

This material may be protected by copyright law (Title 17 U.S. Code).

# Assessing Treatment Effects in Caries Trials Using Ordered Categorical Data

J. L. FLEISS

Division of Biostatistics, Columbia University School of Public Health, 600 W. 168<sup>th</sup> Street, New York, New York 10032

Several methods are available for analyzing the results of caries trials when response to treatment is measured on an ordered categorical scale, such as Grainger's Severity Index and modifications of it. These methods are applied to data from a randomized clinical caries trial. As the methods become more informative, they also become more difficult computationally. The phenomenon of reversals and the selection of patients for study are also discussed.

J Dent Res 63(Spec Iss):778-782, May, 1984

Even though most clinical and epidemiological studies of dental caries continue to rely on the DMFS score, dissatisfaction with it has led to the development of categorical indices or rating scales for measuring the prevalence and severity of dental caries. One of the more important of these is Grainger's Severity Index<sup>1</sup>, which is used as follows:

Five mutually exclusive zones of the permanent dentition are examined for evidence of caries. These zones are as follows:

**Zone 1:** Occlusal pit and fissure surfaces of molars and pre-molars, and buccal pits and lingual grooves of molars.

**Zone 2:** Approximal surfaces of posterior teeth, including the distal surfaces of canines.

**Zone 3:** Approximal surfaces of the maxillary anterior teeth, excluding the distal surfaces of canines.

**Zone 4:** Labial surfaces of the anterior teeth.

**Zone 5:** Approximal surfaces of the mandibular teeth, excluding the distal surfaces of canines.

The examiner begins by examining the surfaces in Zone 5 for evidence of caries, and stops with a classification of the patient into category 5 if such evidence is found. If no evidence of caries is found in Zone 5, the examiner proceeds to examine the surfaces in Zone 4, and continues until some evidence of caries is found, with the patient being classified into the category corresponding to the zone in which caries was first detected. Patients with no evidence of caries are classified into category 0.

Kingman<sup>2</sup> proposed a modification of Grainger's Severity Index, denoted the MGSI, which calls for an examination of the surfaces in each of the zones and for assigning as a score to a patient the number of zones in which evidence of caries was found. It is the MGSI that serves in this paper as an example of an ordered categorical scale for the measurement of dental caries. Some results of a comparative study of mouthrinses<sup>3</sup> will be used to illustrate a number of the methods available for analyzing ordered categorical data.

## Clinical methods.

A total of 608 children aged 10-12 in Biddeford, Maine, was randomized to one of three treatment groups: daily mouthrinse with 0.05% NaF (n=199), weekly mouthrinse with 0.2% NaF (n=200), and placebo rinse with 0.1% NaCl (n=209). Each subject was classified on the MGSI at baseline and after 24 months of treatment by an examiner kept blind to the subject's treatment. For simplicity, only the data from the weekly and the placebo

This work was supported in part by Grant DE 04068 from the National Institute of Dental Research.

mouthrinse groups will be analyzed. The results are presented in Table 1.

## Analysis of post-treatment data only.

If the pre-treatment data were not available (or if the study protocol did not call for baseline measurements), the two treatment groups would be compared with respect to their post-treatment distributions across the categories of the MGSI (i.e., the bottom row of Table 1a would be compared with the bottom row of Table 1b: See Table 2). The formula for the standard chi-square statistic for comparing the frequencies in a general 2xk table may be shown<sup>4</sup> to simplify to

$$\chi^2 = \frac{n_1 n_2}{n_1 + n_2} \sum_{i=1}^k \frac{(p_{i1} - p_{i2})^2}{\bar{p}_i} \quad (\text{Eq. 1})$$

with  $k - 1$  degrees of freedom, where  $n_1$  and  $n_2$  are the two sample sizes;  $\{p_{i1}: i=1, \dots, k\}$  and  $\{p_{i2}: i=1, \dots, k\}$  are the two relative frequency distributions; and  $\bar{p}_i = (n_1 p_{i1} + n_2 p_{i2}) / (n_1 + n_2)$ , the overall proportion of patients assigned to category  $i, i=1, \dots, k$ .

