This article presents a unifying approach to the analysis of repeated univariate categorical (ordered) responses based on the application of the generalized log-linear modeling framework proposed by Lang and Agresti. It is shown that three important research questions in longitudinal studies can be addressed simultaneously. These questions are the following: What is the overall dependence structure of the repeated responses? What is the structure of the change between consecutive time points? and What is the structure of the change in the marginal distributions? Each of these questions involves specifying log-linear models for different marginal distributions of the multiway cross classification of the responses. The proposed approach is illustrated by means of two real data examples.

Modeling Joint and Marginal Distributions in the Analysis of Categorical Panel Data

JEROEN K. VERMUNT
Tilburg University
MARÍA FLORENCIA RODRIGO
University of Valencia
MANUEL ATO-GARCIA
University of Murcia

1. INTRODUCTION

Consider the four-way cross tabulation presented in Table 1. It contains data on marijuana use taken from four annual waves (1977-1980) of the National Youth Survey (Elliot, Huizinga, and Menard 1989; Lang, McDonald, and Smith 1999). The table reports the information on marijuana use of 237 respondents who were age 14 in 1977. The variable of interest is a trichotomous ordinal variable: marijuana use in the past year measured at four occasions. This is the kind of data that plays a central role in this article.

Longitudinal data obtained via panel studies contain rich information on processes of social and psychological change. The analysis of this kind of data is, however, not straightforward. The most important problem is that we are dealing with dependent observations.

SOCIOLOGICAL METHODS & RESEARCH, Vol. 30 No. 2, November 2001 170-196

TABLE 1: Data on Marijuana Use in the Past Year Taken From Four Yearly Waves of the National Youth Survey (1977-1980)

		1979 (C)								
		1			*2*			*3*		
1977	*1978*	*1980 (D)*			*1980 (D)*			*1980 (D)*		
(A)	*(B)*	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
1	1	115	18	7	6	6	1	2	1	5
1	2	2	2	1	5	10	2	0	0	6
1	3	0	1	0	0	1	0	0	0	4
2	1	1	3	0	1	0	0	0	0	0
2	2	2	1	1	2	1	0	0	0	3
2	3	0	1	0	0	1	1	0	2	7
3	1	0	0	0	0	0	0	0	0	1
3	2	1	0	0	0	1	0	0	0	1
3	2	0	0	0	0	2	1	1	1	6

NOTE: The three levels of the ordinal response variable marijuana use are *never* (1), *no more than once a month* (2), and *more than once a month* (3).

Consequently, when modeling such repeated responses, one has to take into account that one does not have four times 237 independent observations but 237 multivariate observations. There are three main approaches to the analysis of longitudinal data: conditional or transition models, random-effects models, and marginal models (Diggle, Liang, and Zeger 1994; Fahrmeir and Tutz 1994). Transition models like Markov-type models concentrate on changes between consecutive time points. Random-effect and marginal models can be used to investigate changes in univariate distributions. These three approaches not only differ in the questions they address but also in the way they deal with the dependencies between observations. Because of their structure, transition models take the bivariate dependencies between observations at consecutive occasions into account. Random-effects models capture the dependence by introducing a latent variable. In marginal models, the dependency is dealt with in a more ad hoc way in the estimation procedure.

This article presents a unifying approach to the analysis of univariate repeated (ordered) categorical responses that combines elements of the approaches discussed above. Restrictions on cell counts are formulated in the form of log-linear models. Our approach simultaneously addresses three important questions about a panel data set:

1. What is the overall dependence structure of the repeated responses?
2. What is the structure of the change between consecutive time points?
3. What is the structure of the change in the marginal distributions?

To answer the first question, we have to analyze the joint distribution of the responses. We are interested in whether, for example, a first-order Markov model, a Rasch-type model, or a model containing only two-variable interactions describes the associations in the four-way table. It should be noted that these three structures correspond to the three approaches to longitudinal data analysis mentioned above.

The second question is about the gross change between t and $t + 1$, which involves modeling the bivariate marginal distributions formed by responses at consecutive time points. For these margins, we might specify well-known models for the association in squared tables, such as (quasi-)independence, (quasi-)symmetry, and linear-by-linear association models, as well as models that restrict the transition probabilities. In addition, questions concerning the stability of these bivariate marginal associations and transitions over time can be addressed.

The third question concerns the univariate marginal distributions. An important test is, of course, the hypothesis of marginal homogeneity or no net change. In most longitudinal studies, we want to check whether at the aggregate level the situation changed. If the marginal homogeneity hypotheses do not hold, we may want to test certain hypotheses about the observed marginal change.

Using log-linear analysis techniques, it is no problem to address the question related to the joint distribution of the four responses. Methods based on the use of standard log-linear models have also been proposed to address questions concerning bivariate and univariate marginal distributions. These methods make, however, certain assumptions about the joint distribution. The two most important examples are the indirect test for marginal homogeneity assuming a quasi-symmetry model for the joint distribution (Bishop, Fienberg, and Holland 1975; Conaway 1989; Meiser, Von Eye, and Spiel 1997) and the modeling of bivariate margins assuming a Markov structure for the joint distribution (Andersen 1990; Lindsey 1993). If the quasi-symmetry model holds, a conditional test between this model and the symmetry model yields a test for marginal homogeneity (Caussinus 1965). If the first-order Markov assumption holds, the adjacent two-way tables can be analyzed as if they were tables from independent samples. A problem with these approaches is, however, that the validity of the test for the marginal model depends on the validity of assumptions about the joint distribution. Consequently, if the model for the joint distribution does not hold, we no longer have a valid test for the models we are interested in.

We propose an alternative, more direct, log-linear modeling approach to the analysis of multiwave panel data that overcomes the above problems. As long as no contradictory constraints are specified, any kind of model for the joint distribution can be combined with any kind of model for the bivariate and univariate distributions. For example, marginal homogeneity can be tested in conjunction with a first-order Markov model for the joint distribution, and symmetric association in consecutive bivariate margins can be assumed in combination with a second-order Markov model for the joint distribution. Our work is related to the work of Agresti (1997) and Croon, Bergsma, and Hagenaars (2000). They concentrated, however, on the change in the marginal association between two response variables in two-wave panel studies. Here, we focus on models for a single response variable that is observed at three or more occasions.

Rather than using standard log-linear models, we use the generalized log-linear modeling approach proposed by Lang and Agresti (1994), which permits simultaneous modeling of marginal and joint distributions. Besides its flexibility, other potential benefits of this simultaneous modeling approach relative to a separate fitting approach come in terms of model parsimony and more efficient estimators of cell expected frequencies and model parameters. One also obtains a single test that simultaneously summarizes goodness of fit and a single set of fitted values and residuals (Becker, Minick, and Yang 1998; Lang and Agresti 1994). Another advantage of using a simultaneous modeling approach is that it makes it possible to detect that the postulated hypotheses for the various distributions are incompatible with one another. For example, in one of the reported analyses, we found that a model of homogeneous bivariate transition probabilities is incompatible with a constant univariate marginal shift model.

