STATISTICS IN MEDICINE, VOL. 5, 15–20 (1986)

A NEW MODEL FOR ORDINAL PAIN DATA FROM A PHARMACEUTICAL STUDY

CHRISTY CHUANG

Department of Statistics, Division of Biostatistics, The University of Rochester, Rochester, NY 14627, U.S.A.

AND

ALAN AGRESTI

Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.

SUMMARY

For the pain data analysed previously by Cox and Chuang,[1] this paper proposes a new model that assumes monotone scores for ordered response categories. This proposed model possesses several attractive features and allows a stochastic ordering of the drugs under comparison. Such a model also provides insight regarding the ordinal scale used to classify response. Estimation of the parameters in the model is obtained by use of BMDP3R.

KEY WORDS Base comparison model Equally-spaced scores Ordered scores Ordinal pain data RC model

INTRODUCTION

Cox and Chuang[1] compared a chi-square partitioning analysis and two logit analyses of ordinal pain data from a pharmaceutical study. These analyses were designed, more generally, to compare nominal treatments with respect to an ordinal response variable. The chi-square partitioning technique uses continuation ratios and allows one to examine several subtables independently. Cox and Chuang[1] referred to the first class of logit models as base comparison models, because the logits compared each response to a baseline response. The most general base comparison model considered is equivalent to the multiplicative row and column effects (RC) association model considered by Goodman.[2,3] The second logit analysis applied cumulative odds models proposed by McCullagh.[4]

All three approaches were applied to model responses on pain from a single-dose, postoperative analgesic trial[5] that employed four drugs (Z100, EC4, C60, C15) and four ratings (poor, fair, good, very good to excellent). These data appear in Table I. Cox and Chuang[1] presented the results and the advantages and disadvantages of each approach. Both types of logit models are highly parametrized and provide considerable flexibility for describing data. Their parameters, however, are rather difficult to interpret in the general formulation of the models given by Cox and Chuang.

In this paper, we propose a slightly different model to describe the data. This new model is a special case of the base comparison model given by Cox and Chuang, one that generates monotone scores for the response categories and implies a stochastic ordering of the rating distributions of the drugs. This stochastic ordering leads to simpler interpretations of the results for this model than for those obtained with the models considered by Cox and Chuang. The model is a generalization of the log-linear 'row effects' model, which assumes equally-spaced scores for the response

0277–6715/86/010015–06$01.00

Received March 1985
Revised April 1985

Table I. Ratings of four drugs in a single-dose, postoperative analgesic trial[5]

Drug	Rating			
	Poor	Fair	Good	Very good to excellent
Z100	5	1	10	14
EC4	5	3	3	20
C60	10	6	12	3
C15	7	12	8	2

categories. Yet, it is more structured than the RC model, which allows arbitrary (possibly non-monotone) scores. Maximum likelihood estimates of the parameters in this new model are obtained by use of BMDP3R,[6] also used by Cox and Chuang. We feel that the model proposed in this paper, when it fits reasonably well, gives information directed toward the principal goal of analgesic studies, namely to find the most effective drug when data suggest differences in their analgesic effects.

MODEL SPECIFICATION

Let $\{p_{ij} \mid \sum_j p_{ij} = 1, i = 1, \ldots, 4\}$ be the multinomial probabilities of the respone categories under the four drugs. The homogeneity model, which assumes $p_{ij} = p_j$ $(i = 1, \ldots, 4)$, fits the data in Table I poorly (the likelihood ratio statistic $G^2 = 46{\cdot}74$, with d.f. $= 9$, so $p < 0{\cdot}001$). The low p-value suggests that patients did respond differently to the four drugs.

Since the response categories (columns) are ordered but the drugs (rows) are not, one approach is to assign score j to the jth response category and unknown parameter μ_i to the ith drug and consider interaction of the form $j\mu_i$. This interaction structure gives the following model:

$$\log p_{ij} = \mu + \alpha_i + \beta_j + j\mu_i, \quad i, j = 1, \ldots, 4 \tag{1}$$
$$\sum \alpha_i = \sum \beta_j = \sum \mu_i = 0.$$

Model (1), often referred to as a row effects (or R) model, has been discussed by various authors, including Fienberg,[7] Plackett,[8] Haberman[9] and Goodman.[2] Model (1) treats the ordinal response as an equally-spaced numerical scale. Furthermore, it assumes that the explanation of the interaction between drug and response results from the product of a set of drug-specific parameters with these equally-spaced scores. This model, when fitted to Table I, yields $G^2 = 24{\cdot}46$ with d.f. $= 6$ $(p < 0{\cdot}001)$. Though substantially better than the homogeneity model, the fit is still rather poor.

