BIOMETRICS 38, 215–220
March 1982

The Use of Historical Control Information in Testing for a Trend in Proportions

Robert E. Tarone
National Cancer Institute, Bethesda, Maryland 20205, U.S.A.

SUMMARY

A method is developed for utilizing historical control information in testing for a trend in proportions. The resulting test statistic is a modification of the statistic proposed by Cochran (1954, *Biometrics* **10,** 417–451) and by Armitage (1955, *Biometrics* **11,** 375–386). The contribution of the historical data in the analysis depends upon the degree of heterogeneity of the historical rates.

1. Introduction

A problem commonly faced in biometric applications is to test for a trend in proportions in a $2 \times k$ contingency table in which the k columns have a natural ordering. For example, a test compound may be administered at increasing dose levels to groups of test animals to determine if the proportion of animals developing a tumor increases with increasing dose level. Cochran (1954) and Armitage (1955) proposed a chi square statistic for testing for a trend in proportions. For problems with a large amount of historical data, a test for trend incorporating the historical data would be useful. This paper develops such a test, based on a statistic which is closely related to that of Cochran and Armitage. This method for binomial proportions is analogous to the method derived by Pocock (1976) for normally-distributed data.

The problem will be developed in the context of an animal carcinogenesis experiment. Consider an experiment involving $r+1$ groups of test animals. One group serves as a control group, and the remaining r groups are given a test compound at increasing dose levels, $d_1 < d_2 < \cdots < d_r$. The animals are followed for a fixed period of time, and the number of animals which develop a tumor at a particular organ is recorded for each group. Let n_i denote the number of animals in the ith group, and let the number of these animals which develop a tumor during the course of the experiment be denoted by x_i, $i = 0, 1, \ldots, r$. The results of this experiment can be summarized in a $2 \times (r+1)$ table such as Table 1.

To test for an increase in the proportions $\hat{p}_i = x_i/n_i$ with increasing dose level, Cochran and Armitage suggested the test statistic

$$X^2 = \left(\sum_{i=1}^{r} x_i d_i - \hat{p} \sum_{i=1}^{r} n_i d_i\right)^2 \Big/ \left[\hat{p}\hat{q}\left\{\sum_{i=1}^{r} n_i d_i^2 - \left(\sum_{i=1}^{r} n_i d_i\right)^2 \Big/ n_.\right\}\right],$$

where $\hat{p} = x_./n_.$ and $\hat{q} = 1 - \hat{p}$. This statistic is distributed asymptotically as a chi square random variable with 1 df if there are no differences in the probability of developing a tumor among the $r+1$ groups. Cox (1958) showed that this statistic gives the uniformly

Key words: Trend tests for proportions; Beta-binomial distribution; Historical controls; Maximum likelihood estimation.

Table 1
Summary data from an animal carcinogenesis bioassay

Dose level:	0	d_1	$\cdots$	d_r	Total
Animals with tumor	x_0	x_1	$\cdots$	x_r	$x_.$
Animals without tumor	n_0-x_0	n_1-x_1	$\cdots$	n_r-x_r	$n_.-x_.$
Sample size	n_0	n_1	$\cdots$	n_r	$n_.$

most powerful unbiased test against logistic alternatives, and Tarone and Gart (1980) showed that it is asymptotically locally optimal against any alternative which can be expressed as a smooth, increasing function of dose.

2. Use of Historical Data

For many species and strains of animals used in carcinogenesis bioassay experiments there are large amounts of historical control data. Assume that for the animal species and organ represented in Table 1 we have control tumor rates from N previous experiments. Let these historical control rates be denoted by $P_j = Y_j/M_j$, where M_j is the size of the jth control group and Y_j is the number of animals with a tumor at the organ in question in the jth control group, $j = 1, 2, \ldots, N$. The proportion $Y_./M_.$ will be referred to as the pooled historical control rate. For many tumor types the historical control rates are more variable than would be expected if they followed a binomial distribution (Tarone, Chu and Ward, 1981). Thus we shall in the sequel fit a beta-binomial distribution to the historical tumor rates.

