Published in final edited form as:
Psychometrika. 2012 April 1; 77(2): 244–262. doi:10.1007/s11336-012-9248-6.

PREDICTING LATENT CLASS SCORES FOR SUBSEQUENT ANALYSIS

Janne Petersen,
COPENHAGEN UNIVERSITY HOSPITAL, HVIDOVRE

Karen Bandeen-Roche,
JOHNS HOPKINS UNIVERSITY

Esben Budtz-Jørgensen, and
UNIVERSITY OF COPENHAGEN

Klaus Groes Larsen
H. LUNDBECK A/S

Abstract

Latent class regression models relate covariates and latent constructs such as psychiatric disorders. Though full maximum likelihood estimation is available, estimation is often in three steps: (i) a latent class model is fitted without covariates; (ii) latent class scores are predicted; and (iii) the scores are regressed on covariates. We propose a new method for predicting class scores that, in contrast to posterior probability-based methods, yields consistent estimators of the parameters in the third step. Additionally, in simulation studies the new methodology exhibited only a minor loss of efficiency. Finally, the new and the posterior probability-based methods are compared in an analysis of mobility/exercise.

Keywords

classification; latent class regression; latent class scores; three-step procedure; least squares class

1. Introduction

Latent class models (LCM) (Lazarsfeld & Henry, 1968; Goodman, 1974) are widely used in social and psychological research for analyzing constructs thought not to be directly measurable. The LCM aims to identify homogeneous subgroups of persons based on their categorical, observed, surrogate variables. Often the purpose of such an analysis is to characterize persons, for example to make diagnoses (Pepe, 2003). Several methods exist for characterizing persons within an LCM framework; these are all based on the posterior probabilities of class membership (Bolck, Croon, and Hagenaars, 2004).

LCM can be extended to latent class regression (LCR) that allows dependence of latent class membership on explanatory variables. Estimation of both the LCM and regression parameters may be accomplished simultaneously by maximum likelihood, for instance by using the EM-algorithm (Dayton & Macready, 1988; Bandeen-Roche, Miglioretti, Zeger, and Rathouz, 1997; Bartolucci & Forcina, 2006), hereafter called one-step estimation. Although such estimation indeed is employed, researchers often carry out an alternative analysis in three steps

Requests for reprints should be sent to Janne Petersen, Copenhagen University Hospital, Hvidovre, Denmark. petersen.janne@gmail.com.

(Reyes, Henry, Tolan, and Wakschlag, 2009; Steiger, Richardson, Schmitz, Joober, Israel, and Bruce, 2009; Markkula, Jarvinen, Leino-Arjas, Koskenvuo, Kalso, and Kaprio, 2009; Schorr, Ulbricht, Schmidt, Baumeister, Ruge, and Schumann, 2008; Unick, Snowden, and Hastings, 2009; Funderbuck, Maisto, Sugarman, and Wade, 2008). First, the LCM is fitted without considering explanatory variables. Then the resulting parameter estimates are used to "score" the latent classes—often as predictions of latent class memberships. Finally, the effect, $\boldsymbol{\beta}$, of the explanatory variable is estimated by regressing the predicted latent class score on the covariates. Ultimately, the predicted latent scores can be used as responses in various regression analyses addressing different scientific questions with different sets of covariates within a cohort. The use of this three-step estimation procedure may be motivated by the desire to develop an index measuring the construct or because standard statistical programs such as SAS and SPSS do not yet have ready means for fitting an LCR in a one-step procedure.

No existing methods for predicting latent classes yield consistent estimators of the regression coefficients $\boldsymbol{\beta}$ in three-step analysis (Bolck et al., 2004; Croon, 2002). Errors-in-variables theory (Carroll, Ruppert, Stefanski, and Crainiceanu, 2006) can be applied to correct for bias (Croon, 2002; Bolck et al., 2004). Moreover, there has been relevant work on three-step estimation procedures for other types of latent variable models. In factor analysis, Skrondal and Laake (2001) showed that both the so-called "Bartlett" and least squares scores (Johnson & Wichern, 2007) yield consistent estimators of $\boldsymbol{\beta}$. Alternatively, Croon (2002) showed how errors-in-variables theory may be applied to correct the bias. For a latent trait model, with a continuous latent variable and discrete outcomes, Lu and Thomas (2008) recently studied the performance of Bartlett-like scores. They found that a weighted maximum likelihood estimator produces nearly consistent estimates of $\boldsymbol{\beta}$ if the discrete outcomes have sufficiently many categories.

In this paper we study the three-step procedure for LCR. The outline of the paper is as follows: In Section 2 we describe known methods for predicting latent class scores. These are all based on the posterior probabilities; in Section 3 we propose a new method for assigning values to the latent class variable aiming to reduce the bias in three-stage estimation. The method is not based on the posterior probabilities but inspired by Bartlett scoring; in Section 3.1 an analytical argument is provided demonstrating how these scores may be used to consistently estimate the structural parameters, $\boldsymbol{\beta}$; in Section 4 a simulation study, estimation and inference from the posterior probability-based three-step methods and the new approach are compared with results of one-step estimation of the LCR; in Section 5 we apply our method in comparison with one-step estimation and the posterior probability-based methods to data from The Salisbury Eye Evaluation Project (West, Munoz, Rubin, Schein, Bandeen-Roche, and Zeger, 1997), to study the age- and comorbidity-adjusted association between visual acuity and a latent measure of "mobility/exercise tolerance" among older adults; finally Section 6 gives a discussion.

2. Background

2.1. Latent Class Regression

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$ denote a vector of J latent class indicators for person i, $i \in \{1, \ldots, N\}$. For convenience we have assumed Y_{ij}, $j \in \{1, \ldots, J\}$ to be binary indicators. The results may easily be extended to polytomous indicators; when needed the required modifications are noted. Let $\mathbf{x}_i$ be a vector of P covariates, let M be the number of latent classes, and let S_i identify the class membership for person i. We assume for all $i \in \{1, \ldots, N\}$ and $m \in \{1, \ldots, M\}$ that, given class membership, the distribution of $\boldsymbol{Y}_i$ does not depend on the covariates. This is the assumption of non-differential item functioning (non-DIF), also termed a homogeneous model,

$$\Pr(Y_{ij}=1|S_i=m,\mathbf{x}_i)=\Pr(Y_{ij}=1|S_i=m)=\pi_{jm}.$$

Also, local independence is assumed, so that, given the class membership, S_i, the indicators are independent:

$$\Pr(Y_{i1}=y_1,\ldots,Y_{iJ}=y_J|S_i=m)=\prod_{j=1}^{J}\Pr(Y_{ij}=y_j|S_i=m).$$

The LCR model assumes

$$\begin{aligned}\Pr(Y_{i1}=y_1,\ldots,Y_{iJ}=y_J|\mathbf{x}_i)&=\sum_{m=1}^{M}\Pr(S_i=m|\mathbf{x}_i)\Pr(\boldsymbol{Y}_i|S_i=m)\\&=\sum_{m=1}^{M}\eta_m(\mathbf{x}_i)\prod_{j=1}^{J}\pi_{jm}^{y_j}(1-\pi_{jm})^{1-y_j},\end{aligned}\qquad(1)$$

where $\eta_m(\mathbf{x}_i) = \Pr(S_i = m|\mathbf{x}_i)$ is the structural model component. This is often modeled using a generalized logit link:

$$\eta_m(\mathbf{x}_i)=\frac{\exp(\mathbf{x}_i\beta_m)}{\sum_{k=1}^{M}\exp(\mathbf{x}_i\beta_k)}.\qquad(2)$$

Coefficients in this model $\boldsymbol{\beta}_{mp}$ have the interpretation of log odds ratios (by convention, as compared to the last or reference class, M):

$$\beta_{mp}=\log\left[\frac{\Pr(S_i=m|x_{ip}=z+1,\boldsymbol{x}_{i,-p})/\Pr(S_i=M|x_{ip}=z+1,\boldsymbol{x}_{i,-p})}{\Pr(S_i=m|x_{ip}=z,\boldsymbol{x}_{i,-p})/\Pr(S_i=M|x_{ip}=z,\boldsymbol{x}_{i,-p})}\right].$$

Here, the odds ratios is with respect to a one unit difference in x_p, holding all other covariates, denoted as $\boldsymbol{x}_{-p}$, constant. To achieve the desired interpretation the constraint $\boldsymbol{\beta}_M = 0$ is adopted.

2.2. The Posterior-Probability-Based Three-Step Procedures

Step 1—The first step consists of fitting an LCM without any covariates, that is, model (1) with only $M - 1$ intercepts in the linear predictor in model (2):

$$\Pr(Y_{i1}=y_1,\ldots,Y_{iJ}=y_J)=\sum_{m=1}^{M}\eta_m\prod_{j=1}^{J}\pi_{jm}^{y_j}(1-\pi_{jm})^{1-y_j}.\qquad(3)$$

Let $\hat{\boldsymbol{\pi}}_{jm}$ and $\hat{\eta}_m$, $j \in \{1, \ldots, J\}$, $m \in \{1, \ldots, M\}$ be the resulting maximum likelihood estimators.

