Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Grace S. Kim	Date

Flexible Methods to Incorporate Covariates in Latent Class Analysis

By

Grace S. Kim

Doctor of Philosophy

Biostatistics

John J. Hanfelt, Ph.D.
Advisor

Felicia Goldstein, Ph.D.
Committee Member

Robert H. Lyles, Ph.D.
Committee Member

Amita Manatunga, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Flexible Methods to Incorporate Covariates in Latent Class Analysis

By

Grace S. Kim
M.S., Emory University, 2019
B.S., Emory University, 2015

Advisor: John J. Hanfelt, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2021

Abstract

Flexible Methods to Incorporate Covariates in Latent Class Analysis

By

Grace S. Kim

Mild Cognitive Impairment (MCI) is a neurocognitive disorder with a complex structure that sometimes precedes dementia. It is comprised of heterogeneous subclinical entities, which necessitates clinicians to assess different domains of cognitive, functional, neuropsychiatric, and possibly biological features for an accurate diagnosis and early intervention. Latent class analysis (LCA) is a method based on rigorous statistical derivation that can be used to explore heterogeneity of MCI. Latent class regression, an extension of the latent class framework established by Bandeen-Roche et al. (1997), can be used to incorporate covariates as risk factors of class membership. Under the latent class regression model, the population of interest consists of mixture of different subcategories of MCI with unobserved or latent class membership, which is further associated with risk factors of interest.

The first aim of this research is to explore situations when covariates unintentionally influence conceptualization of latent classes, and develop a flexible method for researchers to incorporate covariates without distorting too extensively the clinical interpretation of the latent classes in the maximum likelihood solution. Relative frequencies of latent classes resulting from covariates will be used to help investigate the structure of MCI. The EM algorithm will be used to provide optimal parameter estimates and latent class-specific means of manifest variables.

The second aim expands on the first aim by focusing on high-dimensional and potentially correlated covariates to develop a new method, termed compound LCA, that applies dimension reduction in covariate space simultaneously with dimension reduction in manifest variable space. Compound LCA will effectively avoid uncertainties or "fuzziness" in dimension reduction that are propagated in the LCA by introducing a second set of latent classes that are formulated based on the observed high-dimensional covariate patterns. The EM algorithm will be used to find the prevalence of classes of covariates and features, posterior probabilities of each individual, and latent class-specific means of covariates and feature variables for clinical interpretation of the latent classes. The third aim introduces an extension of compound LCA, which assumes that feature classes are nested within covariate classes. We provide a likelihood ratio test that compares compound LCA and its extension.

Flexible Methods to Incorporate Covariates in Latent Class Analysis

By

Grace S. Kim
M.S., Emory University, 2019
B.S., Emory University, 2015

Advisor: John J. Hanfelt, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Mild Cognitive Impairment (MCI) is defined to be an intermediary neurocognitive disorder that sometimes precedes dementia (Petersen, 2011). While dementia is described to be a debilitating disease that perturbs everyday routine and contributes to patient dependence due to cognitive deficit (Petersen, 2011), MCI is a complex syndrome where patients exhibit cognitive decline that is not as severe as those who suffer from dementia and tend to display relatively normal functional abilities. That is, patients who are diagnosed with MCI do not experience interference in instrumental activities of daily living (IADLs) such as cooking and bill management, and can be distinguished from people who experience normal aging, such as increased forgetfulness. Therefore, diagnosis of MCI can help with early detection of cognitive decline before discernible functional impairment.

There is growing awareness that MCI is a highly heterogeneous syndrome (Hanfelt et al., 2011; Diaz-Mardomingo et al., 2017). Researchers are finding that aside from the degeneration of brain cells, underlying medical conditions such as depression and anxiety disorder may be attributing to cognitive decline (Petersen, 2016). Moreover, by expanding the phenotype of MCI beyond cognitive features to include neuropsychiatric features and IADLs, improved predictions of subsequent cognitive decline and underlying neuropathology is possible (Hanfelt et al., 2018). However, characterization of subcategories in MCI is heavily debated among clinicians.

Various statistical methods have been proposed to help clarify our understanding and evaluation of MCI. Latent class analysis (LCA) is a powerful method based on rigorous statistical derivation that can be used to explore heterogeneity of MCI. The traditional latent class model (Lazarsfeld and Henry, 1968) aims to clarify the relationships among manifest (observed) variables in the model by uncovering a structure of latent (unobserved) variables. This approach adopts a finite mixture model and assumes that the manifest variables are conditionally independent given the latent classes. An important extension is latent class regression (Bandeen-Roche et al., 1997), which additionally incorporates covariates as risk

factors of class membership and assumes that covariates are only associated with manifest variables through latent class frequencies. The application of this method to MCI is founded on the assumption that the MCI syndrome consists of a mixture of different subtypes of MCI with unobserved or latent class membership, which is further associated with different risk factors of interest. Under this rigorous statistical framework, the maximum likelihood method is available to optimally estimate parameters and conduct inference.

