Power and Sample Size Computation for Wald Tests in Latent Class Models

Dereje W. Gudicha
Tilburg University
The Netherlands

Fetene B. Tekle
Janssen Pharmaceutica
Belgium

Jeroen K. Vermunt
Tilburg University
The Netherlands

Abstract

Latent class (LC) analysis is used by social, behavioral, and medical science researchers among others as a tool for clustering (or unsupervised classification) with categorical response variables, for analyzing the agreement between multiple raters, for evaluating the sensitivity and specificity of diagnostic tests in the absence of a gold standard, and for modeling heterogeneity in developmental trajectories. Despite the increased popularity of LC analysis, little is known about statistical power and required sample size in LC modeling. This paper shows how to perform power and sample size computations in LC models using Wald tests for the parameters describing association between the categorical latent variable and the response variables. Moreover, the design factors affecting the statistical power of these Wald tests are studied. More specifically, we show how design factors which are specific for LC analysis, such as the number of classes, the class proportions, and the number of response variables, affect the information matrix. The proposed power computation approach is illustrated using realistic scenarios for the design factors. A simulation study conducted to assess the performance of the proposed power analysis procedure showed that it performs well in all situations one may encounter in practice.

Keywords: Latent Class Models; Sample Size; Statistical Power; Information matrix; Wald test; Design Factor.

1 Introduction

Latent class (LC) analysis was initially introduced in the 1950s by Lazarsfeld (1950) as a tool for identifying subgroups of individuals giving similar responses to sets of dichotomous attitude questions. It took another two decades till LC analysis started attracting the attention of other statisticians. Since then, various important extensions of the original LC model have been proposed, such as models for polytomous responses, models with covariates, models with multiple latent variables, and models with parameter constraints (Goodman, 1974; Dayton and Macready, 1976; Formann, 1982; McCutcheon, 1987; Dayton and Macready, 1988; Formann, 1992; Vermunt, 1996; Magidson and Vermunt, 2004). More recently, statistical software for LC analysis has become generally available – e.g., Latent GOLD (Vermunt and Magidson, 2013b), Mplus (Muthén and Muthén, 2012), LEM (Vermunt, 1997), the SAS routine PROC LCA (Lanza et al., 2007), and the R package poLCA (Linzer and Lewis, 2011) – which has contributed to the increased popularity of this method among applied researchers. Applications of LC analysis include building typologies of respondents based on social survey data (McCutcheon, 1987), identifying subgroups based on health risk behaviors (Collins and Lanza, 2010), identifying phenotypes of stalking victimization (Hirtenlehner et al., 2012), and finding symptom subtypes of clinically diagnosed disorders (Keel et al., 2004). Applications which are specific for medical research include the estimation of the sensitivity and specificity of diagnostic tests in the absence of a gold standard (Rindskopf and Rindskopf, 1986; Yang and Becker, 1997) and the analysis the agreement between raters (Uebersax and Grove, 1990).

Despite the increased popularity of LC analysis in a broad range of research areas, no specific attention has been paid to power analysis for LC models. However, as in the application of other statistical methods, users of LC models wish to confirm the validity of their research hypotheses. This requires that a study has sufficient statistical power; that is, that it is able to confirm a research hypothesis when it is true. Also reviewers of journal publications and research grant proposals often request sample size and/or power computations (Nakagawa and Foster, 2004). However, in the literature on LC analysis, methods for sample size and/or power computation as well as a thorough study on the design factors affecting the power of statistical tests used in LC analysis are lacking.

In this paper, we present a method for assessing the power of tests related to the class-specific response probabilities, which are the parameters of main interest in LC analysis. Relevant tests include tests for whether response

probabilities are equal across latent classes, whether response probabilities are equal to specific values, whether response probabilities are equal across response variables (indicators), and whether sensitivities or specificities are equal across indicators (Goodman, 1974; Holt and Macready, 1989; Vermunt, 2010b). Since the class-specific response probabilities are typically parameterized using logit equations (Formann, 1992; Vermunt, 1997), as in logistic regression analysis, hypotheses about these LC model parameters can be tested using Wald tests (Agresti, 2002). The proposed power analysis method is therefore referred to as a Wald based power analysis.

For logistic regression models, Demidenko (2007, 2008) and Whittemore (1981) described the large-sample approximation for the power of the Wald test. In this paper, we show how to use this procedure in the context of LC analysis. An important difference compared to standard logistic regression analysis is that in a LC analysis the predictor in the logistic models for the responses, the latent class variable, is unobserved. This implies that the uncertainty about the individuals' class memberships should be taken into account in the power and sample size computation. As will be shown, factors affecting this uncertainty include the number of classes, the class proportion, the strength of the association between classes and indicator variables, and the number of indicator variables involved (Collins and Lanza, 2010; Vermunt, 2010a).

The remainder of this paper is organized as follows. Section 2 presents the LC model for dichotomous responses and discusses the relevant hypotheses for the parameters of the LC model. Section 3 discusses power computation for Wald tests in LC analysis and, moreover, shows how LC specific design factors affect the power via the information matrix. Section 4 presents a numerical study in which we assess the performance of the proposed method and illustrates power/sample size computation for different scenarios of the relevant design factors. In the last section, we provide a brief discussion of the main results of our study.

2 The LC model

The LC model is a probabilistic clustering or unsupervised classification model for dichotomous or categorical response variables (Goodman, 1974; McCutcheon, 1987; Hagenaars, 1988; Magidson and Vermunt, 2004; Vermunt, 2010b). Taking the dichotomous case as an example, let y_{ij} be the value of response pattern i for the binary variable Y_j, for $j = 1, 2, 3, ..., p$, where $y_{ij} = 1$ represent a positive response and 0 a negative response. We

denote the full-response vector by $\mathbf{y}_i$. For example, for $p = 3$, $\mathbf{y}_i$ takes on one of the following eights triplets of 0 and/or 1's:

$$\{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}.$$

The three response variables could, for example, represent the answers to the following questions: "Do you support gay marriage?", "Do you support a raise of minimum wages?", and "Do you support the initiative for health care reform?". In a sample of size n persons, a particular person could answer these questions with 'no', 'yes', and 'yes', respectively, in which case the response pattern for this subject becomes $(0,1,1)$. In such an application, the aim of the analysis would be to determine whether one can identify two latent classes with different response tendencies (say republicans and democrats), and subsequently to classify subjects into one of these classes based on their observed responses.

