A comparative analysis of MELODIC and LCA performance

Koman, Friso

Universiteit Leiden

Faculteit der Sociale Wetenschappen

A comparative analysis of MELODIC and LCA performance

Friso Koman

Master's Thesis Methodology and Statistics in Psychology
Methodology and Statistics Unit, Institute of Psychology,
Faculty of Social and Behavioral Sciences, Leiden University
Date: October 12, 2021
Student number: 1283340
Supervisor: Prof. Dr. M.J. de Rooij & Dr. Z. Bakk

Abstract

Introduction: In research with respect to non-communicable diseases (NCDs), multimorbidity and external variables, unobserved latent classes are often presumed. Consequently, latent class analysis (LCA) with predictors is often used to analyze the data. However, it has been shown that LCA is not accurate in all instances. If inaccurate, one still might want to relate predictors to the NCDs without imposing on latent classes. For that purpose, the recently developed MELODIC could be used instead. MELODIC is used to simultaneously estimate binary response variables from a (set of) predictor(s), with a dimensionality reduction. The aim of this study was to evaluate which method has a higher prediction accuracy and whether the accuracy depends on the data generation method.

Method: Two simulations were performed. In Simulation 1, data were generated with MELODIC parameters, on which MELODIC and LCA were performed on various samples. In Simulation 2, data were generated with LCA parameters, on which MELODIC and LCA were performed on various samples. Predictor-response variable relationships were calculated in the populations and in the samples, and the prediction accuracy was assessed in terms of RMSE.

Results: The accuracy of MELODIC and LCA highly depends on the data generation method. When data were generated with MELODIC, MELODIC outperformed LCA. When data were generated with LCA, LCA outperformed MELODIC. The general RMSE difference between MELODIC and LCA was smaller when data were generated with LCA. Larger sample sizes are beneficial for the prediction accuracy of MELODIC and LCA. Modifying method specific parameters, resulting in more frequent NCDs in the populations, is detrimental for the prediction accuracies of the methods.

Discussion: If there are reasons to believe that there are latent classes underlying the data, LCA is the most appropriate method. If unobserved latent classes are not presumed, MELODIC is the most accurate method.

Table of contents

1 Introduction

The worldwide number one cause of mortality are non-communicable diseases (NCDs). The World Health Organization (WHO; 2021) currently estimates that annually 71% of the worldwide mortality - equal to 41 million deaths - is caused by NCDs. According to the WHO (2018), the proportion of the world its population of 60+ years will increase from 12% to 22%, between 2015 and 2050. Due to the population ageing and population growth, the relative and absolute mortality-by-NCDs rates, respectively, are increasing (Murray et al., 2015; Wang et al., 2016). The increasing prevalence of NCDs and especially the increasing prevalence of multimorbidity (i.e., the presence of two or more chronic conditions or NCDs in patients) have become an international research priority, because of the impact on the global health, the high pressure on the health care systems and the enormous global health care costs (Wang et al., 2016; Bloom et al., 2011; Bertram et al., 2018). Therapeutic management of multimorbidity is often complex due to the 'single disease focus' of current clinical guidelines, resulting in poor health outcomes (Wallace et al., 2015). Therefore, there has been a need to identify common clusters of NCDs for better understanding of multimorbidity, to improve the clinical guidelines and treatment of multimorbidity (Islam et al., 2014).

Recently, a lot of research with respect to multimorbidity has been conducted to identify these common clusters of NCDs (cf. Bayes-Marin et al., 2020; Whitson et al., 2016; Olaya et al., 2017). In these studies, it is hypothesized that there are unobserved multimorbid subgroups (clusters) within the populations, for each subgroup a more or less unique multimorbidity pattern. Latent class models (or finite mixture models) are used to analyze the data.

Latent class analysis (LCA), initially introduced by Lazarsfeld (1950), is a technique that is used to analyze relationships between response variables that are measured on a nominal or categorical level. Based on subjects' scores on a set of these observed nominal or categorical response variables, latent classes can be identified (McCutcheon, 1987). In the measurement part of the LCA, the response variables are related to the latent classes (e.g., based on the presence or absence of a set of NCDs in patients, groupings of patients with similar NCD patterns (latent classes) are detected. Based on these patterns, subjects will be assigned to one of these classes). In the structural part of the LCA, relationships between the latent classes and external variables are estimated (e.g., having defined the multimorbid latent classes in the measurement model, the relationship between the latent classes and the external

variable ‘age’ is estimated). These external variables can function as, e.g., predictors or distal outcome variables (Vermunt, 2010). The latent class model will be discussed in more detail in the next section.

In the study of Bayes-Marin et al. (2020), for example, data were used from the WHO’s study on Global Ageing and Adult Health (SAGE; Kowal et al., 2012), the English Longitudinal Study of Ageing (ELSA; Steptoe et al., 2013) and the Survey of Health, Ageing, and Retirement in Europe Study (SHARE; Börsch-Supan et al., 2013). Based on the scores of subjects on eight NCDs, three latent classes had been identified. Consequently, predictor variables were added in the structural part of the model to determine their influence on class membership.

Another example is the study of Olaya et al. (2017) in which data were used from the Collaborative Research on Ageing in Europe (COURAGE in Europa) study (Leonardi et al., 2014). Here, based on the scores of subjects on eleven NCDs, three latent classes have been identified as well. However, in contrast with the study of Bayes-Marin et al. (2020), in this study it was checked, after assigning individuals to the latent classes, whether distal outcome variables could be predicted from class membership.

In these studies, latent classes are presumed and the LCA approaches are well-motivated in light of the need for finding multimorbidity patterns. However, one of the main limitations of LCA is that latent classes are unobserved. One cannot know if the latent classes are real subgroups in the population. The researcher fits a number of models to the data, with an increased number of latent classes for each model. Based on various criteria and judgements of the researcher concerning the substantive usefulness of the different models, an optimal solution is chosen. However, even the optimal solution does not guarantee the existence of real subgroups. It may identify invalid and unreliable latent classes, since the LCA will just seek for the best statistically determined partitioning of subjects into classes and does not take any substantive usefulness into account (Bauer & Curran, 2004; Petersen et al., 2019). Consequently, the usefulness of estimating relationships between potentially ill-defined classes and external variables is questionable.

Bayes-Marin et al. (2020) state, for example, that their three-class solution was forced to different samples, since they aimed to compare the samples of different regions and different age groups. An optimal LCA solution in one region for one age group was applied to all other samples, although there might have been different optimal solutions for these samples. Although forcing the solution to all samples is necessary to do comparisons between samples, the reliability and validity of the latent classes in these samples are questionable, let

alone the usefulness of relating predictor variables to the identified latent classes. Whitson et al. (2016) identified six meaningful latent classes in their study. However, they conclude that there was a high misclassification error, indicating that assigning individual patients to one of the latent classes was problematic. They concluded that simply counting the number of NCDs of individuals for creating multimorbid groups is as informative as using the identified latent classes of the study.

Altogether, identifying latent multimorbid NCD classes with LCA has proven to be troublesome. Consequently, relating external variables to latent classes is not substantively useful, and there is evidence that creating multimorbid groups based on the number of NCDs in patients would be just as meaningful as using ill-defined latent classes. Therefore, performing an LCA on NCD data, for detecting subgroups, should be reconsidered in some cases. Fit statistics for LCA (which will be discussed in more detail later) can be used to indicate whether identified classes are discriminable enough to proceed with LCA. However, if these fit indices indicate that the classes are not very discriminable, LCA should be reconsidered.

However, one still might be interested in linking predictors or distal outcome variables (e.g., demographic variables and/or potential risk- or protecting factors) to the individual NCDs. In the case of predictors, one option would be to predict each individual NCD from one or a set of predictors by performing a logistic regression. In logistic regression, discrete binary response variables are predicted from a (set of) predictor(s) (Agresti, 2003). Since NCDs are discrete binary variables (‘yes’ or ‘no’), fitting many single logistic regressions would be a suitable procedure. Often, however, it has been showed that it is more beneficial to simultaneously predict response variables from the same set of predictors.

For example, Breiman and Friedman (1997) developed a regularization procedure that simultaneously predicts a set of continuous response variables from the same set of predictors. The response variables are correlated, due to their shared dependencies on the predictors, and these dependencies are taken into account by applying a multivariate shrinkage on the coefficients. This procedure yields a substantial improvement of the prediction accuracy, compared to the prediction accuracy when performing many single regressions.

Taking the advantages of simultaneously predicting multiple response variables into account, De Rooij and Groenen (2021) recently developed a distance-based family of logistic models: the MultivariatE LOgistic DIstance to Categories family (the MELODIC family). In MELODIC, multiple binary response variables are predicted from a set of predictors simultaneously. It is a method based on the ideas of multidimensional unfolding, which is a

distance-based approach for dimensionality reduction. Subjects and the two outcome categories for each binary response variable are placed in a low-dimensional Euclidean space. The distances between the subjects and the categories define the probabilities of belonging to the categories. Here, the dependencies of the response variables are taken into account by reducing the dimensionality, resulting in changes in all response variables when a predictor value is changed. The subject's position in the low-dimensional Euclidean space is parametrized as a linear combination of the predictors.

So far, two methods have been mentioned in particular: MELODIC and LCA. In the former, multiple NCDs can be predicted from a set of predictors simultaneously, while reducing the dimensionality. In the latter, latent classes can be identified, and we can predict multiple NCDs from a set of predictors via these latent classes. More specifically, for both methods it is possible to calculate the relationships for each predictor for each response variable: the predictor-response variable relationships.

In this paper, NCD data from SHARE (Börsch-Supan et al., 2013) will be used to compare the prediction accuracy of MELODIC and LCA. From this data, predictor and response variables scores will be extracted and will be analyzed using MELODIC and LCA. The goal is to identify which method more accurately estimates the predictor-response variables relationships. That is, which method performs more accurately when there are true latent classes in the population. And, which method is more accurate when the dimensionality of the Euclidean space is truly reduced in the population.

For that purpose, two simulations will be performed, both based on the SHARE data (Börsch-Supan et al., 2013). In the first simulation, population data will be generated based on MELODIC parameters. In the second simulation, population data will be generated based on LCA parameters. In both simulations MELODIC and LCA will be fitted on many samples of the generated population data. In each of the analyses of samples, the estimated predictor-response variables relationships will be calculated as well as for the populations. The accuracy of the methods will be determined in terms of Root Mean Squared Error (RMSE). Finally, it will be concluded if the prediction accuracy of the two methods depend on the data generation method. The simulation studies will be discussed in more detail in Section 5.

MELODIC has been developed recently and although the technique already has been used in an application study of De Rooij and Groenen (2021), the prediction accuracy of the method has not been compared so far with another statistical method, which can be used for the same purpose. Therefore, this study can be interpreted as an exploratory study to determine whether MELODIC could potentially serve as an alternative to LCA, when

analyzing NCD data. Due to the design of the simulations, it is possible to evaluate how accurate MELODIC is when applied to samples from populations based on LCA parameters and vice versa. Results of these simulations might provide insights in the appropriateness of using MELODIC as alternative to LCA when it concerns NCD data. If the results are promising, the limitations of using LCA on NCDs in the studies mentioned earlier, could be overcome by performing MELODIC analyses instead. More insights in the appropriateness of statistical analyses on data of this kind is necessary to improve the prevention and treatment of multimorbidity.

The rest of this paper is organized as follows. In the next section the MELODIC- and LCA model are discussed in more detail (Section 2), followed by Section 3 in which MELODIC and LCA are applied on the SHARE data. Subsequently, the simulations are discussed more elaborately in Section 4. In Section 5 the results of the simulation studies will be provided, and Section 6 is the Discussion section.

2 The MELODIC and LCA models

In subsections 2.1. and 2.2., the MELODIC- and the latent class model, respectively, will be outlined. Both subsections conclude with the formulation of the predictor-response variable relationships per predictor per response variable that can be derived from both models.

2.1 The MELODIC Model

As mentioned in the previous section, in MELODIC, all subjects and binary response variable outcome categories are placed in a low-dimensional Euclidean space and the distance between a subject and the two outcome categories of a single response variable defines the probabilities of belonging to those categories (De Rooij & Groenen, 2021).
The following notation will be used for displaying the model:

- $i \in \{1, \dots, N\}$ for the subjects.
- $p \in \{1, \dots, P\}$ for the predictor variables.
- $r \in \{1, \dots, R\}$ for the binary response variables.

- $m \in \{1, \ldots, M\}$ for the dimensions in the Euclidean space.
- $c \in \{0,1\}$ for the two categories of the response variables.
- **X** is an $N \times P$ matrix, containing the scores of N subjects on P predictor variables.
- **Y** is an $N \times R$ matrix, containing the scores of N subjects on R response variables.
- **U** is an $N \times M$ matrix containing the coordinates of N subjects in M dimensions.
- **V** is a $2R \times M$ matrix containing the coordinates of the two response variable categories c of R response variables in M dimensions.
- **B** is a $P \times M$ matrix containing the regression weights of P predictors in M dimensions.

A capital bold character represents a matrix (**X**), a small bold character represents a vector ($\mathbf{x}_i$) and a small cursive character represents a single entry (x_{ip}).

The probability for a subject of belonging in one of the response variable outcome categories, given the predictor values of the subject, ($P(y_{ir} = c|\mathbf{x}_i)$) is defined as

$$P(y_{ir} = c|\mathbf{x}_i) = \frac{\exp\left(-\delta(\mathbf{u}_i, \mathbf{v}_{rc})\right)}{\exp\left(-\delta(\mathbf{u}_i, \mathbf{v}_{r0})\right) + \exp\left(-\delta(\mathbf{u}_i, \mathbf{v}_{r1})\right)}, \quad (1)$$

where $\delta(\mathbf{u}_i, \mathbf{v}_{rc})$ is defined as half the squared Euclidean distance between subject *i* and category *c* of response variable *r*, i.e.,

$$\delta(\mathbf{u}_i, \mathbf{v}_{rc}) = \frac{1}{2}\sum_{m=1}^{M}(u_{im} - v_{rcm})^2 = \frac{1}{2}\sum_{m=1}^{M}(u_{im}^2 + v_{rcm}^2 - 2u_{im}v_{rcm}). \quad (2)$$

The log-odds in favor of category 1 of response variable *r* for subject *i*, is defined as

$$\log\frac{P(y_{ir} = 1|\mathbf{x}_i)}{P(y_{ir} = 0|\mathbf{x}_i)} = \delta(\mathbf{u}_i, \mathbf{v}_{r0}) - \delta(\mathbf{u}_i, \mathbf{v}_{r1}), \quad (3)$$

which, consequently, can be rewritten as

$$\log\frac{P(y_{ir}=1|\mathbf{x}_i)}{P(y_{ir}=0|\mathbf{x}_i)} = \frac{1}{2}\sum_{m=1}^{M}(u_{im}^2 + v_{r0m}^2 - 2u_{im}v_{r0m}) - \frac{1}{2}\sum_{m=1}^{M}(u_{im}^2 + v_{r1m}^2 - 2u_{im}v_{r1m})$$

$$= \sum_{m=1}^{M}[\frac{1}{2}(v_{r0m}^2 - v_{r1m}^2) - u_{im}v_{r0m} + u_{im}v_{r1m}] \quad (4)$$

$$= \sum_{m=1}^{M}[\frac{1}{2}(v_{r0m}^2 - v_{r1m}^2) + u_{im}(v_{r1m} - v_{r0m})].$$

Since the coordinates of the subject are parametrized as a linear combination of predictors, u_{im} can be substituted by $\mathbf{x}_i^{\mathrm{T}}\mathbf{b}_m$, i.e.,

$$\log\frac{P(y_{ir}=1|\mathbf{x}_i)}{P(y_{ir}=0|\mathbf{x}_i)} = \sum_{m=1}^{M}[\frac{1}{2}(v_{r0m}^2 - v_{r1m}^2) + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_m(v_{r1m} - v_{r0m})]. \quad (5)$$

De Rooij and Groenen (2021) rewrite this log-odds as

$$\log\frac{P(y_{ir}=1|\mathbf{x}_i)}{P(y_{ir}=0|\mathbf{x}_i)} = a_r^* + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_r^*, \quad (6)$$

in which $a_r^* = \frac{1}{2}\sum_{m=1}^{M}(v_{r0m}^2 - v_{r1m}^2)$ and $\mathbf{b}_r^* = \sum_{m=1}^{M}\mathbf{b}_m(v_{r1m} - v_{r0m})$. From this stage it is becoming more apparent that the model has the shape of a logistic regression model. The log-odds for subject *i* of belonging to category 1 for response variable *r*, given the predictor values of subject *i*, is calculated with a_r^* being the intercept and $\mathbf{b}_r^*$ the standardized slope coefficients vector of length *P*. Matrix $\mathbf{B}^*$ of size $P \times R$, then, contains all the estimated standardized logistic regression slope coefficients of the MELODIC model. which are referred to as the implied coefficients (De Rooij & Groenen, 2021).