The value of chi-square for the data in Table 2 is  $\chi^2 = 4.61$  (d.f. = 4, n.s.). No significant difference was found from this simple analysis that took no account at all of the ordering inherent in the scale. Of the several methods available for taking the ordering into account, the following is the simplest:

Let the categories be numbered from 0 to  $k - 1$  (as with the MGSI), let  $x_0 > x_1 > \dots > x_{k-1}$  be a sequence of numerical constants, and define

$$d = \sum x_i (p_{i1} - p_{i2}) \quad (\text{Eq. 2})$$

TABLE 1  
POST-TREATMENT DISTRIBUTION OF PATIENTS ON  
MODIFIED GRAINGER SEVERITY INDEX BY TREATMENT  
AND BY PRE-TREATMENT CATEGORY

(a) Weekly Rinse Group						
Pre-treatment Category	Post-treatment Category					Total
	0	1	2	3	4	
0	16	7	1	0	0	24
1	3	77	18	2	0	100
2	0	5	43	5	1	54
3	0	0	1	14	2	17
4	0	0	0	1	4	5
Total	19	89	63	22	7	200

  

(b) Placebo Rinse Group						
Pre-treatment Category	Post-treatment Category					Total
	0	1	2	3	4	
0	11	9	3	1	0	24
1	1	71	28	6	0	106
2	0	3	40	7	1	51
3	0	0	2	15	5	22
4	0	0	0	0	6	6
Total	12	83	73	29	12	209

TABLE 2  
OVERALL POST-TREATMENT RELATIVE FREQUENCY DISTRIBUTIONS OF TWO TREATMENT GROUPS

Group	0	1	2	3	4	Total
Weekly ( $p_{11}$ )( $n_1 = 200$ )	0.0950	0.4450	0.3150	0.1100	0.0350	1.0000
Placebo ( $p_{12}$ )( $n_2 = 209$ )	0.0574	0.3971	0.3493	0.1388	0.0574	1.0000
Overall ( $\bar{p}_1$ )( $n_1+n_2 = 409$ )	0.0758	0.4205	0.3325	0.1247	0.0465	1.0000
Difference	0.0376	0.0479	-0.0343	-0.0288	-0.0224	0
$x_i$	0.9621	0.7140	0.3375	0.1089	0.0232	

TABLE 3  
POST-TREATMENT RELATIVE FREQUENCY DISTRIBUTIONS OF TWO GROUPS IN PRE-TREATMENT CATEGORY 2

Group	Post-treatment Category					Total
	0	1	2	3	4	
Weekly ( $n_{21} = 54$ )	0	0.0926	0.7963	0.0926	0.0185	1.0000
Placebo ( $n_{22} = 51$ )	0	0.0588	0.7843	0.1373	0.0196	1.0000
Overall ( $n_2 = 105$ )	0	0.0762	0.7905	0.1143	0.0190	1.0000
Difference	0	0.0338	0.0120	-0.0447	-0.0011	0
$x_i$	0.9621	0.7140	0.3375	0.1089	0.0232	

The quantity  $d$  is a weighted sum of the differences between the relative frequencies with the following properties:

(i) If the two treatments have identical underlying frequency distributions, the expected value of  $d$  is zero.

(ii) If treatment 1 is superior to treatment 2, then relatively more of the patients in group 1 should end up classified in the lower ranking categories (for which  $x_i$  is large), and relatively more of the patients in group 2 should end up classified in the higher ranking categories (for which  $x_i$  is small). The net result is that  $d$  may be expected to be large in the positive direction.

(iii) If treatment 2 is superior to treatment 1, the reverse pattern should emerge, and  $d$  may be expected to be large in the negative direction.

