PSYCHOMETRIKA—VOL. 69, NO. 1, 5–32
MARCH 2004

BUILDING AN IDENTIFIABLE LATENT CLASS MODEL WITH COVARIATE EFFECTS ON UNDERLYING AND MEASURED VARIABLES

GUAN-HUA HUANG

UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN

KAREN BANDEEN-ROCHE

THE JOHNS HOPKINS UNIVERSITY, BALTIMORE, MARYLAND

In recent years, latent class models have proven useful for analyzing relationships between measured multiple indicators and covariates of interest. Such models summarize shared features of the multiple indicators as an underlying categorical variable, and the indicators' substantive associations with predictors are built directly and indirectly in unique model parameters. In this paper, we provide a detailed study on the theory and application of building models that allow mediated relationships between primary predictors and latent class membership, but that also allow direct effects of secondary covariates on the indicators themselves. Theory for model identification is developed. We detail an Expectation-Maximization algorithm for parameter estimation, standard error calculation, and convergent properties. Comparison of the proposed model with models underlying existing latent class modeling software is provided. A detailed analysis of how visual impairments affect older persons' functioning requiring distance vision is used for illustration.

Key words: EM algorithm, finite mixture model, identifiability, multiple discrete indicators, visual functioning.

1. Introduction

In many studies, the conceptually or clinically most meaningful outcome is inaccessible due to cost, time, and difficulty of measurement. A set of multiple indicators is then measured in place of this outcome. For example, psychiatric disorders are often assessed by applying standardized criteria to patients' report of symptoms (Eaton, Dryman, Sorenson, & McCutcheon, 1989). Biomarkers are used very often as substitutes for observing new cases of cancer in testing treatments for cancer prevention, where event rates are low and a long time may be needed to obtain cancer cases (Piantadosi, 1997). Functional disability is commonly quantified as self-reported categorical responses to a series of questions about difficulty performing tasks of routine living (e.g., Katz et al., 1963), because no obvious single measure of disability exists. Statistical methods for analyzing these measured indicators should have the capability to model the relationship between indicators and conceptual outcomes, and to describe the underlying mechanism of the condition under investigation. The present paper investigates an increasingly widespread strategy for analyzing data collected in situations where investigators use multiple discrete indicators to measure the conceptually defined outcome.

Particularly in psychosocial research, latent variable models are recognized as an effective tool for analyzing measured indicators. There are two primary latent variable approaches for situations where multiple categorical indicators are used: latent trait models and latent class models.

This work was supported by National Institute on Aging (NIA) Program Project P01-AG-10184-03 and National Institutes of Mental Health grant R01-MH-56639-01A1. Dr. Bandeen-Roche is a Brookdale National Fellow. The authors wish to thank Drs. Gary Rubin and Sheila West for kindly making the Salisbury Eye Evaluation data available. We also thank the Editor, the Associate Editor, and three referees for their valuable comments.

Requests for reprints should be sent to Guan-Hua Huang, Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, TAIWAN

0033-3123/2004-1/2002-0971-A $00.75/0

Latent trait models express the unobservable conceptual outcome as a continuous score and determine the association with risk factors in a single modeling step (Rasch, 1960; Hambleton, Swaminathan, & Rogers, 1991; Muthén, 1983, 1984; Sammel, Ryan, & Legler, 1997; Legler & Ryan, 1997). Latent class models differ from latent trait models in that the continuous score is replaced by a variable identifying several "classes" that define homogeneous groups of individuals (Green, 1951; Lazarsfeld & Henry, 1968; Goodman, 1974; Haberman, 1974, 1979). An underlying categorical variable is arguably most robust for summarizing data whose basic structure is patterns of categorical responses. It also does not require modeling according to a specific distribution, as does the continuous score. In recent years, latent class modeling has been receiving increasing attention in both psychosocial (e.g., Neuman et al., 2001; Garrett & Zeger 2000; Hudziak et al., 1998) and medical research (e.g., Moustaki, 1996; Sullivan, Kessler, & Kendler, 1998; Bandeen-Roche, Huang, Munoz, & Rubin, 1999).

In this paper, we extend the latent class model to allow both the distribution of the underlying class variable and the within-class distributions of measured indicators to be functionally related to individual-level independent variables (henceforth, regression extension of latent class analysis will be called RLCA). This idea is not new in and of itself. Quite general regression models have been developed to describe the relation between covariates and the underlying variable (Dayton & Macready, 1988; Van der Heijden, Dessens, & Böckenholt, 1996; Bandeen-Roche, Miglioreti, Zeger, & Rathouz, 1997), or the relation between covariates and the measured indicators themselves (Melton, Liang, & Pulver, 1994). The former seeks to estimate the effects of independent variables on the conceptual outcome, whereas the latter aims to adjust for characteristics associated with measurement, hence preventing possible misclassification of underlying variable categories. Models incorporating covariates to predict both the underlying and measured outcomes date to the mid-1980's (Clogg and Goodman, 1984, 1985; Formann, 1985, 1992; Hagenaars, 1993), but these were highly constrained in applying to categorical covariates. Recent methodology and software have very generally incorporated covariates for predicting both underlying variables and measured indicators (Muthén and Shedden, 1999; Roeder, Lynch, & Nagin, 1999; LEM: Vermunt, 1996; Mplus: Muthén and Muthén 1998; Latent GOLD: Vermunt and Magidson, 2000).

Despite this body of modeling research, we believe that at least two important issues remain unresolved by the prior literature. First, simultaneously regressing the latent class variable and measured indicators on covariates raises substantial identification questions. The issue is distinct from identification challenges in models that regress either the latent class variable or their measured indicators, but not both, on covariates. Second, while maximum likelihood procedures have been reported and implemented for models that allow simultaneous regressions, their inferential and convergence properties have not been detailed. It is known that implementing the maximum likelihood procedure to estimate RLCA parameters is time-consuming, does not result in direct variance estimation, and carries no guarantee of finding a (global or local) maximum. A detailed discussion of these issues is extremely valuable.

To address the important issues just identified, this paper: (a) formulates sufficient conditions for model identifiability of RLCA with two types of covariate effects; (b) proposes modeling that guarantees identifiability and confers meaningful parameter interpretation; and (c) details full maximum likelihood inference and convergence properties of the estimating procedure. Our model can be viewed as a latent class analogy of "MIMIC" models (Jöreskog and Goldberger, 1975), and the developed theorem for model identifiability is the analogy of identification findings for MIMIC models with direct effects (Bollen, 1989, p. 328). To outline the remainder of the paper: section 2 proposes our model. Sufficient conditions for the identifiability of the proposed model are provided in section 3. In section 4, we develop an EM algorithm for estimating parameters and calculating their standard errors. We also justify convergent properties of this estimating procedure. Section 5 provides a comparison of our model with models underlying existing latent class software. In section 6, visual functioning data are used to illustrate the model. We also offer a comparison of our results with a published analysis obtained by a closely related model. Discussion is provided in section 7.

2. Regression Extension of Latent Class Models

Latent class analysis (LCA) aims to classify subjects based on their responses to a set of categorical items. To introduce the methodology, let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})^T$ denote a set of M observable polytomous indicators for the ith individual in a study sample of N persons. Y_{im}, $m = 1, \dots, M$ can take values $\{1, \dots, K_m\}$, where $K_m \geq 2$. The basic model postulates an underlying categorical latent variable $S_i = 1, \dots, J$ for individual i; within any category of the latent variable, the measured indicators are assumed to be independent of one another. Therefore, the distribution for $\mathbf{Y}_i$ can be expressed as

$$\Pr(Y_{i1} = y_1, \dots, Y_{iM} = y_M) = \sum_{j=1}^{J} \{\Pr(S_i = j) \prod_{m=1}^{M} \prod_{k=1}^{K_m} [\Pr(Y_{im} = k | S_i = j)]^{y_{mk}}\}, \tag{1}$$

where $y_{mk} = 1$ if $y_m = k$; 0 otherwise. The LCA model assumes that

$$\Pr(Y_{im} = k | S_i = j) = p_{mkj}, \quad \Pr(S_i = j) = \eta_j, \tag{2}$$

$i = 1, \dots, N$; $m = 1, \dots M$; $k = 1, \dots, K_m$; $j = 1, \dots, J$. Thus, the model treats class membership probabilities, η_j, and item response probabilities conditional on class membership, p_{mkj}, as homogeneous over individuals. Heuristically, η_j is the population prevalence of class j, and p_{mkj} is the probability of an individual in class j being at level k of Y_{im}. Goodman (1974) provided an excellent overview of the LCA model, including a maximum likelihood strategy for estimating model parameters, conditions to determine local model identifiability, a strategy to test overall model fit, and the use of constraints to identify models.

The present goal is to extend latent class analysis to allow both the probabilities of latent class membership and the distribution of observed responses given latent class membership to be functionally related to concomitant variables, while preserving model identifiability. By allowing covariate effects on latent class probabilities, we can summarize the effect of risk factors on the underlying mechanism. In the case of incorporating covariates into conditional probabilities, we can adjust for characteristics that determine responses other than underlying classes, hence hopefully improving the accuracy of classifying individuals. For example, in evaluating functional disability, some data have suggested that women tend to rate tasks as “difficult” more readily than men independently of ability (Bandeen-Roche, Huang, Munoz, & Rubin, 1999). Without adjusting for a gender effect, the model might well classify some men and women with identical underlying functioning differently (men as “able”, women as “disabled”).

Let $(\mathbf{x}_i, \mathbf{z}_i)$ be the concomitant covariates of the ith person, where $\mathbf{x}_i = (1, x_{i1}, \dots, x_{iP})^T$ are primary covariates hypothesized to be associated with latent class membership, S_i, and $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{iM})$ with $\mathbf{z}_{im} = (1, z_{im1}, \dots, z_{imL})^T$, $m = 1, \dots, M$, are secondary covariates used to build direct effects on measured indicators. The covariates may include any combination of continuous and discrete measures, and two sets of covariates may be mutually exclusive or overlap. When common covariates are used to predict both underlying and measured variables, our following proposed model can still be identifiable (see section 3 for details).

The regression extension of LCA may then be stated as follows:

$$\Pr(Y_{i1} = y_1, \dots, Y_{iM} = y_M | \mathbf{x}_i, \mathbf{z}_i) = \sum_{j=1}^{J} \left\{ \eta_j(\mathbf{x}_i^T \boldsymbol{\beta}) \prod_{m=1}^{M} \prod_{k=1}^{K_m} p_{mkj}^{y_{mk}}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T \boldsymbol{\alpha}_m) \right\} \tag{3}$$

with $\eta_j(\mathbf{x}_i^T\boldsymbol{\beta})$ and $p_{mkj}(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^T\boldsymbol{\alpha}_m)$ defined as in the generalized linear framework (McCullagh and Nelder, 1989). Various link functions (e.g., probit, ordinal) could be used easily. We specifically propose to use the generalized logit link function (Agresti, 1984):

$$\log\left[\frac{\eta_j(\mathbf{x}_i^T\boldsymbol{\beta})}{\eta_J(\mathbf{x}_i^T\boldsymbol{\beta})}\right]=\beta_{0j}+\beta_{1j}x_{i1}+\cdots+\beta_{Pj}x_{iP}\quad \text{for } i=1,\dots,N;\ j=1,\dots,J-1, \tag{4}$$

and

$$\log\left[\frac{p_{mkj}(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^T\boldsymbol{\alpha}_m)}{p_{mKj}(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^T\boldsymbol{\alpha}_m)}\right]=\gamma_{mkj}+\alpha_{1mk}z_{im1}+\cdots+\alpha_{Lmk}z_{imL}$$
$$\text{for } i=1,\dots,N;\ m=1,\dots,M;\ k=1,\dots,(K_m-1);\ j=1,\dots,J. \tag{5}$$

Three assumptions complete the model (3):

1. Class membership probabilities are associated with $\mathbf{x}_i$ only:
$$\Pr(S_i=j\,|\,\mathbf{x}_i,\mathbf{z}_i)=\Pr(S_i=j\,|\,\mathbf{x}_i). \tag{6}$$
2. Conditioning on class membership, responses are only associated with $\mathbf{z}_i$:
$$\Pr(Y_{i1}=y_1,\dots,Y_{iM}=y_M|S_i,\mathbf{x}_i,\mathbf{z}_i)=\Pr(Y_{i1}=y_1,\dots,Y_{iM}=y_M|S_i,\mathbf{z}_i). \tag{7}$$
3. The multiple measurements are independent given class membership and $\mathbf{z}_i$:
$$\Pr(Y_{i1}=y_1,\dots,Y_{iM}=y_M|S_i,\mathbf{z}_i)=\prod_{m=1}^{M}\Pr(Y_{im}=y_m|S_i,\mathbf{z}_{im}). \tag{8}$$

Some key features of the proposed regression extension of latent class model (3) are: First, by incorporating covariates $(\mathbf{x}_i,\mathbf{z}_i)$ into class prevalences and conditional probabilities, we relax the homogeneous probability assumption (2) in the sense that the probabilities vary with some individual characteristics. Second, there are several useful sub-models of the proposed model. By fixing γ_{mkj} in (5) at positive or negative infinity, we can fit a constrained RLCA with the corresponding conditional probabilities being 1 or 0. If the regression coefficients in (4) or (5) are set as 0, our proposed model (3) reduces to models studied by Melton, Liang, and Pulver (1994), Dayton and Macready (1988) or an ordinary latent class analysis. Third, we allow unrestricted intercepts and level- and item-specific covariate coefficients in the conditional probability model (5), but we do not allow the coefficients to vary across latent classes. This constraint is logical if the primary purpose of modeling conditional probabilities is to prevent possible misclassification by adjusting for characteristics associated with item measurements. As we now discuss, it is also necessary to unambiguously distinguish covariate effects on measured responses from covariate effects on class membership itself.