Step 2—To predict latent class scores, one uses $\hat{\boldsymbol{\pi}}_{jm}$ and $\hat{\eta}_m$, $j \in \{1, \ldots, J\}$, $m \in \{1, \ldots, M\}$ from Step 1 to estimate the posterior probabilities of class membership

$$\widehat{q}_{im}=\Pr(S_i=m|\boldsymbol{Y}_i=\boldsymbol{y}_i)$$
$$=\frac{\widehat{\eta}_m\prod_{j=1}^{J}\widehat{\pi}_{jm}^{y_{ij}}(1-\widehat{\pi}_{jm})^{1-y_{ij}}}{\sum_{k=1}^{M}\widehat{\eta}_k\prod_{j=1}^{J}\widehat{\pi}_{jk}^{y_{ij}}(1-\widehat{\pi}_{jk})^{1-y_{ij}}}. \qquad (4)$$

Arguably the most common latent class scores, $\tilde{S}_i$, are modal assignments in which each person is designated to the class with the highest posterior probability. In pseudo-class assignment a class membership is drawn randomly from the distribution of posterior probabilities (4). Multiple pseudo-class memberships are obtained by multiple random draws from the posterior probabilities (4). Finally, one might retain the fitted posterior probabilities themselves as scores.

Step 3—The regression of predicted class scores on explanatory variables depends on the method used to predict in Step 2. The most common approach is to fit a polytomous regression of modal assignments on covariates (Croon, 2002) as in (2). It is possible to apply the Croon/Bolck correction factor (Croon, 2002; Bolck et al., 2004) in conjunction with this approach. The Croon/Bolck correction uses the fact that the structural part of the model can be rewritten as $\Pr(S|\boldsymbol{x})=\sum_{m=1}^{M}\Pr(S|\tilde{S}=m)\Pr(\tilde{S}=m|\boldsymbol{x})$. The term $\Pr(S|\tilde{S})$ is a measure of the bias in the regression estimation in Step 3 and can be calculated as $\Pr(S|\tilde{S}) = \Sigma_y \Pr(S|y)\Pr(y|\tilde{S})$. This measure of bias is used in the Croon/Bolck correction to correct the bias of the Step 3 regression $\Pr(\tilde{S} = m|\boldsymbol{x})$. Analysis of pseudo-class assignments is analogous.

In the case of multiple pseudo-class assignments there are several class assignments, $\tilde{S}_h$, $h = 1, \ldots, H$ for each individual. For each draw, one would conduct a polytomous regression analysis yielding $\hat{\boldsymbol{\beta}}_h$ with estimated standard error $\mathrm{SE}(\hat{\boldsymbol{\beta}}_h)$. A joint estimate is obtained as $\widehat{\beta}=\frac{1}{H}\sum_{h=1}^{H}\widehat{\beta}_h$, and the associated variance is calculated as $\widehat{\mathrm{var}}(\widehat{\beta})=\frac{1}{H}\sum_{h=1}^{H}[\mathrm{SE}(\widehat{\beta}_h)]^2+\frac{1}{H-1}\sum_{h=1}^{H}(\widehat{\beta}_h-\overline{\widehat{\beta}})^2$ (Wang, Brown, and Bandeen-Roche, 2005) where $\overline{\widehat{\beta}}$ is the average of the estimated $\boldsymbol{\beta}$ values.

Using the posterior probabilities themselves as latent scores leads to an $M - 1$-dimensional response variable for each person, $\hat{\boldsymbol{q}}_i = (\hat{q}_{i1}, \ldots, \hat{q}_{iM-1})$. This response might subsequently be used in a number of ways. Appealing to the asymptotical normality of estimators $\hat{\boldsymbol{q}}_i$ given $\boldsymbol{y}_i$, one approach is to analyze the posterior probabilities in a non-linear multivariate normal model with mean

$$\mathrm{E}_{(q_{im})}=\frac{\exp(\mathbf{x}_i\beta_m)}{\sum_{k=1}^{M}\exp(\mathbf{x}_i\beta_k)}, \qquad m=1,\ldots,M-1 \qquad (5)$$

and an unstructured covariance matrix. The non-linear mean structure (5) is analogous to the structural part of the LCR (2).

3. A New Method for Estimating Latent Class Scores

We were motivated by the Bartlett method for the latent factor model, which estimates an individual's latent variables as if they were fixed parameters in a linear regression of $\boldsymbol{Y}_i$ on the estimated matrix of factor loadings $\boldsymbol{\Lambda}$ (Johnson & Wichern, 2007),

$$\boldsymbol{Y}_i=\boldsymbol{\Lambda}\boldsymbol{f}_i+\boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\varepsilon}_i \sim (0, \boldsymbol{\Sigma})$. Weighted least squares is typically employed yielding Bartlett scores $\tilde{\boldsymbol{f}}_i = (\hat{\boldsymbol{\Lambda}}^t\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Lambda}})^{-1}\hat{\boldsymbol{\Lambda}}^t\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y}_i$, which is the maximum likelihood estimator conditional on $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Sigma}}$ if $\boldsymbol{\varepsilon}_i$ is normally distributed.

Our key insight was to notice that the LCM (3) is a linear model for an individual's item response probabilities in terms of class membership, $\mathrm{E}(Y_{ij}|S_i)=\pi_j^t\mathbf{G}(S_i)$, where $\boldsymbol{\pi}_j = (\boldsymbol{\pi}_{j1}, \ldots, \boldsymbol{\pi}_{jM})^t$, $j = 1, \ldots, J$, and $\mathbf{G}(S_i) = [\mathrm{g}_1(S_i), \ldots, \mathrm{g}_M(S_i)]^t = (1_{S_i=1}, \ldots, 1_{S_i=M})^t$ is the vector of indicators of class membership for individual i. Under the assumption of local independence, the surrogate indicators are independent within each subject, and we have a linear association between $\boldsymbol{Y}_i$ and the conditional probabilities $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_J)^t$:

$$\mathrm{E}(\boldsymbol{Y}_i|S_i)=\pi\mathbf{G}(S_i)=\sum_{m=1}^{M}\pi_m \mathrm{g}_m(S_i), \quad (6)$$

where $\boldsymbol{\pi}_m = (\boldsymbol{\pi}_{1m}, \ldots, \boldsymbol{\pi}_{Jm})^t$. Full analogy to the Bartlett method would have been to implement maximum likelihood estimation by iteratively reweighed least squares (IRWLS) for the likelihood functions,

$$p(\boldsymbol{y}_i|S_i)=\prod_{j=1}^{J}\left[\widehat{\pi}_j^t\mathbf{G}(S_j)\right]^{y_{ij}}\left[1-\widehat{\pi}_j^t\mathbf{G}(S_i)\right]^{1-y_{ij}}. \quad (7)$$

But a direct solution of the likelihood equation is not straightforward, and the method breaks down for patterns with all 0s or all 1s.

However, as we wish to estimate a vector of indicator functions of class memberships for each subject, $(1_{S_i=1}, \ldots, 1_{S_i=M})^t$, and we posit that subjects belong to exactly one class, we have applied this restriction to the vector of class membership $\mathrm{g}_1(S_i)+\cdots+\mathrm{g}_M(S_i)=1 \iff \mathrm{g}_M(S_i)=1-\sum_{m=1}^{M-1}\mathrm{g}_m(S_i)$. Let $\mathbf{g}(S_i)$ be the first $M - 1$ elements of $\mathbf{G}(S_i)$, $\mathbf{g}(S_i) = [\mathrm{g}_1(S_i), \ldots, \mathrm{g}_{M-1}(S_i)]^t$ and let $\boldsymbol{\Pi}$ be a $J \times M - 1$ matrix with $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1 - \boldsymbol{\pi}_M, \ldots, \boldsymbol{\pi}_{M-1} - \boldsymbol{\pi}_M)$. Substituting this into Equation (6) gives

$$E(\boldsymbol{Y}_i|S_i)-\pi_M=\boldsymbol{\Pi}\mathbf{g}(S_i). \quad (8)$$

It follows that scores of the latent variable can be estimated by regressing $\boldsymbol{Y}_i - \boldsymbol{\pi}_M$ on $\boldsymbol{\Pi}$. Note that the response variable of this regression (8) no longer consists of discrete measures as in the model (7) but instead consists of continuous measures. This fact is very useful for persons where $Y_{i1} = \cdots = Y_{iM}$, as the regression (8) then still is possible to fit. We suggest the following three-step procedure:

Step 1—Fit the LCM (3) to the data.

Step 2—Estimate $[\mathrm{g}_1(S_i), \ldots, \mathrm{g}_{M-1}(S_i)]$ in the regression of the outcomes on the conditional probabilities for each person by applying ordinary least squares (OLS) to

$$\mathrm{E}(\boldsymbol{Y}_i-\widehat{\pi}_M|S_i)=\widehat{\boldsymbol{\Pi}}\mathbf{g}(S_i), \quad (9)$$

yielding predicted scores $[\tilde{g}_1(S_i) + \cdots + \tilde{g}_{M-1}(S_i)]$. The score for being in the class M can for each person be calculated as $1 - \tilde{g}_1(S_i) - \cdots - \tilde{g}_{M-1}(S_i)$. In the case of polytomous outcomes the prediction method (9) is slightly modified. Instead of using $\hat{\boldsymbol{\pi}}_{jm}$ in (9), the expected value of Y_{ij} given $S_i = m$ is used. This is given by $\widehat{\mathrm{E}}(Y_{ij}|S_i=m)=\sum_{y_j} y_j \widehat{\mathrm{Pr}}(Y_{ij}=y_j|S_i=m)$, where the sum is over all possible values of Y_{ij}.

Inspired by factor analysis terminology, we shall refer to these new predicted latent class scores, $\tilde{\mathbf{g}}(S_i)$, as least squares class (LSC) scores.