Importantly, the latent class regression model (Bandeen-Roche et al., 1997) assumes that covariates affect the latent class frequencies (i.e., mixture probabilities) but not the conceptualization of the latent classes. Despite this assumption, we explore an application to MCI where the presence of covariates altered not only the mixture probabilities but also the conceptualization of the latent classes. We introduce a covariate activity governor to more flexibly incorporate covariates into the model, allowing investigators to richly explore the impact of covariates on study findings, and limit, or customize, the extent to which covariates alter the clinical interpretation of the latent classes.

In addition, there has been an increasing interest in analyzing high-dimensional and potentially correlated covariates associated with MCI, such as sociodemographic characteristics, health history such as indicators of cerebrovascular disease, coronary artery disease and cardiac dysrhythmias, genetics and biomarkers. The latent class regression model assumes that there is a direct parametric relationship between covariates and the relative frequencies of the latent classes, limiting this approach to covariates that are low-dimensional, i.e., less than 10 covariates. We propose an alternative method to apply LCA on high-dimensional and potentially correlated covariates, i.e., 20-30 covariates, which is a bigger magnitude of what is usually handled by standard LCA. We outline the concept of compound LCA, that introduces a second set of latent classes that are formulated based on the observed high-dimensional covariate patterns. We extend compound LCA to assume that classes of feature variables are nested within covariate classes, and provide a likelihood ratio test to help determine which method to use.

1.2 Motivating Example

1.2.1 Study Sample

We included individuals from the Uniform Data Set (UDS) of the National Alzheimer's Coordinating Center (NACC), which is a longitudinal study that includes patients who have dementia, mild cognitive impairment and who are cognitively normal (National Alzheimer's Coordinating Center, 2021b). We focused on a sample of 6034 participants as of June 2015 freeze date and included standardized evaluations of functional abilities, neuropsychiatric symptoms and assessments of cognitions as manifest variables. We included vascular risk factors as covariates.

1.2.2 Assessments of Functional Abilities

The functional Assessment Questionnaire (FAQ) was used to evaluate functional abilities of patients in the UDS dataset. The FAQ utilizes 10 questionnaires to measure a level of functional ability relating to instrumental activities of daily living (IADL) in areas such writing checks, assembling tax records or remembering appointments (Ito et al., 2012). A patient can have scores ranging from 0=normal to 3=dependent, and the sum of all questions becomes the FAQ score that ranges from 0 to 30 (Marshall et al., 2015).

1.2.3 Assessments of Neuropsychiatric Symptoms

The Neuropsychiatric Inventory Questionnaire (NPI-Q) was used to assess the severity of neuropsychiatric symptoms of participants. It is a self-administered questionnaire that examines 12 neuropsychiatric symptom domains: delusions, hallucinations, agitation/aggression, dysphoria/depression, anxiety, euphoria/elation, apathy/indifference, disinhibition, irritability/lability, aberrant motor behaviors, night-time behavioral disturbances, and appetite/eating disturbances (Musa et al., 2017). The symptoms are evaluated in terms of severity, where 1=mild, 2=moderate and 3=severe, and the sum of NPI-Q scores can be up to 36 (Kaufer

et al., 2000). Geriatric Depression Scale (GDS), which uses 15 questionnaires with yes/no answers and scores of 12-15 indicate severe depression, was used to assess depression among patients (Yesavage and Sheikh, 1986).

1.2.4 Assessments of Cognition

Assessments of cognition can be divided into evaluation of global cognitive status, memory, attention, language, executive function and visuomotor. Mini-Mental State Exam (MMSE) was used to assess an overall cognitive status of patients, where the test covers orientation, memory and attention in the first part and the ability to name, follow verbal and written commands, write a sentence spontaneously and copy a complex polygon in the second part (Folstein et al., 1975). Memory function was evaluated with the Logical Memory test, where patients are asked to recall a story immediately and from memory (immediate and delayed recall) (Gavett et al., 2016). Memory function was additionally evaluated using Category Fluency test, where participants are asked to name as many examples as possible in a specific category, such as animals, within 60 seconds (Rosen, 1980). Attention was evaluated using the Trail Making Test A, which tests a patient's ability to sequentially connect numbers in circles with lines (Tombaugh, 2004), and also using the Digit Span Forward test, which tests a patient's ability to read and recall a number in order (Richardson, 2007). Language was evaluated by using the Boston Naming Test, which asks a patient to name drawing of objects and the score is equivalent to the number of correct responses in the first 20 seconds (Kaplan et al., 1983). After 20 seconds, phonemic and/or semantic cues are provided. Executive function was tested using the Trail Making Test B, which tests a patient's ability to sequentially connect numbers and alphabets in circles (e.g., 1-A-2-B...) (Tombaugh, 2004), and also using the Digit Span Backward test, which tests a patient's ability to read and recall a number in reverse order (Richardson, 2007). Visuomotor skill was evaluated with the Digit Symbol Test, or the Wechsler Adult Intelligence Scale (WAIS), and the test consists of a row of numbers that a patient has to match with a provided key, where the key includes symbols

corresponding to the numbers in the test (David Wechsler, 2008).