Estimation of the models presented in this article cannot be done with standard software for log-linear analysis. For this article, we used an experimental version of the LEM program (Vermunt 1997) that implements the estimation procedure proposed by Bergsma (1997), which is a slightly modified version of the Fisher-scoring method described in Lang and Agresti (1994).

The remainder of this article is organized as follows. Section 2 links the three questions about panel data to the generalized log-linear modeling approach. Sections 3, 4, and 5 describe the most interesting log-linear models for joint, bivariate, and univariate distributions in panel studies. Section 6 discusses the issue of simultaneously modeling joint and marginal distributions. Section 7 gives some details on maximum likelihood estimation. In section 8, we illustrate our approach by means of a second empirical example. The article ends with a short discussion.

2. GENERALIZED LOG-LINEAR MODELS FOR PANEL DATA

Let $m_{ijk\ell}^{ABCD}$ denote an expected cell entry in the four-way table obtained by cross tabulating the measurements of the same variable at four time points. Here, A, B, C, and D serve as variable labels and i, j, k, and ℓ as their indices.

We are interested in modeling the joint distribution of A, B, C, and D, as well as the three two-way margins of adjacent time points and the four one-way margins. The latter two types of margins are obtained by collapsing the cell entries $m_{ijk\ell}^{ABCD}$ over the appropriate indices. Denoting a summation over a certain index by a "+," the marginal cell entries of interest are m_{ij++}^{ABCD}, m_{+jk+}^{ABCD}, $m_{++k\ell}^{ABCD}$, m_{i+++}^{ABCD}, m_{+j++}^{ABCD}, m_{++k+}^{ABCD}, and $m_{+++\ell}^{ABCD}$.

Lang and Agresti (1994) proposed a generalization of the standard log-linear models that allows specifying log-linear models for sums of cell entries (see also Becker 1994; and Bergsma 1997). The model they proposed is of the form

$$\ln \mathbf{A}\,\mathbf{m} = \mathbf{X}\,\mathbf{b}. \tag{1}$$

Here, **m** is the vector of expected cell entries, and **A** is a matrix with ones and zeroes that is used to define the appropriate marginal cell entries. The other two terms, **X** and **b**, have their standard meaning, that is, the model or design matrix and the vector of unknown parameters. The same model can also be described without using matrix notation:

$$\ln \sum_k a_{ik} m_k = \sum_j b_j x_{ij} .$$

As can be seen, this yields a log-linear model for sums of cell entries or, equivalently, for marginal cell entries. This is the basic structure that we use to specify log-linear models for the various types of marginal tables.

It should be noted that the original model proposed by Lang and Agresti (1994) is **C** ln **A m** = **X b**. The matrix **C** can be used to define certain contrasts like log odds or log odds ratios. Because we restrict ourselves to log-linear models for marginal frequencies, we can use the slightly simpler formulation without the **C** matrix.

Below, we describe the most important log-linear models for the cell entries of the joint, the two-way marginal, and the one-way marginal distributions. In each case, we begin with the saturated model and then proceed considering more parsimonious models that are of practical interest in the context of longitudinal data analysis. We also present the results obtained when applying these models to the data reported in Table 1.

3. LOG-LINEAR MODELS FOR THE JOINT DISTRIBUTION

The most general log-linear model for the cell entries in the joint distribution $m_{ijk\ell}^{ABCD}$ is the saturated log-linear model. This model is given by

$$\begin{aligned} \ln m_{ijk\ell}^{ABCD} = {} & \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_\ell^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{i\ell}^{AD} + \lambda_{jk}^{BC} \\ & + \lambda_{j\ell}^{BD} + \lambda_{k\ell}^{CD} + \lambda_{ijk}^{ABC} + \lambda_{ij\ell}^{ABD} + \lambda_{ik\ell}^{ACD} + \lambda_{jk\ell}^{BCD} + \lambda_{ijk\ell}^{ABCD} . \end{aligned} \quad (2)$$

More restricted models can be obtained by setting certain parameters equal to zero and/or imposing certain equality constraints.[1] Here, we will concentrate on restrictions that make sense in the context of the analysis of panel data.

Note that a model for the joint distribution can be obtained from the generalized log-linear model described in equation (1) by setting the matrix **A** equal to the identity matrix. The **X** matrix is a standard design matrix. It should also be noted that for the estimation of models for the joint distribution, we do not need the marginal log-linear modeling framework. However, as will be shown in section 6, this approach makes it possible to combine a model for the joint distribution with models for the bivariate and/or univariate distribution.

A more or less exploratory method to obtain a simpler structure involves including all terms up to a certain order. The most restrictive model of interest is the independence model, which is obtained by omitting all two-, three-, and four-variable terms from the saturated model. Another relatively simple model is obtained by excluding all three- and four-factor terms from equation. This model assumes that there is an association between each pair of points in time:

$$\begin{aligned}\ln m_{ijkl}^{ABCD} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_\ell^D + \lambda_{ij}^{AB} \\ &\quad + \lambda_{ik}^{AC} + \lambda_{i\ell}^{AD} + \lambda_{jk}^{BC} + \lambda_{j\ell}^{BD} + \lambda_{k\ell}^{CD}. \end{aligned} \tag{3}$$

Models that have proved useful for longitudinal data are Markov models. Their underlying assumption is that the (conditional) dependence between responses becomes weaker when time points are farther apart. The most restrictive model of this type is the first-order Markov model, which postulates that there is only an association between adjacent time points, that is,

$$\ln m_{ijk\ell}^{ABCD} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_\ell^D + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{k\ell}^{CD}.$$

As can be seen, this model assumes conditional independence between *A* and *C*, between *A* and *D*, and between *B* and *D*.

Less restrictive is the second-order Markov model, which is obtained by excluding terms involving variables that are more than two time points apart—in this case, *A* and *D*—from the saturated model. This model is defined as

$$\begin{aligned}\ln m_{ijkl}^{ABCD} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_\ell^D + \lambda_{ij}^{AB} \\ &\quad + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{j\ell}^{BD} + \lambda_{k\ell}^{CD} + \lambda_{ijk}^{ABC} + \lambda_{jk\ell}^{BCD}. \end{aligned}$$

Here, the responses at waves 1 and 4 are assumed to be independent given the states at waves 2 and 3. A restricted variant can be obtained by excluding the three-variable terms λ_{ijk}^{ABC} and $\lambda_{jk\ell}^{BCD}$ from the model.

Other kinds of log-linear models that are often used to model dependencies among repeated observations are symmetry and quasi-symmetry models. These are multivariate generalizations of the well-known symmetry and quasi-symmetry models for square tables (Bishop et al. 1975). The multivariate quasi-symmetry model is also known as the Rasch model (Kelderman 1984; Conaway 1989; Agresti 1993).