3. Identifiability

In some statistical models, different parameterizations determine identical distributions. This is referred to as nonidentifiability. Before estimation of RLCA (3) can be meaningfully attempted, model identifiability must be verified.

The latent class analysis literature has focused on checking "local" identifiability (McHugh 1956; Goodman 1974; Formann 1992). By definition, a distribution $F_{\mathbf{Y}}$ is locally identifiable at the parameter $\boldsymbol{\phi}_0$ if there exists some neighborhood $\boldsymbol{\chi}$ of $\boldsymbol{\phi}_0$ such that

$$F_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\phi}_0) = F_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\phi}) \quad \text{for all} \quad \mathbf{y} \in U_{\mathbf{Y}} \Leftrightarrow \boldsymbol{\phi} = \boldsymbol{\phi}_0 \quad \text{for all} \quad \boldsymbol{\phi} \in \boldsymbol{\chi} \subset \boldsymbol{\Phi},$$

where $\boldsymbol{\Phi}$ denotes the parameter space of the model and $U_{\mathbf{Y}}$ denotes the support of $\mathbf{Y}$. McHugh (1956) proposed sufficient conditions for the local identifiability of the LCA model with dichotomous observed variables and Goodman (1974) extended the conditions to polytomous variables. We here modify them for models with prefixed parameters (constrained model) and propose a condition equivalent to the full column rank of the Jacobian matrix (for proof, see Appendix A):

Proposition 1. For $j = 1, \ldots, J$, let $\boldsymbol{\psi}_j$ be a $((\prod_{m=1}^{M} K_m) - 1) \times 1$ vector with hth element

$$\psi_{hj} = \Pr(\mathbf{Y}_i = \mathbf{y}_h | S_i = j) = \prod_{m=1}^{M} p_{my_{hm}j},$$

where $\mathbf{y}_h = (y_{h1}, \ldots, y_{hM})$ is the hth possible among $((\prod_{m=1}^{M} K_m) - 1)$ distinct response patterns, excluding a reference pattern. C is the number of pre-fixed conditional probabilities $p_{mkj} = 0$ or 1. Suppose that

(i) $(\prod_{m=1}^{M} K_m) - 1 \geq J(\sum_{m=1}^{M}(K_m - 1)) + J - 1 - C$;
(ii) $p_{mkj} > 0$ and $\eta_j > 0$ for all free parameters (i.e., parameters that are not prefixed); and
(iii) $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_J$ are linearly independent.

Then, the constrained latent class analysis model (1, 2) is locally identifiable at free parameters of $\{(p_{mkj}, \eta_j);\ \forall m, k, j\}$.

The LCA is constrained by fixing specific conditional probabilities. The proposition aims to determine whether the unknown (i.e., free) parameters in a constrained model are identifiable. Condition (i) states that the number of unique model parameters cannot exceed the number of independent pieces of observed information. Condition (ii) is to ensure that the probability of each possible response pattern is positive, which is (iii) in Theorem 1 of McHugh (1956). Condition (iii) is equivalent to requiring that the Jacobian of the LCA model has full column rank and has the meaning that the probability distributions for possible response patterns are linearly independent across latent classes.

For RLCA models, the Jacobian grows to an unreasonably large row-dimension in continuous covariate applications. In the following, we develop a method for checking the identifiability of the RLCA model (3) by separating out the covariate effects and then applying Proposition 1 to each subject.

Theorem 1. For $j = 1, \ldots, J$, let $\boldsymbol{\tau}_j$ be a $((\prod_{m=1}^{M} K_m) - 1) \times 1$ vector with hth element

$$\tau_{hj} = \prod_{m=1}^{M} \left\{ \frac{e^{\gamma_{my_{hm}j}}}{1 + \sum_{k=1}^{K-1} e^{\gamma_{mkj}}} \right\}; \quad \gamma_{mKj} = 0,$$

with $\mathbf{y}_h = (y_{h1}, \ldots, y_{hM})$ as defined in Proposition 1 and γ_{mkj} as in (5). C is the number of γ_{mkj}'s that tend to $\pm\infty$. Suppose that

(i′) $(\prod_{m=1}^{M} K_m) - 1 \geq J(\sum_{m=1}^{M}(K_m - 1)) + J - 1 - C$;
(ii′) free model parameters γ_{mkj}, α_{qmk}, β_{pj}, and covariate values x_{ip}, z_{imq} are all finite;
(iii′) $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_J$ are linearly independent; and

(iv′) the design matrix of the primary predictors

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^T = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1P} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{NP} \end{bmatrix}$$

and the design matrices of the items' predictors

$$\mathbf{Z}_m = (\mathbf{z}_{1m}, \ldots, \mathbf{z}_{Nm})^T = \begin{bmatrix} 1 & z_{1m1} & \cdots & z_{1mL} \\ \vdots & \vdots & & \vdots \\ 1 & z_{Nm1} & \cdots & z_{NmL} \end{bmatrix}, m = 1, \ldots, M,$$

have full column rank.

Then, the constrained regression extension of latent class model (3, 4, 5) is locally identifiable at free parameters of $\{(\gamma_{mkj}, \alpha_{qmk}, \beta_{pj});\ \forall m, k, j, p, q\}$.

For proof, see Appendix A. Conditions (i′), (ii′), and (iii′) provide sufficient conditions for the local identifiability of all LCA's that result by applying (3) to a population whose members have identical covariate values. Condition (iv′) requires covariates not to be perfectly collinear. Model (3) formulates each subject's probability components as a combination of an LCA model plus variations across individual characteristics $(\mathbf{x}_i, \mathbf{z}_i)$. By requiring that (i′)–(iii′) and (iv′) hold, the LCA model, and the parameters that determine how the model varies with individual characteristics, are both locally identifiable; hence, their combination is also locally identifiable. Notice that condition (iv′) only requires individual $\mathbf{X}$ and $\mathbf{Z}_m$ to have full column rank, not jointly to have full rank. Therefore, in the case where common covariates are used to predict both underlying and measured variables, model (3) can still be identifiable as long as the above conditions are met. The Jacobian matrix of the RLCA (3) can be partitioned into sub-matrices where each sub-matrix is represented as either a combination of $\mathbf{X}$ and the Jacobian of LCA (1) with respect to latent prevalences, or as $\mathbf{Z}_m$ and the Jacobian of LCA (1) with respect to the mth item's conditional probabilities. If the LCA model is identifiable, then only full column rank for each individual $\mathbf{X}$ and $\mathbf{Z}_m$ is needed to obtain an identifiable RLCA (3).

Importantly, model identifiability may fail if covariate effects on the conditional probabilities are not constrained to be equal across classes. To illustrate this, consider a two-class RLCA with five two-level measured indicators and "gender" associated with both the class membership and measured indicators themselves (i.e., $J = 2, M = 5, K_1 = \cdots = K_5 = 2, P = L = 1$). Under the RLCA model with unconstrained covariate effects on the conditional probabilities,

$$\log\left[\frac{p_{m1j}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_{mj})}{p_{m2j}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_{mj})}\right] = \gamma_{mj} + \alpha_{mj}(\text{gender})_{im} \quad \forall i, m, j, \tag{9}$$

where $(\text{gender})_{im} = 1$ if female (F), 0 if male (M). Suppose that $\gamma_{m1} \neq \gamma_{m2}$ for $m = 1, \ldots, 5$, so that for males, the conditional probabilities of responding positively differ across the two classes, for all five items (i.e., $p_{m11}(\text{M}) \neq p_{m12}(\text{M})$, $m = 1, \ldots, 5$). Suppose further that $\alpha_{m1} = 0$ and $\alpha_{m2} = \gamma_{m1} - \gamma_{m2}$ for $m = 2, \ldots, 5$, so that for females, the conditional probabilities of responding positively are identical across the two classes, for all items except the first (i.e., $p_{m11}(\text{F}) = p_{m12}(\text{F}) = p_m(\text{F})$, $m = 2, \ldots, 5$). Then, the likelihood of measured item responses is

$$\text{L}(\mathbf{Y}) = \prod_{i \in \text{male}} \left\{ \eta_1(\text{M}) \prod_{m=1}^{5} [(p_{m11}(\text{M}))^{y_{im}}(1 - p_{m11}(\text{M}))^{1-y_{im}}] \right.$$

$$+ (1-\eta_1(\mathrm{M})) \prod_{m=1}^{5} [(p_{m12}(\mathrm{M}))^{y_{im}} (1-p_{m12}(\mathrm{M}))^{1-y_{im}}] \Bigg\}$$

$$\times \prod_{i \in \text{female}} \left\{ \Theta_i \prod_{m=2}^{5} [(p_m(\mathrm{F}))^{y_{im}} (1-p_m(\mathrm{F}))^{1-y_{im}}] \right\},$$

where

$$\Theta_i = \eta_1(\mathrm{F})(p_{111}(\mathrm{F}))^{y_{i1}}(1-p_{111}(\mathrm{F}))^{1-y_{i1}} + (1-\eta_1(\mathrm{F}))(p_{112}(\mathrm{F}))^{y_{i1}}(1-p_{112}(\mathrm{F}))^{1-y_{i1}}. \quad (10)$$

Notice that (10) imposes two restrictions on parameters (i.e., for $y_{i1} = 1$ or 0), and there are three parameters that we need to consider (i.e., α_{11}, α_{12} and the gender coefficient in η_1). Because the number of restrictions is less than the number of parameters of interest, equation (10), thus the RLCA model with the conditional probability regression (9), is not locally identifiable at α_{11}, α_{12} and the gender coefficient in η_1. This example can be generalized to the J-class model with M dichotomous items, where females have the same conditional probabilities in latent classes, say, 1 and 2 for the last $M_1 (< M)$ items. The sort of nonidentifiability that we have highlighted can occur in practice. Our identifiability finding in class-independent covariate effects on conditional probabilities is particularly important for complex examples (many predictors) and provides protection for this general case.

Another way of evaluating the local model identifiability of LCA models is to examine whether or not the Fisher information matrix (i.e., the negative expected matrix of the second-order partial derivatives of the log likelihood) possesses eigenvalues greater than 0. Formann (1985, 1992) showed that this approach is equivalent to examining the rank of the Jacobian matrix. Under the RLCA model (3), the Fisher information matrix can be expressed as

$$\mathrm{E}[-D^2_{\boldsymbol{\phi}} \log \mathrm{L}] = \mathrm{E}\left[\left(\frac{\partial \log \mathrm{L}}{\partial \boldsymbol{\phi}}\right)\left(\frac{\partial \log \mathrm{L}}{\partial \boldsymbol{\phi}}\right)^T\right] = \sum_i \sum_h \left[\frac{1}{\pi_{ih}}\left(\frac{\partial \pi_{ih}}{\partial \boldsymbol{\phi}}\right)\left(\frac{\partial \pi_{ih}}{\partial \boldsymbol{\phi}}\right)^T\right], \quad (11)$$

where $\log \mathrm{L} = \sum_{i=1}^{N} \log \Pr(\mathbf{Y}_i|\mathbf{x}_i, \mathbf{z}_i)$ is the log likelihood function, $D^2_{\boldsymbol{\phi}}$ is the Hessian operator with respect to $\boldsymbol{\phi} = (\boldsymbol{\gamma}_{mj}, \boldsymbol{\alpha}_m, \boldsymbol{\beta})$, and $\pi_{ih} = \Pr(\mathbf{Y}_i = \mathbf{y}_h; \boldsymbol{\phi})$ is the probability that ith subject has $\mathbf{y}_h$ response pattern. Notice that the Fisher information matrix (11) is equal to $\mathbf{D}^T\mathbf{G}\mathbf{D}$, where $\mathbf{D}$ is the Jacobian matrix of RLCA model (3) with elements described in Appendix (A.7), (A.8) and (A.9), and $\mathbf{G}$ is a diagonal matrix with elements equal to $(1/\pi_{ih})$. Therefore, if $\mathbf{D}$ is of full column rank, the Fisher information matrix (11) has all eigenvalues greater than 0 (Graybill, 1969, p. 318). Theorem 1 (iii′) and the Fisher information matrix provide equivalent information for identifiability. A standard practice for checking identifiability is using multiple sets of initial values for parameter estimation. Different sets of initial values that yield the same likelihood maximum should result in the same final parameter estimates. If not, the model is not identifiable.

Complications often arise from applying Proposition 1, Theorem 1, and the Fisher information matrix to a given analysis. Ideally, one would want to determine those regions of the parameter space in which a given model is locally identifiable. Because this is typically computationally difficult, these methods are often evaluated with respect to estimated parameters to establish local model identifiability at estimated values (Goodman, 1974). When using the Fisher information matrix, there is one more complication. Since the observed Fisher information, $-D^2_{\boldsymbol{\phi}} \log \mathrm{L}$, is typically used to estimate the standard errors of maximum likelihood estimators (Efron and Hinkley, 1978; Louis, 1982), the Fisher information (11) is not always obtained and the observed Fisher information is used for empirical checking. Empirical identifiability checking through the observed Fisher information might cause errors because we use the "single" observation in place of the averaged effect. It needs to be implemented cautiously.

4. Estimation

4.1. Parameter Estimation

We use maximum likelihood (ML) to estimate the parameters in (3) for a fixed number of classes, J. The problem of selecting J empirically is beyond the scope of this paper; rather we proceed as if the number of classes can be selected based on prior knowledge and the scientific objective. Viewing the class membership S_i as unobservable, the RLCA model (3) becomes a typical incomplete-data problem. The Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) is an iterative approach to computing ML estimates when a model can be formulated in terms of quantities that may be viewed as missing data. The EM algorithm maximizes the likelihood by iterating between imputation of missing data from a model parameterized at the most recent estimates and maximization of the "complete-data" likelihood (joint with respect to observable and missing data). Formally, imputation is carried out through an E (expectation) step that calculates the expected complete-data likelihood given observed data, and an M (maximization) step that maximizes the likelihood calculated from the E-step.