Following the definition of (6), the probability for subject *i* of belonging to category 1 of response variable *r*, given the predictor scores of subject *i*, (i.e., $P(y_{ir}=1|\mathbf{x}_i)$, can be rewritten as

$$P(y_{ir}=1|\mathbf{x}_i) = \frac{\exp\big(-\delta(\mathbf{u}_i,\mathbf{v}_{r1})\big)}{\exp\big(-\delta(\mathbf{u}_i,\mathbf{v}_{r0})\big) + \exp\big(-\delta(\mathbf{u}_i,\mathbf{v}_{r1})\big)} = \frac{\exp\big(a_r^* + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_r^*\big)}{1 + \exp\,(a_r^* + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_r^*)}. \quad (7)$$

It can be observed from $\mathbf{b}_r^* = \sum_{m=1}^{M} \mathbf{b}_m(v_{r1m} - v_{r0m})$ that the effect of predictors on a response variable is determined by both the regression weights and the distance between the response variable categories. To estimate the regression weights and the coordinates of the categories of the response variables for the MELODIC model, an iterative majorization algorithm was developed by De Rooij and Groenen (2021). The algorithm constantly updates the parameters **B**, **K** (an $R \times M$ matrix containing the discriminatory power for the response variables) and **L** (an $R \times M$ matrix containing the response variable midpoint locations of the categories), until the global optimum of the deviance function is reached. The initialization values of the three parameters are determined by a generalized singular value decomposition. **K** and **L** together define **V** (from which $\mathbf{v}_{rc}$ will be derived), and thus the result of the optimization procedure provides the optimal regression weights and coordinates of the response variable categories for the MELODIC model.

The model takes the dependencies among the response variables into account, by placing the categories and the subjects in a Euclidean space with reduced dimensionality. Therefore, changes in a subject's predictor scores, will alter the response variable probabilities of all response variables, instead of only one response variable. Due to the dimension reduction, the variance of model is reduced substantially at the cost of a little bias, resulting in a higher prediction accuracy, in contrast to fitting many single logistic regressions for all the response variables.

When fitting a MELODIC analysis to data, three input statements are required: **X**, **Y** and *M*, the number of dimensions in the Euclidean space in which the subjects and the response variable categories will be placed. The researcher is free to decide how many dimensions will be used, as long as the number of dimensions is larger than zero, and not larger dan min(*P*, *R*). The idea is that the researcher performs multiple MELODIC analyses, each analysis with an increasing number of *M*. Based on fit indices as the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), and based on the interpretation of the researcher, the optimal model will be chosen for interpretation and further analysis.

2.1.1. Predictor-response variable relationships in MELODIC

Since it is desired to calculate RMSEs of each individual predictor-response variable relationship for doing comparisons between MELODIC and LCA, the last step of this subsection, is to define the predictor-response variable relationships in MELODIC. The response variables are all binary with response categories $c \in \{0,1\}$. When it concerns

estimating predictor-response variable relationships, predicted response variable probabilities always refer to probabilities for category $c = 1$ (as opposed to $c = 0$) for the remainder of this paper. That is, it is desired to know an individual's probability of suffering from an NCD. Note, however, that response probabilities for category $c = 0$, are easily computed (i.e., $P(y_{ir} = 0|\mathbf{x}_i) = 1 - P(y_{ir} = 1|\mathbf{x}_i)$).

Recall from (7) that the probability for subject *i* to score 1 on response variable *r*, given the predictor scores of subject *i*, is defined as

$$P(y_{ir} = 1|\mathbf{x}_i) = \frac{\exp(a_r^* + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_r^*)}{1 + \exp(a_r^* + \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_r^*)},$$

with both $\mathbf{x}_i$ and $\mathbf{b}_r^*$ being of length *P*. However, instead of $P(y_{ir} = 1|\mathbf{x}_i)$ it is desired to obtain $P(y_{ir} = 1|x_{ip})$, i.e., the probability for subject *i* to score 1 on response variable *r*, given the predictor score *p* of subject *i*. Estimating the predictor-response variable relationships between, for example, $x_{i,p=1}$ and probabilities for $y_{i,r=1} = 1$, is done by constantly changing $x_{i,p=1}$ while holding the $\mathbf{x}_{i,p\neq1}$ fixed, resulting in different probabilities for $y_{ir=1} = 1$ for each value of $x_{i,p=1}$. Since $\mathbf{X}$ is standardized, fixing $\mathbf{x}_{i,p\neq1}$ at their means implies that $\mathbf{x}_{i,p\neq1}$ contains solely 0's. Therefore, $\mathbf{x}_{i,p\neq1}$ and the corresponding $\mathbf{b}_{p\neq1,r}$ drop out of the equation. Hence, $P(y_{ir} = 1|x_{ip})$ can be denoted as

$$P(y_{ir} = 1|x_{ip}) = \frac{\exp(a_r^* + x_{ip}\mathrm{b}_{pr}^*)}{1 + \exp(a_r^* + x_{ip}\mathrm{b}_{pr}^*)}, \tag{8}$$

for all predictors $p \in \{1, \ldots, P\}$ and response variables $r \in \{1, \ldots, R\}$. This formula will be used to calculate the predictor-response variable relationships for MELODIC.

2.2 The latent class model

As mentioned in the Introduction, LCA refers to partitioning subjects in latent classes based on a set of nominal or categorical response variables (McCutcheon, 1987). Earlier, subject *i*'s score on response variable *r* was defined as y_{ir}. Then $\mathbf{y}_i$ is a vector of all response variables for subject *i*. When introducing the latent class variable *H*, with $t \in \{1, \ldots, T\}$ latent classes, the latent class model is defined as

$$P(\mathbf{y}_i) = \sum_{t=1}^{T} P(H = t) P(\mathbf{y}_i | H = t). \tag{9}$$

In latent class models, it is assumed that, given latent class membership, the response variables are locally independent (i.e., the local independence assumption), hence

$$P(\mathbf{y}_i | H = t) = \prod_{r=1}^{R} P(y_{ir} | H = t) = \prod_{r=1}^{R} \prod_{c=0}^{nc_r - 1} \pi_{rct}^{I(y_{ir}=c)}, \tag{10}$$

where nc_r indicates the number of response categories of response variable *r*, $\pi_{rct} = P(y_{ir} = c | H = t)$ indicates the estimated class-conditional response probabilities and where $I(y_{ir} = c)$ is an indicator variable which is equal to 1 if the score of subject *i* on response variable *r* is equal to *c*. Another assumption of the latent class model is that $P(\mathbf{y}_i)$ is a weighted average of the *t* class-specific $P(\mathbf{y}_i | H = t)$. This assumption is referred to as the joint distribution assumption. Together, they form the definition of $P(\mathbf{y}_i)$, the latent class model, as defined in (9).

The model as defined above is the most simple latent class model and it only consists of a measurement part in which the response variables are related to the latent classes. The model can be extended by adding a structural part to the model, e.g., by adding a (set of) predictor(s) to the model (Vermunt, 2010). By doing so, the model can be defined as

$$P(\mathbf{y}_i | \mathbf{x}_i) = \sum_{t=1}^{T} P(H = t | \mathbf{x}_i) P(\mathbf{y}_i | H = t). \tag{11}$$

The conditional distribution of class membership given the predictors can be defined as a multinomial logistic regression model, i.e.,

$$P(H = t | \mathbf{x}_i) = \frac{\exp(a_t + \mathbf{x}_i^{\mathrm{T}} \mathbf{b}_t)}{\sum_{t=1}^{T} \exp(a_t + \mathbf{x}_i^{\mathrm{T}} \mathbf{b}_t)}, \tag{12}$$

in which a_t is the multinomial logistic regression intercept for class *t* and $\mathbf{b}_t$ containing the multinomial logistic regression slopes for class *t*. Both $\mathbf{x}_i$ and $\mathbf{b}_t$ are of length *P*.

When estimating the response variable pattern of subject *i*, given the predictor scores of subject *i* (i.e., $P(\mathbf{y}_i|\mathbf{x}_i)$), it is clear from the model definition that the predictor scores only have an effect on the class membership probabilities, but not on the response variables. This is due to another local independence assumption, i.e., the non-differential measurement assumption, which states that given an individual's class, predictors are not associated with response variables. Therefore, direct effect of the predictors and the response variables are not estimated (although recent work suggests that relaxing this assumption is sometimes beneficial for parameter estimation (cf. Janssen et al., 2019; Reboussin et al., 2008)).

The one-step (or Full Information Maximum Likelihood (FIML)) approach estimates the latent class model parameters by maximizing the loglikelihood function for $P(\mathbf{y}_i|\mathbf{x}_i)$ (Vermunt, 2010), i.e.,

$$\log L_{FIML} = \sum_{i=1}^{N} \log \sum_{t=1}^{T} P(H = t|\mathbf{x}_i) P(\mathbf{y}_i|H = t), \tag{13}$$

with the Expectation Maximum algorithm (EM; Dempster et al., 1977). In the algorithm, the multinomial logistic regression intercepts **a** (a vector of length $T - 1$) and slopes **B** (a matrix of size $P \times (T - 1)$) and the estimated class-conditional response probability $\boldsymbol{\pi}$ (($nc - 1$) × R × T unique class-conditional response probabilities) are updated until a maximum is reached. EM uses random class-conditional response probabilities $\boldsymbol{\pi}$ as starting values, and to make sure that the algorithm is not optimizing into a local maximum, it is often wise to repeat the optimizing procedure multiple times (Linzer & Lewis, 2011).

Although there many alternatives to the one-step approach, (e.g., a two-step approach (Bakk & Kuha, 2017) and three-step approaches (Vermunt, 2010)), for the remainder of this paper, only the one-step approach LCA will be used, and will be referred to as simply LCA.

When fitting an LCA to data, **X** and **Y** are required, as well as the number of classes *T*. The researcher is free to decide how many latent classes should be identified when fitting the LCA to the data, as long as $T > 1$. Similar to MELODIC, the idea is that the researcher performs multiple LCAs, each analysis with an increased number of *T*. Based on fit indices as the AIC and BIC and based on the interpretation of the researcher, the optimal model will be chosen for interpretation and further analysis.

2.2.1 Predictor-response variable relationships in LCA

Similar to the procedure in the MELODIC section, predictor-response variable relationships need to be identified. Due to the non-differential measurement assumption of LCA, no direct relations between predictors and response variables are estimated. However, since predictors and response variables are related to latent classes, it is possible to calculate a predictor-response variable relationship via the latent classes. Recall from (11) that

$$P(\mathbf{y}_i|\mathbf{x}_i) = \sum_{t=1}^{T} P(H = t|\mathbf{x}_i)P(\mathbf{y}_i|H = t).$$

Here, a set of response variables for subject *i* is predicted, given subject *i*'s predictor scores. To estimate a single response variable for subject *i*, given *i*'s score on a single predictor, some changes need to be made to the formula. Instead of estimating class membership conditional on all *R* predictor scores of subject *i*, i.e., $P(H = t|\mathbf{x}_i)$, it is needed to estimate class membership probabilities conditional on a single predictor score of subject *i*, i.e., $P\big(H = t|x_{ip}\big)$. For example, when estimating the latent class membership given $x_{i,p=1}$, and $x_{i,p=1}$ is constantly changed, while holding $\mathbf{x}_{i,p\neq 1}$ fixed, the conditional class membership probability will be different for each different value of $x_{i,p=1}$. Following the same argumentation as in the MELODIC section, when using a standardized **X** in the LCA, $\mathbf{x}_{i,p\neq 1}$ and the corresponding $\mathbf{b}_{p\neq 1,t}$ are removed from the equation, which means that

$$P\big(H = t|x_{ip}\big) = \frac{\exp\big(a_t + x_{ip}b_{pt}\big)}{\sum_{t=1}^{T}\exp\big(a_t + x_{ip}b_{pt}\big)}, \tag{14}$$

for all predictors $p \in \{1, \ldots, P\}$.

Similarly, instead of estimating the complete response variable pattern probability for subject *i*, conditional on class membership, i.e., $P(\mathbf{y}_i|H = t)$, it is needed to estimate the probabilities for $y_{ir} = 1$ for each response variable *r*, conditional on class membership. These probabilities are the estimated class-conditional response probabilities, as described earlier, i.e., $P(y_{ir} = 1|H = t)$.

Altogether, combining the adapted elements of the latent class model for estimation of single predictor-response variable relationships, results in

$$P(y_{ir} = 1|x_{ip}) = \sum_{t=1}^{T} P(H = t|x_{ip})P(y_{ir} = 1|H = t), \quad (15)$$

for all predictors $p \in \{1, \dots, P\}$ and response variables $r \in \{1, \dots, R\}$. This formula will be used to calculate the predictor-response variable relationships for LCA.

3 Application study

In this section, the data used for this study will be introduced. Subsequently, MELODIC and LCA will be performed on the data, so that the reader has an idea how the models (as described in the previous section) are applied on real data.

3.1 SHARE data

In the current study, data are used from the Survey of Health, Ageing, and Retirement in Europe study (SHARE; Börsch-Supan et al., 2013). In the SHARE study, the effects of social, economic, environmental and health policies over the lifespan of Europeans are studied (SHARE, n.d.). SHARE involves participants of 50 years and older from 28 European countries and Israel. From 2004 until now, data are collected and bundled in data waves, of which Wave 8 has become available recently. In this study, only data from Wave 1 is used. Since MELODIC does not allow for missing data on the predictors (yet), cases with missing values were listwise deleted, resulting in data for 29207 subjects.

Bayes-Marin et al. (2020) used SHARE data (among others) to study multimorbidity and ageing. Eight NCDs and a set of predictors were selected to perform their analyses. In line with their study, the same predictor and response variables are used in this study. For computational considerations, however, it was decided to reduce the number predictor variables. Eight NCDs are chosen to serve as response variables, i.e.,

- Diabetes (9.97%);
- Hypertension (31.58%);
- Asthma (4.55%);

- Chronic Lung Disease (4.84%);
- Joint Disorders (18.57%);
- Angina (12.1%);
- Stroke (3.49%);
- Depression (37.82%)

All NCDs are measured on a binary scale with value 0, indicating that a subject does not suffer from the NCD, and 1, indicating that the subject does suffer from the NCD. Percentages of subjects suffering from the NCD are provided between parentheses. It can be observed already that, especially, Asthma, Chronic Lung Disease and Stroke are very infrequent NCDs in this data set. Importantly, there are subjects in the data set that do not suffer from an NCD and there are multimorbid subjects, therefore the NCDs are not mutually exclusive nor collectively exhaustive.