1.2.5 Vascular Risk Factors

The Rosen Modification of Hachinski Ischemic Score (RMHIS) was used to assess cerebrovascular disease status of participants, which is a scale modified from Hachinski Ischemic Score to include 8 features that would increase the accuracy of diagnosis of multi infarct dementia (MID), a vascular disorder (Rosen et al., 1980). Participants whose RMHIS scores were greater than 3 were determined to have cerebrovascular disease. Additional risk factors of cerebrovascular disease such as diabetic status, hypercholesterolemia and hypertension were included. Other correlates of cerebrovascular disease not included in the RHMIS were also assessed including decades of smoking, coronary artery disease, hypercholesterolemia, hypertension, stroke, and diabetes.

1.2.6 Analysis

We fitted models with 1-5 latent classes, 13 manifest variables and 7 covariates using Latent Gold 5.1 software package (Statistical Innovations Inc., 2016). We found that the 5-class model was preferred according to the ICL-BIC criterion. All 7 covariates, except smoking, were significantly associated with the latent classes (Table 1.1, left column). Compared to a simpler 5-class model without covariates, the clinical interpretation of the latent classes in the 7-covariate model was influenced by covariates (Table 1.2). In the 7-covariate model, the five classes consisted of:

1. "amnestic multi-domain i.e., memory impairment plus other cognitive domains, with functional impairment and neuropsychiatric features" (24%)

2. "functional impairment and neuropsychiatric features" (22%)

3. "mildly impaired", comprised of patients who were not cognitively normal as judged by clinical experts but whose cognitive performances were within normal range by tests

used in the UDS (21%)

4. "amnestic i.e., memory impairment without other cognitive domains, with functional impairment and neuropsychiatric features" (19%)

5. "neuropsychiatric features only" (14%).

By contrast, under a model that included no covariates, a different 5-class solution resulted (Table 1.2):

1. "mildly impaired" (27%)

2. "executive function with functional impairment" (24%)

3. "amnestic multi-domain with functional impairment and neuropsychiatric features" (19%)

4. "amnestic with functional impairment and neuropsychiatric features" (15%)

5. "functional impairment and neuropsychiatric features" (15%).

Hence, the 0-covariate model revealed a clinically important class characterized by non-amnestic cognitive features, specifically impairment in executive function, whereas the 7-covariate model failed to detect this non-amnestic class and instead heightened the role of neuropsychiatric features.

This example highlights the need to develop methods to handle covariates in LCA, and simultaneous provide clinical interpretations of latent classes that can explore heterogeneity of MCI.

Table 1.1: Estimated Log Odds Ratios (and Standard Errors) of Covariates

Covariate	Class*	7 Covariates	No Covariates	
Intercept	1	—	—	
	2	−0.03(0.09)	−0.11(0.04)	
	3	0.12(0.08)	−0.36(0.04)	
	4	−0.05(0.09)	−0.58(0.06)	
	5	−0.39(0.09)	−0.60(0.06)	
RMHIS	1	—	—	
	2	−1.05(0.19)	—	
	3	−1.32(0.19)	—	
	4	−1.16(0.22)	—	
	5	−0.57(0.19)	—	
Smoking (Per Decade)	1	—	—	
	2	0.03(0.03)	—	
	3	0.01(0.03)	—	
	4	−0.04(0.03)	—	
	5	0.05(0.03)	—	
Coronary Artery Disease	1	—	—	
	2	0.03(0.09)	—	
	3	−0.20(0.09)	—	
	4	−0.14(0.10)	—	
	5	−0.24(0.10)	—	
Hypercholesterolemia	1	—	—	
	2	0.34(0.09)	—	
	3	0.03(0.08)	—	
	4	0.36(0.09)	—	
	5	0.21(0.10)	—	
Diabetic Status	1	—	—	
	2	−0.21(0.12)	—	
	3	−0.22(0.11)	—	
	4	−0.47(0.13)	—	
	5	−0.24(0.13)	—	
Hypertension	1	—	—	
	2	−0.27(0.09)	—	
	3	−0.07(0.09)	—	
	4	−0.26(0.10)	—	
	5	−0.20(0.10)	—	
Stroke	1	—	—	—
	2	−0.01(0.18)	—	
	3	−0.06(0.18)	—	
	4	−0.39(0.24)	—	
	5	−0.99(0.26)	—	

* Row entries for classes 1-5 correspond with the latent classes in Table 1.2.