The symmetry model states that $m_{ijk\ell}^{ABCD}$ is identical for each permutation of (i, j, k, ℓ). This model can be obtained from the saturated model described in equation (2) by imposing certain equality restrictions on its parameters. More precisely, all effects are assumed to be symmetric, which means equal for each permutation of their indices, and all effects of the same order are assumed to be equal. For the one- and two-variable effects, this means

$$\lambda_i^A = \lambda_i^B = \lambda_i^C = \lambda_i^D \text{, and} \tag{4}$$

$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB} = \lambda_{ij}^{AC} = \lambda_{ji}^{AC} = \lambda_{ij}^{AD} = \lambda_{ji}^{AD} = \lambda_{ij}^{BC} = \lambda_{ji}^{BC} = \lambda_{ij}^{BD} = \lambda_{ji}^{BD} = \lambda_{ij}^{CD} = \lambda_{ji}^{CD} . \tag{5}$$

A similar set of constraints is imposed on the three- and four-variable terms. An important feature of the symmetry model in the context of longitudinal data analysis is that it implies homogeneity of the bivariate and univariate marginal tables, that is,

$$m_{ij++}^{ABCD} = m_{i+j+}^{ABCD} = m_{i++j}^{ABCD} = m_{+ij+}^{ABCD} = m_{+i+j}^{ABCD} = m_{++ij}^{ABCD}$$
$$\text{and } m_{i+++}^{ABCD} = m_{+i++}^{ABCD} = m_{++i+}^{ABCD} = m_{+++i}^{ABCD} .$$

The quasi-symmetry model is a generalization of the complete symmetry model that permits different one-variable effects and, hence, marginal heterogeneity (Conaway 1989; Hagenaars 1990; Meiser et al. 1997). This implies that, in comparison with the complete symmetry model, the restrictions on the first-order effects given in equation (4) are relaxed, whereas the constraints on two-, three- and four-factor terms are still in operation.

Structures of complete symmetry or quasi-symmetry can also be specified for "non-saturated" models (Bishop et al. 1975). For instance, the model of complete symmetry without four-factor interaction is obtained by setting $\lambda_{ijk\ell}^{ABCD} = 0$. Similarly, complete symmetry and quasi-symmetry models without three- and four-variable interactions can be obtained. This involves imposing the constraints described in (4) and (5) on the model given in equation (3) (see, for instance, Meiser et al. 1997).

Although Markov-type models are especially suited for the analysis of longitudinal data, symmetric structures can be used for all kinds of multivariate observations. A disadvantage of the symmetric association models is that they postulate that the strength of the association between each pair of times is the same, irrespective of how far they are apart from one other. This assumption seems to be very unrealistic, especially if the number of time points is larger than, say, three.

When the response variable is an ordinal variable, it makes sense to use the ordering of the categories to gain parsimony. For this purpose, we can use log-linear models for ordinal variables proposed by Goodman (1979) (see also Clogg and Shihadeh 1994). The simplest ordinal model is the uniform association model, which is obtained by using the variable indices as category scores:[2]

$$\lambda_{ij}^{AB} = i \bullet j \bullet \lambda_{..}^{AB}. \tag{6}$$

This yields a two-way interaction between *A* and *B* containing only one parameter, $\lambda_{..}^{AB}$.

The first part of Table 2 presents the values of the likelihood-ratio statistic (L^2) obtained when estimating several of the models outlined above with the data reported in Table 1. It is well known that asymptotic *p* values are unreliable when analyzing sparse frequency tables like the one we have here. To circumvent this problem, we estimated the *p* values by means of parametric bootstrapping (see, for example, Langeheine, Pannekoek, and Van de Pol 1996; or Vermunt 1999).

The quasi-symmetry model, one of the simple structure models that are often applied with repeated categorical responses data, is too

TABLE 2: Goodness-of-Fit Statistics for the Estimated Models for the Data in Table 1

Model	L^2	df	$\hat{p}$
Joint distribution			
1. Independence	403.3	72	.00
2. All two-variable terms	36.9	48	.12
3. First-order Markov	58.7	60	.05
4. Second-order Markov	19.6	36	.30
5. Second-order Markov without three-variable terms	37.9	52	.26
6. First-order Markov + $\lambda^{BD}_{j\ell}$	41.6	56	.30
7. Quasi-symmetry	72.3	60	.00
8. Symmetry	158.2	66	.00
Bivariate marginal distributions			
1. Quasi-symmetry	1.2	3	.77
2. Symmetry	59.0	9	.00
3. Uniform association	31.5	9	.00
4. Homogeneous quasi-symmetry	20.2	9	.02
5. Homogeneous transitions with quasi-symmetry	22.1	13	.05
Univariate marginal distributions			
1. Homogeneity	58.1	6	.00
2. Ordinal nonconstant shift	2.0	3	.59
3. Ordinal constant shift	10.6	5	.08
Simultaneous models			
1. $J(6) \cap B(1) \cap U(S)$	43.0	59	.42
2. $J(6) \cap B(S) \cap U(2)$	43.5	59	.37
3. $J(6) \cap B(1) \cap U(2)$	44.8	62	.47

NOTE: The reported *p* values are estimated by means of parametric bootstrapping using 1,000 replications.

restricted for this data set ($L^2 = 72.3$; $df = 60$; $\hat{p} = .00$). As a result, the multivariate symmetry model also will not fit the data. Another simple structure model is the first-order Markov model, which cannot be rejected at a 5% significance level: $L^2 = 58.7$; $df = 60$; $\hat{p} = .05$. However, it is clear that the model with all two-variable terms ($L^2 = 36.9$; $df = 48$; $\hat{p} = .12$), the second-order Markov model ($L^2 = 19.6$; $df = 36$; $\hat{p} = .30$), as well as the second-order Markov model without three-variable terms ($L^2 = 37.9$; $df = 52$; $\hat{p} = .26$) fit better. Inspection of the results of the latter three models showed that, compared to the first-order Markov model, the only significant terms are the λ^{BD}_{jl} parameters. The model that is obtained by adding the λ^{BD}_{jl} parameters to the first-order Markov model fits the data very well ($L^2 = 41.6$; $df = 56$; $\hat{p} = .32$) and can, therefore, serve as the final model for the joint distribution.

4. LOG-LINEAR MODELS FOR THE BIVARIATE MARGINAL DISTRIBUTIONS

What we are interested in now is the pairwise associations between adjacent time points without conditioning on an individual's response at the other two occasions. This involves the specification of models for the three second-order marginal tables *AB*, *BC*, and *CD*, with cell entries m_{ij++}^{ABCD}, m_{+jk+}^{ABCD}, $m_{++k\ell}^{ABCD}$, respectively. These marginal cell entries have to be specified by the matrix **A** appearing in equation (1), which will contain one row for each relevant marginal cell entry. With four time points and a trichotomous response variable, this will be 27 (3 times 9) rows. For example, an element of the row corresponding to m_{+12+}^{ABCD} will be 1 if $B = 1$ and $C = 2$, and 0 otherwise.