In general, for p dichotomous response variables, we have 2^p tuples of 0 and/or 1's. We denote the number of individuals with response pattern $\mathbf{y}_i$ by n_i, where the total sample size $n = \sum_{i=1}^{2^p} n_i$. The LC model assumes that the response probabilities depend on a discrete latent variable, which we denote by X with categories $t = 1, 2, 3, ..., c$. The probability of having response pattern $\mathbf{y}_i$ is modeled as a mixture of c class-specific probability functions (Dayton and Macready, 1976; Goodman, 1974; McCutcheon, 1987; McLachlan and Peel, 2000; Vermunt, 2010b). That is,

$$P(\mathbf{y}_i, \boldsymbol{\Psi}) = \sum_{t=1}^{c} P(X = t)P(\mathbf{Y} = \mathbf{y}_i/X = t), \tag{1}$$

where $P(X = t)$, which we also denote by π_t, represents the relative size of class t, and $P(\mathbf{Y} = \mathbf{y}_i/X = t)$ is the corresponding class-specific joint response probability. The class-specific probabilities for binary variable Y_j is usually modeled using a logistic parameterization; that is, $\theta_{jt} = P(Y_j = 1/X = t) = \frac{\exp{(\beta_{jt})}}{1+\exp{(\beta_{jt})}}$, where β_{jt} is the log-odds of giving a positive response on item j in class t. Moreover, assuming that the response variables are independent within classes – which is referred to as the local independence assumption – the LC model represented by equation (1) can be rewritten as follows:

$$P(\mathbf{y}_i, \boldsymbol{\Psi}) = \sum_{t=1}^{c} \pi_t \prod_{j=1}^{p} \left(\frac{\exp{(\beta_{jt})}}{1+\exp{(\beta_{jt})}} \right)^{y_{ij}} \left(1 - \frac{\exp{(\beta_{jt})}}{1+\exp{(\beta_{jt})}} \right)^{1-y_{ij}}, \tag{2}$$

4

where π_t is such that $0 < \pi_t < 1$ and $\sum_{t=1}^{c} \pi_t = 1$. The vector of parameter $\boldsymbol{\Psi}$ consists of the sub-vector $\boldsymbol{\pi}$, the class proportions, and the sub-vector $\boldsymbol{\beta}$, the class-specific logits for the indicator variables. For example, for $c = 2$ and $p = 3$, the parameter vector will be: $\boldsymbol{\Psi}' = (\boldsymbol{\pi}', \boldsymbol{\beta}') = (\pi_1, \beta_{11}, \beta_{21}, \beta_{31}, \beta_{12}, \beta_{22}, \beta_{32})$. In the application presented above, these parameters would correspond to the proportion of 'republicans', the log-odds of a republican responds 'yes' instead of 'no' to questions Y_1, Y_2, and Y_3, and the log-odds of a democrat responding 'yes' instead of 'no' questions Y_1, Y_2, and Y_3.

In general, for a LC model having c classes and p binary indicator variables, we have $m = c - 1 + c \cdot p$ free model parameters. These parameters are usually estimated by maximum likelihood (ML) (Dayton and Macready, 1976; Goodman, 1974; McLachlan and Peel, 2000; Vermunt, 2010b), which involves seeking the values of $\boldsymbol{\Psi}$, say $\hat{\boldsymbol{\Psi}}$ which maximize the log-likelihood function:

$$l(\boldsymbol{\Psi}) = \sum_{i=1}^{2^p} n_i \log P(\mathbf{y}_i, \boldsymbol{\Psi}). \tag{3}$$

Maximizing the log-likelihood function in (3) produces a unique estimate for $\boldsymbol{\Psi}$, provided that the LC model in equation (1) is identifiable. Identification is always an issue in LC modeling (Goodman, 1974). For example, the 2-class model with 3 dichotomous response variables we used as an example is exactly identified. With p equal to 3, it is therefore not possible to have more than 2 classes. In fact, one needs 5 dichotomous responses to get an identified model with 3 latent classes. For the scenarios considered in this paper (with p equal to 6 or 10), conditions for local identification as described by Goodman (1974) are satisfied.

Typically, researchers using LC models do not only wish to obtain point estimate for the $\boldsymbol{\Psi}$ parameters, but are also interested in tests concerning these parameters. For simplicity of exposition we will focus here on a single type of test, which in most applications is the test of main interest. That is, the test to determine whether there is a significant association between the latent classes and a particular outcome variable. Inference regarding this association involves testing the null hypothesis stating that the response logit does not differ across latent classes for the indicator variable concerned. This null hypothesis can be formulated as $H_0 : \beta_{j1} = \beta_{j2} = ... = \beta_{jc}$, for $j = 1, 2, 3, ..., p$. An equivalent formulation of this hypothesis is

$$H_0 : \beta_{j1} - \beta_{j2} = 0$$
$$\beta_{j1} - \beta_{j3} = 0$$

$$.$$
$$.$$
$$.$$

$$\beta_{j1} - \beta_{jc} = 0.$$

Or, using matrix notation, as $H_0 : \boldsymbol{H}\boldsymbol{\beta}_j = 0$, where $\boldsymbol{H}$ is a $c-1$ by c design matrix with linear contrasts and $\boldsymbol{\beta}_j$ is a c by 1 column vector with the parameters for Y_j, i.e., $\boldsymbol{\beta}_j' = (\beta_{j1}, \beta_{j2}, ..., \beta_{jc})$. Under the null hypothesis of no association, the difference $\beta_{j1} - \beta_{jt}$ occurs by chance alone, implying that the indicator does not contribute to the definition of classes in a statistically significant way.