Table 1.2: Class-Specific Means or Proportions of Functional, Neuropsychiatric, and Cognitive Features in Five-Class Model. Shown are the results for 7 covariates and no covariates. Cognitive test scores were standardized by demographics so that a cognitive value of -1.5 indicates that an individual with MCI performed 1.5 standard deviations worse than a cognitively normal person of the same age, race, and years of education. Neuropsychologists typically regard cognitive standardized scores of -1.5 or worse as evidence of impairment in a specific cognitive domain.

Model Type	Test Type		Amnestic Multi-Domain With Functional Impairment And Neuropsychiatric Features (Relative Frequency=24%)	Functional Impairment With Neuropsychiatric Features (Relative Frequency=22%)	Mildly Impaired (Relative Frequency=21%)	Amnestic With Functional Impairment And Neuropsychiatric Features (Relative Frequency=19%)	Neuropsychiatric Features Only (Relative Frequency=14%)
7 Covariates	Functional	No. of IADL impaired	3.96	3.07	0	3.73	0
	Neuropsychiatric	% with GDS ≥ 5	28.80%	18.74%	9.84%	15.26%	19.15%
		No. of NPI-Q symptoms present	2.48	2.19	0	2.11	2.13
	Cognitive	Global					
		MMSE	−2.55	−0.77	−1.26	−2.25	−1.19
		Logical Memory					
		Immediate	−1.43	−0.46	−0.95	−2.11	−1.06
		Delayed	−1.50	−0.55	−0.97	−2.38	−1.08
		Semantic Memory					
		Category Fluency	−1.38	−0.56	−0.75	−1.10	−0.76
		Attention					
		Trails A	2.26	0.09	0.50	0.07	0.39
		Digit Span Forward	−0.67	−0.13	−0.27	−0.16	−0.22
		Language					
		Boston Naming	−1.83	−0.33	−1.06	−1.00	−0.92
		Executive Function					
		Trails B	3.32	0.36	1.01	0.37	0.98
		Digit Span Backward	−0.88	−0.22	−0.43	−0.37	−0.45
		Visuomotor					
		Digit Symbol	−1.74	−0.43	−0.60	−0.52	−0.67

Model Type	Test Type		Mildly Impaired (Relative Frequency=27%)	Executive Function With Functional Impairment (Relative Frequency=24%)	Amnestic Multi-Domain With Functional Impairment And Neuropsychiatric Features (Relative Frequency=19%)	Amnestic with Functional Impairment And Neuropsychiatric Features (Relative Frequency=15%)	Functional Impairment And Neuropsychiatric Features (Relative Frequency=15%)
No Covariates	Functional	No. of IADL impaired	0	2.06	3.95	3.90	3.26
	Neuropsychiatric	% with GDS ≥ 5	13.43%	12.13%	31.02%	18.60%	22.84%
		No. of NPI-Q symptoms present	0.99	0	3.15	2.95	3.05
	Cognitive	Global					
		MMSE	−0.89	−2.06	−2.51	−2.11	−0.70
		Logical Memory					
		Immediate	−0.86	−1.30	−1.49	−2.02	−0.38
		Delayed	−0.89	−1.41	−1.57	−2.27	−0.45
		Semantic Memory					
		Category Fluency	−0.63	−1.01	−1.43	−1.05	−0.55
		Attention					
		Trails A	0.14	0.98	2.24	0.07	0.13
		Digit Span Forward	−0.18	−0.36	−0.68	−0.16	−0.14
		Language					
		Boston Naming	−0.65	−1.44	−1.93	−0.80	−0.28
		Executive Function					
		Trails B	0.61	1.60	3.31	0.36	0.39
		Digit Span Backward	−0.34	−0.57	−0.89	−0.37	−0.22
		Visuomotor					
		Digit Symbol	−0.46	−0.89	−1.75	−0.54	−0.48

* Higher scores on Trail A and Trail B indicate worse performance.

1.3 Scope of Research

The goal of this dissertation research is to address problems that arise from incorporating covariates within the latent class framework. A motivating example (Chapter 2) highlights a unique problem in the application of latent class analysis to MCI, where risk factors related to vascular comorbidity pose a challenge in exploring heterogeneity of MCI. The first aim of this research (Chapter 3) explores situations when the presence of covariates alters not only the mixture probabilities but also the conceptualization of the latent classes by conducting different simulation studies. A new method is formulated to incorporate covariates without potentially distorting the interpretations of the latent classes. The second aim (Chapter 4) expands on the first aim by developing a method that can incorporate high-dimensional and potentially correlated covariates by introducing a second set of latent classes that are formulated based on the observed high-dimensional covariate patterns. The third aim (Chapter 5) is an extension of the second aim which additionally assumes that there is an underlying nested structure between the latent classes of feature variables and covariates. Chapter 6 provides a summary of three main topics and recommendations for future research in handling covariates within the latent class framework.