We start again with a saturated log-linear model, in this case for each of the three bivariate marginal tables. These are given by

$$\begin{aligned}
\ln m_{ij++}^{ABCD} &= \alpha^{(1)} + \alpha_i^{A(1)} + \alpha_j^{B(1)} + \alpha_{ij}^{AB(1)}, \\
\ln m_{+jk+}^{ABCD} &= \alpha^{(2)} + \alpha_j^{B(2)} + \alpha_k^{C(2)} + \alpha_{jk}^{BC(2)}, \\
\ln m_{++kl}^{ABCD} &= \alpha^{(3)} + \alpha_k^{C(3)} + \alpha_l^{D(3)} + \alpha_{kl}^{CD(3)}.
\end{aligned} \tag{7}$$

Here, the α parameters denote marginal log-linear parameters. Note that we added a superscript, say *t*, to denote the time point, where $t = 1$, $t = 2$, and $t = 3$ refer to the pairs *AB*, *BC*, and *CD*, respectively. This saturated model can also be specified with the **X** matrix (see equation 1), which will contain nine columns per bivariate table. The **X** matrix is a block-diagonal matrix of the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{AB}^{sat} & 0 & 0 \\ 0 & \mathbf{X}_{BC}^{sat} & 0 \\ 0 & 0 & \mathbf{X}_{CD}^{sat} \end{bmatrix}.$$

Here, $\mathbf{X}_{AB}^{sat}$, $\mathbf{X}_{BC}^{sat}$, and $\mathbf{X}_{CD}^{sat}$ refer to the three bivariate margins. Each of these submatrices has the form of a design matrix of a saturated model for a two-way table.

Several meaningful restrictions can be used to simplify these saturated models. One type are the widely discussed models for the analysis of square turnover tables (see Bishop et al. 1975; Andersen 1990; Hout, Duncan and Sobel 1987; Hagenaars 1990). For example, we might consider the model of marginal symmetry for $t = 1$, $t = 2$, and $t = 3$, which implies imposing the following restrictions on the parameters of equation (7):

$$\alpha_i^{A(1)} = \alpha_i^{B(1)}, \alpha_j^{B(2)} = \alpha_j^{C(2)}, \alpha_k^{C(3)} = \alpha_k^{D(3)},$$

in combination with

$$\alpha_{ij}^{AB(1)} = \alpha_{ji}^{AB(1)}, \alpha_{jk}^{BC(2)} = \alpha_{kj}^{BC(2)}, \alpha_{k\ell}^{CD(3)} = \alpha_{\ell k}^{CD(3)}.$$

It should be noted that the model of marginal symmetry implies homogeneity of the univariate distributions; that is, $m_{i+++}^{ABCD} = m_{+i++}^{ABCD} = m_{++i+}^{ABCD} = m_{+++i}^{ABCD}$. The model of marginal quasi-symmetry is obtained by relaxing the constraints on the one-variable terms. Another class of marginal log-linear models is obtained by taking the ordinality of variables into account. The model of marginal uniform association for $t = 1$, $t = 2$, and $t = 3$ is defined by

$$\alpha_{ij}^{AB(1)} = i \bullet j \bullet \alpha_{..}^{AB(1)}, \alpha_{jk}^{BC(2)} = j \bullet k \bullet \alpha_{..}^{BC(2)}, \alpha_{kl}^{CD(3)} = k \bullet \ell \bullet \alpha_{..}^{CD(3)}.$$

Note that these restrictions are similar to the one described in equation (6). In addition, it should be noted that these restrictions yield a symmetric association structure or, equivalently, a restricted quasi-symmetry model.

Specifying restrictions for the separate bivariate marginal tables is just the first step in the simplification of the bivariate marginal association structure. A second step will generally consist of testing hypotheses with respect to the homogeneity of certain parameters over time. Croon et al. (2000) discussed several types of homogeneity hypotheses in the context of generalized log-linear models. In our case, there are three kinds of interesting across-time equality constraints. The first involves

$$\alpha_{ij}^{AB(1)} = \alpha_{ij}^{BC(2)} = \alpha_{ij}^{CD(3)}, \tag{8}$$

which yields a homogenous marginal association model. Note that such a constraint can be combined with any of the above within-margin constraints. For example, in combination with a uniform association structure, restriction (8) yields a homogenous uniform marginal association model.

The second type of homogeneity constraint is obtained by combining restrictions described in (8) with

$$\alpha_j^{B(1)} = \alpha_j^{C(2)} = \alpha_j^{D(3)}.$$

It may not be immediately clear, but this yields time-homogenous transition probabilities. Note that the transition probabilities for the first time point equal

$$\pi_{ji}^{B|A} = \frac{m_{ij++}^{ABCD}}{m_{i+++}^{ABCD}} = \frac{\exp(\alpha_j^{B(1)} + \alpha_{ij}^{AB(1)})}{\sum_{j'} \exp(\alpha_{j'}^{B(1)} + \alpha_{ij'}^{AB(1)})}.$$

As can be seen, the main effect and the parameter for the first of the two time points—in this case, $\alpha^{A(1)}$ and $\alpha_i^{A(1)}$—cancel from this expression. Consequently, by restricting the other effects to be time homogenous, one obtains time-homogeneous transition probabilities. The last homogeneity model involves the two above constraints in combination with

$$\alpha_i^{A(1)} = \alpha_i^{B(2)} = \alpha_i^{D(3)}.$$

This yields complete bivariate marginal homogeneity,[3] that is, $m_{ij++}^{ABCD} = m_{+ij+}^{ABCD} = m_{++ij}^{ABCD}$, and, as a result, also univariate marginal homogeneity: $m_{i+++}^{ABCD} = m_{+i++}^{ABCD} = m_{++i+}^{ABCD} = m_{+++i}^{ABCD}$. If none of the above homogeneity assumptions holds, we might want to investigate whether some structure can be detected in the change of the marginal association over time. An option could be to test whether the strength of the marginal association changes linearly over time. This implies imposing the following constraint on the two-way interactions:

$$\alpha_{ij}^{..(t)} = t \bullet \alpha_{ij}^{..(.)}.$$

This is similar to constraints used in log-linear models for two-way tables with a third so-called layer variable. Here, the variable "time" serves as layer.

The second part of Table 2 reports the goodness-of-fit statistics obtained when applying some of the above bivariate models to the data in Table 1. From the three models that do not impose homogeneity constraints, only the marginal quasi-symmetry model fits the data well: $L^2 = 1.2$; $df = 3$; $\hat{p} = .77$. The homogeneity constraints across time points seem to be too restrictive for this data set.