Step 3—Regress $\tilde{\mathbf{g}}(S_i)$ on $\mathbf{x}_i$ in an $M - 1$ dimensional normal model with the non-linear mean function (2) and an unstructured covariance matrix $\boldsymbol{\Sigma}$. For the time being we obtain the standard error for $\hat{\boldsymbol{\beta}}$ as the naive estimator derived from this multivariate normal model, without taking into account that $\boldsymbol{\pi}$ is estimated.

Note that the distribution of $\tilde{\mathbf{g}}(S_i)$ given $\mathbf{x}_i$ is difficult to specify and it is unwieldy to apply maximum likelihood estimation. However, it is possible to achieve consistent estimates of $\boldsymbol{\beta}$. The estimation function from the Step 3 regression can be written as a generalized estimating equation (GEE),

$$\sum_{i=1}^{N}\boldsymbol{D}_i^t\sum\nolimits_i^{-1}\left[\tilde{\mathbf{g}}(S_i)-\mu_i(\underset{\sim}{\beta})\right]=0,$$

where $\boldsymbol{D}$ is a vector of partial derivatives $\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$. If the mean of the left hand side is 0, then the solution to the equation will be consistent under minor regularity conditions even when $\boldsymbol{\Sigma}$ is misspecified (Liang & Zeger, 1986). A consequence of this is that the Step 3 regression results in consistent estimates of $\boldsymbol{\beta}$ under the latent class modeling assumption because the mean of the scores in the estimation function is specified correctly for such a model: $\mathrm{E}[g_m(S_i)|\mathbf{x}] = \mathrm{E}(1_{(S_i=m)}|\mathbf{x}) = \eta_m(\mathbf{x})$. That $\mathrm{E}[\tilde{g}_m(S_i)|\mathbf{x}]$ has the same limiting mean is shown in next section.

In the two-class case the GENMOD procedure in SAS, assuming a normal distribution and a logit link, can be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in the Step 3 regression. For models with more classes, we have developed a SAS program (see the Appendix) that maximizes the Step 3 estimation equation.

3.1. Analytical Properties of LSC

Setup: To enhance clarity we denote $\tilde{\mathbf{g}}(S_i)$ explicitly in terms of its parameter inputs, $\tilde{\mathbf{g}}(S_i; \boldsymbol{\pi})$. The LSC score is then given by $\tilde{\mathbf{g}}(S_i; \hat{\boldsymbol{\pi}})$. Let $\mathbf{A}$ be an $M - 1$ dimensional square matrix given by

$$\mathbf{A}=\boldsymbol{\Pi}^t\boldsymbol{\Pi}, \quad (10)$$

and $\boldsymbol{b}_i$ be an $M - 1$ dimensional vector given by

$$\boldsymbol{b}_i=\boldsymbol{\Pi}(\boldsymbol{Y}_i-\pi_M). \quad (11)$$

Notice that **A** can be written as $\sum_{j=1}^{J}\boldsymbol{c}_j\boldsymbol{c}_j^t$, where $\boldsymbol{c}_j = (\boldsymbol{\pi}_{j1} - \boldsymbol{\pi}_{jM} \cdots \boldsymbol{\pi}_{j,\,M-1} - \boldsymbol{\pi}_{jM})^t$, and thus a sufficient condition for **A** to be positive definite is that $\text{span}\{\boldsymbol{c}_j\}_{j\,\in\,\{1,\,\ldots,\,J\}} = R^{M-1}$. If (1) is identifiable the condition must hold, and thus $\mathbf{A}^{-1}$, exist.

Lemma 1—If the LCM (1) is identifiable such that $\mathbf{A}^{-1}$ is elementwise bounded in absolute value by a constant K, and $\mathrm{g}_m(S_i)$ in model (6) is estimated by OLS under the assumption that $\mathrm{g}_1(S_i) + \cdots + \mathrm{g}_M(S_i) = 1$, then

$$\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\left[\tilde{\mathbf{g}}(S_i;\pi)|\mathbf{x}_i\right]=\mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathbf{g}(S_i)|\mathbf{x}_i\right]\ and\lim_N\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\left[\tilde{\mathbf{g}}(S_i;\widehat{\pi})|\mathbf{x}_i\right]=\mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathbf{g}(S_i)|\mathbf{x}_i\right].$$

Proof: Assume $\boldsymbol{\pi}$ to be known. Then the OLS estimator for $\mathbf{g}(S_i)$ from (9) is given by $\tilde{\mathbf{g}}(S_i;\boldsymbol{\pi}) = \mathbf{A}^{-1}\boldsymbol{b}_i$. Now,

$$\begin{aligned}\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\{\tilde{\mathbf{g}}(S_i;\pi)|\mathbf{x}_i\} &= \mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\{\tilde{\mathbf{g}}(S_i;\pi)|\mathbf{x}_i,\mathbf{g}(S_i)\}\,|\mathbf{x}_i\right]\\ &= \mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\left\{\mathbf{A}^{-1}\boldsymbol{b}_i|\mathbf{x}_i,\mathbf{g}(S_i)\right\}|\mathbf{x}_i\right].\end{aligned}$$

The assumption of non-DIF implies that $\boldsymbol{Y}_i$ is independent of $\mathbf{x}_i$ given S_i, and therefore that

$$\begin{aligned}&= \mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathbf{A}^{-1}\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}\{\boldsymbol{b}_i|\mathbf{g}(S_i)\}\,|\mathbf{x}_i\right]\\ &= \mathrm{E}_{S_i|\mathbf{x}_i}\left(\mathbf{A}^{-1}\begin{bmatrix}\sum_{j=1}^{J}[\sum_{m=1}^{M-1}\mathrm{g}_m(S_i)(\pi_{jm}-\pi_{jM})](\pi_{j1}-\pi_{jM})\\ \vdots\\ \sum_{j=1}^{J}[\sum_{m=1}^{M-1}\mathrm{g}_m(S_i)(\pi_{jm}-\pi_{jM})](\pi_{j,M-1}-\pi_{jM})\end{bmatrix}|\mathbf{x}_i\right)\\ &= \mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathbf{A}^{-1}\mathbf{A}\mathbf{g}(S_i)|\mathbf{x}_i\right]\\ &= \mathrm{E}_{S_i|\mathbf{x}_i}\left[\mathbf{g}(S_i)|\mathbf{x}_i\right].\end{aligned}$$

For the last statement of Lemma 1, we must account for the fact that $\boldsymbol{\pi}$ is estimated rather than known. The marginalization property (Bandeen-Roche et al., 1997), states that the joint distribution of outcomes sampled according to (1) has the latent class form when ignoring covariates, in conjunction with the consistency of maximum likelihood estimation, provides consistent estimates of $\boldsymbol{\pi}$ estimated in Step 1, $\widehat{\pi}\xrightarrow{p}\pi$. Let $\hat{\mathbf{A}}$ and $\hat{\boldsymbol{b}}_i$ be estimates of **A** and $\boldsymbol{b}_i$ obtained by plugging in $\hat{\boldsymbol{\pi}}$ in Equations (10) and (11); because **A** and $\boldsymbol{b}_i$ are continuous in $\boldsymbol{\pi}$ we have $\widehat{\mathbf{A}}\xrightarrow{p}\mathbf{A}$ and $\widehat{\boldsymbol{b}_i}\xrightarrow{p}\boldsymbol{b}_i$. Because all elements of $\boldsymbol{\pi}$ are bounded between 0 and 1, $|\mathbf{A}^{-1}| < K < \infty$ and $\mathrm{E}|\boldsymbol{b}_i| < \infty$, and all elements of both $\hat{\mathbf{A}}$ and $\hat{\boldsymbol{b}}_i$ are uniformly integrable, yielding $\lim_{N\to\infty}\mathrm{E}[\hat{\mathbf{A}}^{-1}\hat{\boldsymbol{b}}_i|\mathbf{g}(S_i)] = \mathrm{E}[\mathbf{A}^{-1}\boldsymbol{b}_i|\mathbf{g}(S_i)]$ (Serfling, 1980).

Theorem 1—Under the conditions of Lemma 1, the proposed LSC method consistently estimates $\boldsymbol{\beta}$.

Proof: In Step 3, $\boldsymbol{\beta}$ is estimated in an $M - 1$ dimensional multivariate normal distribution with a mean function, $\underset{\sim}{\mu(\mathbf{x}_i;\beta)}$, given by (2). The function does not correspond to the full likelihood of the LCR and we do not know the distribution of $\tilde{\mathbf{g}}(S_i)$. However, the differentiated log of the Step 3 regression can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}\left[\tilde{\mathbf{g}}(S_i;\widehat{\pi})-\mu(\mathbf{x}_i;\underset{\sim}{\beta})\right]=0. \quad (12)$$

This estimation function can be rewritten as the sum of two parts, one not accounting for that $\boldsymbol{\pi}$ is estimated and the other reflecting just that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}\left[\tilde{\mathbf{g}}(S_i;\pi)-\mu(\mathbf{x}_i;\underset{\sim}{\beta})\right]+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}\left[\tilde{\mathbf{g}}(S_i;\widehat{\pi})-\tilde{\mathbf{g}}(S_i:\pi)\right]=0,$$

that is, $A_N + B_N = 0$. If B_N is $o_p(1)$, then the estimating equation is mean-asymptotically equivalent to A_N, which by Lemma 1 and theory elucidated by Liang and Zeger (1986) is a consistent GEE under standard regularity conditions bounding $\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}$ and the likelihood. To this end,

$$\begin{aligned}B_N&=\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}\left(\widehat{\mathbf{A}}^{-1}\widehat{\boldsymbol{b}}_i-\mathbf{A}^{-1}\widehat{\boldsymbol{b}}_i+\mathbf{A}^{-1}\widehat{\boldsymbol{b}}_i-\mathbf{A}^{-1}\boldsymbol{b}_i\right)\\&=C_N+D_N.\end{aligned}$$

$C_N=(\widehat{\mathbf{A}}^{-1}-\mathbf{A}^{-1})\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}\widehat{\boldsymbol{b}}_i$ is $o_p(1)$ because $\hat{\mathbf{A}}^{-1}$ is a continuous function of $\hat{\boldsymbol{\pi}}$, $\widehat{\pi}\xrightarrow{p}\pi$, and $|\hat{b}_{im}| \le J$ for each i, m, N.