Let S_{ij} indicate whether subject i belongs to latent class j, Y_{imk} indicate whether subject i's mth measurement belongs to level k, and $\boldsymbol{\phi} = (\boldsymbol{\gamma}_{mj}, \boldsymbol{\alpha}_m, \boldsymbol{\beta})$ be the parameter in (3). If S_{ij} were directly observable, the complete-data log likelihood of (3) would be

$$\log \mathrm{L_c}(\boldsymbol{\phi}; \mathbf{Y}, \mathbf{S}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\{S_{ij}[\log \eta_j(\mathbf{x}_i^T\boldsymbol{\beta})]\} + \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{m=1}^{M}\sum_{k=1}^{K_m}\{S_{ij}Y_{imk}[\log p_{mkj}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m)]\}. \tag{12}$$

We introduce a new function

$$Q(\boldsymbol{\phi}|\boldsymbol{\phi}') = \mathrm{E}[\log \mathrm{L_c}(\boldsymbol{\phi}; \mathbf{Y}, \mathbf{S})|\mathbf{Y} = \mathbf{y}, \boldsymbol{\phi}', \mathbf{x}, \mathbf{z}], \tag{13}$$

which is the expected log likelihood function conditional on the observed data $\mathbf{y}$, $\mathbf{x}$, $\mathbf{z}$ and provisional estimates $\boldsymbol{\phi}'$. Then, the EM algorithm taking $\boldsymbol{\phi}^{(p)}$ to $\boldsymbol{\phi}^{(p+1)}$ is:

E-step: Compute $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(p)})$.

M-step: Find $\boldsymbol{\phi}$ which maximizes $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(p)})$.

Since there is no closed form solution for above maximization process, we use the one iteration Newton–Raphson method (Lange, 1995) to approximate the maximum values in the M-step (Appendix B). This single step approximation has been shown to have a convergence rate that is almost identical to the EM algorithm rate and hence saves time over repeatedly performing Newton's method. The E- and M-steps are alternated repeatedly until the difference in log likelihood $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y}) = \sum_{i=1}^{N} \log \Pr(\mathbf{Y}_i|\mathbf{x}_i, \mathbf{z}_i)$ between $\boldsymbol{\phi}^{(p+1)}$ and $\boldsymbol{\phi}^{(p)}$ is arbitrarily small (McLachlan and Krishnan, 1996).

To ensure reasonable convergence properties in practice, values to initialize the EM procedure for estimating latent class model parameters must be chosen with some care. One reasonable set of initial estimates for the $\boldsymbol{\gamma}_{mj}$ and $\boldsymbol{\alpha}_m$ may be obtained by fitting M separate polytomous logistic regressions for $(Y_{i1}, \mathbf{z}_{i1}), \ldots, (Y_{iM}, \mathbf{z}_{iM})$. To obtain initial estimates for $\boldsymbol{\beta}$, we first fit an LCA whose initial parameters are determined by dividing subjects into J groups according to the most common response patterns in the population. Then, we randomly assign each person i to a class $C_i \in \{1, \ldots, J\}$ with posterior probabilities of class membership $\{\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iJ}\}$ of the fitted LCA, where $\hat{\theta}_{ij} = \mathrm{E}(S_{ij}|\mathbf{Y}_i, \hat{\boldsymbol{\phi}})$. The coefficient estimates in the polytomous logistic regression of C_i versus $\mathbf{x}_i$ then give reasonable initial estimates of $\boldsymbol{\beta}$.

4.2. Variance Estimation

Since the EM algorithm is a method for ML estimation in incomplete-data problems, the observed Fisher information matrix based on the incomplete-data likelihood $\mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$ can be used to estimate standard errors of the parameter estimates conditioning on the number of classes. However, analytically evaluating the second-order derivatives of the incomplete-data log likelihood may be difficult, or at least tedious. Here, we consider methods that calculate the observed Fisher information of incomplete data within the EM framework. Louis (1982) showed that the observed information matrix of incomplete data can be computed in terms of the conditional moments of the first- and second-order partial derivatives of the complete-data log likelihood function introduced within the EM framework. We therefore implement Louis' approach for calculating the variance-covariance matrix of the parameter estimates. Details of variance estimation can be found in Appendix B.

4.3. Convergence of the Estimating Procedure

Implementing the EM algorithm to estimate parameters in finite-mixture models is typically time-consuming. In this section, we aim to investigate the convergence properties of the EM sequence under the proposed RLCA (3).

Let $\boldsymbol{\Phi} = \{(\boldsymbol{\gamma}_{mj}, \boldsymbol{\alpha}_m, \boldsymbol{\beta}); j = 1, \ldots, J; m = 1, \ldots, M\}$ be the parameter space of model (3). We assume that it is a finite subset of r-dimensional Euclidean space $\mathcal{R}^r$, where r is the number of parameters in (3). Consider settings in which the incomplete-data log likelihood, $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$, is bounded above for all $\boldsymbol{\phi}$ in $\boldsymbol{\Phi}$, and $Q(\boldsymbol{\phi}|\boldsymbol{\phi}')$ is continuous in both $\boldsymbol{\phi}, \boldsymbol{\phi}' \in \boldsymbol{\Phi}$. Then, it has been shown (Wu, 1983) that, for any EM sequence $\{\boldsymbol{\phi}^{(p)}\}_{p\geq 0}$, $\log \mathrm{L}(\boldsymbol{\phi}^{(p)}; \mathbf{Y})$ converges to $\log \mathrm{L}^* = \log \mathrm{L}(\boldsymbol{\phi}^*; \mathbf{Y})$ for some stationary point $\boldsymbol{\phi}^*$, i.e., a point $\boldsymbol{\phi}^* \in \boldsymbol{\Phi}$ such that $D^1_{\boldsymbol{\phi}} \log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*} = 0$, where $D^1_{\boldsymbol{\phi}}$ is the gradient operator with respect to $\boldsymbol{\phi}$. In our RLCA model (3), it is easy to verify that $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y}) \leq 0$ for all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ and that Q in (13) satisfies the continuity condition for all $\boldsymbol{\phi}, \boldsymbol{\phi}' \in \boldsymbol{\Phi}$. Therefore, any EM sequence of parameter estimates of RLCA (3) ends in a stationary point under the stopping criterion of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$ convergence. Since the purpose of the EM algorithm is to provide iterative computation of the maximum likelihood estimates of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$, convergence of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$ to stationary values is all we need. We therefore use the convergence of the incomplete-data log likelihood as a stopping criterion in section 4.2. The same criterion is also suggested by McLachlan and Krishnan (1996, pp. 22–23).

There is no guarantee that $\log \mathrm{L}^*$ is a (global or local) maximum of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$ over $\boldsymbol{\Phi}$. To decide whether the stationary value $\log \mathrm{L}^*$ is a local maximum, we can examine the observed Fisher information of the incomplete data, $\mathrm{I}(\boldsymbol{\phi}^*) = -D^2_{\boldsymbol{\phi}} \log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*}$, where $D^2_{\boldsymbol{\phi}}$ is the Hessian operator with respect to $\boldsymbol{\phi}$. If $\mathrm{I}(\boldsymbol{\phi}^*)$ is positive definite, then $\log \mathrm{L}^*$ corresponds to a local maximum of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$. If $\mathrm{I}(\boldsymbol{\phi}^*)$ is positive semi-definite, there is a probable lack of local identifiability or a boundary solution for γ_{mkj} at $\boldsymbol{\phi}^*$ (Formann, 1992). The parameters involved in the lack of local identifiability can be empirically identified from their extremely large asymptotic standard errors (Formann, 1992). The estimators $\hat{\gamma}_{mkj}$, which are the boundary solutions, tend to $\pm\infty$; as a consequence, one or more conditional probabilities tend to 1 or 0. Constraining conditional response probabilities appropriately can solve both identifiability and boundary problems. If $\mathrm{I}(\boldsymbol{\phi}^*)$ is indefinite or negative (semi-) definite, the solution corresponds to a saddle point or a (local) minimum of the incomplete data likelihood.

To establish global maxima, Wu (1983) proved that if the incomplete data likelihood $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$ is unimodal, then $\{\boldsymbol{\phi}^{(p)}\}_{p\geq 0}$ converges to the unique MLE of $\log \mathrm{L}(\boldsymbol{\phi}; \mathbf{Y})$. The unimodality condition on $\log L(\boldsymbol{\phi}; \mathbf{Y})$ generally does not hold in LCA and RLCA models. Rather, the log likelihood $\log L(\boldsymbol{\phi}; \mathbf{Y})$ often has several (local or global) maxima and stationary values, and the convergence to either type of value depends on the choice of starting point. We therefore recommend that several EM iterations be performed using different sets of starting points representative of the parameter space. If there appear to be multiple maxima, known scientific theory of the investigated questions may guide the choice between solutions with similar likelihoods. To obtain unimodality, one can also adjust the number of classes to prevent partitioning ambiguously major response patterns into distinct classes, or impose theoretically reasonable constraints on $\boldsymbol{\Phi}$.

5. Latent Class Modeling Software

A computer module to implement the proposed latent class model (3) is created using statistical package S-PLUS (Statistical Sciences, Inc., 1995) and programming language C. The module needs to be operated under the S-PLUS environment. It provides initial values for the estimation, parameter and variance estimates, model identifiability checking using both the proposed method (section 3) and the observed Fisher information matrix, the number of latent classes selection (Huang, 2004: in press), and graphical displays for model diagnosis.

Several computer programs are also available for estimating various types of latent class models. A web page created by John Uebersax (http://ourworld.compuserve.com/homepages/jsuebersax/index.htm) provides much useful information about currently available programs, which we will not repeat. Instead we will explicitly compare our model with the models underlying these existing statistical programs to help the reader to position the role of our proposed model as a latent variable modeling tool. We will describe eight existing programs for estimating latent class models. They fall into three categories, which we will detail in ascending order of capability.

The two programs in the first category are LLCA, a program for located latent class analysis (Uebersax, 1993) and WINMIRA, which can estimate latent class models, Rasch models, and Rasch mixture models (Rost, 1990, 1991). LLCA requires ordinal observed variables, while WINMIRA can analyze any kind of categorical measured variable. Neither of the programs in this first category is able to model covariate effects or perform identifiability checking. Both calculate Akaike's and Bayesian Information Criteria (AIC and BIC) (Akaike, 1987; Schwartz, 1978).

The second category of software includes MLLSA, LCAP, and LCAG, which formulate the latent class model in terms of loglinear modeling and use the modified LISREL approach to include categorical covariates (Clogg and Goodman, 1984; McCutcheon, 1987; Hagenaars, 1993). All three programs analyze the effect of categorical covariates on latent class and observed variables. MLLSA evaluates the Jacobian matrix at estimated parameters for checking identifiability, but does not calculate the AIC or BIC, while LCAP does not do identifiability checking but does provide the information criteria.

The third and most powerful category of software is made up of LEM, Latent GOLD, and Mplus. These programs are very flexible in the specification of model structure and can model latent class models, latent trait models, and a mixture of continuous and categorical observed-variables. For Mplus, categorical data must be scaled ordinally. All three programs model the effect of categorical or continuous covariates on latent class and observed variables. The programs in this category use the observed Fisher information matrix to do identifiability checking and all provide the AIC and BIC.

6. Example

To illustrate our model, we use data from the Salisbury Eye Evaluation project (SEE), a population-based, prospective study ($N = 2520$) of how vision affects older adults' functioning ability (West et al., 1997). Several studies have demonstrated that people aged more than 65 years report difficulty in performing their daily activities, and visual impairment is associated with difficulty in these activities (Rubin et al., 2001; Jette and Branch, 1985). The analysis reported here aims to describe the associations between ability in activities requiring distance vision and various visual impairment.

Several studies have analyzed the SEE data, using different statistical methods (Rubin et al., 2001; Huang et al., 2002). Particularly relevant here, Bandeen-Roche et al. (1999) fit a regression extension of latent class model that did not allow direct effects of covariates on item responses (henceforth, B-R). This paper fits a model (3) for self-reported visual disability that includes direct covariate effects on item responses. In the following analysis, we highlight the comparison of our model via the B-R model, which provides a unique opportunity to look at what insight gains from our approach.

6.1. Data

In the SEE project, vision-related disability was assessed using the Activities of Daily Vision Scale (ADVS), a standardized instrument that has been described elsewhere (Mangione et al., 1992; Valbuena et al., 1999). Disability related to distance vision was determined via self-reports of difficulty in five tasks comprising the "far vision" subscale of the ADVS: reading street signs at night (signs-night), reading street signs in daylight (signs-day), walking down steps during daylight (steps-day), walking down steps in dim light (steps-dim), and watching TV (watch TV). Here, we measured difficulty as a binary indicator ($1 =$ having difficulty; $2 =$ no difficulty) on signs-day, steps-day, steps-dim and watch TV, and as a three-level categorical indicator ($1 =$ extreme or moderate difficulty; $2 =$ a little difficulty; $3 =$ no difficulty) on signs-night. The frequency distributions of far vision subscale items of the whole study population are shown in Figure 1 ($N = 2520$); all are severely skewed, with most participants reporting no difficulty.