Chapter 2

Literature Review

2.1 Latent Class Analysis

2.1.1 Overview

Lazarsfeld and Henry (1968) developed the latent class framework, which is concerned with measurement of characteristics that are not directly observable, and included indicator or manifest variables to cluster individuals and measure discrete "subpopulations". This traditional latent class model is used to clarify the relationships among discrete manifest (observed) variables in the model by uncovering a structure of latent (unobserved) classes.

2.1.2 Methods

Latent class analysis is a statistical method with unique properties that makes it an ideal tool for identifying heterogeneity within a study population. It is considered to be a method of model-based clustering based on finite mixture models.

Finite mixture models are derived under the assumption that the overall population is a mixture of C components. Using notations by McLachlan and Peel (2000), we first define a vector of M dichotomous features for the ith individual in a sample size of n as

$$Y_i = (y_{i1}, ..., y_{iM})^T, \quad i = 1, ..., n.$$

Assuming that the components of the vector Y_i are conditionally independent given their class membership within the mixture model, we can write the density function of an observation Y_i as

$$f(y_i) = \sum_{j=1}^{C} \pi_j f(y_i; \theta_j)$$

where $\pi_1, ..., \pi_C$ are mixing proportions of C components with $\sum_{j=1}^{C} \pi_j = 1$. Then the jth

component density is given by

$$f(y_i;\theta_j) = \prod_{m=1}^{M} \theta_{jm}^{y_{i1}}(1-\theta_{jm})^{1-y_{im}}$$

where θ_{jm} is a conditional probability of a response, i.e., $y_{im} = 1, m = 1, ..., M$, given its membership within the mixture and $\theta_j = (\theta_{j1}, ..., \theta_{jM})^T$. Then the log-likelihood function of latent class models can be derived as

$$l(\beta,\theta) = \sum_{i=1}^{n} log\Big\{ \sum_{j=1}^{C} \pi_{ij} f(y_i;\theta_j) \Big\}.$$

More generally, this idea can be expanded to a mixture of latent classes of where component densities can be Poisson, normal distribution, etc.

2.1.3 EM Algorithm

Using the Expectation-Maximization algorithm, we can maximize the log-likelihood function to find estimates of (β, θ). In order to implement the EM algorithm, we first derive the score equations for (β, θ):

$$\begin{aligned} S(\beta) = \frac{l(\beta,\theta)}{\beta} &= \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\partial log \pi_{ij}}{\partial \beta} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\partial log \pi_j(x_i;\beta)}{\partial \beta} \end{aligned}$$

$$S(\theta) = \frac{l(\beta,\theta)}{\theta} = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\partial log f(y_i;\theta_j)}{\partial \theta}$$

where τ_{ij} is the posterior probability that subject i belongs to latent class j

$$\begin{aligned}\tau_{ij} &= E(z_{ij}|x_i, y_i; \beta, \theta) \\ &= P(z_{ij} = 1|x_i, y_i; \beta, \theta) \\ &= \frac{\pi_{ij} f(y_i; \theta_j)}{\sum_{k=1}^{C} \pi_{ik} f(y_i; \theta_k)}\end{aligned}$$

and $z_{ij} = 1$ for some class j, and 0 otherwise. Starting with randomized initial values of posterior probabilities, the EM algorithm is implemented by iteratively solving for joint score equations and updating posterior probabilities until estimates of (β, θ) converge.

One limitation of using the EM algorithm is that it is easy to run into a multiple roots problem when solving for score functions. Bandeen-Roche et al. (1997) advises usage of multiple starting points to accurately detect global maximum of the log-likelihood function.

2.1.4 Information Matrix Under EM Algorithm

Efron and Hinkley (1978) provided theoretical and empirical evidence of inference of single parameter problems using the observed Fisher's information matrix. Using the EM algorithm, computation of the information matrix is straightforward in cases of complete data. In cases of incomplete data, however, it is necessary to include computation of the gradient and second derivative matrix within the frame of EM algorithm. That is, the gradient and second derivative matrix will go through iterations until convergence by the EM algorithm.

Consider the probability density $f(x|\theta)$ on a sample space χ, but we observe values of a measurable function $Y(x) = y \in Y$ rather than x. Defining $R = \{x : y(x) = y\}$, Louis (1982) showed that when regularity conditions hold, the information matrix of an incomplete data Y is expressed as

$$I_Y(\theta) = I(\hat{\theta}) = I_X - I_{X|Y}$$

where the first term is the conditional expected full data observed information matrix and the

last term is expected information for the conditional distribution of X given $X \in R$. These terms can easily be computed using the gradient and second derivative matrix obtained from EM algorithm, since

$$I_X = E_\theta\{B(X,\theta)|X \in R\}$$

where $B(X,\theta)$ is negative of second derivative matrix and

$$I_{X|Y} = E_\theta\{S(X,\theta)S^T(X,\theta)|X \in R\} - S^*(X,\theta)S^{*T}(X,\theta)$$

where $S(X,\theta)$ and $S^*(X,\theta)$ are gradient vectors of full data and incomplete data, respectively. I_Y is also the observed Fisher's information matrix defined by Efron and Hinkley (1978), and it can be used to approximate true standard error values of parameter estimates.