As already indicated in the introduction section, various other types of hypotheses concerning the class-specific response logits may be of interest. Examples include tests for whether β_{jt} is equal to a particular value (e.g., $\beta_{11} = 1$), whether the β_{jt} parameters are equal across two or more items (e.g., $\beta_{1t} - \beta_{2t} = 0$), and whether the value is the opposite of the value for another class (e.g., $\beta_{11} + \beta_{12} = 0$) (Goodman, 1974). In medical research, we may be interested in comparing the sensitivity and specificity of diagnostic tests (see, for example Yang and Becker (1997)), yielding hypotheses such as $\beta_{11} - \beta_{21} = 0$ and $\beta_{12} - \beta_{22} = 0$, respectively. Note that all these hypothesis are special cases of the general form $\boldsymbol{H}\boldsymbol{\beta} = 0$.

3 Wald based power analysis for LC models

3.1 The Wald statistic and its asymptotic properties

One of the properties of the ML estimator is that, under certain regularity conditions (McHugh, 1956; Birch et al., 1964; White, 1982; Hausman and McFadden, 1984), the estimator $\hat{\boldsymbol{\Psi}}$ converges in probability to $\boldsymbol{\Psi}$ as the sample size tends to infinity. That is, for any sequence $\hat{\boldsymbol{\Psi}}_n$ we have $\hat{\boldsymbol{\Psi}}_n \xrightarrow{a.s.} \boldsymbol{\Psi}$. The other interesting property of the ML estimator is that it has a limiting normal distribution. More specifically, for large sample size n,

$$\sqrt{n}(\hat{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}) \longrightarrow N(\boldsymbol{0}, \mathbf{V}), \qquad (4)$$

where $\longrightarrow$ denotes convergence in distribution, $\mathbf{V} = \mathbf{I}^{-1}(\boldsymbol{\Psi})$ is the asymptotic co-variance of $\sqrt{n}\hat{\boldsymbol{\Psi}}_n$, and $\mathbf{I}(\boldsymbol{\Psi})$ is the m by m information matrix (McHugh, 1956; Redner, 1981; Rencher, 2000; Wald, 1943; Wolfe, 1970). The latter has the following block structure:

$$\mathbf{I}(\boldsymbol{\Psi}) = \begin{bmatrix} \mathbf{I}_1 = \{(\pi_t, \pi_s)\} & \mathbf{I}_2 = \{(\pi_t, \beta_{js})\} \\ \mathbf{I}_3 = \{(\beta_{jt}, \pi_s)\} & \mathbf{I}_4 = \{(\beta_{jt}, \beta_{ks})\} \end{bmatrix},$$

for $t, s = 1, 2, 3, ..., c$ and $k, j = 1, 2, 3, ..., p$. The sub-matrices $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}_3$, and $\mathbf{I}_4$ are of dimensions $c - 1$ by $c - 1$, $c - 1$ by $c \cdot p$, $c \cdot p$ by $c - 1$, and $c \cdot p$ by $c \cdot p$, respectively. The terms between braces indicate the parameters involved in the sub-matrix concerned.

Using the algebraic properties of block matrices, it follows that

$$\mathbf{V} = \mathbf{I}^{-1}(\boldsymbol{\Psi}) = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{I}_1^{-1}\mathbf{I}_2\mathbf{B}^{-1} \\ -\mathbf{I}_4^{-1}\mathbf{I}_3\mathbf{A}^{-1} & \mathbf{B}^{-1} \end{bmatrix}, \tag{5}$$

where $\mathbf{A} = \mathbf{I}_1 - \mathbf{I}_2\mathbf{I}_4^{-1}\mathbf{I}_3$ and $\mathbf{B} = \mathbf{I}_4 - \mathbf{I}_3\mathbf{I}_1^{-1}\mathbf{I}_2$. A necessary condition for $\mathbf{A}$ to be invertible, which is a requirement to obtain the covariance matrix of $\hat{\boldsymbol{\Psi}}_n$, is that both $\mathbf{I}_1$ and $\mathbf{I}_4$ are non-singular matrices (Rencher, 2000). In the Appendix, we provide the expressions for the $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}_3$, and $\mathbf{I}_4$ for the LC models of interest.

The consistency and multivariate normality discussed above apply to the estimators of the component parameters as well. That is, using the property of multivariate normal random variables which states that the sub-vectors of a multivariate normal are also normal, the limiting distributions of $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\beta}}$ become

$$\hat{\boldsymbol{\pi}}_n \longrightarrow N(\boldsymbol{\pi}, \frac{1}{n}\mathbf{A}^{-1}) \tag{6}$$

$$\hat{\boldsymbol{\beta}}_n \longrightarrow N(\boldsymbol{\beta}, \frac{1}{n}\mathbf{B}^{-1}). \tag{7}$$

Also sub-vector $\hat{\boldsymbol{\beta}}_j$ of $\hat{\boldsymbol{\beta}}$ is normally distributed, with mean $\boldsymbol{\beta}_j$ and with co-variance $\mathbf{V}_j$, being a c by c sub-matrix of $\frac{1}{n}\mathbf{B}^{-1}$. In the remaining part of the paper, we focus on this $\boldsymbol{\beta}_j$.

Using the Continuous Mapping Theorem (Mann and Wald, 1943), for a design matrix $\mathbf{H}$ that defines the contrasts on the null hypothesis, one can show that $\mathbf{H}\hat{\boldsymbol{\beta}}_j \longrightarrow N(\mathbf{H}\boldsymbol{\beta}_j, \mathbf{H}\mathbf{V}_j\mathbf{H}')$. The quadratic form of the test for the hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta}_j = \mathbf{0}$ yields the well-known Wald statistic W; i.e.,

$$W = n\left((\mathbf{H}\hat{\boldsymbol{\beta}}_j)'(\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1}(\mathbf{H}\hat{\boldsymbol{\beta}}_j)\right). \tag{8}$$

Under the null hypothesis, that is, if $H_0 : \mathbf{H}\boldsymbol{\beta}_j = \mathbf{0}$ holds, the Wald statistic W has an asymptotic (central) chi-square distribution with $c - 1$ degrees of freedom (Rencher, 2000; Wald, 1943). That is,

$$W = n\left((\mathbf{H}\hat{\boldsymbol{\beta}}_j)'(\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1}(\mathbf{H}\hat{\boldsymbol{\beta}}_j)\right) \longrightarrow \chi^2_{(c-1)}. \tag{9}$$

Under the alternative hypothesis, W follows a non-central chi-square distribution with $c - 1$ degrees of freedom and non-centrality parameter λ. That is,

$$W = n \left((\mathbf{H}\hat{\boldsymbol{\beta}}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\hat{\boldsymbol{\beta}}_j) \right) \longrightarrow \chi^2_{(c-1,\lambda)} \tag{10}$$

where $\lambda = n(\mathbf{H}\boldsymbol{\beta}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\boldsymbol{\beta}_j)$.