The variables we used to measure visual impairment have been described elsewhere (Rubin et al., 1997). In brief, these include: (a) visual acuity, which measures the ability to resolve images clearly; (b) contrast sensitivity of better eye, which measures the ability to distinguish shading; (c) glare sensitivity, which measures the ability to cope with glare in distinguishing shading; (d) stereoacuity, which measures depth perception; and (e) central visual field, which measures range of peripheral vision as well as the presence of blind spots. In this analysis, all the measures were re-scaled so that a higher score indicated worse vision.

6.2. Model Fitting, Model Identifiability, and Diagnosis

Because there was no one adequate measure of ability in far vision functioning, the SEE project used five self-reported visual disability measurements as quantities that imperfectly determined this unobserved, theoretical object. Analyzing these data posed two challenges: First, five measurements were designed to jointly describe far vision functioning. These measurements needed to be combined appropriately to derive summary statements about far vision functioning. Second, a self-reported questionnaire was used to quantify visual disability. This method was advantageous because the questionnaire was easy to administer; however, a potential drawback was that individual variation in defining levels of difficulty may yield a response with substantial error.

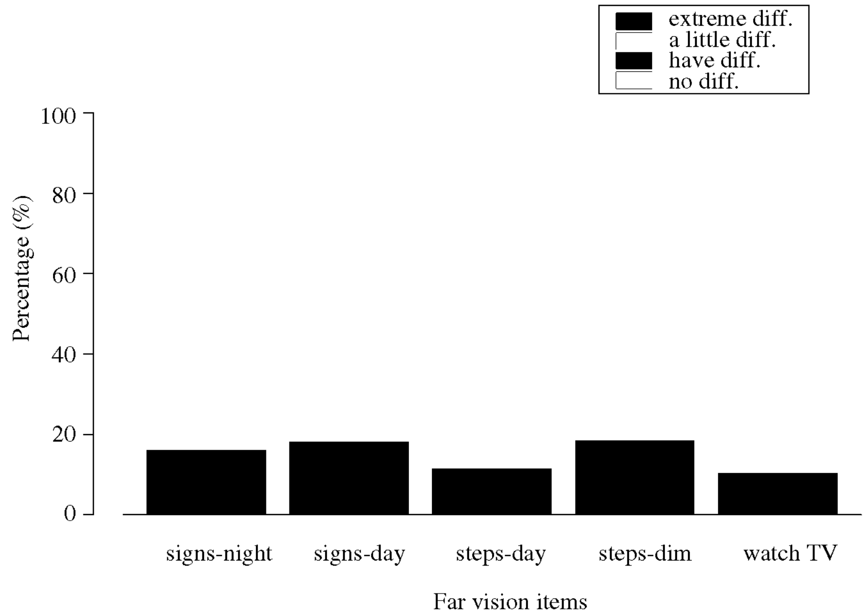

FIGURE 1.
Frequency distributions of far vision difficulty items: SEE project.

To overcome these challenges, we proposed to use the regression extension of latent class model (3) for self-reported visual disability. We modeled latent class memberships as depending on visual impairment and the number of reported comorbid diseases. Comorbidities included arthritis, broken hip, cardiovascular disease, hypertension, diabetes, emphysema, asthma, Parkinson's disease, cancer, and stroke. This would help us obtain the effect of visual impairment on the underlying far vision functioning. The following personal characteristics were identified or hypothesized as extraneous influences (other than the underlying far vision functioning) that could affect individual's reporting in the questionnaire: age at clinic exam, cognitive status assessed with the MMSE score (Folstein, Folstein, & McHugh, 1975), years of education, gender, race, and General Health Questionnaire (GHQ) depression subscale score (Goldberg, 1972). We modeled reporting of the measured indicators themselves as varying with these personal characteristics and hopefully could yield a more accurate latent class. It is arguable that age, MMSE, and GHQ score seem good predictors of the underlying latent class. Further analyses that allow these variables to affect both class membership and measured indicators themselves will be performed to judge the possibility.

The B-R model did not allow direct covariate effects on measured variables, but included all personal characteristics, vision and disease variables in predicting latent class memberships. Bandeen-Roche et al. (1999) justified that four classes were adequate to describe the SEE far vision data in their model, although they opted for a five-class solution for hypothesis-based reasons. For the proposed RLCA (3), AIC's under three, four, and five latent classes are 6153.87, 6061.47, and 6064.29, respectively; BIC's are 6521.28, 6499.11, and 6572.18, respectively. Based on above model selection and B-R's results, our analysis assumed four classes.

In our analysis, LCA and RLCA (3) models were fit to the sub-sample of participants who rated each far vision item and also had no missing covariates ($N = 1641$). To check the local identifiability of the two models at estimated values, we first saw that the number of unique parameters in the saturated LCA ($= 47$) is greater than the number of unique model parameters in LCA ($= 27-3$). Conditions (ii) and (ii$'$) in Proposition 1 and Theorem 1 are clearly satisfied. The

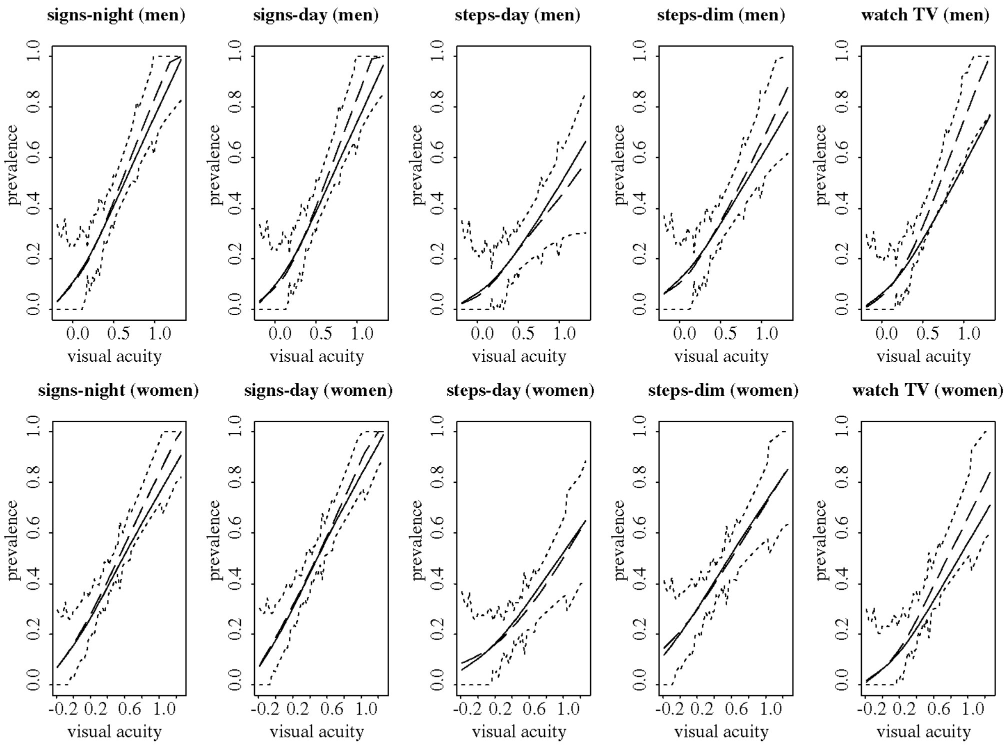

FIGURE 2.
Observed and fitted task difficulty prevalences, by gender and visual acuity: SEE project. In each plot, the solid line represents the predicted curve based on the RLCA model (3), the dashed line represents the observed curve, and dotted lines are 95% confidence bands of the observed curve.

matrices $\hat{\boldsymbol{\psi}} = [\hat{\boldsymbol{\psi}}_1, \ldots, \hat{\boldsymbol{\psi}}_J]$ and $\hat{\boldsymbol{\tau}} = [\hat{\boldsymbol{\tau}}_1, \ldots, \hat{\boldsymbol{\tau}}_J]$ both have full column ranks: The eigenvalues of $(\hat{\boldsymbol{\psi}}^T \hat{\boldsymbol{\psi}})$ are 0.283, 0.122, 0.07, 0.031; and the eigenvalues of $(\hat{\boldsymbol{\tau}}^T \hat{\boldsymbol{\tau}})$ are 0.236, 0.16, 0.055, 0.031. Also, all the design matrices have full column ranks. Therefore, the RLCA (3) are locally identifiable at the corresponding estimates.

To compare the fit of our chosen RLCA model with the fit of the B-R analysis, we reproduced Figure 4 of Bandeen-Roche et al. (1999) for our analysis (Figure 2). The resulting display plots the proportions reporting difficulty in each of the five self-reported vision activities and those predicted by the RLCA model (3) as a function of gender and visual impairment. The observed (dashed line) and predicted (solid line) proportions agreed closely for most items. Although our RLCA model under-predicted the proportions reporting difficulty in watching TV among men with substantial acuity loss, it has greatly improved upon the under-prediction that resulted in the B-R fit.

6.3. Analysis Results

Table 1 displays the estimated LCA conditional probabilities $\hat{p}_{mkj}$ and latent prevalences $\hat{\eta}_j$. Class 1 was an able group who rarely reported any difficulty; class 2 appeared to represent a group who frequently reported difficulties *reading signs* in both daylight and at night but rarely reported other difficulties; class 3 was a group who frequently reported problems *reading signs at night* and difficulties in *descending steps*, but less often reported difficulty reading signs in daylight or watching TV; and class 4 was a severely far vision disabled population. The estimated

TABLE 1.
Estimated conditional probabilities and latent prevalences from LCA (1, 2) and RLCA (3) for far vision difficulty: SEE project

Self-Reported		Class 1 (none)		Class 2 (signs)		Class 3 (steps)		Class 4 (severe)	
Difficulty	Level	LCA	RLCA	LCA	RLCA	LCA	RLCA	LCA	RLCA
signs-night	extreme diff.	0.030	0.012	0.508	0.424	0.168	0.260	1*	1*
	a little diff.	0.197	0.166	0.474	0.519	0.606	0.517	0*	0*
	no diff.	0.773	0.822	0.018	0.056	0.226	0.223	0*	0*
signs-day	have diff.	0.007	0.002	0.648	0.496	0.240	0.293	1*	1*
	no diff.	0.993	0.998	0.352	0.504	0.760	0.707	0*	0*
steps-day	have diff.	0.001	0.002	0.001	0.001	0.596	0.743	0.780	0.860
	no diff.	0.999	0.998	0.999	0.999	0.404	0.257	0.220	0.140
steps-dim	have diff.	0.021	0.021	0.276	0.235	0.836	0.830	0.912	0.877
	no diff.	0.979	0.979	0.724	0.765	0.164	0.170	0.088	0.123
watch TV	have diff.	0.011	0.007	0.184	0.136	0.171	0.158	0.573	0.808
	no diff.	0.989	0.993	0.816	0.864	0.829	0.842	0.427	0.192
latent prevalence		0.736	0.700	0.120	0.177	0.095	0.090	0.049	0.033

*Values equal to 1 or 0 were pre-set to uniquely identify the model.

latent prevalences show that 73% of participants rarely reported any difficulty (class 1), roughly 10% of participants were in each class 2 and 3, and only 5% of participants reported severe far vision difficulty (class 4).

The likelihood ratio test (LRT) comparing RLCA (3) with LCA indicated that the addition of covariates significantly improved the model fit (LRT = 448.122, df = 54). Table 1 displays the RLCA conditional probabilities evaluated at the sample means of the incorporated covariates $\hat{p}^*_{mkj} = p_{mkj}(\hat{\boldsymbol{\gamma}}_{mj} + \bar{\mathbf{z}}_m^T \hat{\boldsymbol{\alpha}}_m)$, $\bar{\mathbf{z}}_m = \frac{1}{N}\sum_{i=1}^N \mathbf{z}_{im}$, and the sample averages of the RLCA prevalences $\hat{\eta}^*_j = \frac{1}{N}\sum_{i=1}^N \eta_j(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})$. Here, we evaluated the conditional probabilities at the sample averages of covariates to reflect the underlying latent structure adjusting for possible confounding. As expected, the latent class prevalence estimates were quite similar across approaches. LCA and RLCA estimates of conditional probabilities were similar in class 1; they were different but "nested" in classes 2, 3, and 4. RLCA analysis estimated a more modest percentage of reporting difficulty in *reading signs in daylight* for class 2 members, a higher percentage of reporting difficulty in *walking down steps during daylight* for class 3 members, and a higher percentage of reporting difficulty in *watching TV* for class 4 members. The difference in conditional probabilities comparing LCA and RLCA suggests differential reporting by personal characteristics.

The B-R analysis fit a five-class model. Its class compositions (Table 4 of Bandeen-Roche et al., 1999) had a basic structure similar to the class compositions of LCA, with a new class that included participants who had difficulty in *reading signs at night* but rarely reported other difficulties and were originally classified into class 1 or 2 under the LCA model.