A test for the significance of the difference between the two distributions that is sensitive to one treatment's being superior to the other may be based on the magnitude of

$$\chi^2 = \frac{n_1 n_2}{n_1 + n_2} \cdot \frac{d^2}{\sum \bar{p}_i (x_i - \bar{x})^2} \quad (\text{Eq. 3})$$

with one degree of freedom, where  $\bar{x} = \sum \bar{p}_i x_i$ . For the data in Table 2, take the values of  $x_0, \dots, x_4$  to be the overall ridsits<sup>5</sup>:  $x_0 = \bar{p}_0/2 + \bar{p}_1 + \dots + \bar{p}_4$ ;  $x_1 = \bar{p}_1/2 + \dots + \bar{p}_4$ ;  $\dots$ ;  $x_4 = \bar{p}_4/2$ .

Then

$d = 0.9621 \times 0.0376 + 0.7140 \times 0.0479 + \dots + 0.0232 \times (-0.0224) = 0.0551$ , which is positive, reflecting the larger relative frequencies for the weekly than the placebo mouthrinse group in categories 0 and 1, and the larger relative frequencies for the placebo than the weekly mouthrinse group in the remaining categories. The value of  $d$  may be interpreted as follows (see reference 5): When the constant 0.50 is added to  $d$ , the resulting value is an estimate of the probability that a randomly selected member of group 1 will end up after treatment as well as or better off than a randomly selected member of group 2. For the current data,  $d + 0.50 = 0.5551$ , so the chances are 55.51% that a typical subject using the weekly mouthrinse will, after two years, have carries the same as or less severe than a typical subject using the placebo mouthrinse.

The value of  $\bar{x}$  is 0.50, and the value of the chi-square statistic in (3) is  $\chi^2 = 4.20$  (d.f. = 1,  $p < 0.05$ ). Thus, by taking the ordering into account in a relatively simple way, a statistically significant difference between the treatments

has been demonstrated: Weekly mouthrinse is significantly superior to placebo rinse after 24 months with respect to the MGSI. Similar values of chi-square would be obtained with other choices for the constants  $x_0, \dots, x_4$ , such as the integers +2, +1, 0, -1, and -2, but the value of  $d$  would no longer be easily interpretable.

### Taking the pre-treatment data into account.

Methods have been developed for comparing different treatments on categorical scales when the samples have been stratified; these methods are analogous to the two-way analysis of variance that is appropriate when the response variable is continuous.<sup>6,7</sup> One of the more important is a generalization by Mantel<sup>8</sup> of the test statistic given in Equation 3.

Suppose the patients have been stratified into  $s$  levels of a prognostic variable (in the present example,  $s=5$ ). Within stratum  $j$  ( $j=0, \dots, s-1$ ), let  $n_{j1}$  and  $n_{j2}$  denote the numbers of patients from treatment groups 1 and 2 falling into that stratum, and let  $n_j = n_{j1} + n_{j2}$  denote the total number of patients in it. Let  $p_{j10}, p_{j11}, \dots, p_{j1,k-1}$  represent the relative frequency distribution of the  $n_{j1}$  patients from treatment 1 across the  $k$  levels of the response variable, let  $p_{j20}, p_{j21}, \dots, p_{j2,k-1}$  represent the same for the  $n_{j2}$  patients from treatment 2, and let  $\bar{p}_{j0}, \bar{p}_{j1}, \dots, \bar{p}_{j,k-1}$  represent the overall frequency distribution of all  $n_j$  patients in stratum  $j$ :  $\bar{p}_{ji} = (n_{j1} p_{j1i} + n_{j2} p_{j2i})/n_j$ ,  $i=0, \dots, k-1$ . Finally, let  $d_{ji} = (p_{j1i} - p_{j2i})$ ,  $i=0, \dots, k-1$ . Table 3 presents the values of these statistics for the patients in pre-treatment category 2.