5. LOG-LINEAR MODELS FOR THE UNIVARIATE MARGINAL DISTRIBUTIONS

The last question about the ordinal repeated responses in Table 1 refers to the four first-order marginal distributions. Saturated log-linear models for these univariate marginal distributions are

$$\begin{aligned} \ln m_{i+++}^{ABCD} &= \gamma^{(1)} + \gamma_i^{A(1)}, \\ \ln m_{+j++}^{ABCD} &= \gamma^{(2)} + \gamma_j^{B(2)}, \\ \ln m_{++k+}^{ABCD} &= \gamma^{(3)} + \gamma_k^{C(3)}, \\ \ln m_{+++\ell}^{ABCD} &= \gamma^{(4)} + \gamma_\ell^{D(4)}. \end{aligned} \tag{9}$$

These marginal cell entries have to be specified by the matrix **A** appearing in equation (1), which will contain one row for each relevant marginal cell entry. With four time points and a trichotomous response variable, this will be 12 (4 times 3) rows. For example, an element of the row corresponding to m_{++2+}^{ABCD} will be 1 if $C = 2$ and 0 otherwise. The block diagonal **X** matrix for the saturated model will contain three columns per marginal table; that is,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_A^{sat} & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{X}_B^{sat} & 0 & 0 \\ \mathbf{0} & 0 & \mathbf{X}_C^{sat} & 0 \\ \mathbf{0} & 0 & 0 & \mathbf{X}_D^{sat} \end{bmatrix}.$$

Here, $\mathbf{X}_A^{sat}$, $\mathbf{X}_B^{sat}$, $\mathbf{X}_C^{sat}$, and $\mathbf{X}_D^{sat}$ refer to the four univariate margins. Each of these submatrices has the form

$$\mathbf{X}_A^{sat} = \mathbf{X}_B^{sat} = \mathbf{X}_C^{sat} = \mathbf{X}_D^{sat} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

When modeling univariate margins, the most interesting types of hypotheses concern constraints across time points. The most restricted variant is the model of marginal homogeneity, that is,

$$m_{i+++}^{ABCD} = m_{+i++}^{ABCD} = m_{++i+}^{ABCD} = m_{+++i}^{ABCD} .$$

This model is obtained by restricting the one-variable terms to be equal across time points:[4]

$$\gamma_i^{A(1)} = \gamma_i^{B(2)} = \gamma_i^{C(3)} = \gamma_i^{D(4)},$$

which involves using a design matrix like

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}.$$

As can be seen, imposing equality constraints across log-linear parameters involves adding up the corresponding columns of the design matrix.

Less restricted models than this marginal homogeneity model can be specified when the categories of the response variable are ordered. Lang and Eliason (1997) proposed what they called marginal shift models, which they applied for modeling differences in marginal distributions in square social mobility tables. A constant marginal shift model is obtained by imposing the following structure on the one-variable terms appearing in equation (9):

$$\gamma_i^{(t)} = i \bullet t \bullet \gamma_{.}^{(.)}.$$

As can be seen, the time-specific one-variable terms are restricted by means of a kind of uniform association model. This yields a marginal shift that is constant across time points and across categories of the response variable. We call this model the ordinal constant shift model. Rather than equal-interval scores, we could also use other sets of category scores for the response variable and/or time. Furthermore, we could use the less restricted model

$$\gamma_i^{(t)} = i \bullet \gamma_{.}^{(t)},$$

which relaxes the assumptions that the change is constant over time. This model could be labeled an ordinal nonconstant marginal shift model. Similarly, we could relax the assumption that the shift is constant across levels of the response variable:

$$\gamma_i^{(t)} = t \bullet \gamma_i^{(.)}.$$

This model could be labeled the nominal constant shift model.

The goodness-of-fit statistics for some of the above models applied to data in Table 1 are displayed in the third part of Table 2. The marginal homogeneity model does not hold for this data set ($L^2 = 58.1$; $df = 6$; $\hat{p} = .00$). The ordinal nonconstant marginal shift model fits very well: $L^2 = 2.0$; $df = 3$; $\hat{p} = .59$. Even though the more restricted ordinal constant shift model cannot be rejected at a 5 percent significance level ($L^2 = 10.6$; $df = 5$; $\hat{p} = .08$), it fits significantly worse than the ordinal nonconstant marginal shift model.

6. SIMULTANEOUS LOG-LINEAR MODELS

What we have been doing so far is restricting either the joint, the bivariate, or the univariate distributions, assuming a saturated model for the other distributions. With the generalized log-linear modeling approach, it is straightforward to specify the restrictions on the three kinds of margins simultaneously.

Using a nomenclature that is similar to the one introduced by Lang and Agresti (1994), let $J(.)$ denote the model for the joint distribution, $B(.)$ the model for the bivariate marginal distributions, and $U(.)$ the model for the univariate marginal distributions. With $J(X) \cap B(Y) \cap U(Z)$, we can denote the model that specifies simultaneously model X for the joint distribution, model Y for the bivariate marginal distributions, and model Z for the univariate marginal distributions.

Actually, the models described in the previous sections are special cases of the general class of simultaneous models for joint and marginal distributions. Let S denote the saturated model. Fitting a standard log-linear model is equivalent to fitting the simultaneous model $J(X) \cap B(S) \cap U(S)$, that is, we model the joint association structure of the responses without making assumptions about the bivariate or univariate marginal distributions. Similarly, a model that restricts the bivariate marginal associations without restricting the first-order marginal distributions and the joint distribution—for example, marginal quasi-symmetry or marginal uniform association—can be denoted by $J(S) \cap B(Y) \cap U(S)$. Models for the univariate marginal distributions, such as marginal homogeneity, which involve imposing restrictions on the univariate marginal distributions without restricting the joint and bivariate marginal distributions, can be denoted by $J(S) \cap B(S) \cap U(Z)$.

Lang (1996) showed that in situations in which a joint and a marginal model are asymptotically separable, the chi-squared statistic for the simultaneous model can be asymptotically partitioned into two components, which implies that the fit of the simultaneous hypothesis can be assessed by separately testing the subhypotheses regarding the joint and the marginal distributions. In such cases, the likelihood-ratio test of the simultaneous model is asymptotically equivalent to the sum of the values of the separate models. The models are asymptotically separable if the marginal frequencies constrained by the marginal model are a linear combination of the sufficient statistics of the log-linear model for the joint distribution. This implies that the model for the joint distribution, $J(X)$, must contain the unrestricted association terms corresponding to the bivariate and univariate marginal distributions modeled by $B(Y)$ and $U(Z)$. For example, if we assume a first-order Markov model for $m_{ijk\ell}^{ABCD}$, the marginal model for the bivariate tables with cell entries m_{ij++}^{ABCD}, m_{+jk+}^{ABCD}, and $m_{++k\ell}^{ABCD}$ is asymptotically separable from the joint distribution model. But, the asymptotic separability condition would not be satisfied if we would omit one of the λ_{ij}^{AB}, λ_{jk}^{BC}, or λ_{kl}^{CD} terms from $J(X)$. According to Lang, separability is especially important when analyzing sparse tables (see also Bergsma 1997). In that case, the goodness of fit of a marginal model can be assessed by taking the difference between the L^2 value of the joint distribution model and the L^2 value of the simultaneous model.