For a given experiment, assume that the number of control animals which develop a tumor will follow a binomial distribution with parameter p, but that the spontaneous tumor probability varies from experiment to experiment according to a beta distribution with density

$$g(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \qquad 0<p<1.$$

It will be convenient to utilize a logistic model, and if we let $p = e^a/(1+e^a)$ be the probability of a spontaneous tumor in a control group, then the corresponding probability density function for a will be

$$f(a) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{a\alpha}(1+e^a)^{-(\alpha+\beta)}, \qquad -\infty<a<\infty.$$

If we model the probability of a tumor corresponding to dose d_i by the logistic model,

$$p_i = \exp(a+bd_i)/\{1+\exp(a+bd_i)\}$$

then the conditional likelihood, given a, for the data represented in Table 1 will be

$$L(b \mid a) = \binom{n_0}{x_0} \frac{\exp(ax_0)}{\{1+\exp(a)\}^{n_0}} \prod_{i=1}^{r} \binom{n_i}{x_i} \frac{\exp\{(a+bd_i)x_i\}}{\{1+\exp(a+bd_i)\}^{n_i}}.$$

The full likelihood, therefore, is

$$\begin{aligned} L(b, a) &= L(b \mid a) f(a) \\ &= K(\mathbf{n}, \mathbf{x}, \alpha, \beta) \frac{\exp\{a(x_0+\alpha)\}}{\{1+\exp(a)\}^{(n_0+\alpha+\beta)}} \prod_{i=1}^{r} \frac{\exp\{(a+bd_i)x_i\}}{\{1+\exp(a+bd_i)\}^{n_i}}, \end{aligned} \tag{1}$$

where the function K does not depend on a or b. Note that if we ignore terms not involving a and b, this likelihood differs from the binomial, conditional likelihood only in that $x_0+\alpha$ replaces x_0 and $n_0+\alpha+\beta$ replaces n_0 in the term corresponding to the control group. Thus, if we assume that α and β are known, it follows (cf. Tarone and Gart, 1980) that the score test for $H_0: b=0$, treating a as a nuisance parameter, will be based on the statistic

$$\tilde{X}^2=\left(\sum_{i=1}^{r} x_i d_i-\tilde{p}\sum_{i=1}^{r} n_i d_i\right)^2 \Big/ \left[\tilde{p}\tilde{q}\left\{\sum_{i=1}^{r} n_i d_i^2-\left(\sum_{i=1}^{r} n_i d_i\right)^2 \Big/ \tilde{n}\right\}\right],$$

where $\tilde{n}=n_{\cdot}+\alpha+\beta$, $\tilde{p}=(x_{\cdot}+\alpha)/\tilde{n}$ and $\tilde{q}=1-\tilde{p}$. In the Appendix it is shown that this statistic is asymptotically equivalent to the score statistic obtained for testing $H_0: b=0$ after elimination of a from $L(b, a)$ by integrating to obtain the marginal likelihood for b. This statistic will be approximately distributed as a chi square random variable with 1 df under H_0, provided that the control rates follow a beta-binomial distribution.

Since α and β are unknown, it is recommended that α and β be estimated from the N historical control rates using maximum likelihood estimation [see Kleinman (1973) and references therein], and that $\tilde{X}^2$ be calculated by substituting the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$. Because $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimators of α and β, the resulting test statistic will be asymptotically distributed as a chi square random variable with 1 df under H_0. The validity of the chi square approximation when estimates of α and β are based on a small number of historical control rates has not been investigated, and thus it is recommended that the proposed method be used only in cases for which there is a large amount of historical data.

Note that $\tilde{X}^2$ is simply the test statistic of Cochran and Armitage, with an adjustment to the concurrent control data depending upon the historical tumor information. If the historical tumor rates are extremely variable (relative to binomial variation) then $\hat{\alpha}+\hat{\beta}$ will be small, and the historical data will make relatively little contribution to the test statistic. If, however, the historical tumor rates are homogeneous, then $\hat{\alpha}+\hat{\beta}$ will be large, and the historical data may have a considerable impact on the test statistic. In either case, the influence of the historical data will be greatest if the concurrent control group has a tumor rate which is unusually high or low.

3. Examples

The method developed above will be applied to two examples from the Carcinogenesis Bioassay Program of the National Cancer Institute, one case for which the historical rates are fairly homogeneous and a second case for which the historical rates are extremely heterogeneous. The first example concerns the incidence of lung tumors in female F344 rats in the bioassay of nitrilotriacetic acid (National Cancer Institute, 1977a). The data are summarized in Table 2. Since the dose levels were equally spaced, the doses are presented

Table 2
Summary of lung tumor incidence in female F344 rats in the bioassay of nitrilotriacetic acid

Dose level:	0	1	2	Total
Animals with tumor	0	3	7	10
Sample size	15	49	46	110
Percent animals with tumor	0	6	15	9