Define

$$d_j = \sum_{i=0}^{k-1} x_i d_{ji},$$

$$w_j = \frac{n_{j1} n_{j2}}{n_j}$$

and

$$v_j = \sum_{i=0}^{k-1} \bar{p}_{ji} (x_i - \bar{x}_j)^2,$$

where

$$\bar{x}_j = \sum_{i=0}^{k-1} \bar{p}_{ji} x_i.$$

For the data in Table 3 (for which  $j=2$ ), again taking the  $x_i$ 's to be the overall ridits,

$$d_2 = 0.7140 \times 0.0338 + 0.3375 \times 0.0120 + \dots + 0.0232 \times (-0.0011) = 0.0233,$$

$$w_2 = \frac{54 \times 51}{105} = 26.2286,$$

$$\bar{x}_2 = 0.3341,$$

and

$$v_2 = 0.0186.$$

One might, if desired, calculate the chi-square statistic in Equation 3 for each of the  $s$  strata, but more informative is a summary test statistic that assesses the significance of the average of the  $d_j$ 's across all strata. The test statistic proposed by Mantel is

$$\chi^2 = \frac{\sum_{j=0}^{s-1} w_j d_j^2}{\sum_{j=0}^{s-1} \frac{n_j}{n_j - 1} w_j v_j} \quad (\text{Eq. 4})$$

with one degree of freedom.

The values of the relevant quantities for each of the five pre-treatment categories are presented in Table 4. It may be checked that

$$\sum w_j d_j = 5.0146$$

and

$$\sum \frac{n_j}{n_j - 1} w_j v_j = 2.9841$$

so that

$$\chi^2 = \frac{5.0146^2}{2.9841} = 8.43$$

(d.f. = 1,  $p < 0.01$ ). The value of Mantel's chi-square from the stratified analysis is larger than, and suggests greater statistical significance than, the values of chi-square found in the preceding section where the stratification by pre-treatment MGSI category was not taken into account. More powerful comparisons between treatments may usually be expected when the patient samples are stratified on prognostic variables.

### A logistical parametric analysis of the data.

A major deficiency in the above analyses — as well as in the application of  $t$  tests, analyses of variance, or analyses of covariance in which the numerals designating the categories are taken as *bona fide* measurements — is that no easily interpretable measure is provided of just how different the treatments are. In order for an analysis to pro-

duce estimates of parameters describing differences between treatments as well as tests of their statistical significance, a parametric statistical model should be postulated. Especially useful for categorical data are logistical models<sup>9</sup>, which have recently been applied to dental data by Wallenstein *et al.*<sup>10</sup>.

Following McCullagh<sup>11</sup>, a weighted least-squares logistical analysis was applied to the data under analysis. Let  $P_{j1}$  denote the underlying probability that a patient initially in category  $j$ , after receiving treatment 1, will end up worse off than initially, and let  $P_{j2}$  denote the same for treatment 2. The ratio

$$\Omega_{j1} = P_{j1}/(1-P_{j1})$$

is the odds for worsening under treatment 1 for patients in pre-treatment category  $j$ , and

$$\Omega_{j2} = P_{j2}/(1-P_{j2})$$

is the odds for worsening under treatment 2 for patients in pre-treatment category  $j$ . The ratio of these two odds,

$$\omega_j = \frac{\Omega_{j1}}{\Omega_{j2}} = \frac{P_{j1}(1-P_{j2})}{P_{j2}(1-P_{j1})},$$

is the odds ratio associating treatment with worsening in pre-treatment category  $j$  [see (4) and (10) for discussion of the interpretation and uses of the odds ratio].

The analysis of odds ratios is more easily carried out in terms of their logarithms. Define

$$\lambda_j = \ln(\omega_j),$$

and let  $L_j$  denote the estimated log odds ratio for pre-treatment category  $j$ . The Appendix describes how the estimates and their standard errors, displayed in Table 5, are calculated.