It should be noted that Lang's (1996) sufficient conditions for separability concern the joint model and a single marginal model. This implies for our case that $J(X)$ and the combination of $B(Y)$ and $U(Z)$ are separable under the conditions mentioned above. It can be expected that similar conditions yield a mutual separability of $B(Y)$ and $U(Z)$, which is confirmed by the test results we obtained for the estimated simultaneous models (see below). However, further study that is outside the scope of this article is needed for a formal proof.

Another issue in simultaneous modeling is that in some cases a simultaneous model for the joint and marginal distributions may be equivalent to a more restricted model for the joint distribution. For instance, the complete symmetry model for the joint distribution is equivalent to the simultaneous model that specifies quasi-symmetry for the joint distribution and marginal homogeneity for bivariate marginal distributions (Lang and Agresti 1994). This means that we have to be cautious not to impose redundant constraints. As explained in the next section, there is a way to detect such redundant constraints.

For the data in Table 1, separate fitting suggests that the first-order Markov model with the *BD* interaction, $J(6)$, provides a good description of the joint distribution; the model of marginal quasi-symmetry, $B(1)$, fits best for the bivariate marginal distributions; and the ordinal nonconstant shift model, $U(2)$, describes well the change in the univariate marginal distributions. The L^2 values for three simultaneous models are given in Table 2. Both model $J(6) \cap B(1) \cap U(S)$ and $J(6) \cap B(S) \cap U(2)$ fit well. The same applies to the model that combines the three best separate sets of constraints, $J(6) \cap B(1) \cap U(2)$: $L^2 = 44.8$; $df = 62$; $\hat{p} = .47$. Note that this L^2 value is exactly the same as the sum of the L^2 values of the separate models ($41.6 + 1.2 + 2.0 = 44.8$), which is an indication that mutual separability holds.

The most important parameters of the final model $J(6) \cap B\ (1) \cap U(2)$ are the ordinal marginal shift parameters. These take on the values –0.47, –0.01, 0.17, and 0.32, which shows that the use of marijuana increases with age as well as that the increase is largest between the first two time points. Another interesting result is that the strength of the (symmetric) association in the bivariate tables declines over time, which is an indication that, controlling for the marginal shift, more changes occur at the later ages than at the earlier ages. This is confirmed by the two-way associations in the model for the joint distribution.

7. MAXIMUM LIKELIHOOD ESTIMATION

Lang and Agresti (1994) showed how to estimate models of the form described in equation (1) by means of maximum likelihood.[5] For that purpose, the model described in equation (1) has to be reformulated as follows:

$$\mathbf{U}'(\ln \mathbf{A}\ \mathbf{m}) = 0, \tag{10}$$

where the matrix $\mathbf{U}$ is the orthogonal complement of $\mathbf{X}$. This means that $\mathbf{U}'\mathbf{X} = 0$. Actually, we replace log-linear models by their implied constraints on the logs of the (marginal) cell frequencies.

Assuming a Poisson sampling scheme, maximum likelihood estimation of the cell entries **m** involves finding the saddle point of the following Lagrange log-likelihood function:

$$L = \mathbf{n}'(\ln \mathbf{m}) - 1'\mathbf{m} + \lambda'\mathbf{U}'\,(\ln \mathbf{A}\,\mathbf{m}).$$

Here, **n** is the vector of observed cell entries and λ a vector of Lagrange multipliers. Thus, what we are doing is estimating the cell counts **m** under the constraints formulated in equation (10). Bergsma (1997) showed that with a nonsaturated model for the joint distribution, it is more efficient to treat the joint model and the marginal models differently (see also Lang et al. 1999). More precisely, he proposed using the orthogonal complement transformation only for the marginal part of the model while retaining the log-linear parameterization for the joint distribution. This yields the Lagrange log-likelihood function

$$L = \mathbf{n}'(\ln \mathbf{m}) - 1'\mathbf{m} + \lambda'_{BU}\mathbf{U}'_{BU}\,(\ln \mathbf{A}_{BU}\,\mathbf{m}),$$

with $\ln \mathbf{m} = \mathbf{X}_J\mathbf{b}_J$. The subscripts refer to the three parts of the simultaneous model. To find the restricted maximum likelihood solution, we used the version of the Fisher-scoring algorithm proposed by Bergsma (1997:119-22), which is implemented in an experimental version of the LEM program (Vermunt 1997). Let

$$\mathbf{1} = \mathbf{n}'(\ln \mathbf{m}) - 1'\mathbf{m} \text{ and } \mathbf{h} = \mathbf{U}'_{BU}\,(\ln \mathbf{A}_{BU}\,\mathbf{m});$$

that is, the kernel of the Poisson log-likelihood and the constraints on the bivariate and univariate margins, respectively. Bergsma's algorithm needs the following derivatives with respect to $\mathbf{b}_J$:

$$\mathbf{k} = \frac{\partial \mathbf{l}}{\partial \mathbf{b}_J}, \quad \mathbf{B} = -\frac{\partial^2 \mathbf{l}}{\partial \mathbf{b}_J \partial \mathbf{b}_J'}, \quad \mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{b}_J}.$$

The two-step iteration scheme can now be defined as

$$\lambda_{BU}^{new} = -\,(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1}\,(\mathbf{H}'\mathbf{B}^{-1}\mathbf{k} + \mathbf{h}),$$

TABLE 3: Data From a Five-Wave Consumer Panel

			Wave 4 (D)			
			1		*2*	
Wave 1	*Wave 2*	*Wave 3*	*Wave 5 (E)*		*Wave 5 (E)*	
(A)	*(B)*	*(C)*	*1*	*2*	*1*	*2*
1	1	1	464	31	26	12
1	1	2	28	9	6	5
1	2	1	49	5	7	2
1	2	2	12	5	12	10
2	1	1	79	11	10	8
2	1	2	12	8	3	12
2	2	1	31	5	7	9
2	2	2	25	12	15	58

$$\mathbf{b}_J^{new} = \mathbf{b}_J^{old} + step\mathbf{B}^{-1}(\mathbf{k} + \mathbf{H}\lambda_{BU}^{new}).$$

As can be seen, at each iteration cycle, first new estimates for the Lagrange multipliers λ_{BU} are obtained. Subsequently, the log-linear parameter $\mathbf{b}_J$ are updated using the new estimates of λ_{BU}. The parameter *step* is a step size that has to be adjusted to guarantee convergence.[6] The number of degrees of freedom corresponding to a model is equal to the rank of the information matrix, which is a by-product of the Fisher-scoring algorithm. Thus, by using this estimation method, one is automatically warned if a model with redundant constraints is specified: In that case, the rank of the information matrix will be less than the number of constraints. In addition, one can see which of the constraints is redundant. This proved very useful in the analyses reported in this article.