Table 3
Control lung tumor rates (Y_i/M_i) *from* 70 *experiments with female F*344 *rats*

Y_i/M_i	Frequency	Y_i/M_i	Frequency	Y_i/M_i	Frequency
0/50	3	0/19	6	1/23	2
0/49	3	0/18	4	1/20	14
0/47	2	0/10	1	1/19	1
0/25	2	1/53	1	1/18	1
0/24	2	1/50	2	2/20	6
0/22	1	1/49	2	2/19	1
0/20	14	1/47	1	2/18	1

as 0, 1 and 2. The Cochran–Armitage test statistic yields a value of 4.04 with an associated P-value of .044. While this result is significant at the 5% level, this significant finding would not be regarded as strong evidence of carcinogenicity because of adjustment for the multiple comparisons involved in the analysis of a carcinogenesis bioassay experiment (Gart, Chu and Tarone, 1979).

The historical control lung tumor rates from 70 experiments with female F344 rats are given in Table 3. The pooled historical control rate based on the examination of lungs from 1805 rats is $40/1805 = .022$. The lung tumor rates for female F344 rats are fairly homogeneous and the maximum likelihood estimates for the beta-binomial parameters are $\hat{\alpha} = 11.52$ and $\hat{\beta} = 501.93$. Thus $\tilde{p} = .035$ and $\tilde{n} = 623.45$, and we find that $\tilde{X}^2 = 21.96$. Incorporation of historical control information provides evidence of a highly significant ($P < 10^{-5}$) tumor increase associated with the administration of nitrilotriacetic acid.

As a second example, consider the incidence of endometrial stromal polyps in female F344 rats in a bioassay of the drug phenformin (National Cancer Institute, 1977b). The data are summarized in Table 4. The Cochran–Armitage test statistic yields a value of 4.28 with an associated P-value of .038. Thus this analysis suggests the possibility of a dose-related decrease in endometrial stromal polyp incidence associated with the administration of phenformin.

The historical control rates for endometrial stromal polyp from 70 experiments with female F344 rats are given in Table 5. The pooled historical control rate based on the necropsy of 1725 rats is $262/1725 = .15$. The historical control rates for endometrial stromal polyp are quite variable, yielding beta-binomial maximum likelihood estimates of $\hat{\alpha} = 2.26$ and $\hat{\beta} = 13.86$. Thus $\tilde{p} = .125$ and $\tilde{n} = 98.12$, and we find that $\tilde{X}^2 = 3.24$ with an associated P-value of .072. The evidence of an association between administration of phenformin and the incidence of endometrial stromal polyp has been weakened by the consideration of historical control information.

Table 4
*Summary of endometrial stromal polyp incidence in female F*344 *rats in the bioassay of phenformin*

Dose level:	0	1	2	Total
Animals with tumor	4	4	2	10
Sample size	14	34	34	82
Percent animals with tumor	29	12	6	12

Table 5
Control rates (Y_i/M_i) for endometrial stromal polyp from 70 experiments with female F344 rats

Y_i/M_i	Frequency	Y_i/M_i	Frequency	Y_i/M_i	Frequency
0/20	7	5/49	1	5/22	1
0/19	4	2/19	1	11/46	1
0/18	2	5/46	1	12/49	1
0/17	1	2/17	1	5/20	2
1/20	4	7/49	1	6/23	1
1/19	2	7/47	1	5/19	1
1/18	2	3/20	2	6/22	1
2/27	1	2/13	1	6/20	3
2/25	1	9/48	1	16/52	1
2/24	1	10/50	1	15/47	1
2/23	1	4/20	7	15/46	1
2/20	6	10/48	1	9/24	1
1/10	1	4/19	3		

4. Discussion

The test statistic suggested above provides a method of incorporating historical control tumor information in testing for a trend in proportions. In two examples we have seen that the test statistic provides a compromise between the analysis which uses only the concurrent control group and an analysis which uses the overall pooled control rate (pooling concurrent and historical data). Extremely heterogeneous historical control data will have little impact on the analysis; however, homogeneous historical data can receive more weight than concurrent control data. It should be noted that if the historical control rates are less variable than expected from a binomial distribution, then $\hat{\alpha}+\hat{\beta}$ can exceed $M_.$, the total number of animals represented in all of the historical control groups (see Bailey, 1975, §14.42). In such a case, it is recommended that $Y_.$ be substituted for $\hat{\alpha}$ and $M_.-Y_.$ be substituted for $\hat{\beta}$ in the computation of $\tilde{X}^2$.