The optimally weighted average of the log odds ratios uses as weights the reciprocals of the squared standard errors (denoted  $w_j$  in Table 5),

$$\bar{L} = \sum w_j L_j / \sum w_j = 12.5473/20.2301 = 0.6202.$$

The hypothesis that the log odds ratios for the several pre-treatment categories are equal (*i.e.*, that there is no interaction between treatment and the pre-treatment category) may be tested by referring the value of

$$\chi^2 = \sum w_j (L_j - \bar{L})^2$$

to tables of chi-square with degrees of freedom equal to one less than the number of categories. For the data in Table 5,  $\chi^2 = 0.74$  (d.f. = 4, n.s.). A common log odds ratio, whose estimate is  $\bar{L} = 0.6202$ , apparently distinguishes the two treatments.

The antilogarithm of  $\bar{L}$ , say  $\bar{o}$ , is

$$\bar{o} = e^{\bar{L}} = e^{0.6202} = 1.86;$$

TABLE 4  
QUANTITIES REQUIRED FOR CALCULATION OF  
MANTEL'S SUMMARY CHI-SQUARE STATISTIC

Pre-treatment Category	$d_j$	$w_j$	$v_j$	$n_j$
0	0.1084	12.0000	0.0449	48
1	0.0590	51.4563	0.0365	206
2	0.0233	26.2286	0.0186	105
3	0.0021	9.5897	0.0053	39
4	0.0171	2.7273	0.0006	11

TABLE 5  
ESTIMATED LOGARITHM OF RATIO OF ODDS FOR  
WORSENING UNDER PLACEBO RINSE TO ODDS FOR  
WORSENING UNDER WEEKLY FLUORIDE RINSE

Pre-treatment Category	Log Odds Ratio	Standard Error	$w_j (=1/s.e.^2)$
0	0.8874	0.5645	3.1381
1	0.6604	0.3072	10.5964
2	0.3922	0.4763	4.4080
3	0.3788	0.7328	1.8622
4	1.4663	2.1063	0.2254

on the average, no matter what the initial level, the odds that a patient ends up worse off than initially if he or she uses a placebo rinse are nearly twice the odds for worsening if he or she rinses with fluoride once a week. An odds ratio of this magnitude suggests that the effect of treatment over placebo is only modest.<sup>10</sup>

The estimated standard error of  $\bar{L}$  is

$$\text{s.e.}(\bar{L}) = 1/\sqrt{\sum w_j},$$

equal to  $1/\sqrt{20.2301} = 0.2223$  in the present case. The statistical significance of  $\bar{L}$  may be tested by referring the value of

$$\chi^2 = \left(\frac{\bar{L}}{\text{s.e.}(\bar{L})}\right)^2$$

to tables of chi-square with one degree of freedom. A log odds ratio of zero (equivalently, an odds ratio of unity) signifies no differential odds between the two treatments — i.e., the independence of treatment and response. For the current data,

$$\chi^2 = \left(\frac{0.6202}{0.2223}\right)^2 = 7.78$$

(d.f. = 1,  $p < 0.01$ ). Weekly fluoride mouthrinse is therefore significantly superior to placebo rinse.

A 95% confidence interval for the underlying odds ratio may be constructed by first taking

$$\bar{L} - 1.96 \times \text{s.e.}(\bar{L}) \leq \lambda \leq \bar{L} + 1.96 \times \text{s.e.}(\bar{L})$$

as a 95% confidence interval for the log odds ratio and then taking antilogarithms of the limits. For the data under analysis,  $0.6202 \pm 1.96 \times 0.2223$  yields

$$0.1845 \leq \lambda \leq 1.0559$$

as a 95% confidence interval for the log odds ratio, so that

$$e^{0.1845} \leq \omega \leq e^{1.0559}$$

or

$$1.20 \leq \omega \leq 2.87$$

is a 95% confidence interval for the underlying parameter. The upper limit of nearly 3 represents no better than a moderate differential in odds.