2.1.5 Model Selection

Assessing the number of components in a mixture model is a difficult task, especially when we aim to do model-based clustering without a prior knowledge of the number of components. One way to approach this issue is to use a classification-based information criteria such as the integrated classification likelihood-classification likelihood criterion (ICL-BIC) to overcome shortcomings of the Bayesian information criterion (BIC) of Schwarz (1978) and the classification likelihood criterion (CLC).

The BIC is obtained by applying Laplace's method of approximation,

$$-2\log L(\hat{\boldsymbol{\psi}}) + d\log n$$

where the penalty term $d\log n$ penalizes models that are not parsimonious. It is a reliable model selection criterion since the number of components are not underestimated asymptotically (Leroux, 1992). However, the BIC fits too few components when the model for the

component densities is valid and the sample size is not very large (Celeux and Soromenho, 1996).

Biernacki and Govaert (1997) developed the CLC by incorporating the likelihood and the complete likelihood obtained within the structure of EM algorithm. That is, denoting a vector of parameters as $\boldsymbol{\psi} = (\beta, \theta)$, the log-likelihood can be expressed as

$$\log L(\boldsymbol{\psi}) = \log L_c(\psi) - \log k(\boldsymbol{\psi})$$

where $L_c(k)$ is a complete likelihood and

$$\log k(\boldsymbol{\psi}) = \sum_{i=1}^{n} \sum_{j=1}^{C} z_{ij} \log \tau_{ij}$$

where z_{ij} is the vector of component indicator variables for the observed data Y_i and τ_{ij} is the posterior probability that subject i belongs to latent class j (Hathaway, 1986). This concept can be expanded to define the entropy of fuzzy classification matrix $C = ((\tau_{ij}))$,

$$EN(\tau) = \sum_{i=1}^{n} \sum_{j=1}^{C} \tau_{ij} \log \tau_{ij}$$

where $-EN(\tau)$ is derived from the conditional mean of $\log k(\boldsymbol{\psi})$ given a vector of observations Y_i. Combining these concepts, the CLC can be defined as

$$-2 \log L(\hat{\boldsymbol{\psi}}) + 2EN(\hat{\tau})$$

where ψ and τ are replaced by their maximum likelihood estimation (MLE) values. In this information criterion, the entropy term penalizes models for their complexity. Biernacki et al. (1999) noted that when the number of clusters within a mixture model were distinct, $EN(\hat{\tau})$ was close to 0. However, when the clusters were not definitive within a population, the entropy term was heavily penalized, resulting in an overestimated model.

To compensate for the disadvantages of BIC and CLC, the ICL-BIC model selection criterion was derived as a following objective function

$$-2\log L(\hat{\boldsymbol{\psi}}) + 2EN(\hat{\tau}) + d\log n$$

where d is the number of unknown parameters in $\boldsymbol{\psi}$ and n is the number of subjects. The number of latent classes, C, can be selected to minimize the objective function. Although this is a useful objective guide to model selection, it is also important to consider clinical interpretation of the latent classes.

2.1.6 Latent Class Regression Models

Let $Y_i = (Y_{i1}, ..., Y_{iM})'$ denote a vector of M observed features for the ith individual in a sample of size n. Define x_i to be a vector of covariates for the ith person. Assume that there are C latent classes and let Z_i be the true latent class of individual i. Bandeen-Roche et al. (1997) defines the probability of an individual i belonging to class j, given covariates x_i, as

$$P(Z_i = j|x_i) = \eta_j(\alpha, x_i) = \frac{\exp(x_i^T\alpha_j)}{\sum_{k=1}^{C}\exp(x_i^T\alpha_k)}, j = 1, ..., C$$

where $\alpha_1 = 0$ and $\sum_{j=1}^{C}\eta_j(\alpha, x_i) = 1$. Adopting a finite mixture model framework (McLachlan and Peel, 2000), the latent class regression likelihood can be written as

$$f(y_{i1}, ..., y_{im}|x_i) = \sum_{j=1}^{C}\eta_j(\alpha, x_i)f(y_{i1}, ..., y_{im}|x_i, Z_i = j)$$

where $f(y_{i1}, ..., y_{im}|x_i, Z_i = j)$ is the joint response probability given class j. Importantly, Bandeen-Roche et al. (1997) assume a non-differential measurement condition,

$$f(y_{i1}, ..., y_{im}|x_i, Z_i = j) = f(y_{i1}, ..., y_{im}|Z_i = j) = f_j(y_{i1}, ..., y_{im}).$$

Under the conditional independence assumption, we can write this class-specific joint probability as a product:

$$f_j(y_{i1}, ..., y_{im}) = \prod_{m=1}^{M} f_j(y_{im})$$

where $f_j(y_{im})$ is either a univariate probability density function if y_{im} is continuous or a univariate probability mass function if y_{im} is a discrete feature.