A natural generalization of (1) is to use a set of monotone non-decreasing (not necessarily equally-spaced) scores $\{v_j\}$ for the response categories. In the absence of any obvious choice for $\{v_j\}$, we suggest that one treat them as unknown parameters and estimate them using maximum likelihood. This consideration gives rise to the following model:

$$\log p_{ij} = \mu + \alpha_i + \beta_j + \mu_i v_j, \quad i, j = 1, \ldots, 4 \tag{2}$$
$$\sum \alpha_i = \sum \beta_j = \sum \mu_i = 0$$
$$1 = v_1 \leqslant v_2 \leqslant v_3 \leqslant v_4 = 4.$$

The scaling given here for the $\{v_j\}$ is arbitrary, and we suggest it simply to make the score parameters more comparable to the fixed scores in the R model (1). This model reduces to (1) in the special case when all $v_{j+1} - v_j$ are constant. Goodman's RC model has the same structural form as (2), but in it neither set of scores has an ordering constraint. The addition of this constraint for the ordinal response variable produces a stochastic ordering among the response distributions of the drugs. To see this, note that for a pair of drugs a and b, the log odds ratio for adjacent responses is

$$\log (p_{aj}p_{b,\,j+1}/p_{a,\,j+1}\,p_{bj}) = (\mu_b - \mu_a)\,(v_{j+1} - v_j). \tag{3}$$

If $\mu_b > \mu_a$, then it follows that these log odds ratios are non-negative for $j = 1, 2, 3$, and hence

$$\sum_{j=1}^{t} p_{bj} \leqslant \sum_{j=1}^{t} p_{aj} \quad \text{for} \quad t = 1, 2, 3 \tag{4}$$

where strict inequality holds for at least one t. That is, the response distribution for the bth drug is stochastically greater than that for the ath drug. Hence, when this model fits, we can order the drugs according to the parameters $\{\mu_i\}$. The higher μ_i is, the more effective the associated drug is as an analgesic. More specifically, the odds of making response $j+1$ instead of response j are $\exp[(\mu_b - \mu_a)(v_{j+1} - v_j)]$ times higher for drug b than for drug a.

Another implication of (4) is that if $\mu_b > \mu_a$, then $\sum_j v_j P_{bj} > \sum_j v_j p_{aj}$; that is, the mean score for drug b is greater than for drug a. The mean score under a set of assigned (monotone) response scores is the criterion for comparing drugs in many studies, with use of analysis of variance techniques. This practice, however, has a potential pitfall, since the ordering of the mean scores may depend on the response scores chosen. The ordering of means is necessarily free from the choice of (monotone) scores only when the sample response distributions are stochastically ordered. Since model (2) implies a stochastic ordering, when it fits we avoid the subjective choice of scores and also the possibility of improper ordering due to inappropriate score assignment.

In related research, Agresti, Chuang and Kezouh[10] propose an adapted RC model for two-way tables where *both* variables are ordinal and *both* $\{\mu_i\}$ and $\{v_j\}$ are required to be monotone. Anderson[11] proposed another related class of models for ordinal response variables.

RESULTS

We fit the monotone-scores model (2) to the data in Table I using BMDP3R. An explanation of the use of BMDP3R to carry out maximum likelihood appears both in the BMDP manual and in Reference 1. In this application, we reparametrized the ordered scores $\{v_j\}$ as $v_j = \sum_{k=1}^{j} \Delta_k$ and converted the ordering constraint on $\{v_j\}$ to $\Delta_1 = 0$ and $\Delta_k \geqslant 0, k = 2, 3, 4$. Since BMDP3R allows minimum values to be specified for the parameters, we can easily handle the constraints on $\{\Delta_k\}$. After preliminary fitting, we can rescale the $\{\hat{\mu}_i\}$ and $\{\hat{v}_j\}$ so that the $\{\hat{v}_j\}$ fall in some predetermined range, such as $1 = \hat{v}_1 \leqslant \ldots \leqslant \hat{v}_4 = 4$. A sample program for applying the model to these data is available from the authors.

The general base comparison logit model considered by Cox and Chuang is exactly equivalent to Goodman's RC model, that is model (2) without the ordering constraint. Cox and Chuang noted that this model fits Table I fairly well, with $G^2 = 6{\cdot}69$ based on d.f. $= 4$. Hence most of the lack of fit of the R model is caused by restriction of the response scores to equal intervals. The monotone-scores model yields a G^2 of 9·91. This model also gives a much better fit than that obtained with use of the equal-interval scoring of the R model. Counterbalancing the slight deterioration in fit

compared to the more general RC model is the insight and simpler interpretation obtained with this model. The estimates for the association parameters are $\hat{\mu}_1 = 0{\cdot}37$, $\hat{\mu}_2 = 0{\cdot}59$, $\hat{\mu}_3 = -0{\cdot}38$, $\hat{\mu}_4 = -0{\cdot}58$; $\hat{v}_1 = 1, \hat{v}_2 = 1, \hat{v}_3 = 1{\cdot}46$ and $\hat{v}_4 = 4$. Based on the $\{\hat{\mu}_i\}$, a sample ordering among the four drugs is

$$\text{EC4} \geqslant \text{Z100} \geqslant \text{C60} \geqslant \text{C15}. \tag{5}$$

In other words, drug EC4 tended to induce the best responses in this sample.