$D_N=\mathbf{A}^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\sum^{-1}(\widehat{\boldsymbol{b}}_i-\boldsymbol{b}_i)$ is $o_p(1)$ provided regularity conditions bounding $\frac{\partial\mu(\mathbf{x}_i;\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}$ and $\boldsymbol{\Sigma}^{-1}$. Then, $|\hat{b}_{im}| \le J$ for all i, m, N so that $\sup_i|\widehat{b}_{im}-b_{im}|\xrightarrow{p}0$, and hence $D_N\xrightarrow{p}0$.

It can be seen that using the posterior probabilities $q_{im} = \Pr(S_i = m|\boldsymbol{Y}_i = \boldsymbol{y}_i)$ as scores will lead to biased estimation in Step 3 as the conclusion of Lemma 1 is not true in that case. For the case with two classes we have

$$\mathrm{E}_{Y_i|\mathbf{x}_i}\left[\tilde{q}_{im}|\mathbf{x}_i\right]=\mathrm{E}_{Y_i|S_i}\left[\tilde{q}_{im}|S_i=1\right]\Pr\left[S_i=1|\mathbf{x}_i\right]+\mathrm{E}_{Y_i|S_i}\left[\tilde{q}_{im}|S_i=0\right]\Pr\left[S_i=0|\mathbf{x}_i\right]$$

and

$$\mathrm{E}_{S_i|\mathbf{x}_i}\left[S_i|\mathbf{x}_i\right]=\Pr\left[S_i=1|\mathbf{x}_i\right].$$

Thus, to obtain $\mathrm{E}_{\boldsymbol{Y}_i|\mathbf{x}_i}[\tilde{q}_{im}|\mathbf{x}_i] = \mathrm{E}_{S_i|\mathbf{x}_i}[S_i|\mathbf{x}_i]$ it is necessary that $\mathrm{E}_{\boldsymbol{Y}_i|S_i}[\tilde{q}_{im}|S_i = 1] = 1$ and $\mathrm{E}_{\boldsymbol{Y}_i|Si}[\tilde{q}_{im}|S_i = 0] = 0$. As posterior probabilities are restricted to be between zero and one, this is only true if $\tilde{q}_{i0} = 1$ conditional on $S_i = 0$, and $\tilde{q}_{i1} = 1$ conditional on $S_i = 1$, which is a measurement model without error. In summary, the posterior probability-based methods will be increasingly biased the closer the posterior probabilities are to 0.5.

4. Simulation Study

A first set of simulation scenarios was designed to elucidate the small-sample performance of posterior probability-based three-step procedures for estimating the structural parameter, $\boldsymbol{\beta}$, and to compare such with our LSC procedure and with the one-step maximum likelihood method. We evaluated all the posterior probability-based three-step procedures presented above: use of modal or pseudo assignments as responses in a logistic regression on the covariates; correcting the results using modal assignments by the Croon and Bolck method; and use of the posterior probabilities themselves as responses in a multivariate normal model with mean function given by (5). A second set was designed to evaluate the performance of the LSC procedure in more detail. Table 1 describes the models used.

In the first set of studies, we simulated data from four different models and studied how the performance of the methods varied with the number of items, J, number of observations, N, and the precision with which items measured class membership. Precision is high if the conditional probabilities, $\boldsymbol{\pi}$, are clearly distinguished between classes; we designated models with $\boldsymbol{\pi}_{j1} = 0.1$ and $\boldsymbol{\pi}_{j2} = 0.9$, $j = 1, \ldots, J$, as "precise" in measurement, and models with $\boldsymbol{\pi}_{j1} = 0.3$ and $\boldsymbol{\pi}_{j2} = 0.7$, $j = 1, \ldots, J$, as "imprecise", see Models *a–d*, Table 1. These three model features were studied as these have been shown to be important for the factor analysis model (Skrondal & Laake, 2001). We expected that the posterior probability-based methods would be biased for less precise measurement models as described above. In all four studies there were two dichotomous orthogonal covariates, each with a coefficient of 0.5 corresponding to an odds ratio of 1.65; and an intercept $\beta_0 = 0$, which corresponds to an equal number of persons in each class. When applying multiple pseudo-class assignment, 20 draws were used (Wang et al., 2005).

In the second set of studies, we aimed to challenge the naive standard error for the LSC method, and thus compared only the three-step procedure based on the LSC with the one-step procedure. This was done by decreasing the sample size, which influences the standard error of $\hat{\boldsymbol{\pi}}$; adding a class, which raises the risk of misclassification; complicating the measurement component, which again tends to increase the standard error of $\hat{\boldsymbol{\pi}}$; complicating the structural component, which tends to increase the variability of the overall and Step 3 model fits; and finally complicating both the measurement and structural components at the same time. For a detailed description see Models *e–j*, Table 1. When planning the study we had intended to have N = 500 for all scenarios. However, for some models with imprecise measurement, model fitting proved unstable; this occurred for both the one-step procedure and the three-step procedure. To ensure stable convergence of the estimating algorithm, we increased the number of observations for these models.

For each model, 500 data sets were simulated. In Step 3 the standard error used for the parameter estimates is the naive standard error from the Step 3 regression, which does not account for the measurement error in the estimation of the scores of the latent variable. Let $\hat{\boldsymbol{\beta}}$ denote the structural parameter estimator generically with respect to method. For each of the models and methods, six different performance measures are provided:

$\mathrm{E}(\widehat{\beta}) = \frac{1}{500}\sum_{k=1}^{500}\widehat{\beta}_k$; $\mathrm{SD}(\widehat{\beta}) = \sqrt{\frac{1}{499}\sum_{k=1}^{500}(\widehat{\beta}_k - \overline{\widehat{\beta}})^2}$; $\frac{\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})]}{\mathrm{SD}(\widehat{\beta})}$, where $\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})] = \sqrt{\frac{1}{500}\sum_{k=1}^{500}\widehat{\mathrm{var}}(\widehat{\beta}_k)}$ and

$\widehat{\mathrm{var}}(\widehat{\beta})$ is the standard model-based variance estimator from the third step; MSE = $(\mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})^2 + \mathrm{SD}(\hat{\boldsymbol{\beta}})^2$; *Power* = the proportion of Wald confidence intervals excluding zero; and the *Coverage* = the proportion of Wald confidence intervals including the true value.

Generation of data and estimation of the LCMs were done in Mplus (Muthén & Muthén, 2007). Scoring (Step 2) and regression of scores on covariates (Step 3) were carried out in SAS 9.1. For the procedures employing posterior probabilities and LSC scores as responses, Step 3 was carried out using the IML procedure (see the Appendix). In the special case of a two-class model, the GENMOD procedure in SAS was used assuming a normal distribution and a logit link.

4.1. Simulation Study Results

In Table 2 we see the results from the first set of the simulation studies. We report only the results pertaining to the coefficient of $\mathbf{x}_1$; results for the $\mathbf{x}_2$ coefficient and the intercept were similar except as noted. For all four scenarios the LSC method produced nearly unbiased estimates. The Croon and Bolck correction (Modal with correction) also nearly eliminated the bias. For the LSC there was only a minor loss of efficiency. The naive LSC-based standard error from Step 3 only slightly underestimated the true variability, so coverage and power for the LSC results were similar to those from one-step LCR analysis. However, for the intercept, the naive LSC-based standard errors underestimated the empirically observed standard error on average by 45% in the model with imprecise measurement and lowest sample size (Model *b*) (data not shown). The inferential bias of the intercept lessened in scenarios with more items or higher sample size, for instance $\frac{\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})]}{\mathrm{SD}(\widehat{\beta})}=0.885$ for Model *d*.

For all the posterior probability-based methods, we observed the anticipated underestimation of regression parameters. Among these methods, scoring by modal assignment performed best. In the precise scenario (Model *a*), all methods yielded coverage and power that were similar to one-step LCR. In such settings, correct inferences may be drawn even though the covariate effect will be underestimated. With less precise measurement, the underestimation associated with the posterior probability-based methods worsened appreciably, as did coverage and power. Upon changing the number of items from four to eight in the model with imprecise measurement, the coverage and power of the posterior probability-based methods improved slightly, but the estimation remained substantially biased. Increasing the sample size from 500 to 1000, modestly improved power but eroded coverage. The results from the multiple pseudo-class and posterior probability scoring methods were strikingly similar.

In Table 3 results are shown for the first four models from the second set of simulation studies. The LSC method performed very well for all of these models with respect to both estimation and inference. Performance was comparable to LCR except for Models *g* and *h*, where the LSC method showed a slightly larger sampling variability.