Three variables are selected to serve as predictor variables, i.e., Gender, Age and Self-rated health (SRH). Gender is measured on a binary scale, Age is measured on a continuous scale and SRH is measured on an ordered categorical scale. These predictors are standardized to have mean 0 and standard deviation (*SD*) 1. The predictor values of interest are:

- for Gender: the categories 0 (‘Male’) and 1 (‘Female’);
- for Age: the mean, +/- 1 *SD* from the mean and +/- 2 *SD* from the mean;
- for SRH: the categories 1 (‘Excellent’), 2 (‘Very Good’), 3 (‘Good’), 4 (‘Fair’) and 5 (‘Poor’).

In the next two subsections MELODIC and LCA will be applied to the SHARE data consisting of the variables mentioned above.

3.2 Application MELODIC

In a MELODIC analysis, it is decided first what the optimal number of dimensions is. For that purpose many models are fitted, each with an increased number of dimensions. As mentioned in the previous section on MELODIC, the researcher is only allowed to choose 1 to min(P,R) dimensions to fit the model with. In this case that would be one to three

dimensions, since $P = 3$ and $R = 8$. AIC and BIC are evaluated to find the optimal solution. These are provided in Table 1.

Table 1

AIC and BIC for models with 1 to 3 dimensions

Dim	Deviance	#param	AIC	BIC
1	162941.3	18	162977.3	163126.4
2	160697.9	27	160751.9	160975.5
3	160421.4	35	**160491.4**	**160781.3**

From the AICs and BICs presented in Table 1, it is clear that the fitted MELODIC model with three dimensions would be the optimal solution. However, the solutions with two and three dimensions do only differ very slightly. A model with two dimensions it is a more parsimonious solution and a biplot can be provided for a two-dimensional solution. For the sake of the example, it is decided to continue with the two-dimensional solution.

The subsequent step in a MELODIC analysis is to evaluate the influence of the predictor variables. Three two-dimensional models are fitted, each of them excluding one predictor. If the AIC and/or BIC is improved when excluding a predictor, it might be decided to exclude the predictor for further analysis. When there are a lot of predictors, this is an efficient procedure to reduce redundant variables. Since there are only three predictors in the model, it is questionable to execute this step. However, again, for the sake of the example, the influence of the predictors is checked. Table 2 provides the AICs and BICs of the models.

Table 2

AIC and BIC for the one-predictor-left-out models

Left out *p*	Deviance	#param	AIC	BIC
Gender	162359.7	25	162409.7	162616.8
Age	162076.3	25	162126.3	162333.4
SRH	158160.8	25	**158210.8**	**158417.8**

The results in Table 2 indicate that removing SRH from the model would lead to a substantial better fit. On the contrary, excluding Gender or Age would lead to a worse fit. However, due to reasons described above, it is decided to leave SRH in the model.

In Table 3, $\mathbf{a}^*$ and $\mathbf{B}^*$ are presented. Recall that these are the standardized logistic regression slope coefficients, one for each predictor-response variable relationship, and an intercept for each response variable.

Table 3

Implied coefficients and Q_r for the two-dimensional model

	Diabetes	Hypertension	Asthma	CLD	JD	Angina	Stroke	Depression
Intercept	-2.394	-0.830	-3.155	-3.313	-1.708	-2.380	-3.846	-0.552
Gender	-0.055	-0.004	0.095	-0.114	0.290	-0.377	-0.199	0.421
Age	0.254	0.186	0.117	0.334	0.144	0.488	0.437	-0.025
SRH	0.596	0.488	0.442	0.726	0.783	0.769	0.880	0.518
Quality	0.961	0.966	0.897	0.958	0.980	1.000	0.998	0.991

The coefficients can be interpreted as standardized logistic regression slope coefficients in single logistic regressions. Since standardized, they represent the change in log-odds when the predictor increases with 1 *SD*. These coefficients are used to calculate the estimated predictor-response variable relationships. When evaluating the coefficients, it becomes clear that, although probably not very surprising, getting older results in higher probabilities of suffering from an NCD (except for Depression). Also higher scores on SRH results in higher probabilities of suffering from an NCD (recall that higher scores represent lower rates of healthiness). Lastly, Females have higher probabilities of suffering from Asthma, Joint Disorders and Depression, whereas men have higher probabilities to suffer from the other NCDs (Female is dummy-coded 1).

A Quality of Representation (Q_r; De Rooij & Groenen, 2021) is presented in the last row of Table 3. It is a measure of how well the response variables are represented in the low-dimensional space, defined as

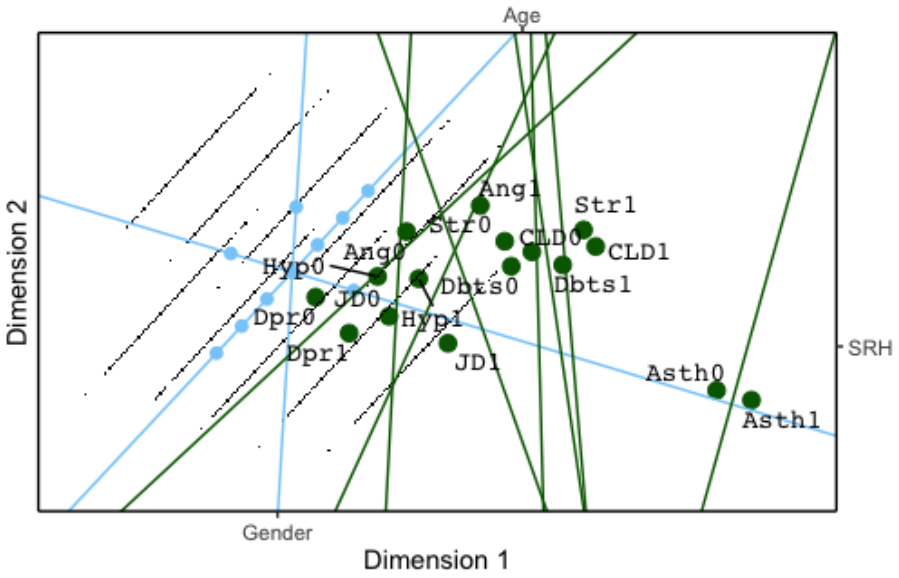
$$Q_r = \frac{L_{(0,r)} - L_r}{L_{(0,r)} - L_{lr}}, \tag{16}$$

in which $L_{(0,r)}$ is the deviance of the logistic regression model with only the intercept, for response variable *r*, L_r is the deviance part of the deviance function (to be minimized with the IM algorithm) for response variable *r*, and L_{lr} is the deviance of the logistic regression with all predictors involved, for response variable *r*. The measure is a proportion of loss in deviance for each response variable, by comparing the deviance part for response variable *r* when analyzing with MELODIC, with the deviance of a single logistic regression for response variable *r*.

It can be observed that Asthma is represented worst. The other NCDs are very well represented. Additionally, a biplot of the results is presented in Figure 1.

Figure 1

Biplot of the two-dimensional model



Note. Dbts = Diabetes, Hyp = Hypertension, Asth = Asthma, CLD = Chronic Lung Disease, JD = Joint Disorders, Ang = Angina, Str = Stroke and Dpr = Depression. NCD categories 'yes' are denoted with an additional 1. NCD categories 'no' are denoted with an additional 0.

In Figure 1, the green dots represent the response categories for the NCDs. The small black dots represent the locations of the subjects. Each NCD has a green line running orthogonally between its two response categories, the so-called decision lines. If a subject's position is on one of the green lines, it has an equal probability of suffering from that NCD. The blue lines are the predictor lines. The labels of the lines are indicative for the direction of the predictors, e.g., SRH is labeled at the right-side of the plot, indicating that higher scores of SRH suggests that subjects are placed more at the right-hand in the plot. Since **X** is standardized and the subject's position is defined as a linear combination of the predictors, subjects with mean scores on all predictors are placed at the position where the predictor lines cross each other.

From the figure, it is clear that all 1 categories are at the right of the 0 categories. Therefore, it can be concluded that, for Dimension 1, subjects suffering from NCDs will be more on the right-side of the figure. Furthermore, the labels of the predictor lines (which are at the right-side of the plot) suggest that higher scores of Age and SRH indicate higher probabilities of suffering from NCDs, whereas Gender behaves differently. The predictor line of Gender is practically orthogonal to the first dimension, indicating that the different values for Gender hardly affect the subject's position on the first dimension.

The interpretation of Dimension 2 is less clear. Since the response categories of the NCDs do not discriminate well in this dimension, it is difficult to determine what the dimension represents. Only the categories for Depression and Angina are relatively well discriminated, and the different values for Gender have a large effect on the subject's position on dimension 2, indicating that females are more likely of suffering from Depression, whereas males are more likely to suffer from Angina (recall that 'Female' is dummy-coded 1). Lastly, both categories for Asthma are placed at the very extreme to right of the mean. It might be due to the very infrequently scored category 1, and due to the relatively poor Q_r.

The predictor-response variable relationships are calculated using (8).

3.3 Application LCA

The first step in LCA is to decide how many latent classes is most optimal. Therefore, six models are fitted, from a model with one latent class to a model with six latent classes. Models with more than six classes could have been chosen, which makes this choice somewhat arbitrary. However, due to computational considerations, it was decided to estimate maximally six latent classes. Choosing the model with the optimal number of latent classes is solely based on the measurement part of the model. After deciding on how many latent classes will be used, the structural part will be included to re-estimate the model. The AICs and BICs for the models are presented in Table 4.

Table 4

AICs, BICs, classification accuracy and entropies for the LCA models

	LL	AIC	BIC	Chisq	npar	cl. acc.	entropy
1cl model	-87344.5	174705.0	174771.2	14293.45	8	1	1
2cl model	-85815.0	171663.9	171804.7	1505.6	17	0.852	0.361
3cl model	-85555.9	171163.8	171379.2	729.5232	26	0.834	0.373
4cl model	-85432.9	170935.8	171225.7	461.073	35	0.683	0.362
5cl model	-85370.2	170828.4	**171192.8**	318.2145	44	0.645	0.396
6cl model	-85332.2	**170770.3**	171209.3	242.8878	53	0.625	0.383

Note. cl. acc. = classification accuracy

According to AIC and BIC, a model with five or six latent classes would be most optimal. However, when evaluating the class sizes or class proportions of the fitted models, presented in Table 5, the six class solution might be reconsidered.

Table 5

Latent class proportions of the LCA models

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1cl model	1.000					
2cl model	0.763	0.237				
3cl model	0.059	0.192	0.748			
4cl model	0.350	0.041	0.420	0.189		
5cl model	0.337	0.077	0.174	0.370	0.043	
6cl model	0.050	0.312	0.078	0.325	0.027	0.208

It is clear from Table 5 that the six class solution contains estimated latent classes with a very low proportion of the subjects, indicating that only some of the subjects are assigned to these classes. Although the best fit based on AIC and BIC, it is probably not the most useful model. The five class solution has the same problem. Also in Table 4, the classification accuracy and entropy for each model are provided. The entropy is an index indicating how good the latent classes in the model are discriminating. It is considered a pseudo R-squared for separation of classes (Magidson & Vermunt, 2004). An index of 0.8 is suggested to indicate highly discriminating classes (Tein et al., 2013).

In Table 4, it can be observed that the four-, five- and six-class solution have a relatively low classification accuracy rate. Suggested classification accuracy cut-off values for acceptable classification are 0.8 and 0.9 (resp., Weden & Zabin, 2005; Muthén & Muthén, 2000). Therefore, the four-, five- and six-class solutions are excluded from consideration. In addition, it can be concluded that the entropies of all models indicate that the classes in all models are not highly discriminating. At this point, the conclusion might be drawn that LCA is not a suitable method for analyzing the SHARE data. However, for the sake of the example, it is decided to proceed. The one-class solution was taken into consideration, because if AIC and BIC would indicate that this is the best solution, again, LCA would not be the best suitable method of analysis. AIC and BIC indicate that the one-class solution has the worst fit, and therefore it is excluded from consideration.

The two-class solution has higher AIC and BIC values, compared to the three-class solution, the entropies and classification accuracies do not differ substantially. Therefore it is decided to select the three class solution as optimal model.

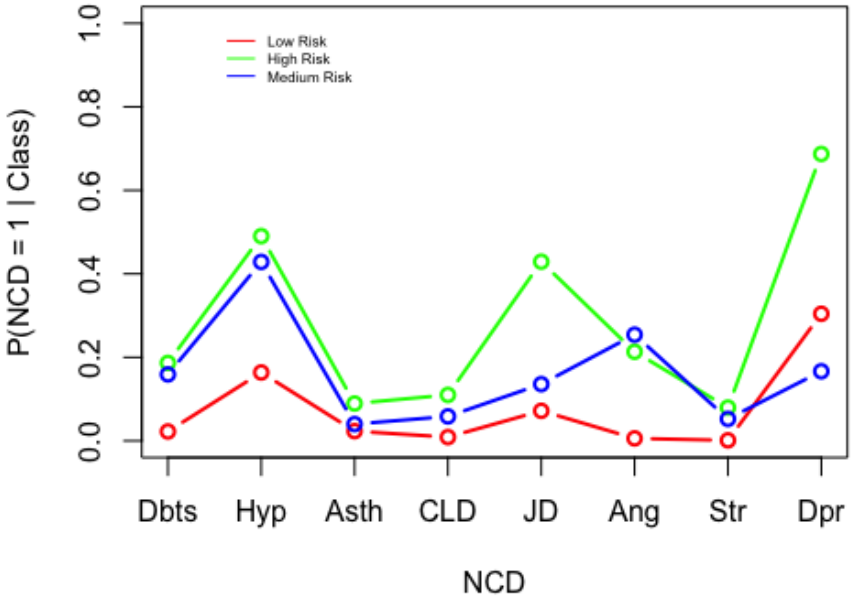
This example demonstrates that choosing the optimal model is based on fit measures, but also on personal preferences of the researcher. Note that it could be well-motivated to select the two-class solution, for example, as the best solution.

Having decided on the most optimal solution, the model with three classes is re-estimated, by including the structural part of the model. The re-estimated model has a classification accuracy of 0.813 (somewhat lower than when only estimating the measurement part of the three-class model) and an entropy of 0.566, which is a substantial increase, when compared to the model with only the measurement part (although still not satisfying).

The estimated class-conditional response probabilities for each NCD are presented in Figure 2 where each line represents one of three classes.

Figure 2

Estimated class-conditional probabilities per NCD for the three class model



Note. Dbts = Diabetes, Hyp = Hypertension, Asth = Asthma, CLD = Chronic Lung Disease, JD = Joint Disorders, Ang = Angina, Str = Stroke and Dpr = Depression.

The first latent class is the group with the lowest probability of suffering from NCDs, which will be named the Low Risk class. This class is presented in red in Figure 2 and has the highest class size proportion (0.490). The second class, presented as the green line in Figure 2, is the class with relatively high probabilities of suffering from NCDs (compared to the other two classes). This class will be referred to as the High risk class, and has a class size proportion of 0.277. The third class, presented in blue, is more or less in-between the Low Risk and High Risk classes, with respect to the class-conditional response probabilities. This class will be referred to as the Medium Risk class, and has a class proportion size of 0.233.

As can observed, the NCDs are not well discriminated between the classes. Especially for Asthma, Chronic Lung Disease and Stroke, the probabilities of suffering from them, hardly differ between the three classes. It represents the low entropy found in this model. In the SHARE data, these three NCDs were the least frequent NCDs, which could explain the low discriminability between the classes. As can be observed from Figure 2, the probabilities of suffering from NCDs given class membership are not high. Only the probability of suffering from Depression for the High Risk class is higher than 50%.

The multionomial logistic regression coefficients are presented in Table 6.