Table 2 contains the association estimation between latent class membership and risk factors. The odds ratios are obtained by exponential transformation of regression coefficients in equation (4) [i.e., $\exp(\beta_{pj})$]. Summarizing, we derived several important findings: (a) Different impairments independently predicted far vision disability. (b) Visual acuity was not significantly associated with steps disability but stereoacuity was only significantly associated with steps disability. This is consistent with our previous findings and the theory underlying the vi-

TABLE 2.
Latent prevalence regression from RLCA (3) for the relationship between underlying visual ability and risk factors: SEE project

Disease and Vision Variables	Comparison	OR*†	95% CI
number of comorbid diseases (1 disease)	signs vs. none	1.064	0.986, 1.148
	steps vs. none	**1.153**	1.047, 1.269
	severe vs. none	**1.422**	1.197, 1.690
visual acuity (0.3 logMAR)	signs vs. none	**3.055**	2.335, 3.997
	steps vs. none	1.063	0.715, 1.580
	severe vs. none	**4.534**	2.921, 7.039
contrast sensitivity (6 letters)	signs vs. none	**1.302**	1.005, 1.687
	steps vs. none	**1.597**	1.135, 2.246
	severe vs. none	**1.855**	1.132, 3.040
glare sensitivity (6 letters)	signs vs. none	**1.696**	1.279, 2.248
	steps vs. none	**1.645**	1.116, 2.426
	severe vs. none	**2.130**	1.142, 3.972
stereoacuity (0.3 log arcsec)	signs vs. none	1.010	0.935, 1.091
	steps vs. none	**1.110**	1.003, 1.228
	severe vs. none	1.173	0.976, 1.411
visual field ($\sqrt{2}$ letters)	signs vs. none	**1.162**	1.034, 1.305
	steps vs. none	**1.183**	1.016, 1.378
	severe vs. none	**1.620**	1.255, 2.090

*Values in bold are significantly different from 1 at the 0.05 level.
†Interpretation = odds ratios with unit identified in parentheses, and "odds" specific to the two classes under comparison category.

sion measures (Bandeen-Roche et al., 1999; Rubin et al., 2001; Valbuena et al., 1999). (c) In the B-R analysis, various visual impairments were shown to be significantly associated with signs or steps disabilities (Bandeen-Roche et al., 1999, Table 5). However, none of the impairments were significantly associated with the severe visual disability, because of large variances of estimators. RLCA (3) showed not only significant associations with signs or steps disabilities, but also a significant association with the severe disability. A more "accurate" ("consistent") underlying disability was created after adjusting for characteristics that determine responses other than underlying classes.

The direct relationships between self-reported difficulty and confounding variables (i.e., exponential transformation of α_{qmk}'s in the equation (5)) are shown in Table 3. Results can be summarized as:

1. People who had higher GHQ depression scores were more likely to report difficulty in each far vision activity.
2. Women and highly educated people were more likely to report difficulty in performing each far vision activity except watching TV.
3. Less cognitively intact people (lower MMSE scores) were less likely to report difficulty reading signs.
4. Older persons were more likely to report steps and reading signs at night difficulty.
5. Race was not significantly associated with differential-reporting.

TABLE 3.
Conditional probability regression from RLCA (3) for the direct relationship between self-reported far vision difficulty and confounding variables: SEE project

Confounding Variables	Self-Reported Difficulty	Difficulty Level	OR*†	95% CI
age	signs-night	extreme diff. vs no diff.	**1.058**	1.015, 1.102
		a little diff. vs no diff.	1.020	0.992, 1.049
	signs-day	have diff. vs no diff.	1.006	0.966, 1.047
	steps-day	have diff. vs no diff.	**1.142**	1.070, 1.220
	steps-dim	have diff. vs no diff.	**1.072**	1.033, 1.113
	watch TV	have diff. vs no diff.	0.987	0.941, 1.035
MMSE score	signs-night	extreme diff. vs no diff.	**1.117**	1.014, 1.232
		a little diff. vs no diff.	**1.101**	1.031, 1.175
	signs-day	have diff. vs no diff.	**1.109**	1.007, 1.220
	steps-day	have diff. vs no diff.	1.015	0.890, 1.159
	steps-dim	have diff. vs no diff.	0.996	0.915, 1.085
	watch TV	have diff. vs no diff.	0.944	0.853, 1.044
years of education	signs-night	extreme diff. vs no diff.	**1.139**	1.065, 1.218
		a little diff. vs no diff.	**1.047**	1.001, 1.094
	signs-day	have diff. vs no diff.	**1.096**	1.026, 1.171
	steps-day	have diff. vs no diff.	**1.101**	1.000, 1.213
	steps-dim	have diff. vs no diff.	**1.086**	1.021, 1.155
	watch TV	have diff. vs no diff.	1.031	0.958, 1.111
female	signs-night	extreme diff. vs no diff.	**5.344**	3.520, 8.112
		a little diff. vs no diff.	**3.450**	2.614, 4.554
	signs-day	have diff. vs no diff.	**4.054**	2.716, 6.051
	steps-day	have diff. vs no diff.	**8.401**	4.477, 15.763
	steps-dim	have diff. vs no diff.	**4.055**	2.738, 6.005
	watch TV	have diff. vs no diff.	1.154	0.731, 1.821
African-American	signs-night	extreme diff. vs no diff.	0.972	0.587, 1.609
		a little diff. vs no diff.	1.376	0.999, 1.880
	signs-day	have diff. vs no diff.	0.818	0.500, 1.336
	steps-day	have diff. vs no diff.	0.927	0.456, 1.887
	steps-dim	have diff. vs no diff.	0.752	0.476, 1.187
	watch TV	have diff. vs no diff.	1.700	0.990, 2.919
GHQ score	signs-night	extreme diff. vs no diff.	**1.891**	1.492, 2.397
		a little diff. vs no diff.	**1.322**	1.089, 1.605
	signs-day	have diff. vs no diff.	**2.217**	1.709, 2.875
	steps-day	have diff. vs no diff.	**2.439**	1.851, 3.213
	steps-dim	have diff. vs no diff.	**2.012**	1.647, 2.458
	watch TV	have diff. vs no diff.	**1.622**	1.322, 1.989

*Values in bold are significantly different from 1 at the 0.05 level.
†Interpretation = odds ratios with unit equal to one, and "odds" specific two levels under difficulty level category.

It is worth noticing that age and MMSE scores were highly associated with steps and signs variables, respectively. A further analysis showed that participants who were not classified as having steps disability based on RLCA (3) but were classified as having steps disability based on LCA were older than the general study population (mean ages: 74.9 versus 72.9, p-value $= 0.01$); participants who were not classified as the able group based on RLCA (3) but were classified as the able group based on LCA had lower MMSE scores than the general study population (mean MMSE scores: 26.5 versus 27.5, p-value < 0.001). An LCA model that did not adjust for confounding characteristics might group older people into the steps disability class and group people with low MMSE score into the able class too frequently. This plausibly underlies the different response patterns for classes 2 and 3 from the LCA and RLCA model.

The models reported so far only allowed *age* effects on conditional probabilities. A model allowing *age* to affect both conditional probabilities and class membership probabilities showed no significant improvement of the model fit (LRT comparing the model with *age* in conditional probabilities only versus the model with *age* in both conditional probabilities and class membership probabilities = 2.493, df = 3). This result has an important implication: Age is not significantly associated with far vision disability once visual impairments were taken into account. Similar analyses for MMSE and GHQ score were also performed. No statistically significant improvement was found.

7. Discussion

Latent class analysis provides a probabilistic model that links observations to idealized concepts that cannot be directly measured. It is thereby able to account for association among the observed items. This paper has studied a regression extension that incorporates two sets of covariates: risk factors that are hypothesized to influence the underlying latent classes, and covariates that may influence observed items directly, hence possibly causing misclassification of the class membership. We provided theoretical justification and systematic methods for model identifiability and parameter estimation.

In the example provided in the section 6, five measured indicators can be divided into three categories: SIGN activities (reading street signs at night and reading street signs in daylight), STEP activities (walking down steps during daylight and walking down steps in dim light), and watching TV. In contrast with the proposed latent class (LC) model with one latent variable containing four classes, an LC model with three dichotomous latent variables might be more appropriate. In fact, an unconstrained LC model with three dichotomous latent variables can be reparameterized as a single-latent-variable LC model with $2^3 = 8$ classes (Magidson and Vermunt, 2001). Based on prior knowledge and reasonable model assumptions, we can fix some parameters in an LC model with several latent variables to increase the model's degree of freedom, while maintaining the capability of a single-latent-variable model with eight classes. As discussed in Hagenaars (1993) and Magidson and Vermunt (2001), constrained LC models with several latent variables might provide a more parsimonious model, fit the data better, and give results that are easier to interpret than the corresponding single-latent-variable model. An extension of the proposed LC model (3) allowing several latent variables and covariates effects on them will greatly increase the flexibility of modeling and provide a useful alternative of describing the underlying structure.

Proposition 1 and Theorem 1 demonstrate that local identifiability for regression extension of latent class models is determined by the response distribution within each class. The more similar the p_{mkj}'s (or γ_{mkj}'s) for different classes, the weaker the local identifiability. This is due to approaching violation of conditions (iii) and (iii$'$). This fact has the important consequence of limiting the number of classes that can be fit: If the number is too large, one risks creating classes with similar response distributions. Moreover, this explains why it is desirable to constrain direct covariate effects on indicators to be equal across classes; the alternative allows the p_{mkj}'s to telescope toward one another for certain covariate values.

Incorporating covariates can sometimes make an otherwise nonidentified LCA model identified. For example, Goodman (1974) analyzed the data of Table 1 in his paper using a three-class

LCA model with four dichotomous indicators. He showed that this was not identifiable because the 15×14 Jacobian matrix had rank 13. Consider an RLCA model that includes covariate "gender" in predicting both latent prevalences and conditional probabilities of the above LCA model. Notice that this RLCA model has $2 \times (2^4 - 1)$ distinct response patterns and 20 unknown parameters. If half the participants who have the same response pattern are females and the other half are males, the Jacobian matrix of the RLCA model has rank 19. However, if females are more likely to give negative answers in items 2 and 3 than males, the Jacobian matrix of the RLCA model then has full column rank. The former gender covariate results in nondifferential conditional probabilities between females and males, while the latter implies differential measurement conditions. Detailed characterization of covariate structures that improve model identifiability is useful in building identifiable RLCA models.

The number of classes is usually pre-selected in practice, either theoretically or empirically. When prior scientific knowledge does not provide an appropriate choice of the number of classes, choosing the number of classes becomes an analytic challenge. Standard practice is to fix the number at the lowest number of classes that yields acceptable fit based on either the likelihood ratio goodness of fit test (Goodman, 1974; Formann, 1992), AIC, or BIC. One common feature of the above methods is that they all must fit the model repeatedly under different numbers of classes. Huang (2004: in press) has proposed a new selection process that was motivated by an analogous method used in factor analysis and does not require repeated fitting. Summarizing, his proposed method calculates the sample correlation matrix of residuals from fitting $\mathbf{Y}_i$ on $\mathbf{z}_i$, then sets the number of classes equal to one plus the number of eigenvalues of the sample correlation matrix of residuals that are greater than or equal to one.

Missing item responses are common in medical studies that generate multiple responses. The standard practice of restricting analysis to persons with complete data may bias findings. If data are missing at random (i.e., the missing mechanism solely depends on subject's observed data; Little and Rubin, 1987), the proposed RLCA can be easily modified to describe all complete and incomplete item responses without additional modeling (Weiner, 1998: unpublished master thesis, Department of Biostatistics, the Johns Hopkins University). When outcomes are subject to nonignorable missing (i.e., the nonresponse is related to values of the missing variables), one needs to construct a model that correctly represents the missing mechanism. Baker and Laird (1988) developed a regression model for categorical responses when missing data are nonignorable. Under the assumption that only outcomes are missing, they used two different regressions to describe the model: A marginal regression for outcomes on covariates, and a nonresponse regression for missing indicators on outcomes and covariates. This formula may provide a workable approach for RLCA modeling (3) when missingness is not random.

Many statisticians are skeptical of latent variable models despite a long tradition of application in the social sciences. A predominant concern is that potential nonidentifiability of latent variable models is a well-known problem. Without identifiability, standard inferences are meaningless. In this paper, we have focused on providing a locally identifiable regression extension of latent class model. To reach global identifiability, appropriate constraints that incorporate scientific knowledge and theory are needed. A second concern is that latent variable model-based scientific findings are likely to be driven by the statistical assumptions rather than by the data. We acknowledge this danger, but we maintain that it can be minimized by diagnosing *whether* and *how* our models fit or may fail to appropriately describe a given dataset. In summary, regression extension of latent class models give well-summarized inferences on theory underlying the choice of multiple indicators and their relationships with covariates of interest in a single step. When model assumptions are at odds with the observed data, a great deal can be learned from identifying the aspects of one's theory that are not borne out in analysis.

Appendix A: Proofs

For simplicity, we assume there are no prefixed conditional probabilities and $K_1 = \cdots = K_M = K$ (i.e., the levels of items are all the same) in the following proofs. For the constrained model, the proofs are based on free parameters only. Extension to allow the levels being different is straightforward.

Proof of Proposition 1. Let $\boldsymbol{\phi}$ denote true parameters of a given LCA. As discussed in the paper, we only need to show that (iii) is equivalent to having full column rank of the Jacobian of an LCA model. Let $\mathbf{A}$ be the LCA's differential matrix w.r.t. $\boldsymbol{\phi}$. Then $\mathbf{A}$ can be partitioned into sub-matrices

$$\mathbf{A} = [\mathbf{A}_1|\cdots|\mathbf{A}_{J-1}|\underbrace{\mathbf{B}_{111}|\cdots|\mathbf{B}_{1(K-1)1}}_{m=1;j=1}|\cdots|\underbrace{\mathbf{B}_{M1J}|\cdots|\mathbf{B}_{M(K-1)J}}_{m=M;j=J}],$$

where $\mathbf{A}_j$ is the $(K^M - 1) \times 1$ vector of the partial derivative of the likelihood function w.r.t. η_j with the hth element equal to

$$\frac{\partial \Pr(\mathbf{Y} = \mathbf{y}_h)}{\partial \eta_j} = \psi_{hj} - \psi_{hJ}. \tag{A1}$$

$\mathbf{B}_{mkj}$ is the $(K^M - 1) \times 1$ vector of the partial derivative of the likelihood function w.r.t. p_{mkj} with the hth element equal to

$$\frac{\partial \Pr(\mathbf{Y} = \mathbf{y}_h)}{\partial p_{mkj}} = \eta_j \psi_{hj} \left(\frac{y_{hmk}}{p_{mkj}} - \frac{y_{hmK}}{p_{mKj}} \right), \tag{A2}$$

where $y_{hmk} = 1$ if $y_{hm} = k$; 0 if $y_{hm} \neq k$, $m = 1, \ldots, M$, $k = 1, \ldots, K$.