3.2 Power and sample size computation

With the establishment of the distribution of the test statistic under the null and alternative hypothesis and the availability of a closed form expression for the non-centrality parameter λ, it becomes possible to compute the power of the test for a given sample size or the sample size for a given power. As in any power analysis, we first have to define the population model. In our case, this involves defining the number of classes and the number of response variables, and, moreover, specifying the values for the class proportions $\boldsymbol{\pi}$ and the class-specific logits $\boldsymbol{\beta}$. For the assumed population model, we can compute the inverse information matrix $\mathbf{V}$ which appears in the formula of the non-centrality parameter.

Once the population parameters are set and $\mathbf{V}$ is computed, power computation for a given sample size and required sample size computation for a given power proceeds along the steps described below.

Steps for power computation

Power computation proceeds as follows:

1. Compute the non-centrality parameter λ for the specified sample size n (use the expression in equation 10).

2. For a given value of type I error α, read the $100(1-\alpha)$ percentile value from the (central) chi-square distribution. That is, find $\chi^2_{(1-\alpha)}(c - 1)$ such that under the null hypothesis, $P\left(W > \chi^2_{(1-\alpha)}(c - 1)\right) = \alpha$. This value is referred as a critical value of a test.

3. Compute the power as the probability that a random variable W from the non-central chi-square distribution (with non-centrality parameter λ given in step 1) will assume a value greater than the critical value obtained under step 2.

Steps for sample size computation

Sample size computation proceeds as follows:

1. For a given value of α, read the $100(1-\alpha)$ percentile value from the (central) chi-square distribution (see the second step for power computation).

2. For a given power and the critical value obtained in step 1, find the non-centrality parameter λ such that, under the alternative hypothesis, the condition that power is equal to $P\left(W > \chi^2_{(1-\alpha)}(c-1)\right)$ is satisfied.

3. From the expression for λ, solve for the sample size as follows: $n = \lambda \left((\mathbf{H}\boldsymbol{\beta}_j)'(\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1}(\mathbf{H}\boldsymbol{\beta}_j) \right)^{-1}$.

Software implementation

The above procedure for power computation can be applied using existing software for LC analysis that allows defining starting values or fixed values for the logit parameters and that provides the (inverse) information matrix as output, for example, using LEM (Vermunt, 1997), Mplus (Muthén and Muthén, 2012), or Latent GOLD (Vermunt and Magidson, 2013b). More specifically, with a LC analysis software package, one can obtain the inverse information matrix $\mathbf{V}$. This will typically require the following two steps:

A. Create a data set containing all possible data patterns and with the expected frequencies according to the LC model of interest as weights. This can be achieved by running the LC software with the population parameters specified as fixed values and with the estimated frequencies as requested output. The created output is, in fact, a data set which is exactly in agreement with the population model. Such a data set is sometimes referred as an 'exemplary' data set (ÓBrien, 1986).

B. Analyze the (exemplary) data set create in step A with the LC model of interest and request the variance-covariance matrix of the parameters (the inverse information matrix) as output. Note that when analyzing a data set which is exactly in agreement with the model, the observed information matrix is identical to the expected information matrix. The same applies to the approximate observed information matrix based on the outer-product of the gradient contributions of the data patterns (see Appendix).

The above two steps provide us with the inverse information matrix $\mathbf{V}$. The actual power or sample size computations using the steps described above can subsequently be performed using software that allows performing matrix computations and that has functions for obtaining the critical value from the chi-squared distribution and the non-centrality value from the non-central chi-squared distribution. For this purpose, one can use R. An R script is available from the first author.

The procedure described above is fully automated in version 5.0 of the Latent GOLD program (Vermunt and Magidson, 2013b). Users define the population model and specify either the sample size or the required power. The program computes the power or the required sample size for the Wald tests it reports by default, as well as for other Wald tests defined by the user. In the Appendix, we give an example of the Latent GOLD syntax for power computation.

3.3 Design factors affecting the power of a Wald test in LC models

Now let us look in more detail at the factors affecting the power of the Wald test in LC models. It should be noted that the power is determined by the value of the type I error and the value of the noncentrality parameter λ. The larger the type I error and the larger λ, the larger the power. As can be observed from equation (10), λ is a function of the sample size n, the precision of the estimator $(\mathbf{V}_j)$, and the effect size $\mathbf{H}\boldsymbol{\beta}_j$. Note that in our case the effect size is the difference between the class-specific β parameters or, equivalently, the strength of the association between the classes and the response variable concerned.

Specific for LC models is that the precision of the estimator is affected by the fact that class membership is unobserved; that is, that we are uncertain about a person's class membership. Recall from equation (5) that the block of $\mathbf{V}$ concerning the β parameters is obtained as the inverse of $\mathbf{B} = \mathbf{I}_4 - \mathbf{I}_3\mathbf{I}_1^{-1}\mathbf{I}_2$. This means that $\mathbf{B}$ becomes larger when $\mathbf{I}_4$ and $\mathbf{I}_1$ become larger and when $\mathbf{I}_2$ and $\mathbf{I}_3$ become smaller. To show how the uncertainty about the class membership affects $\mathbf{B}$, let us have a closer look at $\mathbf{I}_4$, which is the most important term in $\mathbf{B}$. Its elements are obtained as follows:

$$\mathbf{I}_4(\beta_{jt}, \beta_{ks}) = \sum_{i=1}^{2^p} P(X = t/\boldsymbol{y}_i)P(X = s/\boldsymbol{y}_i)(y_{ij} - \theta_{jt})(y_{ik} - \theta_{ks})P(\boldsymbol{y}_i), \ (11)$$

where $\theta_{jt} = \exp(\beta_{jt})/(1+\exp(\beta_{jt}))$. As can be seen, specific for a LC analysis is that the elements of the information matrix are not only a function of the

model parameters, but also of the posterior class membership probabilities $P(X = t/\boldsymbol{y}_i)$. For example, the contribution of response pattern i to the information on parameter β_{jt} equals $P(X = t/\boldsymbol{y}_i)^2(y_{ij} - \theta_{jt})^2 P(\boldsymbol{y}_i)$. In other words, response pattern i contributes with "weight" $P(X = t/\boldsymbol{y}_i)^2$ to the information on a parameter of class t. The contribution to total of the parameters of all c classes equals $\sum_{t=1}^c P(X = t/\boldsymbol{y}_i)^2$. This shows that the information is maximum when $P(X = t/\boldsymbol{y}_i)$ equals 1 for one class and 0 for the other classes, in which case the total contribution equals 1. This occurs when the classes are perfectly separated or when the class membership is observed rather than latent.