Table 6

Multionomial logistic regression coefficients (and standard errors (SEs)) for the three class model

	High Risk	(SE)	Medium Risk	(SE)
Intercept	-1.276	(0.074)	-0.684	(0.075)
Gender	0.438	(0.055)	-0.801	(0.049)
Age	1.046	(0.051)	1.272	(0.054)
SRH	3.106	(0.090)	1.354	(0.067)

The Low Risk class is the reference class, and therefore the coefficients in Table 6 are relative to the Low Risk class. It can be observed that the multinomial log-odds for High Risk and Medium Risk, relative to Low Risk, increase if values of Age and SRH increase. Increase in Gender (i.e., being Female instead of Male), results in an increase of the log-odds for High Risk, but in a decrease of the log-odds for Medium Risk.

On a global level, these findings are in accordance with the findings of the MELODIC analysis. That is, higher levels of Age and SRH correspond with higher probabilities of suffering from an NCD, whereas belonging to either one of the two categories for Gender does not seem to affect the NCD probabilities.

The predictor-response variable relationships are calculated using (15).
In Table 7, predictor-response variable relationships for both MELODIC and LCA are presented for Gender, by means of an example of the output.

Table 7

MELODIC- and LCA estimated predictor-response variable relationships for Gender.

	Diabetes	Hypertension	Asthma	CLD	JD	Angina	Stroke	Depression
MELODIC								
Male	0.088	0.305	0.037	0.040	0.116	0.124	0.026	0.264
Female	0.080	0.303	0.044	0.032	0.190	0.062	0.018	0.456
LCA								
Male	0.104	0.323	0.037	0.042	0.130	0.148	0.033	0.261
Female	0.084	0.284	0.042	0.041	0.170	0.094	0.028	0.379

The similarity of the estimated probabilities of the two methods is probably more remarkable than the differences between them. It can be observed that the direction of the predictor is the same between the two methods, e.g., if Female instead of Male, the probability of suffering from Diabetes decreases, for both methods. This is the case for all NCDs. Furthermore, it was concluded in the MELODIC analysis that Gender hardly influences the probabilities of suffering from NCDs, except for Angina and Depression. More or less the same effect of Gender is found in the LCA. Both methods estimate low probabilities of suffering from Asthma, Chronic Lung Disease and Angina, the infrequent NCDs.

4 Simulation studies

To compare the prediction accuracy of MELODIC and LCA, two Monte Carlo simulations are designed and conducted. In both simulations, variables are used as described in the previous section, i.e., eight response variables (the NCDs) and three predictor variables (Gender, Age, SRH). The simulations will be discussed in more detail in the next subsections.

4.1 Simulation 1

In the first simulation, the SHARE data and the model parameters from the two-dimensional MELODIC model (as defined in the application study section) are used to generate a large population data set (N = 500,000). Consequently, predictor-response variable relationships are estimated in the population, for each NCD for each predictor value, as

described in Section 3.1. From the population data, 100 random samples are drawn on which a MELODIC analysis with two dimensions and LCA with three classes are performed. Model estimates of MELODIC are used to calculate the MELODIC predictor-response variable relationships and model estimates of LCA are used to calculate the LCA predictor-response variable relationships, for each NCD for each predictor value as described in Section 3.1.

From this point it is possible to calculate the difference between the probabilities of suffering from one of each NCD given a predictor value in the population and the probabilities of suffering from one of each NCD given a predictor value in the sample analyzed with MELODIC or with LCA. This is done for all randomly drawn samples from the population. Consequently, all difference values are squared and averaged over the samples. Lastly, the square root is taken over these values, resulting in RMSEs for each predictor-response variable relationship. This RMSE measure is defined as

$$RMSE = \sqrt{\frac{\sum_{sam=1}^{SAM}\left(P\left(y_{ir}=1|x_{ip}\right)_{pop} - P\left(y_{ir}=1|x_{ip}\right)_{sam}\right)^2}{SAM}}, \qquad (17)$$

in which the $P\left(y_{ir}=1|x_{ip}\right)_{pop}$ denotes the probability for subject *i* to suffer from NCD *r*, given subject *i*'s score on predictor *p* in the population. $P\left(y_{ir}=1|x_{ip}\right)_{sam}$ denotes the probability for subject *i* to suffer from NCD *r*, given subject *i*'s score on predictor *p* in one of the randomly drawn samples from the population. These population and sample probabilities are compared with each other for each randomly drawn sample $sam \in \{1, \dots, SAM\}$. Using (17), RMSEs can be calculated for each response variable *r*, given each value (as described in Section 3.1) of each predictor *p*.

The simulation concerns a *2*x*2*x*2* design. That is, the procedure described above, was repeated eight times, one for each unique combination of varying parameters. The following parameters are varied:

- The discriminatory power of the response variables (**K**):
    - **K** from the real data.
    - **K** from the real data, Dimension 2 multiplied by 2.
- The response variable midpoint locations of the response categories (**L**):
    - **L** from the real data.
    - **L** from the real data divided by 3.

- The sample size (n):
    - n = 300.
    - n = 1000.

To be more precise, four times a large population data set (N = 500,000) is generated, one for each unique combination of conditions for **K** and **L**. From each of these four data sets, 100 random samples of size n = 300 are drawn, and 100 random samples of size n = 1000 are drawn. Hence, there are eight unique combinations of conditions.

As can be observed in Figure 1, Dimension 1 discriminates relatively well between the response variable categories. However, Dimension 2 hardly discriminates between the categories. Therefore, it was decided to add the condition in which the discriminatory power of the response variables (i.e., **K**) with respect to Dimension 2 is multiplied by two. It is interesting to see what the effect would be the effect of multiplying the second dimension of **K** by two on the prediction accuracy of MELODIC and LCA, in terms of RMSE, since this dimension hardly discriminates between the response variable categories.

As can be observed from Figure 1 as well, is that the locations of the midpoints of the response categories are all mildly to substantially to the right of the mean subject location. Recall that the mean is the location where the blue predictor lines cross each other, since **X** is standardized. Since all ‘yes’ categories are more to the right, compared to the ‘no’ categories, it is more likely in general that subjects do not suffer from NCDs than that they do suffer from them, especially with respect to Asthma. This is not very surprising since it was illustrated in Section 3.1 that subjects in general score ‘yes’ quite infrequently. Therefore it was decided to add a condition in which the midpoint location coordinates (i.e., **L**) are divided by three, so that the proportions of subjects scoring ‘yes’ and ‘no’ on the NCDs will be more equal to some extent. It is interesting to see what the effect of dividing **L** by three would have on the prediction accuracy of MELODIC and LCA in terms of RMSE, since the midpoints will be closer to the subject mean.

With respect to sample size, it is expected that a larger sample size will lead to better results in terms of RMSE compared to a smaller sample size, irrespective of which **K** and **L** condition is used. A sample size of n = 300 is chosen as being a minimal sample size. Therefore, the goal of adding this condition is to evaluate to what extent the prediction accuracy in terms of RMSE when using samples of size n = 1000 will outperform the prediction accuracy when using samples of size n = 300.

Important note on the conditions where the second dimension of the discriminatory power of the response variables (**K**) is multiplied by two; changing only one dimension results in a tilted decision line (whereas changing all dimensions with the same magnitude would not change the decision line). The midpoint locations of the response variable categories (**L**) need to be corrected for this, when generating population data. Each midpoint coordinate of each response variable is updated by

$$l_{rm}^{+} = \frac{(\sum_{m=1}^{M} k_{rm} l_{rm}) * k_{rm}}{\sum_{m=1}^{M} k_{rm}^{2}}. \quad (17)$$

This update gives the correct midpoint locations of the response variable categories when the second dimension of the discriminatory power of the response variables is multiplied by two. Therefore, when changing **K** by multiplying its second dimension, **L** is also adapted to some extent.

4.2 Simulation 2

In the second simulation, the same procedure is followed as in the first simulation, with respect to calculating the predictor-response variable relationships and the RMSEs, i.e., generate population data, calculate the predictor-response variable relationships in the population, draw samples, perform MELODIC with two dimensions and LCA with three classes on the samples, calculate the predictor-response variable relationships in the samples, and lastly, calculate the RMSEs (using (17)). The only aspect that differs, is that the population data are generated with parameters from the three-class LCA solution, as described in the application study section.

This simulation also concerns a *2*x*2*x*2* design, in which some of the latent class model parameters vary. The following parameters are varied:

- The class conditional probabilities ($\boldsymbol{\pi}$):
    - $\boldsymbol{\pi}$ from the real data.
    - Modified $\boldsymbol{\pi}$ (i.e., Probabilities of 0.2 to score 1 (‘yes’) on all NCDs for the Low Risk class, probabilities of 0.8 to score 1 (‘yes’) on all NCDs for the High Risk class, probabilities of 0.8 to score 1 (‘yes’) on the first four NCDs and

probabilities of 0.2 to score 1 ('yes') on the last four NCDs for the Medium Risk class).

- The latent class sizes:
    - Latent class sizes from the real data.
    - Equal latent class sizes, each class containing 1/3 of the subjects.
- The sample size (*n*):
    - *n* = 300.
    - *n* = 1000.

Also in this simulation, four large population data sets are generated (*N* = 500,000), one for each unique combination of conditions of π and the latent class sizes. From each of the population data sets, 100 random samples are drawn of size *n* = 300 and 100 random samples are drawn of size *n* = 1000. Hence, there are eight unique combinations of conditions.

As was clear from Figure 2, most of the items are not well discriminated between the classes, resulting in a low entropy. Therefore it was decided add a condition to pull the lines further away from each other. By using modified conditional class probabilities as described above (modified π), the classes become highly discriminating, resulting in a substantially higher entropy. It will be evaluated what the effect will be of modifying π on the prediction accuracy of MELODIC and LCA, in terms of RMSE.

In general, latent class models with equal class proportions also have a higher entropy, in contrast to latent class models with unequal class proportions (Nylund et al., 2007; Gudicha, 2015). Therefore it was chosen to add the condition in which there are equal class sizes. It will be evaluated what the effect of this condition is on the prediction accuracy of LCA and MELODIC, in terms of RMSE.

Similar to the first simulation, it was chosen to draw samples with sample size *n* = 300 and *n* = 1000. It has been suggested that the minimal sample size for performing LCA is *n* = 500 (cf. Finch & Bronk, 2011), for obtaining accurate estimates. However, since sample sizes *n* = 300 and *n* = 1000 were used in Simulation 1, and due to the exploratory nature of the study, it was decided to use these sample sizes. It is expected that the prediction accuracy in terms of RMSE is higher when using a sample size of *n* = 1000, compared to using a sample of size *n* = 300.

4.3 Hypotheses

Two simulation studies were illustrated. One in which population data are generated with MELODIC parameters and one in which population data are generated with LCA parameters.

In both the simulations, two sets of model parameters are varied to create four population data sets. The predictor-response variable relationships are estimated in the population, and consequently, 100 random samples of size $n = 300$ and 100 random samples of size $n = 1000$ are drawn from the populations, resulting in data for eight conditions. MELODIC and LCA are performed on these samples and the predictor-response variable relationships are calculated in the samples. Consequently, the prediction accuracy of both methods is determined in terms of RMSE, by comparing the predictor-response variable relationships in the populations with the predictor-response variable relationships in the samples.

It is hypothesized that the prediction accuracy of the method is dependent on the data generation method. For example Simulation 1, where population data are generated with MELODIC parameters, performing MELODIC on the samples is hypothesized to result in a higher prediction accuracy than when LCA is performed on the samples, irrespective of the condition. The prediction accuracies of MELODIC and LCA are hypothesized to be higher, in terms of RMSE, when analyzing samples of size $n = 1000$, compared to analyzing samples of size $n = 300$.

With respect to the method specific parameters (i.e., $\mathbf{K}$ and $\mathbf{L}$), modifications of these parameters from the real SHARE data have been chosen to evaluate what their effects will be on the prediction accuracy of both MELODIC and LCA. It might be interesting to get a grasp of how these changes affect the RMSEs for both methods. Due to the exploratory nature of the study, no hypothesis is specified, concerning their effect.

For the second simulation in which population data sets are generated with LCA parameters, the hypotheses are the exact opposite of the hypotheses described above, since in this simulation, LCA is expected to outperform MELODIC in terms of RMSE. Recall, in Simulation 2, the method specific parameters are $\boldsymbol{\pi}$ and the latent class sizes, instead of $\mathbf{K}$ and $\mathbf{L}$.

4.4 Statistical analysis

To test the hypotheses, for both simulations a Mixed model Analysis of Variance (ANOVA) will be conducted to evaluate the influence of the method (i.e., MELODIC or LCA), the parameters and predictor and response variables on the RMSEs that are collected after doing the simulation studies. The mixed ANOVA concerns a *2*x*2*x*2*x*2*x*3*x*8* design. For Simulation 1, the factors are (respectively):

- Method: MELODIC and LCA,
- *n*: 1) 300, 2) 1000,
- *K*: 1) *K* from the real data, 2) *K* from the real data, Dimension 2 multiplied by 2,
- *L*: 1) *L* from the real data, 2) *L* from the real data divided by 3,
- *X*: 1) Gender, 2) Age, 3) SRH,
- *Y*: 1) Diabetes, 2) Hypertension, 3) Asthma, 4) Chronic Lung Disease, 5) Joint Disorders, 6) Angina, 7) Stroke, 8) Depression.

For Simulation 2, the same factors are used, except for *K* and *L*. Instead, π (π from the real data and modified π) and Latent class sizes (unequal and equal), respectively, are used. Importantly, the levels of all these factors denote the involvement of a condition, rather than the actual values of that condition (e.g., the RMSEs for which $X = 2$ are all the RMSEs in which Age was involved as predictor variable to estimate the predictor-response variable relationships. It is solely a level of the factor *X*). Therefore, to avoid confusion, when referring to output of the simulations and factors in the Mixed ANOVA, the no-bold cursive notation *K*, *L*, *X*, *Y* and π will be used.

The design is a mixed design, since it consists of one within-subjects factor variable (i.e., Method), and between-subjects factors (i.e., the other factors). Method is a within-subjects factor, since both MELODIC and LCA were performed on each sample that was drawn from a population. The main interest of these analyses is to evaluate the effects of the between-subjects factors, as specified above, and their interaction with the within-subjects factor Method on the RMSEs. The effects will be evaluated using the *p*-values and the generalized eta squared (η_G^2; Olejnik & Algina, 2003). Generalized eta squared is a measure of effect size that is commonly used for Repeated Measures and Mixed ANOVAs. These effect sizes of factors can be compared more easily between different research designs, irrespective

of the factor being a within-subjects or a between-subjects factor (Bakeman, 2005). The magnitude of an effect for eta squared (η^2), as suggested by Cohen (1988) are assumed to represent the same magnitude as the generalized eta squared, i.e., small effect: $\eta^2 = \eta_G^2$ = 0.010, medium effect: $\eta^2 = \eta_G^2$ = 0.060, and large effect: $\eta^2 = \eta_G^2$ = 0.140 (Olejnik, Algina, 2003).

Since it is aimed to compare the prediction accuracy of MELODIC and LCA, it is interesting to evaluate whether the performance of both methods depend on the different levels of the factors.

The simulations and mixed model ANOVAs were performed using R version 4.1.0 (R Core Team, 2021). Code developed by De Rooij and Groenen (2021) was obtained and used to perform the MELODIC analyses in R. The poLCA package (Linzer & Lewis, 2011) in R was used to perform the LCAs and the stats package in R was used to perform the Mixed model ANOVAs (R Core Team, 2021).

5 Results

In the next subsections, the results of the simulations and the Mixed ANOVAs will be discussed.

5.1 Results Simulation 1

In the first simulation, RMSEs for predictor-response variable relationships, in varying conditions are calculated, after performing MELODIC with two dimensions and LCA with three classes on samples from populations that were created with MELODIC parameters.