To prove $\mathbf{A}$ is of full column rank, we need to show

$$\sum_{j=1}^{J-1} [a_j(\boldsymbol{\psi}_j - \boldsymbol{\psi}_J)] + \sum_{m=1}^{M} \sum_{k=1}^{K-1} \sum_{j=1}^{J} \left\{ b_{mkj} \eta_j \left[\left(\frac{\mathbf{y}_{mk}}{p_{mkj}} - \frac{\mathbf{y}_{mK}}{p_{mKj}} \right) \# \boldsymbol{\psi}_j \right] \right\} = 0$$
$$\Leftrightarrow a_j = 0\ \forall j;\ b_{mkj} = 0\ \forall m, k, j, \tag{A3}$$

where $\mathbf{y}_{mk}$ is the $(K^M - 1) \times 1$ vector of all possible y_{hmk} over h, and # denotes elementwise multiplication. The left-hand side of (A.3) can be written as

$$\sum_{j=1}^{J} \{\boldsymbol{\psi}_j \# \mathbf{P}_j\} \mathbf{a}_j^* = \sum_{j=1}^{J} \{\mathbf{P}_j \mathbf{a}_j^*\} \# \boldsymbol{\psi}_j. \tag{A4}$$

Here, $\mathbf{P}_j = [\mathbf{1}, \mathbf{P}_{1j}, \mathbf{P}_{2j}, \ldots, \mathbf{P}_{Mj}]$ is a $(K^M - 1) \times (M(K-1)+1)$ matrix with $\mathbf{1}$ as a $(K^M - 1) \times 1$ vector of 1 and

$$\mathbf{P}_{mj} = \left[\left(\frac{\mathbf{y}_{m1}}{p_{m1j}} - \frac{\mathbf{y}_{mK}}{p_{mKj}} \right), \ldots, \left(\frac{\mathbf{y}_{m(K-1)}}{p_{m(K-1)j}} - \frac{\mathbf{y}_{mK}}{p_{mKj}} \right) \right], \quad m = 1, \ldots, M.$$

Notice that $\mathbf{P}_j$ can be expressed as a simple matrix with elements $(1/p_{mkj})$'s. If the response pattern of $(y_{h1}, y_{h2}, \ldots, y_{hM})$ is $(1, 1, \ldots, 1)$, $(1, 1, \ldots, 2)$, …, $(1, 1, \ldots, K)$, …, $(K, K, \ldots, K)$, then

$$\mathbf{P}_j = \begin{bmatrix} 1 & \frac{1}{p_{11j}} & 0 & \cdots & 0 & \cdots & \frac{1}{p_{M1j}} & 0 & \cdots & 0 \\ 1 & \frac{1}{p_{11j}} & 0 & \cdots & 0 & \cdots & 0 & \frac{1}{p_{M2j}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \frac{1}{p_{11j}} & 0 & \cdots & 0 & \cdots & \frac{-1}{p_{MKj}} & \frac{-1}{p_{MKj}} & \cdots & \frac{-1}{p_{MKj}} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \frac{-1}{p_{1Kj}} & \frac{-1}{p_{1Kj}} & \cdots & \frac{-1}{p_{1Kj}} & \cdots & 0 & 0 & \cdots & \frac{1}{p_{M(K-1)j}} \end{bmatrix}.$$

$$\mathbf{a}_j^* = \left[a_j, \overbrace{b_{11j}\eta_j, \ldots, b_{1(K-1)j}\eta_j}^{m=1}, \ldots, \overbrace{b_{M1j}\eta_j, \ldots, b_{M(K-1)j}\eta_j}^{m=M} \right]^T, \quad j = 1, \ldots, (J-1),$$

and

$$\mathbf{a}_J^* = \left[-\sum_{j=1}^{J-1} a_j, \overbrace{b_{11J}\eta_J, \ldots, b_{1(K-1)J}\eta_J}^{m=1}, \ldots, \overbrace{b_{M1J}\eta_J, \ldots, b_{M(K-1)J}\eta_J}^{m=M} \right]^T.$$

If $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_J$ are linearly independent, then (A.4) $= 0 \Leftrightarrow (\mathbf{P}_1\mathbf{a}_1^*) = \cdots = (\mathbf{P}_J\mathbf{a}_J^*) = 0$. Since $p_{mkj} > 0$ for all m, k, j plus the structure of the matrix, $\mathbf{P}_j$ $j = 1, \ldots, J$ has full column rank. Since $\eta_j > 0\ \forall j$, $(\mathbf{P}_1\mathbf{a}_1^*) = \cdots = (\mathbf{P}_J\mathbf{a}_J^*) = 0 \Leftrightarrow \mathbf{a}_1^* = \cdots = \mathbf{a}_J^* = 0 \Leftrightarrow a_1 = \cdots = a_{J-1} = b_{111} = \cdots = b_{M(K-1)J} = 0$. Therefore, if (iii) holds, $\mathbf{A}$ is of full column rank. Conversely, suppose $\mathbf{A}$ has full column rank, $(\mathbf{P}_1\mathbf{a}_1^*) = \cdots = (\mathbf{P}_J\mathbf{a}_J^*) = 0$. Thus, $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_J$ are linearly independent.

Proof of Theorem 1. We first utilize the following two propositions.

Proposition 2. For an LCA model with

$$p_{mkj}^0 = p_{mkj}(\boldsymbol{\gamma}_{mj}^0 + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m^0) = \frac{\exp(\gamma_{mkj}^0 + \alpha_{1mk}^0 z_{im1} + \cdots + \alpha_{Lmk}^0 z_{imL})}{1 + \sum_{s=1}^{K-1}\exp(\gamma_{msj}^0 + \alpha_{1ms}^0 z_{im1} + \cdots + \alpha_{Lms}^0 z_{imL})}$$

and

$$\eta_j^0 = \eta_j(\mathbf{x}_i^T\boldsymbol{\beta}^0) = \frac{\exp(\beta_{0j}^0 + \beta_{1j}^0 x_{i1} + \cdots + \beta_{Pj}^0 x_{iP})}{1 + \sum_{l=1}^{J-1}\exp(\beta_{0l}^0 + \beta_{1l}^0 x_{i1} + \cdots + \beta_{Pl}^0 x_{iP})}$$

for some fixed $\boldsymbol{\gamma}_{mj}^0$, $\boldsymbol{\alpha}_m^0$ and $\boldsymbol{\beta}^0$, the model is locally identifiable at $(\mathbf{p}^0, \boldsymbol{\eta}^0)$ if the following assumptions hold:

(a) $K^M > JM(K-1) + J$;
(b) all $\boldsymbol{\alpha}^0$s, $\boldsymbol{\beta}^0$s, $\mathbf{x}$s and $\mathbf{z}$s are finite; and
(c) $\boldsymbol{\tau}_1^0, \ldots, \boldsymbol{\tau}_J^0$ with γ_{mkj} in $\boldsymbol{\tau}_j$ evaluated at γ_{mkj}^0 are linearly independent, where $\boldsymbol{\tau}_j$ is defined as in Theorem 1.

Proof of Proposition 2. Since assumptions (i) and (ii) in Proposition 1 are true under the above model, we only need to show that (c) implies (iii) in Proposition 1. Let $\boldsymbol{\psi}_j^0$ be a $(K^M-1)\times 1$ vector with hth element

$$\psi_{hj}^0 = \Pr(\mathbf{Y}_i = \mathbf{y}_h | S_i = j, \mathbf{z}_i) = \prod_{m=1}^{M} p_{my_{hm}j}^0$$
$$= \prod_{m=1}^{M} \frac{\exp(\gamma_{my_{hm}j}^0 + \alpha_{1my_{hm}}^0 z_{im1} + \cdots + \alpha_{Lmy_{hm}}^0 z_{imL})}{1 + \sum_{s=1}^{K-1} \exp(\gamma_{msj}^0 + \alpha_{1ms}^0 z_{im1} + \cdots + \alpha_{Lms}^0 z_{imL})}.$$

Therefore,

$$\sum_{j=1}^{J} a_j \boldsymbol{\psi}_j^0 = 0 \Leftrightarrow \sum_{j=1}^{J} \{\boldsymbol{\tau}_j^0 \# \mathbf{Z}_i^*\} a_j^* = 0 \Leftrightarrow \left[\sum_{j=1}^{J} a_j^* \boldsymbol{\tau}_j^0\right] \# \mathbf{Z}_i^* = 0,$$

where

$$\mathbf{Z}_i^* = \begin{bmatrix} \prod_{m=1}^{M} \exp(\alpha_{1my_{1m}}^0 z_{im1} + \cdots + \alpha_{Lmy_{1m}}^0 z_{imL}) \\ \vdots \\ \prod_{m=1}^{M} \exp(\alpha_{1my_{(K^M-1)m}}^0 z_{im1} + \cdots + \alpha_{Lmy_{(K^M-1)m}}^0 z_{imL}) \end{bmatrix},$$

and

$$a_j^* = a_j \prod_{m=1}^{M} \frac{1 + \sum_{s=1}^{K-1} \exp(\gamma_{msj}^0)}{1 + \sum_{s=1}^{K-1} \exp(\gamma_{msj}^0 + \alpha_{1ms}^0 z_{im1} + \cdots + \alpha_{Lms}^0 z_{imL})}.$$

Since $\boldsymbol{\tau}_1^0, \ldots, \boldsymbol{\tau}_J^0$ are independent, $a_1^* = \cdots = a_J^* = 0 \Rightarrow a_1 = \cdots = a_J = 0$. Hence the proof.

Proposition 3. Suppose a given latent class analysis model is locally identifiable at $(\mathbf{p}, \boldsymbol{\eta})$. Then, it is locally identifiable in the transformed parameters $(\boldsymbol{\epsilon}, \boldsymbol{\omega})$, where $\boldsymbol{\epsilon} = (\epsilon_{111}, \ldots, \epsilon_{1(K-1)1}, \ldots, \epsilon_{M1J}, \ldots, \epsilon_{M(K-1)J})$ is defined as $p_{mkj} = \exp(\epsilon_{mkj})/[1 + \sum_{s=1}^{K-1} \exp(\epsilon_{msj})]$, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{J-1})$ as $\eta_j = \exp(\omega_j)/[1 + \sum_{l=1}^{J-1} \exp(\omega_l)]$.

Proof of Proposition 3. Since a given latent class analysis model is locally identifiable at $(\mathbf{p}, \boldsymbol{\eta})$,

$$\sum_{j=1}^{J-1} [a_j(\boldsymbol{\psi}_j - \boldsymbol{\psi}_J)] + \sum_{m=1}^{M} \sum_{k=1}^{K-1} \sum_{j=1}^{J} \left\{ b_{mkj} \eta_j \left[\left(\frac{\mathbf{y}_{mk}}{p_{mkj}} - \frac{\mathbf{y}_{mK}}{p_{mKj}} \right) \# \boldsymbol{\psi}_j \right] \right\} = 0$$
$$\Leftrightarrow a_j = 0 \,\forall j;\, b_{mkj} = 0 \,\forall m, k, j. \tag{A5}$$

To prove this latent class analysis model is also locally identifiable at $(\boldsymbol{\epsilon}, \boldsymbol{\omega})$, we need to show

$$\sum_{j=1}^{J-1} \left[c_j \frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h)}{\partial \omega_j} \right] + \sum_{m=1}^{M} \sum_{k=1}^{K-1} \sum_{j=1}^{J} \left[d_{mkj} \frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h)}{\partial \epsilon_{mkj}} \right] = 0$$
$$\Leftrightarrow c_j = 0 \,\forall j;\, d_{mkj} = 0 \,\forall m, k, j. \tag{A6}$$

Notice that

$$\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h)}{\partial \omega_j} = \eta_j \left[\psi_{hj} - \sum_{l=1}^{J} (\eta_l \psi_{hl}) \right] \quad \forall j,$$

and

$$\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h)}{\partial \epsilon_{mkj}} = \eta_j \psi_{hj} [y_{hmk} - p_{mkj}] \quad \forall m, k, j.$$

Therefore, the left-hand side of (A.6) can be expressed as

$$\sum_{j=1}^{J-1} \left\{ \left[c_j \eta_j - \left(\sum_{l=1}^{J-1} c_l \eta_l \right) \eta_j \right] [\boldsymbol{\psi}_j - \boldsymbol{\psi}_J] \right\}$$
$$+ \sum_{m=1}^{M} \sum_{k=1}^{K-1} \sum_{j=1}^{J} \left\{ \left[d_{mkj} p_{mkj} - \left(\sum_{s=1}^{K-1} d_{msj} p_{msj} \right) p_{mkj} \right] \eta_j \left[\left(\frac{\mathbf{y}_{mk}}{p_{mkj}} - \frac{\mathbf{y}_{mK}}{p_{mKj}} \right) \# \boldsymbol{\psi}_j \right] \right\} = 0.$$

By (A.5), we can get

$$\begin{cases} c_j \eta_j - \left(\sum_{l=1}^{J-1} c_l \eta_l \right) \eta_j = 0 & \text{for } j = 1, \ldots, J-1 \\ d_{mkj} p_{mkj} - \left(\sum_{s=1}^{K-1} d_{msj} p_{msj} \right) p_{mkj} = 0 & \text{for } m = 1, \ldots, M; k = 1, \ldots, K-1; j = 1, \ldots, J, \end{cases}$$

which imply $c_j = 0\ \forall j$; $d_{mkj} = 0\ \forall m, k, j$. Hence the proof.