Also the entries of $\mathbf{I}_1$ become larger when the posterior class membership probabilities get closer to either 0 or 1. The matrices $\mathbf{I}_2$ and $\mathbf{I}_3$ capture the overlap in information between the class proportions and the β parameters. The elements of this matrix are 0 when separation is perfect and become larger with lower class separation. The Appendix provides the exact formulae for these matrices.

The implication of the above is that the power can be increased by increasing the separation between the classes; i.e., by influencing the factors affecting the posterior class membership probabilities. The posterior class membership probabilities depend on the number of classes, the class proportions, the class-specific conditional response probabilities, and the number of response variables (Collins and Lanza, 2010; Vermunt, 2010a). More specifically, class separation is better with less latent classes, a more uniform class distribution, response variables which are more strongly related to the classes, and a larger number of response variables.

Note that the conditional response probabilities have a dual role. The more the conditional response probabilities θ_{jt} or the logit parameters β_{jt} differ across latent classes, the larger the effect size and thus also the higher the power of the test for the parameters of indicator variable Y_j. However, a larger difference between classes in the response on Y_j also increases the class separation, and thus the power of all tests, also the ones for the other response variables.

4   Numerical study

In this section, we present a numerical study that illustrates the Wald based power analysis for different configurations of design factors. As was shown in section 3, in addition to the usual factors (i.e., sample size, level of significance, power, and effect size), power analysis in LC models involves the

specification of design factors such as the number of classes, the number of observed response variables, the class sizes, and the class-specific probabilities (or logits) for the response variables, which we refer to as LC-specific design factors.

As already indicated in section 3.3, LC-specific design configurations yielding better separated classes, or posterior class membership probabilities which are closer to either zero or one, yield more precise estimators, and as a result larger power of the Wald tests. Therefore, in order to be able to compare different design configuration, it is important to have a measure for class separation. For this purpose, we use the entropy based R-square. The entropy of the posterior class membership probabilities for data pattern i, denoted by E_i, equals $\sum_{t=1}^{c} -P(X = t/\boldsymbol{y}_i) \log P(X = t/\boldsymbol{y}_i)$. Note that E_i gets closer to 0 when the posteriors are closer to 0 and 1. The average entropy across data patterns, denoted by E, equals $\sum_{i=1}^{2^p} E_i P(\boldsymbol{y}_i)$. The entropy based R-square can now be obtained as follows:$R^2_{\text{entropy}} = 1 - E/E(0)$. Here, $E(0)$ is the maximum entropy given the class proportions; that is, $E(0) = \sum_{t=1}^{c} -P(X = t) \log P(X = t)$.

4.1 Manipulation of the design factors

The LC-specific design factors that were varied are the number of classes, the number of indicator variables, the class-specific conditional probabilities, and the class proportions. The number of classes varied from 2 to 4 (i.e., $c = 2, 3, 4$). The number of indicator variables was set to $p = 6$ and $p = 10$. The class-specific conditional probabilities θ_{jt} were 0.7, 0.8, and 0.9 (or, depending on the class, 1-0.7, 1-0.8, and 1-0.9), corresponding to a weak, medium, and strong association between classes and indicator variables. The θ_{jt} were high for class 1, say 0.8, and low for class c, say 1-0.8; with $c = 3$, class 2 had high θ_{jt} values for the first half of the items and low values for the other items; with $c = 4$, class 2 had low θ_{jt} values for the first half of the items and high values for the other items, and class 3 had high θ_{jt} values for the first half of the items and low values for the other items. The class sizes were equal or unequal, where for the unequal conditions we used class proportions of (0.75, 0.25), (0.5, 0.3, 0.2), (0.6, 0.3, 0.1), and (0.4, 0.3, 0.2, 0.1), respectively.

In addition to the four LC-specific design factors, we have to define the sample size, power, level of significance, and effect size(Cohen, 1988). For power computation, the sample size was set to 75, 100, 200, 300, 500, 700, 1000, and 1500, whereas for sample size determination, the power was set to .8, .9, and .95. The type I error was fixed to 0.05. The effect size is

already specified via the response probabilities θ_{jt}, where it should be noted that the logit coefficients β_{jt} for which the Wald tests are performed equal $\beta_{jt} = \log \theta_{jt}/(1 - \theta_{jt})$.

4.2 Effects of design factors on power and sample size

Table 1: Entropy based R-square values for different combinations of LC-specific design factors

		Class size		
		Equal	Unequal	More unequal
Number of classes	$c = 2$	.818	.811	
(for $p = 6$ and θ_{j1}=0.8)	$c = 3$	.627	.624	
	$c = 4$	.594	.589	
Number of indicators	$p = 6$	.627	.624	
(for $c = 3$ and θ_{j1}=0.8)	$p = 10$	.790	.788	
	θ_{j1}=0.7	.332	.330	.314
Class-indicator associations	θ_{j1}=0.8	.627	.624	.607
(for $c = 3$, and $p = 6$)	θ_{j1}=0.9	.880	.879	.871

Note: the 'unequal' and 'more unequal' class size conditions refer to the level of deviation from uniform class distribution. For example, for $c = 3$, we used (0.5, 0.3, 0.2) and (0.6, 0.3, 0.1) to represent a smaller and larger deviation from a uniform class distribution, respectively.