## Discussion.

A price obviously has to be paid for taking full advantage of the ordering inherent in Grainger's Severity Index<sup>1</sup>, in Kingman's modification of it<sup>2</sup>, and in other classifications of dental caries: The analysis becomes ever more complicated as one seeks to extract more information from a set of data. The logistical analysis presented in the preceding section is more informative than the analyses that preceded it. Its inordinate complexity, however, and the unfamiliarity of most dental researchers with the odds ratio, would appear to preclude its widespread adoption, at least until the major packages of statistical programs (BMDP, SAS, and SPSS) are updated to permit implementation on the computer.

Mantel's test procedure<sup>8</sup> is not too difficult to carry out, and generally compares favorably with the more complicated procedure with respect to statistical power (the ability to detect as significant a difference between treatments). The value of Mantel's chi-square statistic is sensitive to the numerical values of the  $x_i$ 's assigned to the categories of the response variable. Any set of  $x_i$ 's that are equally spaced will yield the same value for chi-square,

but other sets may change the value of chi-square appreciably. Absent quantitative knowledge about how much more serious one outcome category is than another, the use of the overall ridits is recommended because of its relative simplicity and because it yields results interpretable as probabilities.

Kingman<sup>2</sup> examined the usefulness of the pre-treatment MGSI as a stratification factor in comparative caries clinical trials in which the increment in DMFS score was the response variable. He found that, in each of the six studies he analyzed, stratifying on the pre-treatment MGSI produced more efficient analyses than stratifying on the pre-treatment DMFS score or applying an analysis of covariance.

The present study illustrates the use of the MGSI as the response variable as well as the stratification variable. Its use in this way was successful in that a significant treatment effect was found. A problem exists, however, with respect to the selection of patients for study.

Patients who, at baseline, are already in the highest (most severe) category should probably be excluded from the study, because they cannot possibly be recorded as worsening. Some of them may, it is true, be assigned to a lower ranking category after treatment, but such an occurrence seems to be more informative about the rate of random misclassification than about the effect of treatment.

All told, 16 patients out of the total of 409 analyzed, fewer than 4%, exhibited a "reversal", i.e., a change from a given category on the MGSI to another indicative of less severe caries. This low rate of reversal compares favorably with the rates typically found for the DMFS score.<sup>12</sup> In no instance was there more than a one-category reversal.

## Acknowledgment.

Dr. Albert Kingman of the National Institute of Dental Research made many helpful comments on an earlier draft of this paper.

## REFERENCES

1. World Health Organization: International Dental Epidemiological Methods Series, Manual No. 3. Geneva: WHO, 1967, pp. 26-29.
2. KINGMAN, A.: A Method of Utilizing the Subject's Initial Caries Experience to Increase Efficiency in Caries Clinical Trials, *Community Dent Oral Epidemiol* 7:87-90, 1979.
3. HEIFETZ, S.B.; MEYERS, R.; and KINGMAN, A.: A Comparison of the Anti-caries Effectiveness of Daily and Weekly Rinsing with Sodium Fluoride Solutions: Findings after Two Years, *Pediatr Dent* 3:17-20, 1981.
4. FLEISS, J.L.: *Statistical Methods for Rates and Proportions*, 2nd ed. New York: Wiley, 1981.
5. FLEISS, J.L.; CHILTON, N.W.; and WALLENSTEIN, S.: Ridit Analysis in Dental Clinical Studies, *J Dent Res* 58:2080-2084, 1979.
6. ARMITAGE, P.: *Statistical Methods in Medical Research*. New York: Wiley, 1971.
7. FERTIG, J.W. and CHILTON, N.W.: A Dental Local Anaesthetic Study: Fixed Model, Two-way Layout Design, *Arch Oral Biol* 13:1477-1489, 1968.
8. MANTEL, N.: Chi-square Tests with One Degree of Freedom: Extensions of the Mantel-Haenszel Procedure, *J Amer Statist Assoc* 58:690-700, 1963.
9. COX, D.R.: *Analysis of Binary Data*. London: Methuen, 1970.
10. WALLENSTEIN, S.; FLEISS, J.L.; and CHILTON, N.W.: Logistic Analysis of Categorical Data from Dental and Oral Experiments, *Pharmacol Thera Dent* 6:65-78, 1981.