The estimated response scores $\{\hat{v}_j\}$ have identical values for the poor and fair categories, and these two scores are quite close to the one for the good category. Analysis of the log odds ratios (3) for these response scores indicates that differences among the drugs occur primarily in regard to whether one judges them in the top category (very good to excellent) rather than in another category. In other words, given that a response occurs in one of the first three categories, there is little difference in the drugs in the distribution among these responses. In this study, it seems that the three least favourable categories are not very helpful in telling the drugs apart, and it is the most favourable category that plays the major role of differentiating between the drugs. We consider this observation useful not only in helping us to oder the drugs, but also in understanding the mechanism leading to the result.

INFERENCES

So far we have used the monotone-scores model in a purely descriptive way. Among the inferential questions that arise with this model are the following: does this model or the model that allows arbitrary response scores (i.e. the RC model) fit significantly better than the model that assumes fixed equal-interval scores (the R model)? If the monotone-scores model fits, how can we make pairwise comparisons of the $\{\hat{\mu}_i\}$ to judge whether significant differences exist among the drugs with respect to their response distributions?

Given that the RC structural form holds, we can use the difference between the G^2 values for the R model and for the RC model to test the null hypothesis that the scores are equal-interval. For our example, this difference equals $24{\cdot}46 - 6{\cdot}69 = 17{\cdot}77$, based on d.f. $= 6 - 4 = 2$, and indicates a substantial improvement in fit when we allow for arbitrary scores. Comparisons with the monotone-scores model are a bit more complex, because the asymptotic distribution of G^2 for that model depends on whether the true response score parameters lie on the boundary of the parameter space. If the model holds with true response scores that are strictly monotone, then G^2 has the same asymptotic distribution as it does for the RC model. On the other hand, suppose that the model holds with, say, strictly monotone scores except for an adjacent pair that are identical. Then it follows from results for similar models discussed by Agresti, Chuang and Kezouh[10] that G^2 has an asymptotic distribution that is an equal mixture of the one for the RC model and one having an additional degree of freedom. In either case, for our example, the model having monotone response scores fits substantially better than the one having equal-interval scores, since the reduction in G^2 is $24{\cdot}46 - 9{\cdot}91 = 14{\cdot}55$.

Our next analysis involves checking the extent to which we can order the drugs with respect to the pain rating. The estimates $\{\hat{\mu}_i\}$ fall roughly into two groups, namely (Z100, EC4) and (C60, C15). When we refit the monotone-scores model under the constraint $\mu_1 = \mu_2$, we get $G^2 = 11{\cdot}28$, and we get $G^2 = 10{\cdot}30$ under the constraint $\mu_3 = \mu_4$. Under the joint constraints $\mu_1 = \mu_2$ and $\mu_3 = \mu_4$, we obtain $G^2 = 11{\cdot}76$. These G^2 values are useful for summarizing goodness of fit under various conditions. Since the parameters may lie on the boundary of the parameter space, usual chi-square large-sample results are invalid, and we cannot use these G^2 values formally in hypothesis tests. None the less, the increase in G^2 over the value of 9·91 obtained with only the ordering

constraint on the $\{v_j\}$ is not severe, suggesting that there is little evidence that Z100 and EC4 have different effects or that C60 and C15 have different effects. However, these two groups differ substantially from one another, since an equating of their effects leads to the homogeneity model that has $G^2 = 46{\cdot}74$. This is the same conclusion obtained with use of the more complex models discussed by Cox and Chuang.

DISCUSSION

We formulated the monotone-scores model in (2) for two-way tables having an ordered response classification. The same methodology applies to multi-way tables having an ordinal response variable. For instance, we could compare the four drugs with regard to the pain rating while controlling for other factors, using adaptations of standard log-linear models in which the drug–pain association term has the structure $\mu_i v_j$ for monotone $\{v_j\}$. The other explanatory variables could be nominal or ordinal and one should take the measurement scales into account (as in Chapter 5 of Reference 12) in constructing the other association or interaction terms.