Proof of Theorem 1. Let $\mathbf{D}$ be the differential for the RLCA model (3). Then $\mathbf{D}$ can be partitioned into sub-matrices

$$\mathbf{D} = \big[\mathbf{D}_1 | \cdots | \mathbf{D}_{J-1} | \underbrace{\mathbf{E}_{111} | \cdots | \mathbf{E}_{1(K-1)1}}_{m=1; j=1} | \cdots | \underbrace{\mathbf{E}_{M1J} | \cdots | \mathbf{E}_{M(K-1)J}}_{m=M; j=J} |$$
$$\underbrace{\mathbf{F}_{11} | \cdots | \mathbf{F}_{1(K-1)}}_{m=1} | \cdots | \underbrace{\mathbf{F}_{M1} | \cdots | \mathbf{F}_{M(K-1)}}_{m=M} \big].$$

Here, $\mathbf{D}_j$ is the $((K^M - 1)N) \times (P + 1)$ dimensional matrix of partial derivatives of $\Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)$ w.r.t. $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{Pj})$ with the $((K^M - 1)(i - 1) + h)$th element in the $(p + 1)$th column equal to

$$\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)}{\partial \beta_{pj}} = x_{ip} \eta_{ij} [\psi_{ihj} - \sum_{l=1}^{J} (\eta_{il} \psi_{ihl})], \tag{A7}$$

where $\eta_{ij} = \eta_j(\mathbf{x}_i^T \boldsymbol{\beta})$, $\psi_{ihj} = \Pr(\mathbf{Y}_i = \mathbf{y}_h | S_i = j, \mathbf{z}_i)$. $\mathbf{E}_{mkj}$ is the $((K^M - 1)N) \times 1$ dimensional matrix of partial derivatives of $\Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)$ w.r.t. γ_{mkj} with the $((K^M - 1)(i - 1) + h)$th element equal to

$$\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_{mkj}} = \eta_{ij} \psi_{ihj} [y_{hmk} - p_{imkj}], \tag{A8}$$

where $p_{imkj} = p_{mkj}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T \boldsymbol{\alpha}_m)$. $\mathbf{F}_{mk}$ is the $((K^M - 1)N) \times L$ dimensional matrix of partial derivatives of $\Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)$ w.r.t. $\boldsymbol{\alpha}_{mj} = (\alpha_{1mk}, \dots, \alpha_{Lmk})$ with the $((K^M-1)(i-1)+h)$th element equal to

$$\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)}{\partial \alpha_{pmk}} = z_{imp} \left\{ \sum_{j=1}^{J} [\eta_{ij} \psi_{ihj} (y_{hmk} - p_{imkj})] \right\}. \tag{A9}$$

Each $\mathbf{D}_j$ can be decomposed into $\mathbf{D}_j = \mathbf{d}_j \# \mathbf{X}^D$. Here, $\mathbf{X}^D$ is the Kronecker product of $\mathbf{X}$ and $(K^M - 1)N \times 1$ dimensional vector of ones—that is, the design matrix for latent prevalences with each row repeated $(K^M - 1)$ times. The $((K^M - 1)(i - 1) + h)$th element of $\mathbf{d}_j$ is defined by (A.7) setting the leading x_{ip} equal to one, which is equivalent to the partial derivative of $\Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)$ w.r.t. ω_{ij}, where $\eta_{ij} = \exp(\omega_{ij})/[1 + \sum_{l=1}^{J-1} \exp(\omega_{il})]$.

Each $\mathbf{E}_{mkj}$ can be decomposed into a $(K^M - 1)N \times 1$ dimensional vector of ones—$\mathbf{1}_{(K^M-1)N}$—and a vector $\mathbf{e}_{mkj}$ such that $\mathbf{E}_{mkj} = \mathbf{e}_{mkj} \# \mathbf{1}_{(K^M-1)N}$. The $((K^M-1)(i-1)+h)$th element of $\mathbf{e}_{mkj}$ is defined by (A.8), which is equivalent to the partial derivative of $\Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)$ w.r.t. ϵ_{imkj}, where $p_{imkj} = \exp(\epsilon_{imkj})/[1 + \sum_{s=1}^{K-1} \exp(\epsilon_{imsj})]$.

Each $\mathbf{F}_{mk}$ can be decomposed into $\mathbf{F}_{mk} = \mathbf{f}_{mk} \# \mathbf{Z}_m^F$. Here, $\mathbf{Z}_m^F$ is the Kronecker product of $\mathbf{Z}_m$ and $(K^M - 1)N \times 1$ dimensional vector of ones—that is, the design matrix for the conditional probabilities of the mth item excluding $\mathbf{1}_N$ with each row repeated $(K^M - 1)$ times. The $((K^M - 1)(i - 1) + h)$th element of $\mathbf{f}_{mk}$ is defined by (A.9) setting the leading z_{imp} equal to one, which is equivalent to

$$\sum_{j=1}^{J} \left(\frac{\partial \Pr(\mathbf{Y}_i = \mathbf{y}_h | \mathbf{x}_i, \mathbf{z}_i)}{\partial \gamma_{mkj}} \right).$$

Therefore, $\mathbf{f}_{mk} = \sum_{j=1}^{J} \mathbf{e}_{mkj},\ m = 1, \dots, M,\ k = 1, \dots, K - 1$, which are the exclusively linear combination of vectors $\mathbf{e}_{mkj}$.

The RLCA model (3) will be locally identifiable if $\mathbf{Du} = 0 \Leftrightarrow \mathbf{u} = 0$. Since $\mathbf{f}_{mk} = \sum_{j=1}^{J} \mathbf{e}_{mkj}\ \forall m, k$, $\mathbf{Du}$ can be expressed as

$$\begin{aligned}
\mathbf{Du} &= (\mathbf{X}^D \mathbf{u}_1) \# \mathbf{d}_1 + \cdots + (\mathbf{X}^D \mathbf{u}_{J-1}) \# \mathbf{d}_{J-1} \\
&\quad + (\mathbf{1}_{(K^M-1)N} v_{111}) \# \mathbf{e}_{111} + \cdots + (\mathbf{1}_{(K^M-1)N} v_{M(K-1)J}) \# \mathbf{e}_{M(K-1)J} \\
&\quad + (\mathbf{Z}_1^F \mathbf{w}_{11}) \# \mathbf{f}_{11} + \cdots + (\mathbf{Z}_M^F \mathbf{w}_{M(K-1)}) \# \mathbf{f}_{M(K-1)} \\
&= \sum_{j=1}^{J-1} [(\mathbf{X}^D \mathbf{u}_j) \# \mathbf{d}_j] + \sum_{m=1}^{M} \sum_{k=1}^{K-1} \sum_{j=1}^{J} [(\mathbf{Z}_m^G \mathbf{z}_{mkj}) \# \mathbf{e}_{mkj}],
\end{aligned}$$

where $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_{J-1}^T, v_{111}, \dots, v_{M(K-1)J}, \mathbf{w}_{11}^T, \dots, \mathbf{w}_{M(K-1)}^T]^T$ with $\mathbf{u}_j$ being a $(P + 1) \times 1$ vector; v_{mkj} a constant; and $\mathbf{w}_{mk}$ a $L \times 1$ vector, $\mathbf{Z}_m^G = [\mathbf{1}_{(K^M-1)N}, \mathbf{Z}_m^F]$, and $\mathbf{z}_{mkj} = [v_{mkj}, \mathbf{w}_{mk}^T]^T$. From Proposition 2, 3, and assumption (ii′), the $\mathbf{d}_j$'s $(j = 1, \dots, J - 1)$ and $\mathbf{e}_{mkj}$'s $(j = 1, \dots, J,\ m = 1, \dots, M,\ k = 1, \dots, K - 1)$ are linearly independent. $\mathbf{Du} = 0$ if and only if $(\mathbf{X}^D \mathbf{u}_j) = 0\ \forall j$ and $(\mathbf{Z}_m^G \mathbf{z}_{mkj}) = 0\ \forall m, k, j$. Moreover, since $\mathbf{X}^D$ and $\mathbf{Z}_m^G$ are of full column rank (assumption (iv′)),

$$\mathbf{Du} = 0 \Leftrightarrow \mathbf{u}_j = 0\ \forall j,\ v_{mkj} = 0\ \forall m, k, j, \quad \text{and} \quad \mathbf{w}_{mk} = 0\ \forall m, k \Leftrightarrow \mathbf{u} = 0.$$

Thus, the model is locally identifiable.

Appendix B: EM Estimation of Parameters and Variances

M-step calculation

From equation (12), the parameter space of $\boldsymbol{\phi}$ separates into $(M + 1)$ subsets: $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_P)$ with $\boldsymbol{\beta}_p = (\beta_{p1}, \ldots, \beta_{p(J-1)})$, $p = 0, \ldots, P$; and, for $m = 1, \ldots, M$, $\boldsymbol{\omega}_m = (\boldsymbol{\gamma}_{m1}, \ldots, \boldsymbol{\gamma}_{mJ}, \boldsymbol{\alpha}_{1m}, \ldots, \boldsymbol{\alpha}_{Lm})$ with $\boldsymbol{\gamma}_{mj} = (\gamma_{m1j}, \ldots, \gamma_{m(K_m-1)j})$ and $\boldsymbol{\alpha}_{qm} = (\alpha_{qm1}, \ldots, \alpha_{qm(K_m-1)})$, $j = 1, \ldots, J$, $q = 1, \ldots, L$. Thus, maximization of an EM algorithm can be implemented separately for each subset, saving substantial computing time. The M-step can be written as: Find $\boldsymbol{\beta}$, which maximizes

$$Q_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\boldsymbol{\phi}^{(p)}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\{\theta_{ij}(\boldsymbol{\phi}^{(p)})[\log \eta_j(\mathbf{x}_i^T\boldsymbol{\beta})]\},$$

and find $\boldsymbol{\omega}_m$, $m = 1, \ldots, M$,which maximizes

$$Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_m}\{\theta_{ij}(\boldsymbol{\phi}^{(p)})y_{imk}[\log p_{mkj}(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m)]\},$$

where

$$\begin{aligned}\theta_{ij}(\boldsymbol{\phi}^{(p)}) &= \mathrm{E}(S_{ij}|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\phi}^{(p)}, \mathbf{x}_i, \mathbf{z}_i)\\ &= \frac{\eta_j(\mathbf{x}_i^T\boldsymbol{\beta}^{(p)})\prod_{m=1}^{M}\prod_{k=1}^{K_m} p_{mkj}^{y_{imk}}(\boldsymbol{\gamma}_{mj}^{(p)} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m^{(p)})}{\sum_{l=1}^{J}\eta_l(\mathbf{x}_i^T\boldsymbol{\beta}^{(p)})\prod_{m=1}^{M}\prod_{k=1}^{K_m} p_{mkl}^{y_{imk}}(\boldsymbol{\gamma}_{ml}^{(p)} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m^{(p)})}\end{aligned}$$

is the posterior probability of class membership evaluated at $\boldsymbol{\phi}^{(p)}$. The maximizing process is carried out using the one iteration Newton–Raphson method. The first and second derivatives used in Newton's methods for deriving maximum parameter estimates of $Q_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\boldsymbol{\phi}^{(p)})$ and $Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})$ are as follows:

$$\frac{\partial Q_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\boldsymbol{\phi}^{(p)})}{\partial\beta_{pj}} = \sum_{i=1}^{N}\{x_{ip}[\theta_{ij}^{(p)} - \eta_{ij}]\};$$

$$\frac{\partial^2 Q_{\boldsymbol{\beta}}(\boldsymbol{\beta}|\boldsymbol{\phi}^{(p)})}{\partial\beta_{pj}\partial\beta_{ul}} = -\sum_{i=1}^{N}\{x_{ip}x_{iu}\eta_{ij}[\delta_{jl} - \eta_{il}]\};$$

$$\frac{\partial Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})}{\partial\gamma_{mkj'}} = \sum_{i=1}^{N}\{\theta_{ij'}^{(p)}[y_{imk} - p_{imkj'}]\};$$

$$\frac{\partial Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})}{\partial\alpha_{qmk}} = \sum_{i=1}^{N}\sum_{j'=1}^{J}\{z_{imq}\theta_{ij'}^{(p)}[y_{imk} - p_{imkj'}]\};$$

$$\frac{\partial^2 Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})}{\partial\gamma_{mkj'}\partial\gamma_{msl'}} = -\sum_{i=1}^{N}\{\theta_{ij'}^{(p)}\delta_{j'l'}p_{imkj'}[\delta_{ks} - p_{imsl'}]\};$$

$$\frac{\partial^2 Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})}{\partial\alpha_{qmk}\partial\alpha_{rms}} = -\sum_{i=1}^{N}\sum_{j'=1}^{J}\{z_{imq}z_{imr}\theta_{ij'}^{(p)}p_{imkj'}[\delta_{ks} - p_{imsj'}]\};$$

$$\frac{\partial^2 Q_{\boldsymbol{\omega}_m}(\boldsymbol{\omega}_m|\boldsymbol{\phi}^{(p)})}{\partial\gamma_{mkj'}\partial\alpha_{rms}} = -\sum_{i=1}^{N}\{z_{imr}\theta_{ij'}^{(p)}p_{imkj'}[\delta_{ks}-p_{imsj'}]\},$$

where $\theta_{ij}^{(p)}=\theta_{ij}(\boldsymbol{\phi}^{(p)})$, $\eta_{ij}=\eta_j(\mathbf{x}_i^T\boldsymbol{\beta})$, $p_{imkj}=p_{mkj}(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^T\boldsymbol{\alpha}_m)$, $\delta_{jl}=\mathrm{I}(j=l)$, $x_{i0}=1$, and $i=1,\ldots,N; m=1,\ldots,M; k,s=1,\ldots,(K_m-1); j,l=1,\ldots,J-1; j',l'=1,\ldots,J; p,u=0,1,\ldots,P; q,r=1,\ldots,L$.