8. A SECOND EMPIRICAL EXAMPLE

Consider Table 3, taken from an article of Langeheine and Van de Pol (1994) on latent and mixed Markov models. The data stem from a five-wave consumer panel study. The dichotomous response variable indicates whether a family purchased the product of the brand under study (level 2) or whether it purchased another brand (level 1).[7]

TABLE 4: Goodness-of-Fit Statistics for the Estimated Models for the Data in Table 3

Model	L^2	df	*p*
Joint distribution			
1. Independence	883.8	26	.00
2. All two-variable terms	8.7	16	.92
3. First-order Markov	187.2	22	.00
4. Second-order Markov	42.2	16	.00
5. All two-variable terms except λ_{jm}^{BE}	8.8	17	.95
6. Quasi-symmetry	48.3	22	.00
7. Symmetry	116.8	26	.00
Bivariate marginal distributions			
1. Homogeneous association	7.4	3	.06
2. Homogeneous transitions	11.8	6	.07
3. Homogeneous margins	60.6	7	.00
Univariate marginal distributions			
1. Homogeneity	55.2	4	.00
2. Constant shift	4.2	3	.24
Simultaneous models			
1. $J(5) \cap B(2) \cap U(S)$	20.1	23	.64
2. $J(5) \cap B(S) \cap U(2)$	13.0	20	.88
3. $J(5) \cap B(2) \cap U(2)$	68.9	24	.00
4. $J(5) \cap B(1) \cap U(2)$	19.9	23	.65

Each of the three questions described in the introduction of this article is of interest in this application. First, we are interested in the overall association structure: Does a Markov- or a Rasch-type model provide an adequate description of the data? Second, we want to study the adjacent bivariate tables giving information on the net change from one occasion to the next: Are the transitions time homogeneous? Third, we are interested in the gross change or the change in the univariate distributions: Is there marginal homogeneity or a constant marginal shift?

We start with the modeling of the joint distribution. From the test results reported in Table 4, it can be seen that the first- and second-order Markov do not fit the data. The same applies to the multivariate symmetry and quasi-symmetry models. A model that performs well is the model that contains all two-variable associations. From this model, we can exclude the λ_{jm}^{BE} term, which is the only two-variable interaction term that is not significant. It will be clear that the association structure in the joint distribution is quite complicated in this example.

The second part of Table 4 reports the estimated models for the bivariate margins. As can be seen, the model with homogeneous associations, as well as the more restricted model with homogeneous transition probabilities, fit the data at a 5 percent significance level. The assumption of complete homogeneity of the bivariate margins does not hold. The test results for the models for the univariate margins show that marginal homogeneity does not hold. The constant marginal shift model fits very well.

The first two simultaneous models combine model $J(5)$ with either model $B(2)$ or model $U(2)$. The test results show what could be expected: The L^2 values are near to the sum of the ones of the two separate models. However, if we combine the three sets of constraints, $J(5) \cap B(2) \cap U(2)$, we get an L^2 value that is much higher that the sum of the separate models. What happens is that the only constant marginal shift that is in agreement with homogeneous transitions is a marginal shift of zero, which yields marginal homogeneity. That is the reason why this model has the same L^2 value as model $J(5) \cap B(3) \cap U(S)$. This shows the importance of simultaneous modeling: It prevents ending up with submodels that are incompatible. In this example, it turns out that the hypothesis of constant marginal shift is incompatible with homogeneous transitions. An alternative is to combine the constant shift model with the homogenous marginal association model, which yields $J(5) \cap B(1) \cap U(2)$. This model fits the data very well: $L^2 = 19.9$; $df = 23$; $p = .65$.

The most interesting parameter estimates of the last model are the homogeneous bivariate association parameter (.54) and the constant marginal shift parameter (–.15). The value of the association parameter shows that there is quite a strong association between the responses at adjacent occasions: The odds ratio equals exp $(4 \times .54)$, or 8.7. The negative value of the constant marginal shift parameter indicates that there is a shift from level 2 of the response variable to level 1. So, the popularity of the brand under study declined during the observation period.

9. DISCUSSION

This article described a general approach to the analysis of univariate (ordinal) categorical panel data based on applying the generalized log-linear model proposed by Lang and Agresti (1994). The presented approach overcomes the most important limitations of standard log-linear approaches for modeling marginal distributions of repeated responses, which only yield valid results if a certain restricted log-linear model holds for the joint distribution. Our approach makes it possible to test a large variety of hypotheses about the general association structure between responses, as well as about the net and gross change that occurs over time. There are several possible extensions of the approach proposed here. One important extension is the inclusion of (possibly time-varying) explanatory variables in the model. This is straightforward within the presented generalized log-linear modeling framework, especially if we switch to the slightly more general $\mathbf{C} \ln \mathbf{A}\, \mathbf{m} = \mathbf{X}\mathbf{b}$.

Another extension is the inclusion of latent variables to deal with measurement error in the recorded states and with the problem of unobserved heterogeneity. The approach described in this article could, for example, be used in latent and/or mixed Markov models (see, for instance, Langeheine and Van de Pol 1994). Recently, Becker and Yang (1998) showed how to combine generalized log-linear models with latent class models using an EM algorithm. A third important extension is the possibility of dealing with partially missing data, a problem that often occurs in panel studies. For this purpose, we could use the same type of EM algorithm.

The last possible extension that we would like to mention is the possibility of working with a more general class of restrictions than the log-linear restrictions described in this article. The generalized log-linear modeling approach makes it possible to specify, for instance, restrictions on cumulative and global odds ratios, which could be an alternative to our models for local odds ratios. By means of the recursive exp-log models proposed by Bergsma (1997), an even more general class of constraints can be specified. An example is the model $\mathbf{E} \ln(\mathbf{D} \exp(\mathbf{C}(\ln \mathbf{A}\, \mathbf{m}))) = 0$, which allows defining restrictions on association measures like Kendall's tau, gamma, Somer's *d*, and Cohen's kappa.

NOTES

1. Of course, it is also necessary to impose identifying constraints on the parameters. Here, we use ANOVA-type constraints for identification.

2. Note that we use a dot in a parameter to denote that the concerning index is no longer active.

3. Note that complete bivariate marginal homogeneity is actually a linear constraint on cell entries. Therefore, it can also be specified with the approach proposed by Haber and Brown (1986).