Table 4 displays results for Models *i* and *j*. Beginning with Model *i*, there continued to be little bias, but the observed sampling variability was somewhat larger for LSC than for LCR. Moreover, the naive LSC-based standard error tended to underestimate the empirically observed standard error of the intercept, resulting in low associated coverage, 0.856. For the covariates the coverage was similar but the power was slightly smaller for the LSC method compared to LCR. For Model *j*, where both the structural and measurement components of the model were made challenging, there was still no evidence of bias; but LSC-based standard error mis-estimated empirically observed standard error, on average for all covariates. Coverage for the LSC method was appreciably inaccurate.

5. Example: The Salisbury Eye Evaluation Project

The Salisbury Eye Evaluation Project (SEE) was an epidemiological cohort study designed to characterize the contribution of visual impairment to visual and physical disability in older adults. As an aspect of health that is often amenable to intervention and has particular implications for older adults' quality of life, vision is a potentially important risk factor for disability (Rubin, Bandeen-Roche, Prasada-Rao, and Fried, 1994; West, Gildengorin, Haegerstrom-Portnoy, Schneck, Lett, and Brabyn, 2002). Employing a unique combination of state-of-the-art vision assessment with broad-spectrum assessment of disability, SEE is ideally designed to the investigation of vision and physical disability. Here we apply the newly developed methodology to study the association between mobility function in older adults and presenting corrected binocular visual acuity, using LCM. To our knowledge LCM has not previously been applied to study the association at issue; doing so is potentially advantageous because self-report of task difficulty is susceptible to measurement error, and because LCMs acknowledge that mobility function is a comprehensive construct. The analysis we present here simplifies so as to effectively illustrate our methodology.

The SEE study has been previously described (Bandeen-Roche, Huang, Munoz, and Rubin, 1999; West et al., 1997; Munoz, West, Rubin, Schein, Fried, and Bandeen-Roche, 1999). In brief, the SEE cohort is an age and race stratified random sample of Salisbury, Maryland. Inclusion criteria were age between 65 and 84, community residency with better-than-invalid health, and Mini-Mental State Examination score greater than or equal to 18.

We propose to make operational "mobility function" using a construct of "mobility/exercise tolerance" that has previously been studied by LCM in the Women's Health and Aging Study (WHAS) (Bandeen-Roche et al., 1997; Guralnik, Fried, Simonsick, Kasper, and Lafferty, 1995). There, "mobility/exercise tolerance" was measured by five self-reported questions, all of the type, "without help, do you have any difficulty in": walking $\frac{1}{4}$ mile, climbing 10 steps, getting in and out of a bed or chair, doing heavy housework, and lifting up to 10 pounds. All the questions were modeled as binary indicators of any reported difficulty. A three-class model well characterized the empirical data.

In the SEE cohort we have similar questions; climbing 10 steps (*climbing step*), getting in and out of a bed or chair (*in/out chair*), doing heavy housework (*doing hhw*), getting down a 5 pound object (*get down*), walking $\frac{1}{2}$ mile (*difficulty walk*). Using these questions we attempted to reconstruct the latent measure of mobility/exercise tolerance developed in WHAS. The association between this latent mobility/exercise variable and visual acuity was adjusted for age and comorbidity. All independent variables were modeled as binary variables, with visual acuity dichotomized at worse than 0 logMAR (20/20); age, at 75 or greater; and comorbidity, at three or more diseases. All methods described were compared using this model in complete-case data. Due to a high amount of missing values we also compared the LCR and the LSC on data with at most one of the outcomes missing. The LCR was fitted by maximum likelihood under the assumption of missing at random. A similar approach was used for Step 1 in the LSC method; in Step 2, the score was calculated for the available outcomes (four or five); and Step 3 was carried out just as in the case without missing values. To avoid problems with multimodalities we always used 1000 different starting values for the LCM. In case of label-switching between classes, the model was re-fitted with the maximum likelihood estimates of $\boldsymbol{\pi}$ as starting values switched between classes.

5.1. Result of Analysis

The cohort included 2520 persons who agreed to fully participate. There were 2128 (15.56% with missing values) participants with no missing values in the variables to be analyzed, comprising 1201 women and 927 men. Here 33.7% were 75 years of age or older, 42.3% had at least three comorbidities, and 44.3% had visual acuity worse than 0 logMAR. The five mobility/exercise tasks were endorsed as "difficult" for the following percentages: *In/out chair*, 22.2%; *climbing step*, 34.2%; *difficulty walk*, 38.2%; *doing hhw*, 43.1%; and *get down*, 22.1%.

When allowing one of the outcomes to have a missing value there were 2450 eligible participants (2.78% with missing values), comprising 1410 women and 1040 men. Here 35.4% were 75 years of age or older, 42.0% had at least three comorbidities, and 44.8% had visual acuity worse than 0 logMAR. The five mobility/exercise tasks were endorsed as "difficult" for the following percentages: *In/out chair*, 22.1%; *climbing step*, 34.4%; *difficulty walk*, 38.7%; *doing hhw*, 43.0%; and *get down*, 21.9%.

Preliminary analysis persuaded us to model men and women separately. To select the number of classes, we fitted LCMs on the five questions without any covariates (Bandeen-Roche et al., 1997). There exist various methods for adjudicating the number of classes, but a large simulation study recently suggested that the Bayesian information criterion (BIC) (Schwarz, 1978) is superior to frequently employed alternatives (Nylund et al., 2007) and we therefore employed it here. Among women, three- and four-class models resulted in equal BIC. Among men, BIC was smaller for the three-class model than for the four-class model (4238 versus 4261). We therefore present three-class models for both men and women. Both models were empirically identifiable as they satisfied the criteria developed by Huang and Bandeen-Roche (2004): The number of unique parameters in the saturated LCM (=31) is greater than the number of parameters in the LCM (=17); all the parameters fitted were finite; the M vectors $\hat{\boldsymbol{\kappa}}_m$ with probabilities of each possible pattern given class membership were linearly independent; and the design matrix of the covariates had full rank.

Table 5 shows the estimated conditional probabilities, $\hat{\boldsymbol{\pi}}$, from this study (complete-case analysis) and from the WHAS study. In general, the results were similar, with Class 1 exhibiting high probabilities of reporting problems for all tasks; Class 3, low probabilities of reporting any problems for all tasks; and Class 2, probabilities of reporting problems in between Classes 1 and 3. For the men, the conditional probabilities in Class 2 put more weight on difficulties with *in/out chair*, *climbing step* and *difficulty walk* than for women in SEE or WHAS, and less on *get down* and *doing hhw*.

Results from the complete-case analysis are shown in Table 6. As the LCM has three classes the results are interpreted as from a multinomial logit model as described following Equation (2) and given as the log odds ratios for being in Class 1 versus Class 3, and for being in Class 2 versus Class 3. We see that individuals with reduced visual acuity have a significantly higher risk of being in the class with the worst mobility/exercise tolerance (Class 1) compared to the best functioning (Class 3). Both more comorbidities and higher age were associated with a significantly higher risk of being in Classes 1 and 2 compared to Class 3.

Turning to methods comparison: In general the results for the LSC method closely matched those for LCR, with three exceptions that we detail in the next paragraph. The posterior probability-based methods produced estimates that were closer to the null than for LCR and LSC, but yielded inferences that were qualitatively similar. As in the simulations, results from the multiple pseudo-class and posterior probability methods were very similar.

As the exceptions, the LSC coefficient differed from the LCR coefficient for the age estimate both for Class 2 versus Class 3 and for Class 1 versus Class 3 in men, and for the comorbidity estimate for Class 2 versus Class 3 in women. One of the assumptions required for the accuracy of the LSC method is that there is no association between outcome measures and explanatory variables when conditioning on the latent measure, that is, non-DIF. We evaluated this assumption in the LCR by a process of forward selection, allowing for class-specific DIF. A significance level of 1% was used to account for multiple testing. We found that women in Class 1 with more than three comorbidities were more likely to report problems with *doing hhw* and *get down* than women with fewer comorbidities. For the men in Class 2 we found that men 75 years of age or older more often reported problems with *get down* than younger men. The difference between LSC and LCR estimates of the age coefficient could be explained by the differential measurement of *get down*. It should be noted that none of the estimates in Table 6 account for DIF, including LCR.

When analyzing data where participants with one missing outcome were included, class distribution and conditional probabilities were found to be similar to the complete-case analyses both for men and women (data not shown). Again the LCR and the LSC produced similar results. However, coefficients estimates from this analysis differed considerable from those from complete-case analysis in as much as a factor of two, with both attenuation (LCR coefficient for visual acuity in men = 0.39 versus 0.81, Class 1 versus Class 3) and amplification (LCR coefficient for visual acuity in women = 0.91 versus 0.50, Class 1 versus Class 3). The LCR and LSC regression coefficients differed most for women for age (LCR: 0.74 versus LSC: 0.47, Class 2 versus Class 3); an investigation of DIF demonstrated that women 75 years of age or older more often reported problems with *get down* than younger women. There was little difference in the age coefficients between the LCR and the LSC method for men for the analysis with missing values, nor was there evidence of DIF in this case.

6. Discussion

We have proposed a method (least squares class scoring) to assign latent class values for use as responses in subsequent regression analyses. For data distributed according to a latent class regression model, the least squares class method yields consistent estimators for regression coefficients in a three-step procedure. A simulation study found regression coefficients to be approximately unbiased in finite samples. Moreover, there was only a minor loss in efficiency. In general the naive standard error only slightly underestimated the true variability. Therefore, power and coverage for the proposed method were similar to the results from latent class regression analysis. However, for models with cases of low-prevalence classes and models with a number of items with a very skewed outcome distribution, use of naive standard errors in Step 3 mischaracterized sampling variabilities. Application of least squares class scoring in the SEE data set yielded results that were generally close to those of the one-step latent class regression.