First, it can be concluded that, averaged over all conditions, MELODIC has a lower mean RMSE (*M* = 0.027, *SD* = 0.011), compared to LCA (*M* = 0.057, *SD* = 0.032), indicating that irrespective of condition, MELODIC has a higher prediction accuracy. In Table 8 the RMSEs per method per unique combination of parameters are presented, averaged over the predictors and the response variables. A more detailed overview of the RMSEs per method

per unique combination of parameters per predictor per response variable is provided in Appendix A.

Table 8

Averaged RMSEs per method for each combination of varying parameters

			MELODIC	LCA
n=300,	*K*=*K*,	*L*=*L*.	0.032	0.047
n=1000,	*K*=*K*,	*L*=*L*.	0.017	0.030
n=300,	*K*=*K*dim2*2,	*L*=*L*.	0.029	0.059
n=1000,	*K*=*K*dim2*2,	*L*=*L*.	0.015	0.034
n=300,	*K*=*K*,	*L*=*L*/3.	0.039	0.078
n=1000,	*K*=*K*,	*L*=*L*/3.	0.022	0.049
n=300,	*K*=*K*dim2*2,	*L*=*L*/3.	0.040	0.099
n=1000,	*K*=*K*dim2*2,	*L*=*L*/3.	0.022	0.059

In accordance with the first conclusion, MELODIC has a higher prediction accuracy when compared with LCA, for all combinations of parameters. When using a sample size of n = 300, both MELODIC and LCA scored substantially higher average RMSEs, when compared to using n = 1000. In the conditions where the second dimension of K is multiplied by two, the difference between the RMSEs of MELODIC and LCA is larger than when K from the real data were used. The same can be observed with respect to L. This indicates that changing MELODIC parameters in the population makes the contrast between prediction accuracy of MELODIC and LCA larger, in favor of the MELODIC analysis.

From these first observations, it becomes clear that the different conditions of factors have a different effect on the RMSEs for MELODIC than for LCA, suggesting an interaction effect of the factor and method on the RMSEs.

The results of the Mixed ANOVA indicate that there is a large significant main effect of Method on RMSE. The RMSEs of the MELODIC analyses are significantly lower (M = 0.027, SD = 0.011) than the RMSEs of the LCA (M = 0.057, SD = 0.032; $F(1,179) = 1299.66$, $p < 0.001$, $\eta_G^2 = 0.700$).

Before discussing the interaction statistics, the RMSE means and *SD*s for each combination of the two Method levels and the levels from the between-subjects factors are presented in Table 9.

Table 9

Means and SDs of RMSEs for the interaction groups

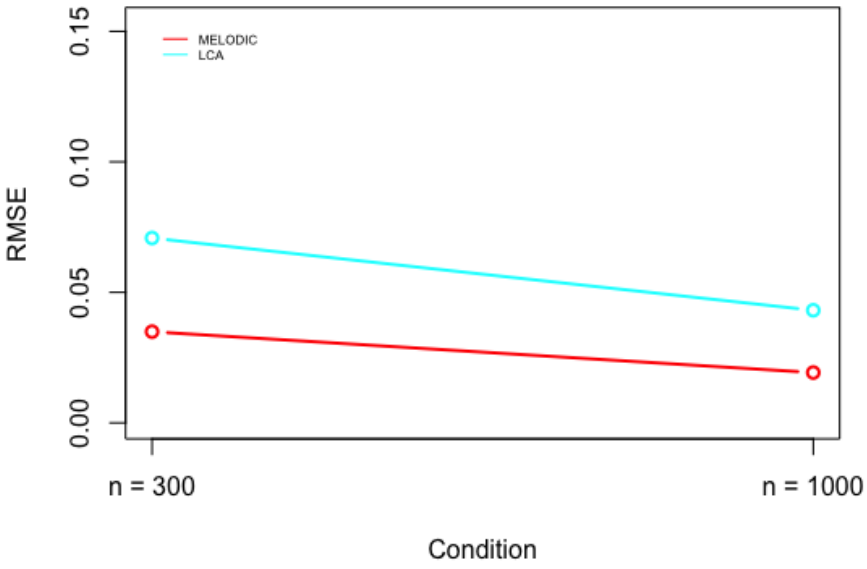
Factor	Factor level	MELODIC		LCA	
		M	*SD*	*M*	*SD*
n	300	0.035	0.009	0.071	0.034
	1000	0.019	0.005	0.043	0.022
K	*K*	0.028	0.010	0.051	0.026
	*K*dim2*2	0.026	0.011	0.063	0.036
L	*L*	0.023	0.010	0.043	0.028
	L/3	0.031	0.009	0.071	0.028
X	Gender	0.026	0.010	0.050	0.026
	Age	0.027	0.011	0.059	0.034
	SRH	0.028	0.010	0.063	0.034
Y	Diabetes	0.026	0.010	0.044	0.020
	Hypertension	0.030	0.009	0.058	0.017
	Asthma	0.022	0.009	0.033	0.021
	CLD	0.023	0.011	0.041	0.026
	JD	0.029	0.009	0.074	0.028
	Angina	0.029	0.009	0.072	0.029
	Stroke	0.022	0.010	0.043	0.030
	Depression	0.035	0.012	0.091	0.031

A first look at Table 9 (being a reformulated extension of Table 8), indicates that, again, MELODIC outperforms LCA, in terms of RMSE, in each condition. Using $n = 1000$ leads to lower RMSEs than $n = 300$ in both MELODIC and LCA conditions. The second *K* condition in which the second dimension of *K* is multiplied by 2, yield more or less the same RMSEs for MELODIC and higher RMSEs for LCA, compared to using *K* from the real data. When the condition of *L* divided by 3 is used, an increase of RMSE is observed for both MELODIC and LCA, compared to using *L* from the real data.

There was a significant interaction effect, with a medium to large effect size, of factor *n* and Method on RMSE ($F(1,179) = 52.53$, $p < 0.001$, $\eta_G^2 = 0.086$). The interaction is displayed in Figure 3.

Figure 3

Mean RMSEs per group in the Method and n interaction

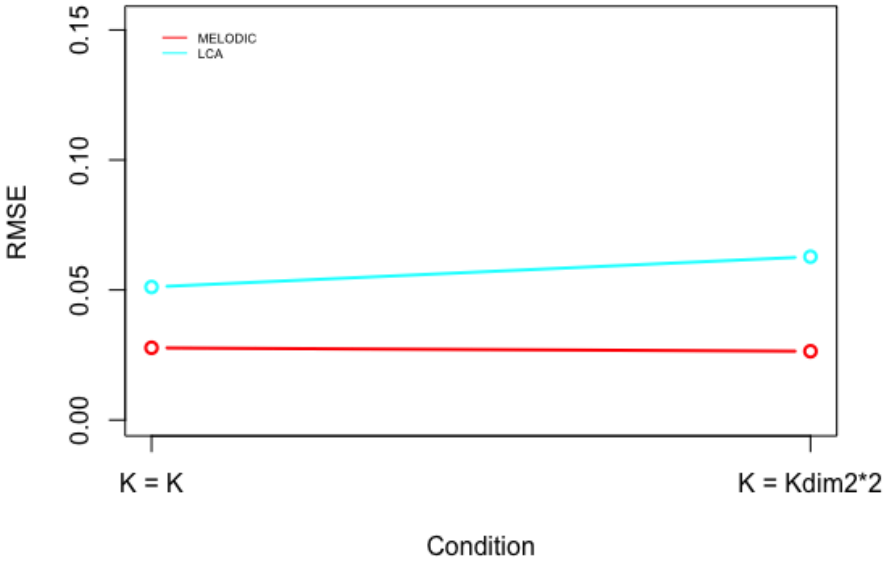


From the figure it is clear that MELODIC outperforms LCA, in terms of RMSE, in each level of *n*. Also, it can be observed that the magnitude of the RMSE differences between MELODIC and LCA is larger for sample size $n = 300$ than for $n = 1000$. A sample size of $n = 1000$ is an improvement for both MELODIC and LCA over a sample size of $n = 300$.

Also, there was a significant interaction effect, with a medium to large effect size, of factor *K* and Method on RMSE ($F(1,179) = 61.45$, $p < 0.001$, $\eta^2_G = 0.100$). The interaction is displayed in Figure 4.

Figure 4

Mean RMSEs per group in the Method and K interaction

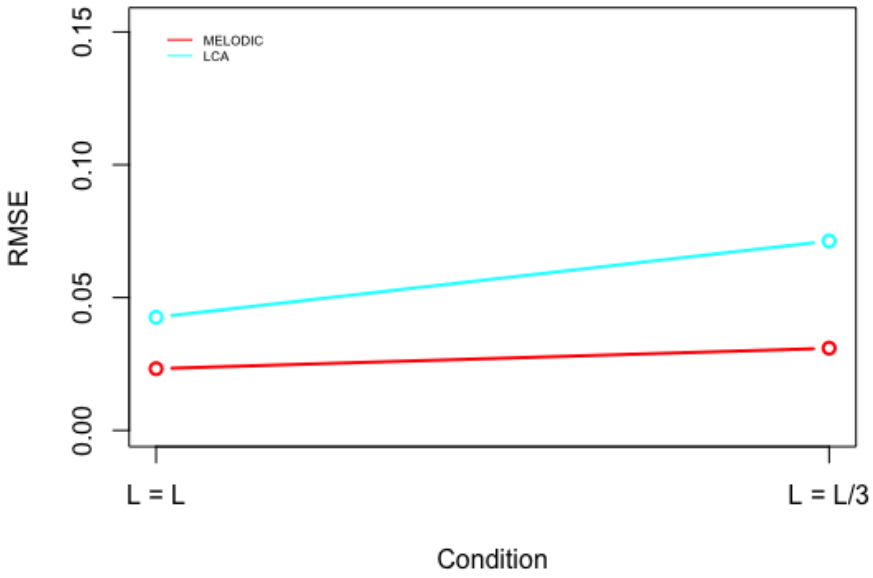


From this Figure it is clear that, also here, MELODIC outperforms LCA in terms of RMSE. The difference between the RMSEs of MELODIC and LCA are larger in the condition where the second dimension of *K* is multiplied by 2, compared to the real data *K* condition. Whereas MELODIC RMSEs seem to be stable over the two *K* conditions, the prediction accuracy of LCA deteriorates in the second *K* condition.

A significant interaction effect, with a large effect size, of factor *L* and Method on RMSE was observed ($F(1,179) = 160.76$, $p < 0.001$, $\eta^2_G = 0.224$). The interaction effect is displayed in Figure 5.

Figure 5

Mean RMSEs per group in the Method and L interaction

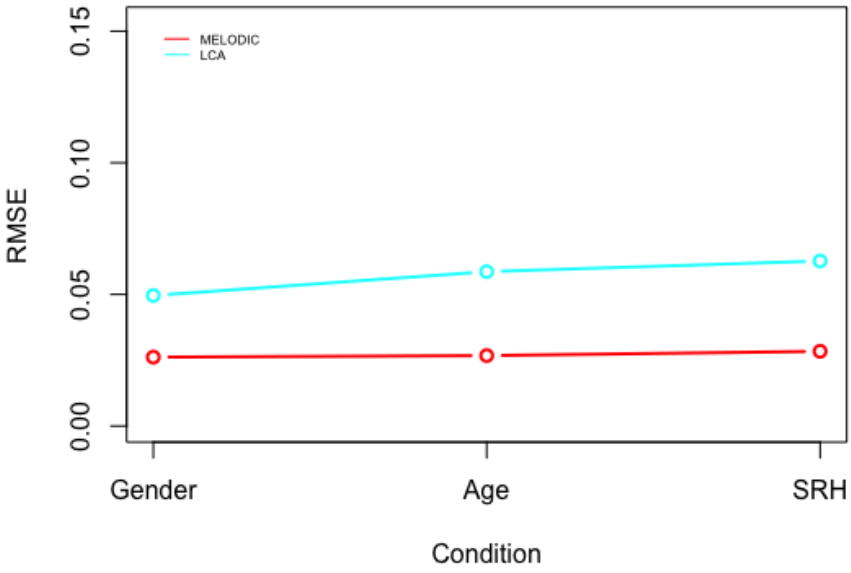


What can be observed from this Figure, is that it is clear that MELODIC outperforms LCA in terms of RMSE. The difference between the RMSEs of MELODIC and LCA are larger in the condition where *L* is divided by 3, and the RMSEs of both MELODIC and LCA are higher in the second *L* condition in contrast to the *L* = *L* condition, although the increase is larger for LCA than for MELODIC

For the factor *X* and Method, a significant interaction effect, with a medium effect size, was observed ($F(2,179) = 15.86$, $p < 0.001$, $\eta^2_G = 0.054$). Figure 6 provides a graphical display of the interaction.

Figure 6

Mean RMSEs per group in the Method and X interaction

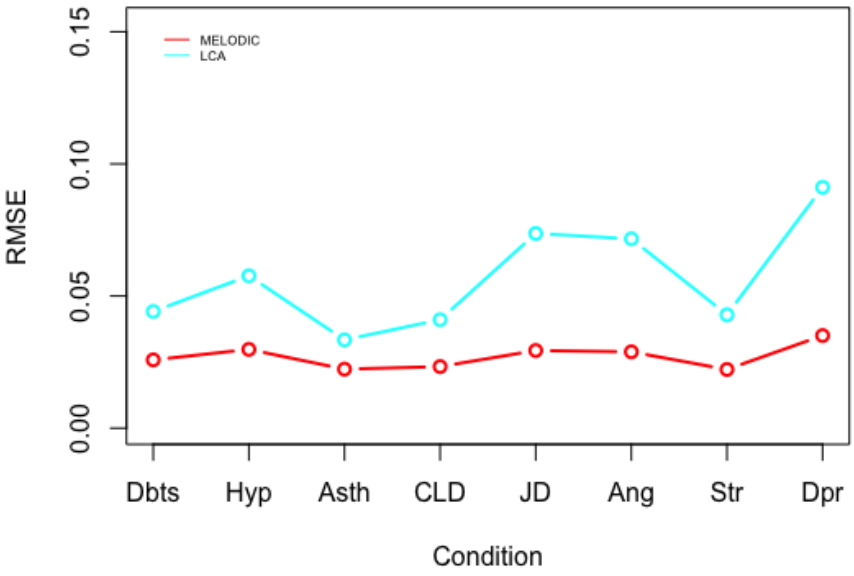


It can be observed that the differences between mean RMSEs for MELODIC and LCA are relatively stable when comparing them for each of the three predictors. Also here, in general, MELODIC has a higher prediction accuracy than LCA. The difference seems to be smallest between MELODIC and LCA when it concerns Gender. Using one of the three levels of X, i.e., Gender, Age and SRH, being respectively, a binary, continuous and ordered categorical variable, does not seem to yield differences in RMSE, when analyzing with MELODIC. When analyzing with LCA, Gender has a slight advantage over Age and SRH in terms of RMSE.

Lastly, a significant interaction effect of factor *Y* and Method on RMSE, with a high effect size, was found ($F(7,179) = 46.44$, $p < 0.001$, $\eta_G^2 = 0.369$), for which the visualization is provided in Figure 7.

Figure 7

Mean RMSEs per group in the Method and Y interaction



Also for *Y*, MELODIC outperforms LCA in terms of RMSE, irrespective of the response variable that was involved in calculating the predictor-response variable relationships. The magnitude of the RMSE differences between MELODIC and LCA differ between the levels of *Y*, e.g., there is a large RMSE difference for Depression, whereas there is a relatively small difference for Asthma. Although the RMSEs are more or less stable over the levels of *Y* for MELODIC, they a more different for LCA. Interestingly, the highest LCA RMSEs are for the levels Hypertension, Joint Disorders, Angina and Depression, which are the most frequent NCDs in the real data.