Altham (1969) has developed a fully Bayesian analysis of multinomial data in a $2\times k$ contingency table based on Dirichlet priors. A. P. Dempster, M. R. Selwyn and B. J. Weeks, in a paper presented at the 1980 Joint Statistical Meetings of the American Statistical Association and the Biometric Society, developed a Bayesian analysis for the problem considered above. They assumed, however, that the logits of the historical control tumor rates were normally distributed. While their analysis and the analysis developed in this paper should give similar results for tumor sites with moderate to high spontaneous rates, for tumor sites such as the female F344 rat lung, where most of the historical control rates are zero, the two methods may give divergent results.

An unstated assumption in the above development has been that the historical control rates used in the analysis come from experiments which are similar to the current experiment in factors known to affect the magnitude of tumor rates. Such factors include the length of time on study, housing conditions (e.g. the number of animals per cage), type of food, method of feeding, and possibly the animal supplier and mean birth date of the test animals (see Gart *et al.*, 1979). Preliminary investigations will be necessary to determine which historical rates are appropriate for use in the analysis of a particular experiment.

ACKNOWLEDGEMENT

The author wishes to thank Virginia Grubar for preparing the manuscript.

RÉSUMÉ

On développe une méthode pour utiliser l'information de contrôle historique lorsque l'on teste une tendance dans des proportions. La statistique de test résultante est une modification de celle proposée par Cochran (1954, *Biometrics* **10,** 417–451) et par Armitage (1955, *Biometrics* **11,** 375–386). La contribution des données historiques dans l'analyse dépend du degré d'hétérogénéité des vitesses historiques.

REFERENCES

Altham, P. M. E. (1969). Exact Bayesian analysis of a 2×2 contingency table, and Fisher's 'exact' significance test. *Journal of the Royal Statistical Society, Series B* **31,** 261–269.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11,** 375–386.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications.* New York: Hafner.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10,** 417–451.

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society, Series B* **20,** 215–242.

Gart, J. J., Chu, K. C. and Tarone, R. E. (1979). Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. *Journal of the National Cancer Institute* **62,** 957–974.

Kleinman, J. C. (1973). Proportions with extraneous variance: Single and independent samples. *Journal of the American Statistical Association* **68,** 46–54.

National Cancer Institute (1977a). Bioassays of nitrilotriacetic acid (NTA) and niltrilotriacetic acid trisodium salt, monohydrate ($Na_3NTA \cdot H_2O$) for possible carcinogenicity. Technical Report No. 6. Bethesda, Maryland: National Cancer Institute.

National Cancer Institute (1977b). Bioassay of phenformin for possible carcinogenicity. Technical Report No. 7. Bethesda, Maryland: National Cancer Institute.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases* **29,** 175–188.

Tarone, R. E., Chu, K. C. and Ward, J. M. (1981). Variability in the rates of some common naturally occurring tumors in F344 rats and $B6C3F_1$ mice. *Journal of the National Cancer Institute* **66,** 1175–1181.

Tarone, R. E. and Gart, J. J. (1980). On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association* **75,** 110–116.

Received November 1980; *revised January* 1981

APPENDIX

The marginal likelihood for b is given by $L(b) = \int_{-\infty}^{+\infty} L(b, a)\, da$, where $L(b, a)$ is given in (1). Although this integral appears intractable at first glance, the score test statistic for $H_0: b = 0$ can be derived without a closed-form expression for $L(b)$. Since the likelihood $L(b, a)$ is in the exponential family, all necessary derivatives of $L(b)$ can be obtained by differentiation under the integral sign. The algebra is tedious; however, it follows that

$$S = \left.\frac{d \ln L}{db}\right|_{b=0} = \frac{L'(0)}{L(0)} = \sum_{i=1}^{r} x_i d_i - \tilde{p} \sum_{i=1}^{r} n_i d_i.$$

Similarly,

$$V = -\left.\frac{d^2 \ln L}{db^2}\right|_{b=0} = \left\{\frac{L'(0)}{L(0)}\right\}^2 - \frac{L''(0)}{L(0)}$$

$$= \frac{\tilde{n}}{\tilde{n}+1} \tilde{p}\tilde{q} \left\{ \sum_{i=1}^{r} n_i d_i^2 - \left(\sum_{i=1}^{r} n_i d_i \right)^2 \Big/ \tilde{n} \right\}.$$

Thus, the score test statistic for $H_0: b = 0$ is given by $S^2/V = (1 + \tilde{n}^{-1})\tilde{X}^2$.