine*, 8, 1455–1468.
- (1992), "Small Sample Performance of Tests of Homogeneity of Odds Ratios in  $K \times 2 \times 2$  Tables," *Statistics in Medicine*, 11, 159–165.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992), *Numerical Recipes: The Art of Scientific Computing* (2nd ed.), Cambridge, UK: Cambridge University Press.
- Sedgewick, R. (1983), *Algorithms*, Reading, MA: Addison-Wesley.
- Tarone, R. E. (1985), "On Heterogeneity Tests Based on Efficient Scores," *Biometrika*, 72, 91–95.
- Thomas, D. G. (1975), "Exact and Asymptotic Methods for the Combination of  $2 \times 2$  Tables," *Computers and Biomedical Research*, 8, 423–446.
- Thomas, D. G., and Gart, J. J. (1992), "Improved and Extended Exact and Asymptotic Methods for the Combination of  $2 \times 2$  Tables," *Computers and Biomedical Research*, 25, 75–84.
- Vollset, S. E., Hirji, K. F., and Elashoff, R. M. (1991), "Fast Computation of Exact Confidence Limits for the Common Odds Ratio in a Series of  $2 \times 2$  Tables," *Journal of the American Statistical Association*, 86, 404–409.
- Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985), "Beta Blockade During and After Myocardial Infarction: An Overview of the Randomized Trials," *Progress in Cardiovascular Diseases*, 27, 335–371.
- Zelen, M. (1971), "The Analysis of Several  $2 \times 2$  Contingency Tables," *Biometrika*, 58, 129–137.