To summarize, there was a large main effect of Method ($\eta_G^2 = 0.700$), a medium to large interaction effect of factor *n* and Method ($\eta_G^2 = 0.086$), a medium to large interaction effect of factor *K* and Method ($\eta_G^2 = 0.100$), a large interaction effect of factor *L* and Method ($\eta_G^2 = 0.224$), a medium interaction effect of factor *X* and Method ($\eta_G^2 = 0.054$) and a high interaction effect of factor *Y* and Method ($\eta_G^2 = 0.369$) on RMSE. For all these effects, RMSEs resulting from MELODIC analyses are lower than RMSEs resulting from LCA

analyses, indicating that has MELODIC has a higher prediction accuracy than LCA in this simulation.

A larger sample size is beneficial for both MELODIC and LCA and the RMSE differences between MELODIC and LCA are smaller in larger sample size conditions. With respect to the MELODIC specific factors, i.e., *K* and *L*, the RMSEs resulting from the MELODIC analyses remain stable over the different levels of these factors, whereas the RMSEs resulting from the LCA analyses become larger in the modified levels of these factors.

No substantial differences in MELODIC RMSEs were found for the different levels of *X* and *Y*, whereas the binary Gender variable was outperforming Age and SRH, for LCA, and infrequent NCDs outperformed the frequent NCDs in terms of RMSE, for LCA.

5.2 Results Simulation 2

As for the second simulation study, RMSEs for predictor-response variable relationships in varying conditions were calculated, after performing MELODIC with two dimensions and LCA with three classes on samples from populations that are created with LCA parameters.

First, averaging over all conditions, it is observed that the RMSEs of LCA are substantially lower ($M = 0.037$, $SD = 0.022$) than the RMSEs of MELODIC ($M = 0.060$, $SD = 0.033$), indicating that, in general, LCA outperforms MELODIC in terms of RMSE, and therefore has a higher prediction accuracy. In Table 10 the RMSEs per method per unique combination of varying parameters are presented, averaged over the predictors and the response variables. A more detailed overview of the RMSEs per method per unique combination of parameters per predictor per response variable is provided in Appendix B.

Table 10

Averaged RMSEs per method for each combination of varying parameters

			MELODIC	LCA
n=300,	π=π,	uneq.class.	0.032	0.034
n=1000,	π=π,	uneq.class.	0.020	0.016
n=300,	π=πmod,	uneq.class.	0.056	0.043
n=1000,	π=πmod,	uneq.class.	0.045	0.024
n=300,	π=π,	eq.class.	0.043	0.042
n=1000,	π=π,	eq.class.	0.035	0.028
n=300,	π=πmod,	eq.class.	0.111	0.076
n=1000,	π=πmod,	eq.class.	0.106	0.067

From Table 10 it becomes clear that in Simulation 2, in contrast to Simulation 1, LCA has lower RMSEs than MELODIC in all unique combinations of varying parameters, except for the first row of the table, where the mean RMSE for MELODIC is slightly smaller than the mean RMSE for LCA. Also, it can be noted that in the first row of Table 10, the RMSEs do not differ that extensively. Conditions in which a sample size of n = 1000 is used, resulted in lower RMSEs for both MELODIC and LCA, compared with conditions in which a sample size of n = 300 is used. The magnitude of the difference between MELODIC and LCA is larger when using n = 1000. When using modified π, the magnitude of the difference between MELODIC and LCA becomes larger in favor of LCA. Finally, with respect to class sizes, in all conditions in which unequal and equal class sizes are used, LCA scores better than MELODIC in terms of RMSE, and the difference in RMSE becomes larger when class sizes are equal.

Also in Simulation 2, a Mixed ANOVA is performed on the RMSEs. From the results it becomes clear that, when averaging over all other conditions, LCA (M = 0.037, SD = 0.022) outperforms MELODIC (M = 0.060, SD = 0.033) in terms of RMSE (F(1,197) = 1119.40, p < 0.001, η_G^2 = 0.328), indicating that LCA in general has a higher prediction accuracy in contrast to MELODIC, in this second simulation.

Before evaluating the interaction effect results, the RMSE means and *SD*s for each combination of the two Method levels and the levels from the between-subjects factors are presented in Table 11.

Table 11

Means and SDs of RMSEs for the interaction groups

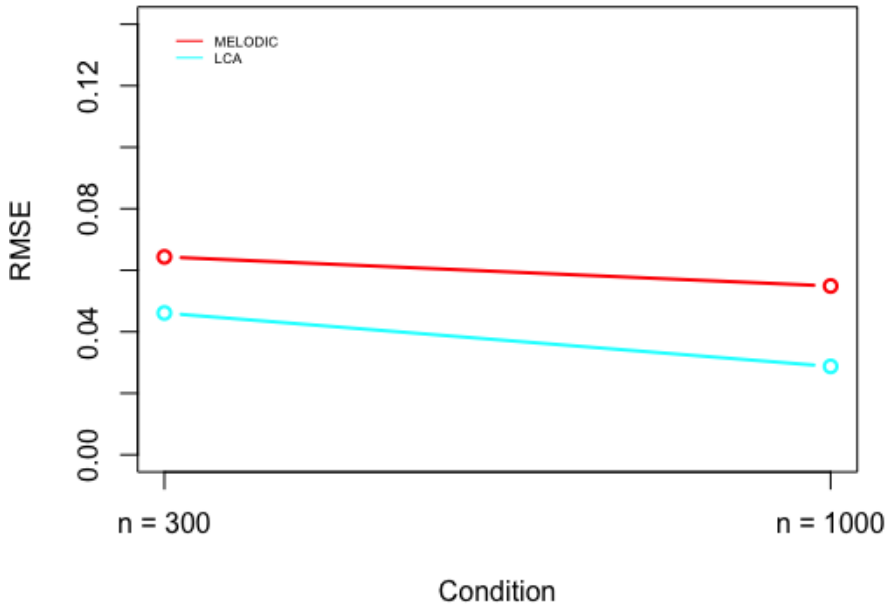
Factor	Factor level	MELODIC		LCA	
		M	*SD*	*M*	*SD*
n	300	0.064	0.032	0.046	0.020
	1000	0.055	0.034	0.029	0.020
π	π	0.037	0.019	0.030	0.018
	modified π	0.082	0.029	0.045	0.022
class size	unequal	0.045	0.020	0.029	0.014
	equal	0.074	0.038	0.045	0.025
X	Gender	0.056	0.035	0.041	0.027
	Age	0.067	0.036	0.037	0.020
	SRH	0.056	0.027	0.034	0.017
Y	Diabetes	0.060	0.040	0.041	0.025
	Hypertension	0.070	0.033	0.051	0.022
	Asthma	0.055	0.044	0.035	0.028
	CLD	0.056	0.044	0.036	0.029
	JD	0.060	0.023	0.036	0.012
	Angina	0.058	0.023	0.035	0.012
	Stroke	0.048	0.031	0.024	0.013
	Depression	0.071	0.014	0.041	0.016

Also from Table 11, it can be noted that in all conditions, LCA has a better performance than MELODIC. As in Simulation 1, using the $n = 1000$ condition yield lower RMSEs for both MELODIC and LCA in contrast to using the $n = 300$ condition. In the modified π condition, RMSEs are higher for both MELODIC and LCA, in contrast to the π from the real data condition. The MELODIC and LCA RMSEs for equal class sizes are higher than for unequal class sizes.

There was a significant interaction effect, with small effect size, of factor *n* and Method on RMSE ($F(1,179) = 35.52$, $p < 0.001$, $\eta_G^2 = 0.015$). The interaction is presented in Figure 8.

Figure 8

Mean RMSEs per group in the Method and n interaction

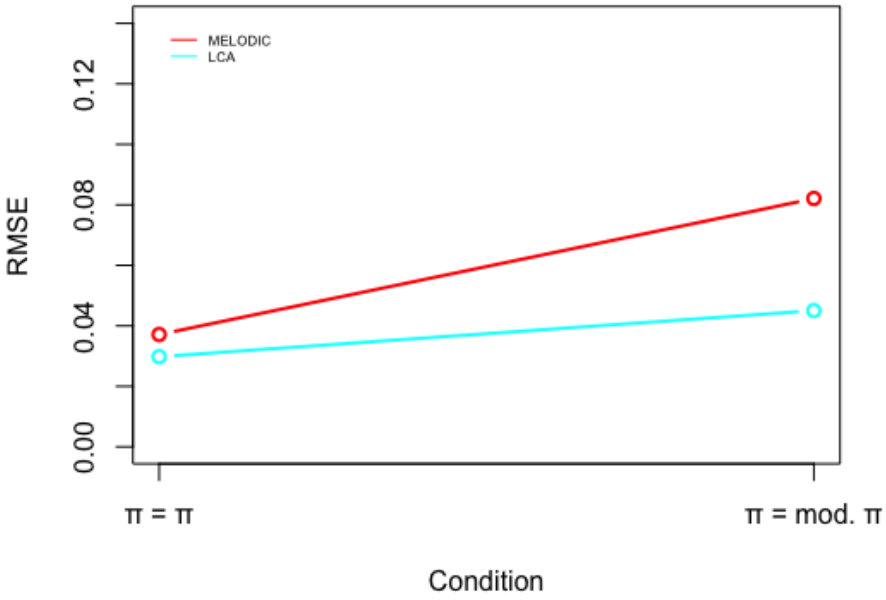


From Figure 8, it can be observed that LCA outperforms MELODIC for each condition of n. The difference between RMSEs is larger for the *n* = 1000 condition, and for both MELODIC and LCA, using *n* = 1000 is an improvement over using *n* = 300.

There also was a significant interaction effect of factor π and Method, with a large effect size, on RMSE ($F(1,179) = 502.54$, $p < 0.001$, $\eta^2_G = 0.180$). The interaction effect is visualized in Figure 9.

Figure 9

Mean RMSEs per group in the Method and π interaction

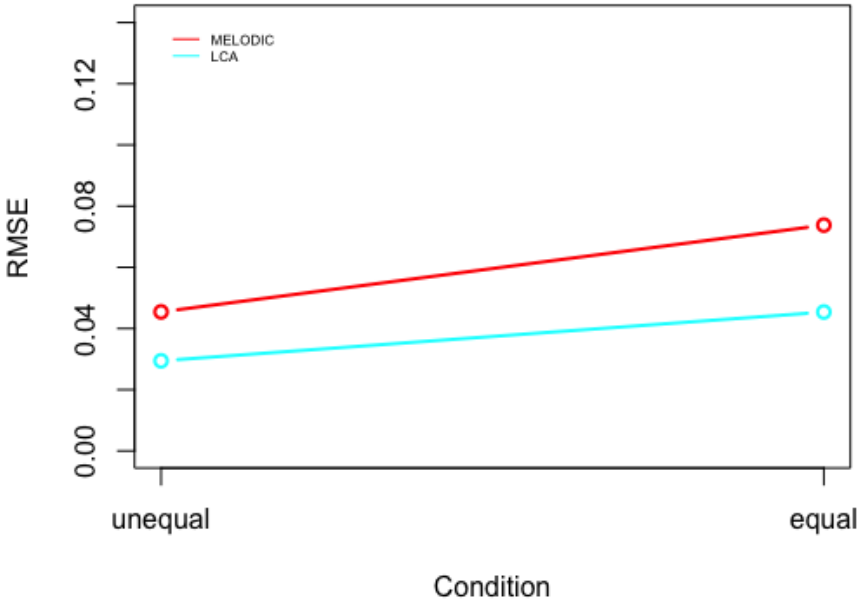


It seems that when using π from the population, the difference in RMSE between MELODIC and LCA is not large, whereas it is larger when using the modified π condition. The π from the real data condition results in lower RMSE, compared to the modified π condition, for both MELODIC and LCA In general, also here, LCA outperforms MELODIC in both conditions.

There was a significant interaction effect, with small to medium effect size, of the class size factor and Method on RMSE ($F(1,179) = 87.46$, $p < 0.001$, $\eta^2_G = 0.037$). This is in line with what was observed from Table 10 and 11. This is illustrated in Figure 10.

Figure 10

Mean RMSEs per group in the Method and class size interaction

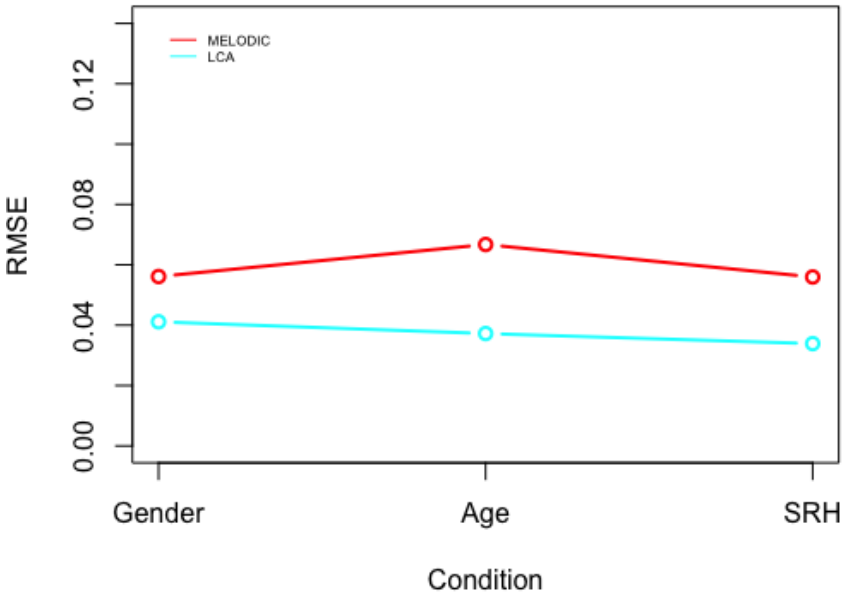


As can be observed from Figure 10, the MELODIC and LCA RMSE differences are larger when there were equal class sizes in the populations. RMSEs in the equal class sizes condition are higher for both MELODIC and LCA, compared to the RMSEs in the unequal class condition from the real data. Overall, it can be observed that also for the class size factor, LCA outperforms MELODIC in terms of RMSE.

A significant interaction effect of factor *X* and Method, with small to medium effect size, on RMSE, was found ($F(2,179) = 39.83$, $p < 0.001$, $\eta^2_G = 0.034$), which is visualized in Figure 11.

Figure 11

Mean RMSEs per group in the Method and X interaction

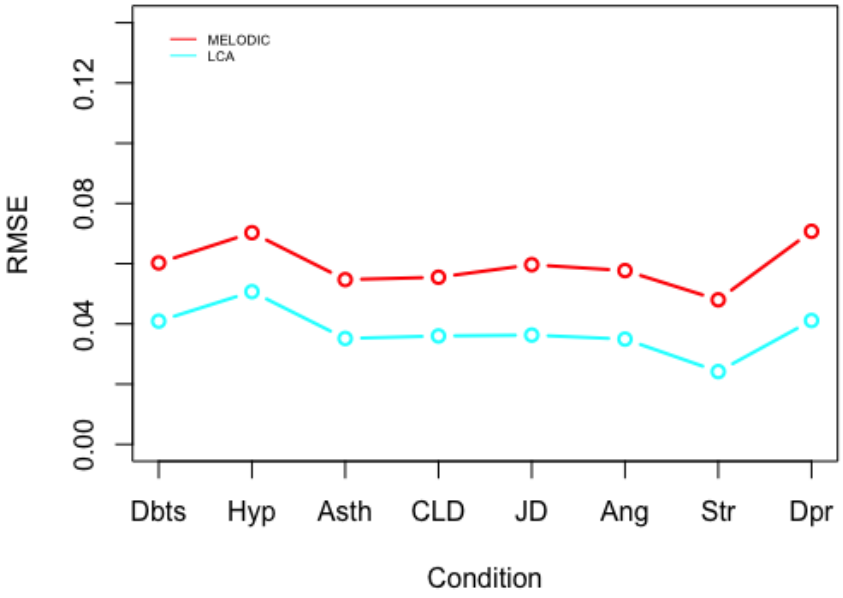


The MELODIC and LCA RMSE differences are larger for Age, than for Gender and SRH. Whereas the mean MELODIC RMSE is the largest for the Age level, the LCA RMSE is largest for Gender. In general, also for *X*, it can be concluded that using LCA results in lower RMSEs than using MELODIC analyses, in this simulation.