2.1.7 High-Dimensional Covariates in Latent Class Analysis

In investigating heterogeneity of MCI subgroups, incorporation of high-dimensional and potentially correlated covariates such as sociodemographic characteristics, health history such as indicators of cerebrovascular disease, coronary artery disease and cardiac dysrhythmias, genetics and biomarkers should be considered. Studies show that African-Americans tend to suffer from health conditions such as diabetes, hypertension, hypercholesterolemia and congestive heart failure, which leads to a higher incidence of non-amnestic MCI with executive dysfunction compared to non-African Americans, leading to the development of vascular dementia (Burke et al., 2018). In addition, APOE $\varepsilon 4$ allele is a well-documented genetic risk factor for Alzheimer's Disease and associated with amnestic MCI (Li et al., 2016).

A common method of incorporating high-dimensional covariates to analyze their relationships with MCI subgroups is using principal component analysis to perform dimension reduction on covariates for further analysis. For a set of variables, principal component analysis is used to explain their variance-covariance structure by using a linear combination of such variables. It is often used as a dimension reduction method, and it can reveal unknown relationships. It is used as a means to reduce dimension of variables, and used as inputs into multivariate analysis, such as multiple linear regression or cluster analysis (Johnson and Wichern, 2007).

Studies have been conducted using principal component analysis on high-dimensional risk factors of MCI subgroups. For instance, PET scans can be used to analyze memory perfor-

mance among patients diagnosed with amnestic MCI, then principal component analysis can be applied for further analysis using ANOVA or discriminant analysis (Nobili et al., 2008). Similarly, principal component analysis has been used on biomarkers of Alzheimer's disease such as triglycerides then further analyzed using linear regression models (Bernath et al., 2020). However, a standard dimension reduction method such as principal component analysis is not a feasible option to apply to latent class analysis, and very little research has been conducted to incorporate high-dimensional covariates within the latent class framework.

Chapter 3

Latent Class Analysis with Covariates

Activity Governor

3.1 Overview

The latent class regression model (Bandeen-Roche et al., 1997) assumes that covariates affect the latent class frequencies (i.e., mixture probabilities) but not the conceptualization of the latent classes. Despite this assumption, Table 1.2 demonstrated a previously unrecognized limitation of latent class regression models, where the presence of covariates altered not only the mixture probabilities but also the conceptualization of the latent classes. The 0-covariate model revealed a clinically important class characterized by non-amnestic cognitive features, specifically impairment in executive function, whereas the 7-covariate model failed to detect this non-amnestic class and instead heightened the role of neuropsychiatric features. To allow researchers the flexibility to incorporate covariates, without making them fully active in the model and potentially distorting the interpretations of the latent classes, we introduce the concept of a covariate activity governor.

3.2 Methods

3.2.1 Relative Frequencies Model

Consider two sets of covariates, x_i and w_i within a latent class model. Define covariates x_i to be ungoverned covariates, which are fully active in the model. Separately define covariates w_i to be governed covariates, whose activity within a latent class model will be adjusted accordingly by an investigator. Denoting an indicator of latent class membership of each subject as $Z_i = (z_{i1}, ..., z_{iC})$ for n subjects and C latent classes, a relative frequencies model of a two-component mixture model with ungoverned and governed components can be written as

$$\begin{aligned} Pr(z_{ij} = 1|x_i, w_i; \alpha, \beta, \gamma, \phi) &= p_j(\alpha, \beta, \gamma, \phi, x_i, w_i) \\ &= (1 - \phi)p_{ju}(\alpha, x_i) + \phi p_{jg}(\alpha, \beta, x_i, w_i), (0 \leq \phi \leq 1) \end{aligned} \quad (1)$$

where we have introduced a covariate activity governor ϕ, which is a constant specified by the investigator to determine contribution of the governed component within a latent class model. The ungoverned component, dependent on the fully active covariates x_i, is the latent polytomous logistic regression model of Bandeen-Roche et al. (1997)

$$p_{ju}(\alpha, x_i) = \frac{exp(x_i^T \alpha_c)}{\sum_{j=1}^{C} exp(x_i^T \alpha_j)}, j = 1, ..., C \tag{2}$$

with $\alpha_1 = 0$ for identifiability and $\alpha = (\alpha_1^T, ..., \alpha_C^T)^T$, and the governed component is extended to include governed covariates w_i