A number of methods exist for predicting latent class scores in latent class models but none of these consistently estimate regression coefficients in a subsequent analysis with covariates. Using the modal-assignment rule and analyzing the scores in a logistic regression model resulted in biased regression coefficients, but this method did seem to be the best of the posterior probability based methods. Assigning latent class membership by pseudo- or multiple pseudo-class assignment resulted in worse bias than modal assignment. Using posterior probabilities as scores is dominated by the other methods and should be avoided. Especially in more precise models, the Croon and Bolck correction method (Bolck et al., 2004) successfully removed the bias in the modal-assignment method. However this method is rarely used, perhaps because it is unclear how to handle continuous covariates.

The least squares class method does not assign subjects to one class but gives an $M - 1$ dimensional score for each person. This complicates the interpretation of the scores. Also, even though the scores are restricted to sum to one for each person, they do not have the probability interpretation and can be outside the unit-interval.

The use of the three-step procedure confers an advantage that one does not need to estimate as many parameters in the first step as when fitting a latent class regression. This could be especially advantageous in studies with few observations. Moreover misspecification in one part of a model may affect conclusions related to another part (Bollen, 1996). Multi-stage regression has been reported to reduce this bias (Sánchez, Budtz-Jørgensen, and Ryan, 2009).

For the three-step method to be fully useful, we believe that further work is needed to extend the procedure in three directions. First, work on how to account for differential item functioning is needed. When analyzing the SEE data, the least squares class method yielded estimates notably different from the latent class regression; concurrently there was evidence for differential item functioning for these parameters. Second, work is needed to accommodate available data for cases having some missing values, which often are prevalent. In the SEE data, we investigated data allowing for at most one missing outcome, but in principle the least squares class score can be calculated for records with a least two outcomes. However, a challenge is that in records with few outcomes a larger variation in the least squares class scores is to be expected. A more efficient analysis maybe achieved by using a weighted regression in Step 3 to account for the larger variation in scores. Third, inference reported herein ignored that parameters from Step 1 were estimated. Our findings showed the expected tendency for such inference, i.e., that the naive standard errors tend to underestimate the true value. However, for the range of simulation designs we considered, such inference was primarily adequate. Still as a next step standard errors that fully account for multi-stage uncertainty should be developed. Use of the sandwich or bootstrapping may provide alternatives, resulting in practical methods to obtain better estimates of the standard errors.

In summary, the proposed method is attractive in large studies where one seeks to conduct multiple regression analyses with a latent variable. Here, one can calculate the least squares class scores and use them as responses in subsequent analyses. Our approach gains advantages of consistency over the posterior probability-based approaches and produces surprisingly accurate inferences as well.

[3] Question different than SEE: Without help, do you have any difficulty in lifting up to 10 pounds? (In SEE: Get down a 5 pound object)

Acknowledgments

The authors wish to thank Dr. Sheila West for kindly making the Salisbury Eye Evaluation data available, and the Johns Hopkins Older Americans Independence Center for funding, NIA: P30AG021334.

References

Bandeen-Roche K, Huang GH, Munoz B, Rubin GS. Determination of risk factor associations with questionnaire outcomes: A methods case study. American Journal of Epidemiology. 1999; 150(11): 1165–1178. [PubMed: 10588077]

Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. Latent variable regression for multiple discrete outcomes. Journal of the American Statistical Association. 1997; 92:1375–1386.

Bartolucci F, Forcina A. A class of latent marginal models for capture-recapture data with continuous covariates. Journal of the American Statistical Association. 2006; 101:786–794.

Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: One-step versus three-step estimators. Political Analysis. 2004; 12(1):3–27.

Bollen KA. An alternative two stage least squares (2sls) estimator for latent variable equations. Psychometrika. 1996; 61:109–121.

Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement error in nonlinear models: a modern perspective. Boca Raton: Chapman & Hall/CRC; 2006.

Croon, M. Using predicted latent scores in general latent structure models. In: Marcoulides, GA.; Moustaki, I., editors. Latent variable and latent structure models. Mahwah: Erlbaum; 2002. p. 195-225.

Dayton CM, Macready GB. Concomitant-variable latent-class models. Journal of the American Statistical Association. 1988; 83:173–178.

Funderbuck JS, Maisto SA, Sugarman DE, Wade M. The covariation of multiple risk factors in primary care: a latent class analysis. Journal of Behavioral Medicine. 2008; 31:525–535. [PubMed: 18800242]

Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika. 1974; 61:215–231.

Guralnik, J.; Fried, L.; Simonsick, E.; Kasper, J.; Lafferty, M. The women's health and aging study: health and social characteristics of older women with disability (Tech. Rep. No. NIH Pub. 95-4009). Bethesda MD: National Institute on Aging; 1995.

Huang GH, Bandeen-Roche K. Building an identifiable latent class model with covariates effects on underlying and measured variables. Psychometrika. 2004; 69(1):5–32.

Johnson, RA.; Wichern, DW. Applied multivariate statistical analysis. New York: Pearson Prentice Hall; 2007.

Lazarsfeld, PF.; Henry, NW. Latent structure analysis. Boston: Houghton Mifflin; 1968.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

Lu IRR, Thomas R. Avoiding and correcting bias in score-based latent variable regression with discrete manifest items. Structural Equation Modeling. 2008; 15:462–490.

Markkula R, Jarvinen P, Leino-Arjas P, Koskenvuo M, Kalso E, Kaprio J. Clustering of symptoms associated with fibromyalgia in a finnish twin cohort. The European Journal of Pain. 2009; 13(7): 744–750.

Munoz B, West S, Rubin GS, Schein OD, Fried LP, Bandeen-Roche K, et al. Who participates in population based studies of visual impairment? The Salisbury eye evaluation project experience. Annals of Epidemiology. 1999; 9(1):53–59. [PubMed: 9915609]

Muthén, LK.; Muthén, BO. Mplus, statistical analysis with latent variables, user's guide. 5. Los Angeles: Muthén & Muthén; 2007.

Nylund KL, Asparouhov T, Muthen BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling. 2007; 14:535–569.

Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press; 2003.

Reyes ADL, Henry DB, Tolan PH, Wakschlag LS. Linking informant discrepancies to observed variations in young children's disruptive behavior. Journal of Abnormal Child Psychology. 2009; 37 (5):637–652. [PubMed: 19247829]

Rubin GS, Bandeen-Roche K, Prasada-Rao P, Fried L. Visual impairment and disability in older adults. Optometry & Vision Science. 1994; 71:750–760. [PubMed: 7898882]

Sánchez BN, Budtz-Jørgensen E, Ryan LM. An estimating equations approach to fitting latent exposure models with longitudinal health outcomes. Annals of Applied Statistics. 2009; 3:830–856.

Schorr G, Ulbricht S, Schmidt CO, Baumeister SE, Ruge J, Schumann A, et al. Does precontemplation represent a homogeneous stage category? A latent class analysis on German smokers. Journal of Consulting and Clinical Psychology. 2008; 76(5):840–851. [PubMed: 18837601]

Schwarz G. Estimating the dimension of a model. Annals of Statistics. 1978; 6:461–464.

Serfling, RJ. Approximation theorems of mathematical statistics. New York: Wiley; 1980.

Skrondal A, Laake P. Regression among factor scores. Psychometrika. 2001; 4:563–576.

Steiger H, Richardson J, Schmitz N, Joober R, Israel M, Bruce KR, et al. Association of trait-defined, eating-disorder sub-phenotypes with (biallelic and triallelic) 5httlpr variations. Journal of Psychiatric Research. 2009; 43(13):1086–1094. [PubMed: 19383563]

Unick GJ, Snowden L, Hastings J. Heterogeneity in comorbidity between major depressive disorder and generalized anxiety disorder and its clinical consequences. The Journal of Nervous and Mental Disease. 2009; 197(4):215–224. [PubMed: 19363376]

Wang CP, Brown CH, Bandeen-Roche K. Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. Journal of the American Statistical Association. 2005; 100(471):1054–1076.