Finally, there was a significant interaction effect, with a small effect size, of factor *Y* and Method on RMSE ($F(7,179) = 3.61$, $p = 0.001$, $\eta_G^2 = 0.011$), indicating that the differences in RMSE between MELODIC and LCA differed for the various levels of *Y*. A visualization is provided in Figure 12.

Figure 12

Mean RMSEs per group in the Method and Y non-interaction



From Figure 11, it is clear that the lines nearly have the same shape and the differences between RMSEs for MELODIC and LCA are more or less the same for each level of factor *Y*. However, the mean MELODIC and LCA RMSE differences are slightly larger for the last four NCDs than for the first four NCDs. LCA outperformed MELODIC in terms of RMSE for all conditions of *Y*.

To summarize, there was a large main effect of Method ($\eta_G^2 = 0.328$), a small interaction effect of factor *n* and Method ($\eta_G^2 = 0.015$), a large interaction effect of factor π and Method ($\eta_G^2 = 0.180$), a small to medium interaction effect of the class size factor and Method ($\eta_G^2 = 0.037$), a small to medium interaction effect of factor *X* and Method ($\eta_G^2 = 0.036$) and a small significant interaction effect of factor *Y* and Method ($\eta_G^2 = 0.011$) on RMSE. For all these interaction effects, RMSEs resulting from LCA analyses are lower than RMSEs resulting from MELODIC analyses, indicating that LCA has a higher prediction accuracy than MELODIC in this simulation.

A larger sample size is beneficial for both MELODIC and LCA and the RMSE differences between MELODIC and LCA are larger in larger sample sizes conditions. When evaluating the effect of the LCA specific factors, i.e., π and class size, the prediction accuracy of both MELODIC and LCA, in terms of RMSE, deteriorated when modified levels of the factors (i.e., modified π and equal class size) were used. In the modified π condition, the differences in MELODIC and LCA RMSEs were larger than in the π from the real data condition, indicating that the prediction accuracy of MELODIC deteriorated to a higher extent than the prediction accuracy of LCA, when using the modified π condition. For the class sizes factor, the same effect was observed; mean MELODIC and LCA RMSE differences were larger when equal classes were used.

For factor *X*, the largest mean MELODIC and LCA RMSE difference was found when Age was involved in estimating the predictor-response variable relationships. The mean MELODIC and LCA RMSEs for the levels of factor *Y*, all follow more or less the same pattern. The RMSE differences between MELODIC and LCA are more or less the same for each level of *Y*. Here, the mean MELODIC and LCA RMSE differences seem to be slightly larger for the last four NCDs.

Since it was observed in both Simulation 1 and Simulation 2 that higher RMSEs seem to correspond with more frequent NCDs, the proportions "no" and "yes" on the NCDs for each generated population data set for each simulation are provided in Table 12.

Table 12

Frequency of NCDs per simulation, per generated population data set

		proportion "no"	proportion "yes"
MELODIC			
K=*K*,	*L*=*L*.	0.847	0.153
K=*K*dim2*2,	*L*=*L*.	0.857	0.143
K=*K*,	*L*=*L*/3.	0.655	0.345
K=*K*dim2*2,	*L*=*L*/3.	0.661	0.339
LCA			
π=π,	uneq.class.	0.846	0.154
π=πmod,	uneq.class.	0.575	0.425
π=π,	eq.class.	0.826	0.174
π=πmod,	eq.class.	0.500	0.500

As can be observed from Table 12, the frequency of NCDs is (more or less) the same for MELODIC and LCA when population is generated with the parameters from the SHARE data (i.e., *K=K, L=L* for MELODIC and $\pi=\pi$, uneq.class. for LCA). It is clear from Table 12 that the proportions "yes" all become larger when population data are generated with modified sets of parameters, except for the conditions in which the second dimension of K is multiplied by 2 in Simulation 1.

6 Discussion

In this study, two simulation studies were performed to evaluate the prediction accuracy of MELODIC and LCA. The first and main conclusion to draw, is that the results indicate that the accuracy of both methods is highly dependent on the data generation method. That is, MELODIC has a higher prediction accuracy when data are generated with MELODIC parameters whereas LCA has a higher prediction accuracy when data are generated with LCA parameters. In addition, the prediction accuracy difference of MELODIC and LCA in Simulation 1 was larger than in Simulation 2, indicating that MELODIC has a better performance in populations generated with LCA parameters, than LCA has in populations generated with MELODIC data.

The results indicate for MELODIC and confirm for LCA, that larger sample sizes (n = 1000) improve the prediction accuracy when compared to smaller sample sizes (n = 300). It was not surprising for LCA that the n = 1000 condition resulted in more accurate estimates, since it is suggested that a minimal sample of n = 500 is required for accurate LCA estimates. However, from Table 8 and 10 it became clear that the mean MELODIC and LCA RMSE differences, in all n = 300 conditions, were substantially larger in Simulation 1 than in Simulation 2. MELODIC even outperformed LCA in one of the n = 300 conditions, when data were generated with LCA parameters. This indicates that when data were generated with MELODIC parameters, the prediction accuracy of MELODIC was less affected by the small sample sizes than the prediction accuracy of LCA, when data generated with LCA parameters. It can be observed from Figure 3 and 8 that irrespective of data generation method, the difference in LCA RMSEs between the two n conditions is larger than the difference in MELODIC RMSEs between the two n conditions. MELODIC has shown to be relatively effective in analyzing n = 300 sample sizes.

As is clear from the results, changing MELODIC specific parameters (i.e., the discriminatory power of the response variables and the midpoint locations of the response variable categories) for data generation, resulted in a stronger difference in prediction accuracy between MELODIC and LCA. When not changing these parameters, the prediction accuracies of both methods were relatively more similar (although still significantly different). The prediction accuracy of MELODIC remained more or less unchanged between the *K* conditions, whereas the prediction accuracy of MELODIC deteriorated when using the modified *L* condition. Per definition the prediction accuracy of LCA deteriorated for both the modified method specific parameters.

In Simulation 2, using the modified method specific parameters for LCA (i.e., class conditional response probabilities and class sizes) also resulted in a stronger difference in prediction accuracy between MELODIC and LCA. Also here, when not changing these parameters, the prediction accuracies of both methods were relatively more similar (although still significantly different). Here, using the modified method specific parameters, both deteriorated the prediction accuracy of LCA.

With respect to the predictors involved in estimating the predictor-response variable relationships, three variables were used, all three from another class. Gender is a binary, Age is continuous and SRH is categorically ordered. The results indicate that when Gender was involved, the MELODIC and LCA RMSE differences were smallest in both simulations. It appears that a binary variable is least dependent on the data generation method, when evaluating prediction accuracy. However, this also might be due to the behavior of this variable in this specific data set.

For the individual NCDs, the largest MELODIC and LCA RMSE differences were found for Hypertension, Joint Disorders, Angina and Depression, in Simulation 1. In general, these NCDs happen to be the most prevalent NCDs in the populations, in Simulation 1. Also, for Simulation 2, the largest differences were found for the Joint Disorders, Angina, Stroke and Depression. Also here, these NCDs happen to be, in general, the most frequent NCDs in the populations of Simulation 2. In the real data, Hypertension, Joint Disorders, Angina and Depression are the most prevalent NCDs. When calculating RMSEs for MELODIC and LCA in conditions where method specific parameters are not modified, RMSEs were highest for these NCDs, for both methods, in both Simulation 1 and 2. The infrequent NCDs all had lower RMSEs. It appears that predictor-response variable relationships in the samples for frequent NCDs are estimated less accurate than predictor-response variable relationships in for the infrequent NCDs, hence higher RMSEs for the frequent NCDs.

This is an interesting finding that can be extrapolated to the rest of the results. As described earlier, it was clear that the MELODIC and LCA RMSE differences were larger when population data were generated with modified method specific parameters, in both Simulation 1 and Simulation 2. However, not only do these RMSE differences become larger, the absolute RMSE values also become larger when data are generated with modified method specific parameters. Implications of modifying method specific parameters for data generation, is that the NCDs will become more frequent, which was shown in Table 12.

For example in Simulation 2, when modifying the class conditional response probabilities, the NCDs are more prevalent in the populations that were generated with these modified class conditional response probabilities. Class conditional probabilities in the real data were relatively low in general, as can be recalled from Figure 2, where there was only one NCD in one class that had a probability higher than 0.5 of suffering from it. In the modified class conditional response probabilities condition, probabilities of 0.8 were introduced, resulting in a population with more frequent NCDs than in the real SHARE data.

The same was observed for the equal latent class sizes condition. In the real SHARE data, there was one large latent class, consisting of subjects with relatively low probabilities of suffering from each NCD. Generating equally sized latent classes implies that the proportion of subjects from the other classes (in which NCDs are more frequent) is larger, at the cost of the proportion of subjects from the class in which the NCDs are relatively infrequent. Therefore, when modifying latent class to sizes to be equal, higher probabilities of suffering from them are introduced.

Also, in Simulation 1, the midpoint locations of the response variable categories in the real SHARE data were all located to the right-side of the subject mean (Figure 1), with all the ‘yes’ categories being at the right-hand of the midpoints and the ‘no’ categories being at the left-hand of the midpoints. In general, the probabilities of suffering from an NCD were relatively low. By modifying the midpoint locations by dividing them by three, the probabilities of suffering from the NCDs become larger, and all NCDs were more frequent in these populations.

In contrast to these findings, in Simulation 1, when the discriminatory power of the response variables was multiplied by two for the second dimension, analyzing samples with MELODIC did not result in higher MELODIC RMSEs (Figure 4). This is the only modification of method specific parameters in which the frequency of NCDs did not become larger, in comparison with the real SHARE data.

These findings provide an explanation why the prediction accuracy of MELODIC hardly differed between the two *K* conditions, whereas it differed between the two *L* conditions, in Simulation 1. It also explains why the prediction accuracy of LCA differed between the two conditions of both the method specific parameters of LCA. More frequent NCDs lead to higher RMSEs for the method with which population data were generated (i.e., MELODIC in Simulation 1 and LCA in Simulation 2). The prediction accuracy of LCA was lower than the prediction accuracy of MELODIC in Simulation 1, and the prediction accuracy of MELODIC was lower than the prediction accuracy of LCA in Simulation 2. When using the modified method specific parameters, the magnitude of the MELODIC and LCA RMSE difference became larger.

In conclusion, the prediction accuracy of MELODIC is higher when data are generated with MELODIC parameters, whereas the prediction accuracy of LCA is higher when data are generated with LCA parameters. The prediction accuracy of the method is therefore highly dependent on the data generation method. MELODIC is more accurate in populations generated with LCA parameters, than LCA is in populations generated with MELODIC parameters. For both MELODIC and LCA, analyzing samples with a higher sample size is beneficial for the prediction accuracy, although it has shown that MELODIC is more accurate when analyzing samples of size $n = 300$ than LCA. When NCDs are more frequent, it will lead to lower prediction accuracy. In the modified method specific parameter conditions (except for the discriminatory power of the response variables, Dimension 2 multiplied by 2), more frequent NCDs are introduced in the population, resulting in higher RMSEs. This is the case for the method with which the populations were generated, and even more for the method with which the data were not generated. Hence, higher MELODIC and LCA RMSE differences are found in data with frequent NCDs than in data with infrequent NCDs.

With respect to NCD and multimorbidity analysis, it can be concluded that LCA is the most appropriate analysis to analyze the data, when there is reason to believe that there are underlying subgroups. If there are underlying subgroups, MELODIC will fail to be as accurate as LCA, as was shown by the results of the simulation studies. If no underlying subgroups are presumed, it is suggested that MELODIC is a more accurate method to analyze the data.

This study is an exploratory study to assess the prediction accuracy of MELODIC and LCA, when applied to NCD data. Since MELODIC has been developed recently, the prediction accuracy has not been assessed a lot in the past. Therefore this study gives more insight in the performance of MELODIC.

There were some limitations of this study. For LCA, the one-step approach was used. Although this method is very effective, one of its limitations is that, when changing, adding or excluding predictor variables from the model, the model needs to be re-estimated in total, which may change the latent classes and the interpretation of the latent classes (Bolck et al., 2004). In this study, three predictors were chosen from a large set of available predictors. Choosing other predictors, might would have led to other latent classes. There are alternatives in which changes on the predictor side do not change the latent classes and the entire model does not have to be re-estimated (e.g., the two-step approach (Bakk & Kuha, 2017) and three-step approaches (Vermunt, 2010)). However, due to the availability of performing the one-step approach LCA in RStudio, it was decided to proceed with the one-step approach. In the future, LATENT GOLD computer software (Vermunt & Magidson, 2005) could be used to do a similar study as this study. Instead of using the one-step approach LCA, alternatives could be used that are available for Latent GOLD.

Also, on every sample that was drawn from the populations, always MELODIC with two dimensions and LCA with three classes was performed. It could have been the case that in some of the samples it did not make sense to perform, for example, an LCA with three classes. In some of the drawn samples, it could have been so that a two- or four-class solution was more optimal.

Another limitation is the number of chosen predictors in the study. The SHARE data has a very wide range of variables to be used for analysis. However, for computational considerations, it was decided to use only three predictors, each being of a different type (i.e., binary, continuous and categorically ordered). Including more variables in the analyses would have resulted in more estimated predictor-response variable relationships, and conclusions could be drawn about the accuracy of the different predictor types. In this study there was only one predictor per type, and therefore, it is difficult to make generalizations of the behavior of the different class types.

All subjects with missing data in Wave 1 from the SHARE data were excluded from analysis. The data that was used is therefore not completely representative for all data that was collected in Wave 1.

Lastly, and importantly, to evaluate the accuracy of the methods, conditional NCD probabilities in the populations were compared with conditional NCD probabilities in the samples. Using (17) the RMSEs were calculated. However, as stated earlier, modifying the parameters lead in the most instances to data sets with more frequent NCDs. In addition, it was concluded that higher RMSEs were estimated for the more frequent NCDs. It is

obviously desired to evaluate the accuracy of a method irrespective of the frequencies of the response variables. Therefore, the use of RMSE as an prediction accuracy measure should be reconsidered for the conditional probabilities as calculated in this study. One possibility could be to use the cross-entropy loss function, which is the RMSE alternative for classification. However, to use such a function, actual values are presumed to be classification values (e.g., 0 or 1), whereas predicted values are (conditional probabilities). However, in this study, the actual values and the predicted values are both conditional probabilities and therefore the cross-entropy function would not be a suitable option eventually. It might be more insightful to calculate RMSEs on the log-odds scale. That is, there is a linear relationship between log-odds ratio in favor of suffering from an NCD, and the predictor variables. Comparing the log-odds ratio in the populations with the log-odds ratio in the samples would, consequently, be less influenced by frequency of the NCDs. In further research, it should be demonstrated whether comparing log-odds, instead of probabilities, would be more suitable evaluation procedure, in terms of RMSE, with respect to the NCD frequency issue.

References

Agresti, A. (2003). *Categorical data analysis* (Vol. 482). John Wiley & Sons.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, *37*(3), 379-384.

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*(4), 871-892.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological methods*, *9*(1), 3.