Table 1 presents the entropy based R-square for several combinations of the LC-specific design factors. It shows how the value of this R-square measure is affected by the number of classes, the class proportions, the number of indicators, and the strength of the class-indicator associations, given specific values of the other design factors. As can be seen, the smaller the number of the classes, the larger the number of indicator variables, and/or the stronger the class-indicator associations, the larger the value of the entropy based R-square. Moreover, the more equal the class sizes, the larger the entropy. It can also be seen that the entropy based R-square may become very low when all conditions are less favorable.

To investigate the effect of class separation on the power of the Wald test for the significance of a class-indicator association, the power is computed for five of the design configurations that were presented in Table 1 under different sample sizes. The results are presented in Table 2. From this table, we can see that the power of a test in LC analysis strongly depends on the class separations. When classes are well separated, a sample size of 100 can be large enough to achieve a power of .8 or more. With a class separation

13

Table 2: Estimated power (%) for different class separation levels and different sample sizes

	Entropy based R-square				
Sample size	.314	.330	.607	.624	.790
75	7	12	22	51	94
100	8	14	28	64	98
200	10	24	52	92	100
300	13	34	71	99	100
500	19	53	91	100	100
700	25	69	98	100	100
1000	34	84	100	100	100
1500	49	96	100	100	100

$H_0 : \beta_{j1} = \beta_{j2} = ... = \beta_{jc}$ for which $j = 1$ and $c = 3$.

of .330, .607, and .624, a sample size of 900, 370, and 140, respectively, is required to achieve such a power. With very badly separated classes as in the worst condition, even a sample size of 1500 is not large enough to achieve a power of .8.

Table 3: Required sample size for different configurations of LC-specific design factors and different power levels

	Number of classes				Number of indicators	
Power	$c = 2$	$c = 3$	$c = 4$		$p = 6$	$p = 10$
.8	33	82	83		82	49
.9	45	108	108		108	64
.95	55	131	130		131	78

	Class-indicator associations			Class sizes		
Power	Low	Medium	High	Equal	Unequal	More unequal
.8	419	82	34	82	141	371
.9	550	108	45	108	185	487
.95	671	131	55	131	226	594

Note: The baseline model is the model with $c = 3$, $p = 6$, equal size classes, and medium association between classes and indicators. One design factor is varied to get the other conditions reported in the table.

Table 3 reports the required sample size for a specified power for various combinations of LC-specific design factors. We use the condition with $c = 3$, $p = 6$, equal size classes, and medium class-indicator associations as the baseline. This condition requires sample sizes of 82, 108, and 131, re-

spectively, to achieve the three reported power levels. The other conditions are obtained by varying one design factor at the time.

The results in Table 3 show that, as expected, the required sample size depends on the number of classes, the number of indicators, the strength of the class-indicator associations, and the class sizes. More specifically, keeping the other LC-specific design factors constant, the larger the number of classes and the fewer the number of indicators, the larger the required sample size to achieve the specified power level. The strength of the class-indicator associations turns out to be one of the key factors affecting the power in LC models; for example, to obtain a power of .80, we need at least 419 observations when these associations are weak, but only 34 observations when these are strong. Moreover, many more observations are required when the class sizes are unequal than when they are equal; for example, to achieve a power of .95, we need approximately 130, 225, and 600 observations for the (0.334, 0.333, 0.333), (0.5, 0.3, 0.2), and (0.6, 0.3, 0.1) condition, respectively.

In summary, these results show that the strength of the class-indicator associations and the class distribution have a much stronger impact on the power than the number of classes and the number of indicator variables. The fact the strength of the class-indicator association is so important can be explained by the fact it affects both the class separation and the effect size. For example, for $P = 6$, $C = 3$, and equal class sizes, when the θ_{jt} value changes from .9 to .7, the class separation drops from .880 to .332 and the difference between classes in their conditional response probabilities drops from .8 to .4. Thus, a θ_{jt} value of .9 yields not only a much larger R-square value but also a much larger effect size than a θ_{jt} value of .7. The class sizes are important because the power of a test regarding difference between groups depends strongly on the size of the smallest group.

4.3 Performance of the power computation procedure

An important question is whether the theoretical power computed using the formulae presented in this paper agrees with the actual power when using the Wald with empirical data. To answer this question, we conducted a simulation study in which the theoretical power is compared with the actual power in data sets generated from the assumed population model. Note that the actual power equals the proportion of simulated data sets in which the null hypothesis is rejected.

The population model was three-class LC model with six indicators and equal class sizes. We varied the strength of the class-indicator associations

15

(same three levels as above) and the sample size (75, 100, 200, 300, 500, 700, and 1000). The actual power was computed using 500 samples from the population under the alternative hypothesis. For each of these samples, the LC model is estimated and it is checked whether the Wald value for the test of interest exceeds the critical value.

Table 4: Theoretical and simulated power of the Wald test

Class-indicator associations	Method	Sample size						
		75	100	200	300	500	700	1000
Weak	Theoretical	.200	.254	.470	.649	.869	.958	.994
	Simulated	.145	.234	.444	.628	.838	.920	.960
Medium	Theoretical	.762	.877	.995	1.000	1.000	1.000	1.000
	Simulated	.714	.848	.944	.992	1.000	1.000	1.000
Strong	Theoretical	.989	.999	1.000	1.000	1.000	1.000	1.000
	Simulated	.986	1.000	1.000	1.000	1.000	1.000	1.000

The power presented here is for the null hypothesis $H_0 : \beta_{j1} = \beta_{j2} = ... = \beta_{jc}$ for which $j = 1$. Moreover, $c = 3$. $p = 6$, and class sizes are equal.

Table 4 presents the theoretical and actual power of the Wald test under the investigated simulation conditions. As can be seen, both measures show the same overall trend, namely that the power increases with increasing sample size and increasing effect size (and class separation). However, the actual power of the Wald test is always slightly lower than its theoretical value, where the differences are larger for the smaller sample size and the weaker class separation conditions. An explanation for these differences is that the estimated asymptotic variance-covariance matrix used in the simulated power computations overestimates the variability of the β_j parameters. On the other hand, substantive conclusions are the same for the simulated and theoretical power levels reported in Table 4. With the small effect size and the corresponding weak class separation condition, a sample size of 500 is needed to achieve a power of .8; with the medium class separation, a sample size of 100 suffices; and with the strong class separation, less than 75 observations are needed.