TABLE A1  
CALCULATION OF LOG ODDS RATIO FOR PATIENTS IN PRE-TREATMENT CATEGORY 2

Category	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	0	0	0	0	4.6913	4.6347	-0.0566	0	0
1	5	3	8	8	2.1972	2.6288	0.4316	70,616	0.3630
2	48	43	91	83	-2.0098	-1.6327	0.3771	121,030	0.6222
3	53	50	103	12	-3.5742	-3.5165	0.0577	2,884	0.0148
4	54	51	105	2	-	-	-	-	-
Total	-	-	-	105	-	-	-	194,530	1.0000

- McCULLAGH, P.: Regression Models for Ordinal Data, *J R Statist Soc* 42(B):109-142, 1980.
- FLEISS, J.L.; FISCHMAN, S.L.; CHILTON, N.W.; and PARK, M.H.: Reliability of Discrete Measurements in Caries Trials, *Caries Res* 13:23-31, 1979.

### Appendix

The calculations required to carry out the weighted least-squares analysis of the data in Table 1 will be illustrated in detail in Table A1 for the patients in pre-treatment category 2.

Columns (1), (2), and (3) contain the cumulative frequency distributions for the weekly mouthrinse group, the placebo rinse group, and the two groups combined.

Column (4) contains the simple frequency distribution for the two groups combined.

Column (5) contains the estimated log odds for worsening for the patients in the weekly mouthrinse group, calculated as follows: Define  $n_{i1}$  to be the entry for category  $i$  in column (1), and  $n_i$  to be the entry for the final category in column (1). The entry for category  $i$  in column (5) is then

$$\ln \frac{n_i - n_{i1} + 0.5}{n_i + 0.5}$$

For example, the entry for category 3 in column (5) is

$$\ln \frac{54 - 53 + 0.5}{53 + 0.5} = \ln \frac{1.5}{53.5} = -3.5742.$$

The entries in column (6) are calculated similarly, using the frequencies in column (2). For example, the entry for category 3 in column (6) is

$$\ln \frac{51 - 50 + 0.5}{50 + 0.5} = \ln \frac{1.5}{50.5} = -3.5165.$$

Each entry in column (7) is simply the difference between the corresponding entry in column (6) and the one in column (5).

The entries in column (8) are based on those in columns (3) and (4). Define  $C_i$  to be the entry for category  $i$  in column (3),  $F_i$  to be the entry for that category in column (4), and  $n$  to be the sum of all entries in column (4). The entry for category  $i$  in column (8) is given by

$$C_i (n - C_i) (F_i + F_{i+1}).$$

For example, the entry for category 3 in column (8) is

$$103 \times (105 - 103) \times (12 + 2) = 2,884.$$

Each entry in column (9), finally, is the corresponding entry in column (8) divided by the sum of all the entries in column (8).

Define  $d_i$  to be the entry for category  $i$  in column (7) and  $w_i$  to be the entry for that category in column (9). The estimated log odds ratio is simply

$$L = \sum w_i d_i.$$

For the data at hand,

$$L = 0 \times (-0.0566) + 0.3630 \times 0.4316 + 0.6222 \times 0.3771 + 0.0148 \times 0.0577 = 0.3922.$$

The standard error of the estimated log odds ratio may be estimated as follows: Define  $W$  to be the sum of all the entries in column (8),  $n_1$  and  $n_2$  to be the numbers of patients in the two treatment groups, and  $n$  to be  $n_1 + n_2$ ; these definitions of  $n_1$ ,  $n_2$ , and  $n$  correspond to the ones above. The estimated standard error of  $L$  is then

$$\text{s.e.}(L) = \frac{n^2}{\sqrt{n_1 n_2 W}}$$

For the data at hand,

$$\text{s.e.}(L) = \frac{105^2}{\sqrt{54 \times 51 \times 194,530}} = 0.4763.$$

The above values for  $L$  and its standard error appear in the row for category 2 in Table 5.