Bayes-Marin, I., Sanchez-Niubo, A., Egea-Cortés, L., Nguyen, H., Prina, M., Fernández, D., ... & Olaya, B. (2020). Multimorbidity patterns in low-middle and high income regions: a multiregion latent class analysis using ATHLOS harmonised cohorts. *BMJ open*, *10*(7), e034441.

Bertram, M. Y., Sweeny, K., Lauer, J. A., Chisholm, D., Sheehan, P., Rasmussen, B., ... & Deane, S. (2018). Investing in non-communicable diseases: an estimation of the return on investment for prevention and treatment services. *The Lancet*, *391*(10134), 2071-2078.

Bloom, D. E., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L. R., Fathima, S., ... & Weiss, J. (2012). *The global economic burden of noncommunicable diseases* (No. 8712). Program on the Global Demography of Aging.

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

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., ... & Zuber, S. (2013). Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *International journal of epidemiology*, *42*(4), 992-1001.

Breiman, L., & Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(1), 3-54.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

De Rooij, M., & Groenen, P. J. F. (2021). The MELODIC family for simultaneous binary logistic regression in a reduced space. *arXiv preprint arXiv:2102.08232*.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using M plus. *Structural Equation Modeling*, *18*(1), 132-151.

Gudicha, D. W. (2015). *Power analysis methods for tests in latent class and latent Markov models*. Tilburg University.

Islam, M. M., Valderas, J. M., Yen, L., Dawda, P., Jowsey, T., & McRae, I. S. (2014). Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. *PloS one*, *9*(1), e83783.

Janssen, J. H., Van Laar, S., De Rooij, M. J., Kuha, J., & Bakk, Z. (2019). The detection and modeling of direct effects in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(2), 280-290.

Kowal, P., Chatterji, S., Naidoo, N., Biritwum, R., Fan, W., Lopez Ridaura, R., ... & SAGE Collaborators. (2012). Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). *International journal of epidemiology*, *41*(6), 1639-1649.

Lazarsfeld, P. F. (1950). The logical and mathematical foundations of latent structure analysis. In S. A. Stouffer (Ed.), *Measurement and prediction* (pp. 362–412). Princeton, NJ: Princeton University Press.

Leonardi, M., Chatterji, S., Koskinen, S., Ayuso-Mateos, J. L., Haro, J. M., Frisoni, G., ... & COURAGE in Europe Project's Consortium. (2014). Determinants of health and disability in ageing population: the COURAGE in Europe Project (collaborative research on ageing in Europe). *Clinical psychology & psychotherapy*, *21*(3), 193-198.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of statistical software*, *42*(1), 1-29.

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

McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Sage.

Murray, C. J., Barber, R. M., Foreman, K. J., Ozgoren, A. A., Abd-Allah, F., Abera, S. F., ... & Del Pozo-Cruz, B. (2015). Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *The Lancet*, *386*(10009), 2145-2191.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, *24*(6), 882-891.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal*, *14*(4), 535-569.

Olaya, B., Moneta, M. V., Caballero, F. F., Tyrovolas, S., Bayes, I., Ayuso-Mateos, J. L., & Haro, J. M. (2017). Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. *BMC geriatrics*, *17*(1), 1-10.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods*, *8*(4), 434.

Petersen, K. J., Qualter, P., & Humphrey, N. (2019). The application of latent class analysis for investigating population child mental health: A systematic review. *Frontiers in psychology*, *10*, 1214.

Reboussin, B. A., Ip, E. H., & Wolfson, M. (2008). Locally dependent latent class models with covariates: an application to under-age drinking in the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(4), 877-897.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

SHARE. (n.d.) SHARE – Survey of Health, Ageing and Retirement in Europe. Retrieved from http://www.share-project.org/home0.html

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: the English longitudinal study of ageing. *International journal of epidemiology*, *42*(6), 1640-1648.

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, *20*(4), 640-657.

Vermunt, J. K. (2010). Latent class models. In: P. Peterson, E. Baker, B. McGaw, (eds.), *International Encyclopedia of Education, Volume 7*, 238-244. Oxford: Elsevier.

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

Vermunt, J. K., & Magidson, J. (2005). Technical guide for Latent GOLD 4.0: Basic and advanced. *Belmont Massachusetts: Statistical Innovations Inc*.

Wallace, E., Salisbury, C., Guthrie, B., Lewis, C., Fahey, T., & Smith, S. M. (2015). Managing patients with multimorbidity in primary care. *Bmj*, *350*.

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., ... & Bell, M. L. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet*, *388*(10053), 1459-1544.

Weden, M. M., & Zabin, L. S. (2005). Gender and ethnic differences in the co-occurrence of adolescent risk behaviors. *Ethnicity and Health*, *10*(3), 213-234.

Whitson, H. E., Johnson, K. S., Sloane, R., Cigolle, C. T., Pieper, C. F., Landerman, L., & Hastings, S. N. (2016). Identifying patterns of multimorbidity in older Americans: application of latent class analysis. *Journal of the American Geriatrics Society*, *64*(8), 1668-1673.

World Health Organization. (2018, February 5). *Ageing and health*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.

World Health Organization. (2021, April 13). *Key facts.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

Appendix A

Simulation 1 RMSEs

Table A1

Simulation 1 RMSEs for each simulation condition for each predictor for each response variable, when analyzed with MELODIC

		Dbts	Hyp	Asth	CLD	JD	Ang	Str	Dpr
n=300,	Gender	0.028	0.037	0.025	0.024	0.033	0.031	0.023	0.041
K=*K*,	Age	0.027	0.037	0.025	0.024	0.034	0.033	0.023	0.050
L=*L*.	SRH	0.032	0.040	0.029	0.029	0.036	0.037	0.028	0.042
n=1000,	Gender	0.015	0.021	0.013	0.013	0.018	0.016	0.013	0.021
K=*K*,	Age	0.015	0.020	0.013	0.013	0.019	0.018	0.013	0.024
L=*L*.	SRH	0.018	0.024	0.015	0.016	0.019	0.020	0.015	0.024
n=300,	Gender	0.023	0.041	0.018	0.015	0.031	0.030	0.012	0.043
K=*K*dim2*2,	Age	0.023	0.039	0.017	0.014	0.034	0.031	0.012	0.052
L=*L*.	SRH	0.027	0.041	0.023	0.020	0.035	0.036	0.019	0.050
n=1000,	Gender	0.012	0.020	0.009	0.009	0.017	0.016	0.007	0.021
K=*K*dim2*2,	Age	0.011	0.020	0.009	0.009	0.020	0.016	0.007	0.025
L=*L*.	SRH	0.014	0.023	0.011	0.011	0.021	0.018	0.009	0.025
n=300,	Gender	0.039	0.038	0.033	0.040	0.041	0.041	0.038	0.041
K=*K*,	Age	0.037	0.034	0.032	0.038	0.045	0.043	0.036	0.047
L=*L*/3.	SRH	0.045	0.037	0.035	0.041	0.039	0.043	0.035	0.046
n=1000,	Gender	0.023	0.021	0.021	0.023	0.024	0.025	0.020	0.023
K=*K*,	Age	0.021	0.020	0.018	0.021	0.025	0.025	0.020	0.027
L=*L*/3.	SRH	0.023	0.023	0.020	0.023	0.020	0.024	0.023	0.024
n=300,	Gender	0.039	0.038	0.038	0.040	0.038	0.039	0.037	0.044
K=*K*dim2*2,	Age	0.038	0.036	0.035	0.039	0.043	0.039	0.037	0.052
L=*L*/3.	SRH	0.041	0.039	0.038	0.039	0.038	0.040	0.041	0.047
n=1000,	Gender	0.023	0.023	0.018	0.019	0.024	0.024	0.022	0.021
K=*K*dim2*2,	Age	0.022	0.020	0.018	0.018	0.027	0.024	0.021	0.027
L=*L*/3.	SRH	0.024	0.023	0.023	0.021	0.024	0.024	0.022	0.024

Table A2

Simulation 1 RMSEs for each simulation condition for each predictor for each response variable, when analyzed with LCA

		Dbts	Hyp	Asth	CLD	JD	Ang	Str	Dpr
n=300,	Gender	0.036	0.062	0.019	0.019	0.051	0.046	0.017	0.074
K=*K*,	Age	0.041	0.075	0.020	0.022	0.065	0.059	0.019	0.092
L=*L*.	SRH	0.043	0.069	0.025	0.030	0.072	0.063	0.025	0.091
n=1000,	Gender	0.020	0.042	0.009	0.012	0.029	0.032	0.009	0.053
K=K,	Age	0.022	0.045	0.012	0.014	0.047	0.034	0.015	0.067
L=*L*.	SRH	0.025	0.045	0.013	0.016	0.046	0.038	0.016	0.062
n=300,	Gender	0.030	0.062	0.022	0.022	0.057	0.056	0.018	0.081
K=*K*dim2*2,	Age	0.039	0.079	0.023	0.027	0.096	0.075	0.023	0.145
L=*L*.	SRH	0.045	0.079	0.028	0.033	0.109	0.090	0.032	0.153
n=1000,	Gender	0.021	0.042	0.011	0.015	0.037	0.040	0.013	0.045
K=*K*dim2*2,	Age	0.021	0.044	0.011	0.015	0.046	0.043	0.014	0.064
L=*L*.	SRH	0.027	0.051	0.015	0.019	0.060	0.058	0.017	0.078
n=300,	Gender	0.055	0.062	0.054	0.062	0.085	0.101	0.069	0.107
K=*K*,	Age	0.069	0.069	0.061	0.073	0.094	0.099	0.080	0.102
L=*L*/3.	SRH	0.073	0.069	0.060	0.073	0.091	0.087	0.073	0.097
n=1000,	Gender	0.033	0.034	0.031	0.037	0.073	0.084	0.044	0.100
K=*K*,	Age	0.038	0.036	0.031	0.043	0.059	0.065	0.046	0.070
L=*L*/3.	SRH	0.039	0.041	0.034	0.039	0.058	0.048	0.041	0.062
n=300,	Gender	0.071	0.063	0.055	0.071	0.089	0.105	0.080	0.095
K=*K*dim2*2,	Age	0.088	0.086	0.074	0.098	0.134	0.135	0.104	0.143
L=*L*/3.	SRH	0.087	0.090	0.079	0.097	0.142	0.134	0.103	0.150
n=1000,	Gender	0.039	0.043	0.032	0.046	0.061	0.072	0.053	0.065
K=*K*dim2*2,	Age	0.042	0.044	0.037	0.043	0.080	0.060	0.049	0.086
L=*L*/3.	SRH	0.056	0.052	0.046	0.059	0.087	0.096	0.070	0.106

Appendix B

Simulation 2 RMSEs

Table B1

Simulation 2 RMSEs for each simulation condition for each predictor for each response variable, when analyzed with MELODIC

		Dbts	Hyp	Asth	CLD	JD	Ang	Str	Dpr
n=300,	Gender	0.027	0.041	0.023	0.021	0.034	0.034	0.020	0.059
π=π,	Age	0.034	0.050	0.024	0.024	0.044	0.045	0.022	0.070
uneq.class.	SRH	0.036	0.049	0.028	0.028	0.046	0.041	0.025	0.075
n=1000,	Gender	0.014	0.026	0.011	0.009	0.020	0.022	0.008	0.053
π=π,	Age	0.021	0.040	0.012	0.012	0.029	0.033	0.011	0.063
uneq.class.	SRH	0.022	0.035	0.014	0.014	0.032	0.030	0.013	0.062
n=300,	Gender	0.055	0.049	0.051	0.054	0.058	0.056	0.063	0.059
π=πmod,	Age	0.073	0.067	0.067	0.070	0.073	0.071	0.077	0.075
uneq.class.	SRH	0.059	0.057	0.057	0.061	0.072	0.076	0.072	0.073
n=1000,	Gender	0.034	0.040	0.037	0.037	0.053	0.051	0.057	0.054
π=πmod,	Age	0.057	0.060	0.059	0.058	0.068	0.065	0.072	0.068
uneq.class.	SRH	0.045	0.046	0.047	0.044	0.065	0.065	0.063	0.065
n=300,	Gender	0.035	0.070	0.025	0.024	0.055	0.040	0.023	0.072
π=π,	Age	0.045	0.082	0.027	0.028	0.066	0.056	0.026	0.084
eq.class.	SRH	0.045	0.071	0.028	0.028	0.051	0.058	0.028	0.067
n=1000,	Gender	0.028	0.057	0.020	0.021	0.047	0.032	0.019	0.059
π=π,	Age	0.036	0.067	0.021	0.024	0.057	0.044	0.022	0.070
eq.class.	SRH	0.035	0.055	0.021	0.023	0.037	0.050	0.022	0.057
n=300,	Gender	0.128	0.127	0.129	0.134	0.097	0.091	0.089	0.094
π=πmod,	Age	0.134	0.134	0.136	0.140	0.112	0.106	0.104	0.110
eq.class.	SRH	0.111	0.111	0.111	0.117	0.070	0.070	0.073	0.070
n=1000,	Gender	0.130	0.123	0.127	0.125	0.086	0.088	0.085	0.082
π=πmod,	Age	0.134	0.128	0.132	0.131	0.101	0.103	0.099	0.098
eq.class.	SRH	0.108	0.102	0.106	0.105	0.059	0.058	0.059	0.059

Table B2

Simulation 2 RMSEs for each simulation condition for each predictor for each response variable, when analyzed with LCA

		Dbts	Hyp	Asth	CLD	JD	Ang	Str	Dpr
n=300,	Gender	0.029	0.052	0.016	0.018	0.044	0.040	0.016	0.058
π=π,	Age	0.033	0.057	0.020	0.019	0.055	0.046	0.017	0.065
uneq.class.	SRH	0.032	0.052	0.021	0.021	0.050	0.036	0.017	0.069
n=1000,	Gender	0.014	0.024	0.008	0.009	0.019	0.018	0.007	0.025
π=π,	Age	0.013	0.027	0.009	0.009	0.020	0.018	0.007	0.033
uneq.class.	SRH	0.015	0.028	0.010	0.011	0.022	0.016	0.009	0.029
n=300,	Gender	0.047	0.045	0.048	0.050	0.038	0.037	0.040	0.038
π=πmod,	Age	0.041	0.041	0.041	0.043	0.042	0.042	0.043	0.043
uneq.class.	SRH	0.038	0.037	0.039	0.042	0.037	0.040	0.039	0.039
n=1000,	Gender	0.025	0.029	0.026	0.026	0.022	0.019	0.022	0.022
π=πmod,	Age	0.023	0.025	0.023	0.023	0.025	0.022	0.024	0.024
uneq.class.	SRH	0.021	0.023	0.022	0.020	0.022	0.021	0.022	0.022
n=300,	Gender	0.039	0.071	0.020	0.021	0.052	0.052	0.019	0.060
π=π,	Age	0.037	0.067	0.023	0.021	0.052	0.054	0.019	0.070
eq.class.	SRH	0.036	0.064	0.024	0.023	0.053	0.048	0.020	0.067
n=1000,	Gender	0.027	0.051	0.010	0.015	0.032	0.041	0.013	0.032
π=π,	Age	0.021	0.038	0.009	0.011	0.026	0.031	0.010	0.032
eq.class.	SRH	0.020	0.034	0.011	0.012	0.025	0.026	0.010	0.029
n=300,	Gender	0.103	0.103	0.104	0.107	0.050	0.049	0.046	0.050
π=πmod,	Age	0.075	0.076	0.077	0.079	0.048	0.046	0.047	0.048
eq.class.	SRH	0.066	0.065	0.064	0.067	0.042	0.043	0.043	0.043
n=1000,	Gender	0.100	0.093	0.097	0.097	0.038	0.038	0.036	0.035
π=πmod,	Age	0.070	0.064	0.068	0.067	0.032	0.032	0.030	0.030
eq.class.	SRH	0.057	0.051	0.054	0.053	0.025	0.024	0.024	0.024