Variance estimation

Let $D_{\boldsymbol{\phi}}^1$ and $D_{\boldsymbol{\phi}}^2$ be the gradient and Hessian operators with respect to $\boldsymbol{\phi}$, and define

$$\mathrm{I}(\boldsymbol{\phi};\mathbf{Y}) = -D_{\boldsymbol{\phi}}^2\log\mathrm{L}(\boldsymbol{\phi};\mathbf{Y})$$

as the observed Fisher information matrix of the incomplete-data likelihood with respect to the elements $\boldsymbol{\phi}$. For the complete-data likelihood $\mathrm{L_c}$, we let

$$\mathrm{I_c}(\boldsymbol{\phi};\mathbf{Y},\mathbf{S}) = -D_{\boldsymbol{\phi}}^2\log\mathrm{L_c}(\boldsymbol{\phi};\mathbf{Y},\mathbf{S}).$$

Louis (1982) showed that

$$\mathrm{I}(\hat{\boldsymbol{\phi}};\mathbf{Y}) = \mathcal{I}_c(\hat{\boldsymbol{\phi}};\mathbf{Y}) - \mathrm{Var}\{\mathrm{S_c}(\hat{\boldsymbol{\phi}};\mathbf{Y},\mathbf{S})|\mathbf{Y}\},$$

where $\mathcal{I}_c(\hat{\boldsymbol{\phi}};\mathbf{Y}) = \mathrm{E}\{\mathrm{I_c}(\boldsymbol{\phi};\mathbf{Y},\mathbf{S})|\mathbf{Y}\}|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$, $\mathrm{S_c}(\hat{\boldsymbol{\phi}};\mathbf{Y},\mathbf{S}) = D_{\boldsymbol{\phi}}^1\log\mathrm{L_c}(\boldsymbol{\phi};\mathbf{Y},\mathbf{S})|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$, and $\hat{\boldsymbol{\phi}}$ is the MLE of $\boldsymbol{\phi}$.

Therefore, the observed information matrix can be computed in terms of the conditional moments of the first- and second-order partial derivatives of the complete-data log likelihood function introduced within the EM framework. The estimator for the variance-covariance matrix of the parameter estimates is the inverse of the observed information matrix, evaluated at the parameter estimates. It is easy to see that $\mathcal{I}_c(\hat{\boldsymbol{\phi}};\mathbf{Y})$ has the same formula as the second derivatives of $Q(\boldsymbol{\phi}|\hat{\boldsymbol{\phi}})$. The elements of $\mathrm{Var}\{\mathrm{S_c}(\hat{\boldsymbol{\phi}};\mathbf{Y},\mathbf{S})|\mathbf{Y}\}$ can be shown as follows:

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\beta_{pj}},\frac{\partial\log\mathrm{L_c}}{\partial\beta_{ul}}\middle|\mathbf{Y}\right\} = \sum_{i=1}^{N}\{x_{ip}x_{iu}\hat{\theta}_{ij}[\delta_{jl}-\hat{\theta}_{il}]\};$$

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\gamma_{mkj'}},\frac{\partial\log\mathrm{L_c}}{\partial\gamma_{vsl'}}\middle|\mathbf{Y}\right\} = \sum_{i=1}^{N}\{[y_{imk}-\hat{p}_{imkj'}][y_{ivs}-\hat{p}_{ivsl'}]\hat{\theta}_{ij'}[\delta_{j'l'}-\hat{\theta}_{il'}]\};$$

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\alpha_{qmk}},\frac{\partial\log\mathrm{L_c}}{\partial\alpha_{rvs}}\middle|\mathbf{Y}\right\} = \sum_{i=1}^{N}\left\{z_{imq}z_{ivr}\sum_{j'=1}^{J}\sum_{l'=1}^{J}\hat{p}_{imkj'}\hat{p}_{ivsl'}\hat{\theta}_{ij'}[\delta_{j'l'}-\hat{\theta}_{il'}]\right\};$$

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\alpha_{qmk}},\frac{\partial\log\mathrm{L_c}}{\partial\gamma_{vsl'}}\middle|\mathbf{Y}\right\} = -\sum_{i=1}^{N}\left\{z_{imq}[y_{ivs}-\hat{p}_{ivsl'}]\sum_{j'=1}^{J}[\hat{p}_{imkj'}\hat{\theta}_{ij'}(\delta_{j'l'}-\hat{\theta}_{il'})]\right\};$$

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\beta_{pj}},\frac{\partial\log\mathrm{L_c}}{\partial\gamma_{vsl'}}\middle|\mathbf{Y}\right\} = \sum_{i=1}^{N}\{x_{ip}[y_{ivs}-\hat{p}_{ivsl'}]\hat{\theta}_{ij}[\delta_{jl'}-\hat{\theta}_{il'}]\};$$

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial\log\mathrm{L_c}}{\partial\beta_{pj}},\frac{\partial\log\mathrm{L_c}}{\partial\alpha_{rvs}}\middle|\mathbf{Y}\right\} = -\sum_{i=1}^{N}\left\{x_{ip}z_{ivr}\sum_{l'=1}^{J}\hat{p}_{ivsl'}\hat{\theta}_{ij}[\delta_{jl'}-\hat{\theta}_{il'}]\right\},$$

where

$$\widehat{\mathrm{Cov}}\left\{\frac{\partial \log \mathrm{L_c}}{\partial\cdot}, \frac{\partial \log \mathrm{L_c}}{\partial\cdot}\middle|\mathbf{Y}\right\} = \mathrm{Cov}\left\{\frac{\partial \log \mathrm{L_c}(\boldsymbol{\phi}; \mathbf{Y}, \mathbf{S})}{\partial\cdot}, \frac{\partial \log \mathrm{L_c}(\boldsymbol{\phi}; \mathbf{Y}, \mathbf{S})}{\partial\cdot}\middle|\mathbf{Y}\right\}|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}},$$

$$\hat{\theta}_{ij} = \theta_{ij}(\hat{\boldsymbol{\phi}}),\ \hat{p}_{imkj} = p_{mkj}(\hat{\boldsymbol{\gamma}}_{mj} + \mathbf{z}_{im}^T\hat{\boldsymbol{\alpha}}_m),\ \delta_{jl} = \mathrm{I}(j = l),\ x_{i0} = 1,$$

and $i = 1, \ldots, N$; $m, v = 1, \ldots, M$; $k, s = 1, \ldots, (K_m - 1)$; $j, l = 1, \ldots, J - 1$; $j', l' = 1, \ldots, J$; $p, u = 0, 1, \ldots, P$; $q, r = 1, \ldots, L$.

References

Agresti, A. (1984). *Analysis of Categorical Data*. New York: John Wiley and Sons.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Baker, S.G., & Laird, N.M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, *83*, 62–69.

Bandeen-Roche, K., Huang, G.H., Munoz, B., & Rubin, G.S. (1999). Determination of risk factor associations with questionnaire outcomes: A methods case study. *American Journal of Epidemiology*, *150*, 1165–1178.

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L., & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*, 1375–1386.

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York: John Wiley and Sons.

Clogg, C.C., & Goodman, L.A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, *79*, 762–771.

Clogg, C.C., & Goodman, L.A. (1985). Simultaneous latent structure analysis in several groups. In *Sociological Methodology 1985*, Tuma N.B. (Ed.), 81–110. San Francisco: Jossey-Bass.

Dayton, C.M., & Macready, G.B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*, 173–178.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Eaton, W. W., Dryman, A., Sorenson, A., & McCutcheon, A. (1989). DSM-III major depressive disorder in the community—A latent class analysis of data from the NIMH epidemiologic catchment-area program. *British Journal of Psychiatry*, *155*, 48–54.

Efron, B., & Hinkley, D.V. (1978). Assessing the acuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*, 457–487.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189.

Formann, A.K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, *38*, 87–111.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, *87*, 476–486.

Garrett, E.S., & Zeger, S.L. (2000). Latent class model diagnosis. *Biometrics*, *56*, 1055–1067.

Goldberg, D. (1972). *GHQ The Selection of Psychiatric Illness by Questionnaire*. London: Oxford University Press.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.

Graybill, F.A. (1969). *Introduction to Matrices with Applications in Statistic*, Belmont, CA: Wadsworth.

Green, B.F. (1951). A general solution of the latent class model of latent structure analysis and latent profile analysis. *Psychometrika*, *16*, 151–166.

Haberman, S.J. (1974). Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *Annals of Statistics*, *2*, 911–924.

Haberman, S.J. (1979). *Analysis of Qualitative Data. Vol. 2: New Developments*. New York: Academic Press.

Hagenaars, J.A. (1993). *Loglinear Models with Latent Variables*. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-094. Newbury Park, CA: Sage.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Huang, G.H., Bandeen-Roche, K., & Rubin, G.S. (2002). Building marginal models for multiple ordinal measurements. *Applied Statistics*, *51*, 37–57.

Hudziak, J.J., Heath, A.C., Madden, P.F., Reich, W., Bucholz, K.K., Slutske, W., Bierut, L.J., Neuman, R.J., & Todd, R.D. (1998). Latent class and factor analysis of DSM-IV ADHD: A twin study of female adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 848–857.

Jette, A.M., & Branch, L.G. (1985). Impairment and disability in the aged. *J Chronic Dis*, *38*, 59–65.

Jöreskog, K.G., & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631–639.

Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., & Jaffer, M.W. (1963). Studies of illness in the age. The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, *185*, 914–918.

Lange, K. (1995). A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *57*, 425–437.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent Structure Analysis*. New York: Houghton-Mifflin.

Legler, J.M., & Ryan, L.M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association*, *92*, 13–20.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *44*, 226–233.

Magidson J., & Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots, and related graphical displays. In *Sociological Methodology 2001*, Stolzenberg R.M. (ed), 223–264. Boston: Blackwell.

Mangione, C.M., Phillips, R.S., Seddon, J.M., Lawrence, M.G., Cook, E.F., Dailey, R., & Goldman, L. (1992). Development of the "activities of daily vision" scale: A measurement of visual functional status. *Medical Care*, *30*, 1111–1126.

McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models, 2nd edition*. London: Chapman and Hall.

McCutcheon, A. C. (1987). *Latent Class Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-064. Beverly Hills, CA: Sage.

McHugh, R.B. (1956). Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21, 331-347.

McLachlan G.J., & Krishnan T. (1996). *The EM Algorithm and Extensions*. New York: John Wiley and Sons.

Melton, B., Liang, K.Y., & Pulver, A.E. (1994). Extended latent class approach to the study of familial/sporadic forms of a disease: Its application to the study of the heterogeneity of schizophrenia. *Genetic Epidemiology*, *11*, 311–327.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, *49*, 313–334.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Muthén, L.K., & Muthén, B.O. (1998). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using EM algorithm. *Biometrics*, *55*, 463–469.

Neuman, R.J., Heath, A., Reich, W., Bucholz, K.K., Madden, P.A.F., Sun, L., Todd, R.D., & Hudziak, J.J. (2001). Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 933–942.

Piantadosi, S. (1997). *Clinical Trials: A Methodologic Perspective*. New York: John Wiley and Sons.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.

Roeder, K., Lynch, K.G., & Nagin, D.S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, *94*, 766–776.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology*, *44*, 75–92.

Rubin, G.S., Bandeen-Roche, K., Huang, G.H., Munoz, B., Schein, O.D., Fried, L.P., & West, S.K. (2001). The association of multiple visual impairments with self-reported visual disability: SEE project. *Investigative Ophthalmology and Visual Science*, *42*, 64–72.

Rubin, G.S., West, S.K., Munoz, B., Bandeen-Roche, K., Zeger, S.L., Schein, O., & Fried, L.P. (1997). A comprehensive assessment of visual impairment in an older American population: SEE study. *Investigative Ophthalmology and Visual Science*, *38*, 557–568.

Sammel, M.D., Ryan, L.M., & Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B*, *59*, 667–678.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Statistical Sciences, Inc. (1995). *S-PLUS User's Manual, Version 3.3 for Windows*. Seattle: Statistical Sciences, Inc.

Sullivan, P.F., Kessler, R.C., & Kendler, K.S. (1998). Latent class analysis of lifetime depressive symptoms in national comorbidity survey. *American Journal of Psychiatry*, *155*, 1398–1406.

Uebersax, J. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association*, *88*, 421–427.

Valbuena, M., Bandeen-Roche, K., Rubin, G.S., Munoz, B., West, S.K., and SEE Project Team (1999). Self-reported assessment of visual functioning in a population based setting. *Investigative Ophthalmology and Visual Science*, *40*, 280–288.

Van der Heijden, P.G.M., Dessens, J., & Böckenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm, *Journal of Educational and Behavioral Statistics*, *21*, 215–229.

Vermunt, J.K. (1996). *Log-linear Event History Analysis: A General Approach with Missing Data, Unobserved Heterogeneity, and Latent Variables*. Tilburg: Tilburg University Press.

Vermunt, J.K., & Magidson, J. (2000). *Latent GOLD 2.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.
West, S.K., Munoz, B., Rubin, G.S., Schein, O.D., Bandeen-Roche, K., Zeger, S., German, P.S., & Fried, L.P. (1997). Function and visual impairment in a population-based study of older adults: SEE project. *Investigative Ophthalmology and Visual Science*, *38*, 72–82.
Wu, C.F. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, *11*, 95–103.

Manuscript received 6 FEB 2002
Final version received 11 FEB 2003