4. Note that the marginal homogeneity model can also be formulated as a model with linear constraints on the cell entries. In that sense, it fits within the framework proposed by Haber and Brown (1986).

5. Another estimation procedure that can be used but has certain disadvantages compared to maximum likelihood is weighted least squares (Grizzle, Starmer, and Koch 1969). In addition, a quasi-likelihood approach known as generalized estimating equations has been proposed for estimating the parameters of marginal models (see, for instance, Diggle, Liang, and Zeger 1994; Fahrmeir and Tutz 1994).

6. We used a step size of 1/4 in the first two iterations, 1/2 in the next two iterations, and 1 in the remaining iterations. In cases in which we had convergence problems with this procedure, we kept the step size of 1/2 until convergence.

7. Note that in this example, we are not able to present models that make use of the ordinal nature of a response variable. These models were, however, already illustrated with the first data set.

REFERENCES

Agresti, Alan. 1993. "Computing Conditional Maximum Likelihood Estimates for Generalized Rasch Models Using Simple Log-Linear Models With Diagonal Parameters." *Scandinavian Journal of Statistics* 20:63-71.

———. 1997. "A Model for Repeated Measurements of a Multivariate Binary Response." *Journal of the American Statistical Association* 92:315-21.

Andersen, Erling B. 1990. *The Statistical Analysis of Categorical Data*. Berlin: Springer-Verlag.

Becker, Mark P. 1994. "Analysis of Cross-Classifications of Counts Using Models for Marginal Distributions: An Application to Trends in Attitudes on Legalized Abortion." Pp. 229-65 in *Sociological Methodology 1994*, edited by Peter V. Marsden. Oxford, UK: Blackwell.

Becker, Mark P. and Ilsoon Yang. 1998. "Latent Class Marginal Models for Cross-Classifications of Counts." Pp. 293-325 in *Sociological Methodology 1998*, edited by Adrien Raftery. Boston: Blackwell.

Becker, Mark P., S. Minick, and Ilsoon Yang. 1998. "Specifications of Models for Cross-Classified Counts: Comparison of the log-Linear Models and Marginal Models Perspectives." *Sociological Methods and Research* 26:511-29.

Bergsma, Wicher. 1997. *Marginal Models for Categorical Data*. Tilburg, the Netherlands: Tilburg University Press.

Bishop, Yvonne M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Caussinus, H. 1965. "Contribution a l'analyse statistique des tableaux de correlation." *Annales de la Faculte des Sciences de l'Universite de Toulouse* 29:77-182.

Clogg, Clifford C. and Edward S. Shihadeh. 1994. *Statistical Models for Ordinal Data*. Thousand Oaks, CA: Sage.

Conaway, M. R. 1989. "Analysis of Repeated Measurements With Conditional Likelihood Methods." *Journal of the American Statistical Association* 84:53-63.

Croon, Marcel, Wicher Bergsma, and Jacques A. Hagenaars. 2000. "Analyzing Change in Discrete Variables by Generalized Log-Linear Models." *Sociological Methods and Research* 29:195-229.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. Oxford, UK: Clarendon.

Elliot, D. S., D. Huizinga, and S. Menard. 1989. *Multiple Problem Youth: Delinquence, Substance Use, and Mental Health Problems*. New York: Springer-Verlag.

Fahrmeir, L. and G. Tutz. 1994. *Multivariate Statistical Modelling Based on Generalised Linear Models*. New York: Springer.

Goodman, Leo A. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 74:537-52.

Grizzle, James E., C. Frank Starmer, and Gary G. Koch. 1969. "Analysis of Categorical Data by Linear Models." *Biometrics* 25:489-504.

Haber, Michael and Morton B. Brown. 1986. "Maximum Likelihood Methods for Log-Linear Models When Expected Frequencies Are Subject to Linear Constraints." *Journal of the American Statistical Association* 81:477-82.

Hagenaars, Jacques A. 1990. *Categorical Longitudinal Data: Log-Linear Analysis of Panel, Trend and Cohort Data*. Newbury Park, CA: Sage.

Hout, Michael, Otis D. Duncan, and Michael E. Sobel. 1987. "Association and Heterogeneity: Structural Models of Similarities and Differences." Pp. 145-84 in *Sociological Methodology 1987*, edited by Clifford C. Clogg. Oxford, UK: Basil Blackwell.

Kelderman, Henk. 1984. "Log-Linear Rasch Models Test." *Psychometrika* 49:223-45.

Lang, Joseph B. 1996. "On the Partitioning of Goodness-of-Fit Statistics for Multivariate Categorical Response Models." *Journal of the American Statistical Association* 91:1017-23.

Lang, Joseph B. and Alan Agresti. 1994. "Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses." *Journal of the American Statistical Association* 89:625-32.

Lang, Joseph B. and Scott R. Eliason. 1997. "The Application of Association-Marginal Models to the Study of Social Mobility." *Sociological Methods and Research* 26:183-213.

Lang, Joseph B., John W. McDonald, and Peter W. F. Smith. 1999. "Association Modeling of Multivariate Categorical Responses: A Maximum Likelihood Approach." *Journal of the American Statistical Association* 94:1161-71.

Langeheine, Rolf, Jeroen Pannekoek, and Frank Van de Pol 1996. "Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis." *Sociological Methods and Research* 24:492-516.

Langeheine, Rolf and Frank Van de Pol. 1994. "Discrete Time Mixed Markov Latent Class Models." Pp. 171-97 in *Analyzing Social and Political Change: A Casebook of Methods*, edited by A. Dale and R. B. Davies. Thousand Oaks, CA: Sage.

Lindsey, J. K. 1993. *Models for Repeated Measurements*. Oxford, UK: Clarendon.

Meiser, Thorsten, Alexander Von Eye, and Chistiane Spiel. 1997. "Log-Linear Symmetry and Quasi- Symmetry Models for the Analysis of Change." *Biometrical Journal* 39:351-68.

Vermunt, Jeroen K. 1997. *LEM: A General Program for the Analysis of Categorical Data. User's Manual*. Tilburg, the Netherlands: Tilburg University.

———. 1999. "A General Class of Non-Parametric Models for Ordinal Categorical Data." Pp. 187-224 in *Sociological Methodology 1999*, edited by M. Sobel and M. Becker. Boston: Blackwell.

Jeroen K. Vermunt is an associate professor in the Department of Methodology and Statistics of the Faculty of Social and Behavioral Sciences and a research associate at the Work and Organization Research Center at Tilburg University, the Netherlands. He specializes in the analysis of categorical data and latent class analysis.

María Florencia Rodrigo is an assistant professor in the Department of Methodology of the Faculty of Psychology at the University of Valencia, Spain. Her research interests include the analysis of categorical data and group decision making.

Manuel Ato-Garcia is a professor in the Department of Experimental Psychology and Methodology of the Psychology Faculty at Murcia University, Spain. His research interest is categorical data analysis.