5   Discussion and Conclusion

In LC analysis, the association between class membership and the response variables is usually modeled using a logistic parametrization. This paper

dealt with power analysis for Wald tests for these logit coefficients, for example, for the hypothesis of no association between class membership and the response provided on one of the indicators. We showed that, in addition to the usual design factors – i.e., effect size, sample size, and level of significance – the power of Wald tests in LC models depends on the amount uncertainty about the subjects' class memberships. More specifically, factors affecting the class separation also affect the power. The most important of these LC-specific design factors are the number of classes, the class proportions, the strength of the class-indicator associations, and the number of indicator variables.

A numerical study was conducted to illustrate the proposed power and sample size computation procedures. More precisely, it was shown how class separation – quantified using the entropy-based R-square – is affected by the number of classes, the class proportions, the strength of the class-indicator associations, and the number of indicator variables, and, moreover, how class separation affects the power. It turned out that under the most favorable conditions a sample size of 100 suffices to achieve a power of .8 or .9. For the situation where the entropy-based R-square is small, a considerably larger sample size is required. It was shown that under the least favorable conditions, even a sample size of 2000 did not suffice to achieve an acceptable power level. This demonstrates the importance of performing a power analysis prior to conducting a study that will make use of LC analysis.

If power turns out to be too low given the planned sample size, instead of increasing the sample size, one may try to increase the class separation, for example, by using a larger number of indicators or by improving the quality of the indicators. Note that improving the quality of indicators has a dual effect on the power of the Wald test for class-indicator associations: It increases both the effect size and the class separation. This dual effect could be seen in our numerical study where we saw a dramatic reduction of the required sample size when the θ_{jt} value increased from .7 to .9.

A simulation study was conducted to evaluate whether the theoretical power corresponds with the actual power of the Wald test. It turns out that the estimated power obtained with the formulae provided in this paper is slightly larger than the actual power, where we see a larger overestimation for smaller sample sizes and lower power levels. This implies that to be on the save side, to achieve the specified power, a slightly larger sample size may be used than the estimated sample size.

In this paper, we restricted ourselves to power computations for Wald tests. However, also likelihood-ratio test are often used in LC models, either for testing the same kinds of hypotheses as discussed here or for comparing

models with different number of latent classes. Future research will focus on power computation for likelihood-ratio tests in LC models.

Another limitation of the current work is that we restricted ourselves to simple LC models. In future work, we will investigate whether the methods discussed in this paper can be extended to more complex LC models, such LC models with covariates, latent Markov models, mixture growth models, and mixture regression models.

6 Appendix

6.1 Elements of the information matrix in a LC model for binary responses

The elements of the information matrix $\mathbf{I}(\boldsymbol{\Psi})$, with $\boldsymbol{\Psi}' = (\boldsymbol{\pi}', \boldsymbol{\beta}')$, equal to minus the expected value of the second-order partial derivatives of the log-likelihood function defined in (3) with respect to the free parameters divided by the sample size. In a LC model, these have the following rather simple from:

$$\mathbf{I}(\psi_l, \psi_q) = -E\left(\frac{\partial^2 l(\boldsymbol{\Psi})}{\partial \psi_l \partial \psi_q}\right)/n = \sum \frac{\partial \log \ P(\boldsymbol{y}_i, \boldsymbol{\Psi})}{\partial \psi_l} \frac{\partial \log \ P(\boldsymbol{y}_i, \boldsymbol{\Psi})}{\partial \psi_q} P(\boldsymbol{y}_i, \boldsymbol{\Psi}).$$

This shows that the computation of the information matrix requires solving the first-order partial derivatives $\frac{\partial \log \ p(\boldsymbol{y}_i)}{\partial \psi_l}$. For a class-proportion π_t and a class-specific response logit β_{jt}, these take on the following form:

$$\frac{\partial \log \ P(\boldsymbol{y}_i, \boldsymbol{\Psi})}{\partial \pi_t} = \frac{P(X = t/\boldsymbol{y}_i)}{\pi_t} - \frac{P(X = c/\boldsymbol{y}_i)}{\pi_c},$$
$$\frac{\partial \log \ P(\boldsymbol{y}_i, \boldsymbol{\Psi})}{\partial \beta_{jt}} = P(X = t/\boldsymbol{y}_i)(y_{ij} - \theta_{jt}).$$

This yields the following forms for the entries of the sub-matrix $\mathbf{I}_1$, $\mathbf{I}_2$,

$\mathbf{I}_3$, and $\mathbf{I}_4$:

$$
\mathbf{I}_1(\pi_t, \pi_s) = \sum_{i=1}^{2^p} \left(\frac{P(X = t/\boldsymbol{y}_i)}{\pi_t} - \frac{P(X = c/\boldsymbol{y}_i)}{\pi_c} \right) \left(\frac{P(X = s/\boldsymbol{y}_i)}{\pi_s} - \frac{P(X = c/\boldsymbol{y}_i)}{\pi_c} \right) P(\boldsymbol{y}_i, \boldsymbol{\Psi}),
$$

$$
\mathbf{I}_2(\pi_t, \beta_{js}) = \sum_{i=1}^{2^p} \left(\frac{P(X = t/\boldsymbol{y}_i)}{\pi_t} - \frac{P(X = c/\boldsymbol{y}_i)}{\pi_c} \right) P(X = s/\boldsymbol{y}_i)(y_{ij} - \theta_{js})P(\boldsymbol{y}_i \boldsymbol{\Psi}),
$$

$$
\mathbf{I}_3(\beta_{jt}, \pi_s) = \sum_{i=1}^{2^p} P(X = t/\boldsymbol{y}_i)(y_{ij} - \theta_{jt}) \left(\frac{P(X = s/\boldsymbol{y}_i)}{\pi_s} - \frac{P(X = c/\boldsymbol{y}_i)}{\pi_c} \right) P(\boldsymbol{y}_i \boldsymbol{\Psi}),
$$

$$
\mathbf{I}_4(\beta_{jt}, \beta_{ks}) = \sum_{i=1}^{2^p} P(X = t/\boldsymbol{y}_i)(y_{ij} - \theta_{jt})P(X = s/\boldsymbol{y}_i)(y_{ik} - \theta_{ks})P(\boldsymbol{y}_i, \boldsymbol{\Psi}).
$$

6.2 An example of the Latent GOLD setup for Wald based power computation

The Latent GOLD 5.0 (Vermunt and Magidson, 2013a) Syntax system implements the power computation procedure described in this paper. In order to perform such a Wald power computation, one should first create a small "example" data set; that is, a data set with the structure of the data one is interested in. With six binary response variables (y1 through y6), this file could be of the form:

```
y1   y2   y3   y4   y5   y6
 0    0    0    0    0    0
```

which is basically a data set with a single observation with a response of 0 on all six variables.