The monotone-scores model is not as parsimonious as the row effects model, since it has parameters instead of fixed scores. However, the number of distinct parameters for this model can be no greater than the number for the RC model, which is the most general base comparison logit model discussed by Cox and Chuang.[1] When the true scores are strictly monotone, the estimates for these two models are asymptotically equivalent. In that case, we can use the information matrix to gauge the precision of parameter estimates, for a fixed sample size. See Cox and Chuang[1] for comments regarding sample size requirements for fitting models of this type.

The reader should note that the response score estimates $\{\hat{v}_j\}$ for model (2) depend on the particular drugs being compared. For instance, different score estimates occur if one or more drugs are deleted from Table I or if others are added. Although this may seem undesirable, it is a feature of all methods in which response scores are data dependent. For instance, if the Kruskal–Wallis statistic (corrected for ties) were used to analyse Table I, the rank scores for the response categories and the sample mean ranks for the drugs would also depend on the particular drugs included in the table.

Of course, in some applications the monotone-scores model will be too simple to give a good fit to the data. In some cases one might find that monotone scores are inappropriate, perhaps because the drugs are not stochastically ordered on the response. For instance, one may obtain a rather poor fit of the monotone-scores model for the cells in two columns that have identical score estimates, but one may obtain a good fit for the RC model because of 'out-of-order' scores for those columns. To test for significance of departures from H_0: $v_a = v_{a+1}$ in the RC model, Goodman[3] suggests the statistic $G^2(\mathrm{I}) + G^2(\mathrm{RC}_{\mathrm{coll}})$-$G^2(\mathrm{RC})$, based on d.f. = 1, where $G^2(\mathrm{I})$ is the G^2 statistic for the independence model fitted to the table consisting of columns a and $a+1$, $G^2(\mathrm{RC})$ is the G^2 statistic for the RC model applied to the complete table, and $G^2(\mathrm{RC}_{\mathrm{coll}})$ is the G^2 statistic for the RC model applied to the table in which columns a and $a+1$ are collapsed into a single column. When the monotone-scores model gives $\hat{v}_1 < \ldots < \hat{v}_a = \hat{v}_{a+1} < \ldots$, it can be shown that Goodman's statistic is identical to the difference between the G^2 statistic for that model and $G^2(\mathrm{RC})$. For instance, to test H_0: $v_1 = v_2$ for the RC model in Table I, the test statistic is $9{\cdot}91 - 6{\cdot}69 = 3{\cdot}21$, based on d.f. = 1. Considering that the data suggested this test, there is no strong evidence of non-monotonicity. When sampling error offers a potential explanation of the non-monotonicity in the RC model, the monotone-scores model has considerable appeal because of the simpler interpretations possible with the stochastic ordering induced by this more severe smoothing of the data.

ACKNOWLEDGEMENTS

This research was partially supported by Grant CA1 1198 of the National Cancer Institute for Dr. Chuang and Grant R01 GM3 3210-01 from the National Institutes of Health for Dr. Agresti.

REFERENCES

1. Cox, C. and Chuang, C. 'A comparison of chi-square partitioning and two logit analyses of ordinal pain data from a pharmaceutical study', *Statistics in Medicine*, **3**, 273–285 (1984).
2. Goodman, L. A. 'Simple models for the analysis of association in cross-classifications having ordered categories', *Journal of the American Statistical Association*, **74**, 537–552 (1979).
3. Goodman, L. A. 'Association models and canonical correlation in the analysis of cross-classifications having ordered categories', *Journal of the American Statistical Association*, **76**, 320–334 (1981).
4. McCullagh, P. 'Regression models for ordinal data (with discussion)', *Journal of the Royal Statistical Society, Series B*, **42**, 109–142 (1980).
5. Calimlim, J. F., Wardell, W. M., Cox, C., Lasagna, L. and Sriwatanakul, K. 'Analgesic efficacy of orally Zomipirac sodium', *Clinical Pharmacology and Therapeutics*, **31**, (2), 208 (1982).
6. Dixon, W. J., Brown, M. B., Engelman, L., Frane, J. W., Hill, M. A., Jennrich. R. I. and Toporek, J. D. *BMDP Statistical Software*, Univeristy of California Press, Los Angeles, CA, 1981.
7. Fienberg, S. E. *The Analysis of Cross-Classified Categorical Data*, 2nd edition, The MIT Press, Cambridge, Mass., 1980.
8. Plackett, R. L. *The Analysis of Categorical Data*, Griffin, London, 1974.
9. Haberman, S. J. *Analysis of Qualitative Data*, Vols I and II, Academic Press, New York, NY, 1979.
10. Agresti, A., Chuang, C., and Kezouh, A. 'Order-restricted score parameters in association models for contingency tables', submitted for publication (1985).
11. Anderson, J. A. 'Regression and ordered categorical variables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **46**, 1–30 (1984).
12. Agresti, A. *Analysis of Ordinal Categorical Data*, Wiley, New York, NY, 1984.