West CG, Gildengorin G, Haegerstrom-Portnoy G, Schneck ME, Lett L, Brabyn JA. Is vision function related to physical functional ability in older adults? Journal of the American Geriatrics Society. 2002; 50:136–145. [PubMed: 12028258]

West SK, Munoz B, Rubin GS, Schein OD, Bandeen-Roche K, Zeger SL, et al. Function and visual impairment in a population-based study of older adults: The SEE project. Investigative Ophthalmology & Visual Science. 1997; 38(1):72–82. [PubMed: 9008632]

Appendix

SAS program to perform Step 3. This program performs maximum likelihood estimation in the Step 3 regression.

```
proc iml;
use date1;
read all var {c_1,..., c_j−1)} into y [colname=varname];
read all var {x_1,..., x_P} into x [colname=covname];
close date1;
/*Make function for calculating −log(likelihood)*/
start lik(th) global(y,x);
apar=(ncol(x)+1)#ncol(y)+ncol(y)#(ncol(y)+1)/2; /*Number of parameters*/
beta=shape(th[1,1:ncol(y)*(ncol(x)+1)],ncol(x)+1,ncol(y));/*beta estimates*/
s=sqrsym(th[(ncol(x)+1)*(ncol(y))+1:apar]); /*read variance matrix*/
/*Mean Structure*/
xcov=j(nrow(y),1)||x;
xbe=j(nrow(y),nrow(beta),.);
xbeta=j(nrow(y),ncol(y),.);
DO i=1 to ncol(y);
xbe=repeat(shape(beta[,i],1,nrow(beta)),nrow(xcov),1)#xcov;
xbeta[,i]=exp(xbe[,+]);
END;
mu=j(nrow(y),ncol(y),.);
mu=xbeta/(1+repeat(xbeta[,+],1,ncol(y)));
/*log likelihood*/
li=(nrow(y)/2)#log(det(s))+0.5#((((y−mu)*inv(s))#(y−mu))[,+])[+];
return(li);
finish;
apar=(ncol(x)+1)#ncol(y)+ncol(y)#(ncol(y)+1)/2; /*Number of parameters*/
th0=shape({0 0.5 0.5 0.3},1,apar); /*Starting values*/
optn={0 2}; /*0 means minimized 2 is what output will be printed*/
call nlpnra(rc,th1,‘lik’,th0,optn); /*Making Newton Raphson,*/
CALL NLPFDD( f, g, h, “lik”, th1); /*Calculates hessian matrix*/
stderr = sqrt(abs(vecdiag(inv(h))))‘;
print th1; print res;
quit;
```

In the line starting with “th0”, one have to give some starting values for the parameters. These have to be given in following order: IS_1, …, IS_{M-1}, x_1S_1, …, x_1S_{M-1}, x_2S_1, …, x_PS_{M-1}, var (S_1), cov(S_1, S_2), var(S_2), cov(S_1, S_3), cov(S_2, S_3), var(S_3), …, var(S_{M-1}), where IS_m is the intercept for class m and x_pS_m is the parameter estimate of covariate p on class m. We recommend that one tries serval different sets of starting values, as local maxima can be reached. As starting values for the covariance matrix for the LSC scores, we suggest the empirical covariance matrix for the LSC predictions.

Table 1

Specifications for each of the models considered in the simulation study. N = number of observations, J = number of items, M = number of classes, $\boldsymbol{\pi}$ = matrix with the conditional probabilities, β_0 and $\boldsymbol{\beta}$ are, respectively, the intercept and the regression coefficients from the regression of the latent classes on covariates. Both β_0 and $\boldsymbol{\beta}$ are given in the log odds scale.

	N	J	M	$\boldsymbol{\pi}^t$	β_0	$\boldsymbol{\beta}$
First set of simulation studies						
Model *a*	500	4	2	$\begin{pmatrix} .1 & .1 & .1 & .1 \\ .9 & .9 & .9 & .9 \end{pmatrix}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *b*	1000	4	2	$\begin{pmatrix} .3 & .3 & .3 & .3 \\ .7 & .7 & .7 & .7 \end{pmatrix}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *c*	500	8	2	$\begin{pmatrix} .3 & \dots & .3 \\ .7 & \dots & .7 \end{pmatrix}_{2\times 8}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *d*	1000	8	2	$\begin{pmatrix} .3 & \dots & .3 \\ .7 & \dots & .7 \end{pmatrix}_{2\times 8}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Second set of simulation studies						
Model *e*	250	4	2	$\begin{pmatrix} .1 & .1 & .1 & .1 \\ .9 & .9 & .9 & .9 \end{pmatrix}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *f*	500	6	3	$\begin{pmatrix} .1 & .1 & .1 & .1 & .1 & .1 \\ .1 & .1 & .1 & .9 & .9 & .9 \\ .9 & .9 & .9 & .9 & .9 & .9 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$
Model *g*	3000	4	2	$\begin{pmatrix} .2 & .2 & .2 & .2 \\ .8 & .8 & .8 & .8 \end{pmatrix}$	2	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *h*	3000	4	2	$\begin{pmatrix} .01 & .01 & .1 & .1 \\ .5 & .5 & .7 & .7 \end{pmatrix}$	0	$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$
Model *i*	3000	4	2	$\begin{pmatrix} .01 & .01 & .1 & .1 \\ .5 & .5 & .7 & .7 \end{pmatrix}$	2	$\begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$

	N	*J*	*M*	$\boldsymbol{\pi}^t$	$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}$
Model *j*	3000	4	2	$\begin{pmatrix} .01 & .01 & .1 & .1 \\ .5 & .5 & .7 & .7 \end{pmatrix}$	2	$\begin{pmatrix} -2 & -0.5 & 1 & 3 \end{pmatrix}$

Table 2

Results from the first set of simulations: Comparison of standard three-step procedures and the LSC procedure with LCR. Results for coefficient of $\mathbf{x}_1$ are shown; results for coefficient of $\mathbf{x}_2$ were similar.

Method	**E($\hat{\beta}$)**	**SD($\hat{\beta}$)**	$\frac{\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})]}{\mathrm{SD}(\widehat{\beta})}$	**MSE**	**Power**	**Coverage**
Model *a* (Precise, 4 items, 500 observations)						
LCR	0.507	0.198	0.988	0.039	0.742	0.954
LSC	0.506	0.209	0.972	0.044	0.708	0.946
Modal	0.484	0.189	1.002	0.036	0.714	0.950
Modal with correction	0.505	0.197	*2*	0.039	*2*	*2*
Pseudo	0.473	0.193	0.976	0.038	0.690	0.952
Multiple pseudo	0.472	*1*	*1*	0.035	0.686	0.960
Posterior probabilities	0.474	0.185	0.988	0.035	0.740	0.946
Model *b* (Imprecise model, 4 items, 1000 observations)						
LCR	0.515	0.219	0.983	0.048	0.680	0.956
LSC	0.509	0.214	0.982	0.046	0.684	0.958
Modal	0.304	0.142	0.961	0.059	0.612	0.680
Modal with correction	0.487	0.229	*2*	0.053	*2*	*2*
Pseudo	0.238	0.137	0.973	0.087	0.428	0.502
Multiple pseudo	0.221	*1*	*1*	0.087	0.164	0.676
Posterior probabilities	0.221	0.093	0.954	0.087	0.690	0.114
Model *c* (Imprecise model, 8 items, 500 observations)						
LCR	0.490	0.242	1.000	0.059	0.520	0.970
LSC	0.485	0.250	0.989	0.063	0.510	0.952
Modal	0.369	0.191	0.990	0.054	0.480	0.888
Modal with correction	0.454	0.236	*2*	0.058	*2*	*2*
Pseudo	0.327	0.191	0.978	0.067	0.410	0.832
Multiple pseudo	0.315	*1*	*1*	0.058	0.262	0.936
Posterior probabilities	0.316	0.155	0.981	0.058	0.544	0.754
Model *d* (Imprecise model, 8 items, 1000 observations)						

Method	**E($\hat{\boldsymbol{\beta}}$)**	**SD($\hat{\boldsymbol{\beta}}$)**	$\frac{\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})]}{\mathrm{SD}(\widehat{\beta})}$	**MSE**	**Power**	**Coverage**
LCR	0.488	0.164	1.022	0.027	0.832	0.954
LSC	0.486	0.174	1.003	0.030	0.794	0.944
Modal	0.373	0.131	1.025	0.033	0.790	0.834
Modal with correction	0.474	0.166	*2*	0.028	*2*	*2*
Pseudo	0.326	0.128	1.028	0.047	0.698	0.760
Multiple pseudo	0.317	*1*	*1*	0.045	0.564	0.856
Posterior probabilities	0.316	0.106	1.006	0.045	0.840	0.596

[1] Not calculated as empirical SD does not take into account that $\boldsymbol{\beta}$ is an average of multiple draws

[2] Not calculated as no standard error is available

Table 3

Results from the first four models of the second set of simulations: Comparison of the LSC method with the LCR. For the model with three classes, results for being in Class 1 versus Class 3 are given.

	Intercept		X_1		X_2	
	LCR	**LSC**	**LCR**	**LSC**	**LCR**	**LSC**
Model *e* (250 observations, precise)						
$E(\hat{\beta})$	−0.003	−0.000	0.536	0.541	0.485	0.475
$SD(\hat{\beta})$	0.229	0.237	0.284	0.292	0.281	0.297
$\frac{E[\widehat{SE}(\widehat{\beta})]}{SD(\widehat{\beta})}$	1.029	0.992	0.984	0.993	0.993	0.975
MSE	0.053	0.056	0.082	0.087	0.079	0.089
Power	0.032	0.032	0.498	0.448	0.382	0.362
Coverage	0.968	0.968	0.950	0.946	0.942	0.952
Model *f* (3 classes, precise)						
$E(\hat{\beta})$	−0.009	−0.008	0.512	0.512	0.515	0.512
$SD(\hat{\beta})$	0.140	0.145	0.181	0.193	0.166	0.174
$\frac{E[\widehat{SE}(\widehat{\beta})]}{SD(\widehat{\beta})}$	1.033	0.978	0.973	0.961	1.064	1.062
MSE	0.020	0.021	0.033	0.037	0.028	0.030
Power	0.060	0.064	0.826	0.788	0.852	0.808
Coverage	0.940	0.936	0.948	0.940	0.970	0.966
Model *g* (Skewed intercept, β_0)						
$E(\hat{\beta})$	2.006	2.009	0.514	0.520	0.504	0.505
$SD(\hat{\beta})$	0.156	0.167	0.177	0.228	0.185	0.238
$\frac{E[\widehat{SE}(\widehat{\beta})]}{SD(\widehat{\beta})}$	1.044	0.863	1.043	1.021	0.997	0.974
MSE	0.024	0.028	0.032	0.053	0.034	0.057
Power	1.000	1.000	0.832	0.640	0.792	0.612

	Intercept		X_1		X_2	
	LCR	**LSC**	**LCR**	**LSC**	**LCR**	**LSC**
Coverage	0.952	0.904	0.960	0.962	0.954	0.958
Model *h* (Skewed marginal distribution of ***Y***)						
$\mathrm{E}(\hat{\beta})$	−0.001	−0.002	0.498	0.497	0.501	0.502
$\mathrm{SD}(\hat{\beta})$	0.080	0.086	0.089	0.096	0.084	0.088
$\frac{\mathrm{E}[\widehat{\mathrm{SE}}(\widehat{\beta})]}{\mathrm{SD}(\widehat{\beta})}$	1.016	0.874	0.979	0.965	1.035	1.042
MSE	0.006	0.007	0.008	0.009	0.007	0.008
Power	0.032	0.082	1.000	1.000	1.000	1.000
Coverage	0.968	0.918	0.944	0.932	0.958	0.958

Table 4

Results from Models *i* and *j* from the second set of simulations: Comparison of LSC method with LCR.