For this small data set, one defines the model of interest and requests the power or the required sample size using the output options. This is done as follows using the Latent GOLD "options", "variables", and "equations" sections:

```
options
  output parameters standarderrors
         WaldPower=<number> WaldTest='fileName';
variables
  dependent y1 2, y2 2, y3 2, y4 2, y5 2, y6 2;
  latent x nominal 2;
equations
  x <- 1;
```

```
y1 - y6 <- 1 | x;
{0.0000000000
1.386294361 -1.386294361
1.386294361 -1.386294361
1.386294361 -1.386294361
1.386294361 -1.386294361
1.386294361 -1.386294361
1.386294361 -1.386294361}
```

In the "variables" section, we define the variables which are in the model and also their number of categories. These are the six response variables and the latent variable "x". The "equations" section specifies the logit equations defining the model of interest, as well as the values of the population parameters. Note that the value 1.386294361 for a logit coefficients corresponds to a conditional response probability of .80.

The "output" line in the "options" section lists the output requested. With WaldPower=<number>, one requests a power or sample size computation. When using a "number" between 0 and 1, the program reports the required sample size for that power, and when using a values larger than 1, the program reports the power obtained with that sample size. The optional statement WaldTest='filename' can be used to define user-specific Wald test in addition to the test which are provided by default. The linear contrasts for the user-defined hypotheses of interest are defined in a text file.

References

Agresti, A. (2002). *Categorical data analysis*, volume 359. Wiley-interscience.

Birch, M. et al. (1964). A new proof of the pearson-fisher theorem. *The Annals of Mathematical Statistics*, 35(2):817–824.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.

Collins, L. M. and Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*, volume 718. Wiley.

Dayton, C. M. and Macready, G. B. (1976). A probabilistic model for validation of behavioral hierarchies. *Psychometrika*, 41(2):189–204.

Dayton, C. M. and Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83(401):173–178.

Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in medicine*, 26(18):3385–3397.

Demidenko, E. (2008). Sample size and optimal design for logistic regression with binary interaction. *Statistics in medicine*, 27(1):36–46.

Formann, A. K. (1982). Linear logistic latent class analysis. *Biometrical Journal*, 24(2):171–190.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87(418):476–486.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2):215–231.

Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators local dependence models. *Sociological Methods & Research*, 16(3):379–405.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, pages 1219–1240.

Hirtenlehner, H., Starzer, B., and Weber, C. (2012). A differential phenomenology of stalking using latent class analysis to identify different types of stalking victimization. *International review of victimology*, 18(3):207–227.

Holt, J. A. and Macready, G. B. (1989). A simulation study of the difference chi-square statistic for comparing latent class models under violation of regularity conditions. *Applied Psychological Measurement*, 13(3):221–231.

Keel, P. K., Fichter, M., Quadflieg, N., Bulik, C. M., Baxter, M. G., Thornton, L., Halmi, K. A., Kaplan, A. S., Strober, M., Woodside, D. B., et al. (2004). Application of a latent class analysis to empirically define eating disorder phenotypes. *Archives of General Psychiatry*, 61(2):192.

Lanza, S. T., Collins, L. M., Lemmon, D. R., and Schafer, J. L. (2007). Proc lca: A sas procedure for latent class analysis. *Structural Equation Modeling*, 14(4):671–694.

Lazarsfeld, P. (1950). The logical and mathematical foundation of latent stricture analysis and the interpretation and mathematical foundation of latent structure analysis. Stouffer, S.A. et al. (eds.). *Measurement and predictions*, pages 362–472.

Linzer, D. A. and Lewis, J. B. (2011). polca: An r package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1–29.

Magidson, J. and Vermunt, J. K. (2004). Latent class models. *The Sage handbook of quantitative methodology for the social sciences*, pages 175–198.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14(3):217–226.

McCutcheon, A. L. (1987). *Latent class analysis*, volume 64. SAGE Publications, Incorporated.

McHugh, R. B. (1956). Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21(4):331–347.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. John Wiley, New York.

Muthén, L. K. and Muthén, B. O. (2012). Mplus. *The comprehensive modelling program for applied researchers: Users guide*, 5.

Nakagawa, S. and Foster, T. M. (2004). The case against retrospective statistical power analyses with an introduction to power analysis. *acta ethologica*, 7(2):103–108.

ÓBrien, R. G. (1986). Using the SAS system to perform power analyses for log-linear models. In *Proc. 11th Annual SAS Users Group Conference*, page 778784.

Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *The Annals of Statistics*, 9(1):225–228.

Rencher, A. C. (2000). *Linear Models in Statistics*. John Wiley, New York.

Rindskopf, D. and Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, 5(1):21–27.

Uebersax, J. S. and Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, 9(5):559–572.

Vermunt, J. K. (1996). *Log-linear event history analysis: A general approach with missing data, latent variables, and unobserved heterogeneity*, volume 8. Tilburg University Press, Tilburg.

Vermunt, J. K. (1997). LEM: A general program for the analysis of categorical data. *Tilburg University*.

Vermunt, J. K. (2010a). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4):450–469.

Vermunt, J. K. (2010b). Latent class models. in: P. Peterson, E. Baker, B. McGaw, (eds.). *International Encyclopedia of Education*, 7:238–244.

Vermunt, J. K. and Magidson, J. (2013a). *LG-Syntax user's guide: Manual for latent GOLD 5.0 syntax module*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K. and Magidson, J. (2013b). Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25.

Whittemore, A. S. (1981). Sample size for logistic regression with small response probability. *Journal of the American Statistical Association*, 76(373):27–32.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5(3):329–350.

Yang, I. and Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics*, pages 948–958.