	Intercept		**X_1**		**X_2**		**X_3**		**X_4**	
	LCR	***LSC***	**LCR**	**LSC**	**LCR**	**LSC**	**LCR**	**LSC**	**LCR**	**LSC**
Model *i*										
$E(\hat{\beta})$	2.004	2.004	0.497	0.512	0.518	0.524	0.509	0.519	0.507	0.514
$SD(\hat{\beta})$	0.202	0.239	0.199	0.243	0.218	0.270	0.208	0.273	0.208	0.258
$\frac{E[\widehat{SE}(\widehat{\beta})]}{SD(\widehat{\beta})}$	0.994	0.774	1.046	1.019	0.959	0.932	1.002	0.922	1.002	0.966
MSE	0.041	0.057	0.040	0.059	0.048	0.073	0.043	0.075	0.043	0.067
Power	1.000	1.000	0.702	0.578	0.678	0.530	0.686	0.564	0.696	0.592
Coverage	0.946	0.856	0.958	0.960	0.954	0.938	0.946	0.932	0.948	0.938
Model *j*										
$E(\hat{\beta})$	2.004	2.016	−2.004	−2.014	−0.508	−0.510	1.017	1.013	3.015	3.030
$SD(\hat{\beta})$	0.177	0.194	0.162	0.188	0.141	0.166	0.143	0.162	0.226	0.281
$\frac{E[\widehat{SE}(\widehat{\beta})]}{SD(\widehat{\beta})}$	0.973	0.871	1.015	0.889	1.002	0.711	1.015	0.756	1.027	1.280
MSE	0.031	0.038	0.026	0.035	0.020	0.028	0.021	0.026	0.051	0.080
Power	1.000	1.000	1.000	1.000	0.932	0.954	1.000	1.000	1.000	1.000
Coverage	0.940	0.926	0.956	0.916	0.952	0.832	0.960	0.876	0.962	0.982

Table 5

WHAS and SEE cohorts: Conditional probabilities ($\boldsymbol{\pi}_1 - \boldsymbol{\pi}_3$) for reporting difficulty with tasks. The estimated class distribution is also given.

	WHAS[1]			SEE, Women			SEE, Men		
	$\hat{\boldsymbol{\pi}}_1$	$\hat{\boldsymbol{\pi}}_2$	$\hat{\boldsymbol{\pi}}_3$	$\hat{\boldsymbol{\pi}}_1$	$\hat{\boldsymbol{\pi}}_2$	$\hat{\boldsymbol{\pi}}_3$	$\hat{\boldsymbol{\pi}}_1$	$\hat{\boldsymbol{\pi}}_2$	$\hat{\boldsymbol{\pi}}_3$
In/out chair	0.60	0.18	0.02	0.67	0.14	0.05	0.77	0.26	0.07
Climbing step	0.95	0.30	0.02	0.93	0.35	0.09	0.96	0.51	0.04
Difficulty walk	0.96[2]	0.65[2]	0.08[2]	0.93	0.45	0.04	1.00	0.76	0.03
Get down	0.84[3]	0.46[3]	0.02[3]	0.68	0.30	0.00	0.66	0.14	0.04
Doing hhw	0.96	0.75	0.07	0.94	0.64	0.06	1.00	0.55	0.14
Class %	15%	28%	57%	27%	37%	35%	8%	28%	63%

[1] Results directly from Bandeen-Roche et al. (1997)

[2] Question different than SEE: Without help, do you have any difficulty in walking $\frac{1}{4}$ mile? (In SEE: $\frac{1}{2}$ mile)

Table 6

SEE cohort: Estimated association expressed as log odds ratios (95% confidence limits) between covariates and the three classes of mobility/exercise tolerance.

Method	Intercept	Visual Acuity	Comorbidity	Age
Women, Class 1 versus Class 3				
LCR	−2.22 (−2.80; −1.64)	0.50 (0.08; 0.91)	2.78 (2.29; 3.26)	1.14 (0.69; 1.58)
LSC	−1.93 (−2.39; −1.48)	0.44 (0.04; 0.83)	2.81 (2.10; 3.53)	1.19 (0.74; 1.65)
Modal	−1.51 (−1.78; −1.24)	0.31 (−0.01; 0.63)	2.11 (1.78; 2.44)	0.92 (0.58; 1.25)
Modal with correction	−1.87[1]	0.40[1]	2.57[1]	1.05[1]
Pseudo	−1.21 (−1.46; −0.96)	0.37 (0.07; 0.67)	1.89 (1.58; 2.21)	0.61 (0.18; 1.04)
Multiple pseudo	−1.50 (−1.81; −1.19)	0.34 (0.00; 0.69)	1.98 (1.63; 2.32)	0.84 (0.45; 1.22)
Posterior probabilities	−1.46 (−1.70; −1.21)	0.35 (0.09; 0.62)	1.93 (1.64; 2.23)	0.78 (0.50; 1.06)
Women, Class 2 versus Class 3				
LCR	−0.83 (−1.28; −0.37)	0.19 (−0.21; 0.59)	1.58 (1.14; 2.02)	1.05 (0.64; 1.47)
LSC	−0.91 (−1.38; −0.43)	0.21 (−0.33; 0.76)	1.98 (1.05; 2.90)	0.88 (0.30; 1.47)
Modal	−0.76 (−0.99; −0.54)	0.11 (−0.19; 0.41)	1.15 (0.84; 1.47)	0.73 (0.42; 1.05)
Modal with correction	−0.72[1]	0.14[1]	1.46[1]	0.97[1]
Pseudo	−0.58 (−0.79; −0.36)	0.16 (−0.13; 0.44)	0.95 (0.64; 1.26)	0.34 (−0.09; 0.76)
Multiple pseudo	−0.76 (−1.03; −0.48)	0.14 (−0.20; 0.47)	1.04 (0.66; 1.41)	0.63 (0.27; 1.00)
Posterior probabilities	−0.76 (−0.91; −0.60)	0.13 (−0.08; 0.35)	1.06 (0.81; 1.32)	0.63 (0.40; 0.87)
Men, Class 1 versus Class 3				
LCR	−3.42 (−4.38; −2.46)	0.81 (0.19; 1.43)	1.81 (1.17; 2.44)	0.72 (0.10; 1.34)
LSC	−3.81 (−4.93; −2.69)	0.73 (0.09; 1.36)	1.86 (0.88; 2.84)	1.15 (0.46; 1.85)
Modal	−2.41 (−2.77; −2.05)	0.41 (0.03; 0.78)	1.29 (0.92; 1.67)	0.58 (0.20; 0.97)
Modal with correction	−3.49[1]	0.66[1]	1.65[1]	0.89[1]
Pseudo	−2.43 (−2.80; −2.07)	0.40 (0.02; 0.78)	1.33 (0.95; 1.72)	0.56 (0.17; 0.95)
Multiple pseudo	−2.47 (−2.85; −2.09)	0.41 (0.00; 0.81)	1.32 (0.91; 1.73)	0.56 (0.14; 0.97)
Posterior probabilities	−2.52 (−2.93; −2.11)	0.45 (0.11; 0.78)	1.31 (0.93; 1.70)	0.61 (0.27; 0.95)
Men, Class 2 versus Class 3				
LCR	−1.53 (−1.98; −1.08)	0.31 (−0.10; 0.72)	1.27 (0.87; 1.67)	0.65 (0.19; 1.11)
LSC	−1.59 (−2.07; −1.11)	0.44 (−0.08; 0.96)	1.19 (0.64; 1.74)	0.07 (−0.57; 0.71)
Modal	−1.89 (−2.20; −1.58)	0.20 (−0.17; 0.56)	0.91 (0.55; 1.26)	0.30 (−0.08; 0.68)
Modal with correction	−1.61[1]	0.27[1]	1.17[1]	0.41[1]
Pseudo	−1.85 (−2.15; −1.55)	0.23 (−0.12; 0.58)	0.88 (0.54; 1.22)	0.50 (0.13; 0.86)
Multiple pseudo	−1.86 (−2.20; −1.51)	0.22 (−0.21; 0.66)	0.90 (0.51; 1.29)	0.35 (−0.08; 0.78)
Posterior probabilities	−1.81 (−2.06; −1.57)	0.23 (−0.04; 0.50)	0.84 (0.58; 1.11)	0.35 (0.07; 0.63)

[1] Confidence limits not calculated as no standard error is available