$$p_{jg}(\alpha, \beta, x_i, w_i) = \frac{exp(x_i^T \beta_c + w_i^T \gamma_c)}{\sum_{j=1}^{C} exp(x_i^T \beta_j + w_i^T \gamma_j)}, j = 1, ..., C \tag{3}$$

with $\beta_1 = 0$, $\gamma_1 = 0$, $\beta = (\beta_1^T, ..., \beta_C^T)^T$ and $\gamma = (\gamma_1^T, ..., \gamma_C^T)^T$. Let ϕ be the covariates activity governor, which an investigator can use to determine contribution of the governed component within a latent class model. Covariates w_i are inactive when $\phi = 0$ and w_i are fully active when $\phi = 1$. Hence, the activity governor provides a continuum of possibilities of latent class models between models with inactive w_i or fully active w_i.

3.2.2 Maximum Likelihood Estimation

Assuming that there are n subjects and C latent classes, log-likelihood for the finite mixture model can be derived as

$$l_\phi(\alpha, \beta, \gamma, \theta) = \sum_{i=1}^{n} log\Big\{ \sum_{j=1}^{C} p_j(\alpha, \beta, \gamma, \phi, x_i, w_i) f_j(y_i; \theta) \Big\}, 0 \leq \phi \leq 1$$

where $f_j(y_i; \theta)$ is the density function of the observed features y_i. Then for a fixed choice of the activity governor ϕ, the EM algorithm can be applied to find estimators of $(\alpha, \beta, \gamma, \theta)$ that maximizes the log-likelihood function. This is derived by jointly solving for following

score equations:

$$S_\phi(\alpha) = \frac{\partial l_\phi}{\partial \alpha} = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{(1-\phi)p_{ju}(\alpha, x_i)}{p_j(\alpha, \beta, \gamma, \phi, x_i, w_i)} \frac{\partial \log p_{ju}(\alpha, x_i)}{\partial \alpha}$$

$$S_\phi(\beta) = \frac{\partial l_\phi}{\partial \beta} = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\phi p_{jg}(\beta, \gamma, x_i, w_i)}{p_j(\alpha, \beta, \gamma, \phi, x_i, w_i)} \frac{\partial \log p_{jg}(\beta, \gamma, x_i, w_i)}{\partial \beta}$$

$$S_\phi(\gamma) = \frac{\partial l_\phi}{\partial \gamma} = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\phi p_{jg}(\beta, \gamma, x_i, w_i)}{p_j(\alpha, \beta, \gamma, \phi, x_i, w_i)} \frac{\partial \log p_{jg}(\beta, \gamma, x_i, w_i)}{\partial \gamma}$$

$$S_\phi(\theta) = \frac{\partial l_\phi}{\partial \theta} = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij} \frac{\partial \log f_j(y_i; \theta)}{\partial \theta}$$

where τ_{ij} is the posterior probability that subject i belongs to latent class j

$$\begin{aligned} \tau_{ij} &= E(z_{ij}|y_i, x_i, w_i; \alpha, \beta, \theta, \phi) \\ &= P(z_{ij} = 1|y_i, x_i, w_i; \alpha, \beta, \theta, \phi) \\ &= \frac{p_j(\alpha, \beta, \gamma, \phi, x_i, w_i) f_j(y_i; \theta)}{\sum_{j=1}^{C} p_j(\alpha, \beta, \gamma, \phi, x_i, w_i) f_j(y_i; \theta)}. \end{aligned} \tag{4}$$

The EM algorithm consists of iteratively solving the joint score equations for a fixed value of τ_{ij} and updating the posterior probability τ_{ij} until convergence, where we use the following steps:

1. Initialize the estimates of (α, β, γ) and values of τ_{ij}.

2. Update α by fitting a polytomous logistic regression model with covariates X_i, and update (β, γ) by fitting a polytomous logistic regression model with covariates (X_i, W_i). Update predictions $p_{ju}(\alpha, x_i)$ and $p_{jg}(\beta, \gamma, x_i, w_w)$.

3. Update the latent class-specific means $\theta_j = (\theta_{j1}, ..., \theta_{jM})$ of M features by

$$\hat{\theta}_{jm} = \frac{\sum_i \tau_{ij} y_{ij}}{\sum_i \tau_{ij}}, \quad m = 1, ..., M$$

4. Update τ_{ij} using equation (4).

and repeat steps 2-4 until $(\alpha, \beta, \gamma, \theta)$ converge.

3.2.3 Model-Averaging

Consider a governor $G \sim Bernoulli(\phi)$ where G is independent of covariates (X_i, W_i), and G determines the activity of covariates w_i, such that $Pr(z_{ij} = 1|x_i, w_i, G = 0)$ is given by Equation (2) and $Pr(z_{ij} = 1|x_i, w_i, G = 1)$ is given by Equation (3). Then we can derive a two-component mixture model with ungoverned and governed components in Equation (1). Moreover, it follows that the model-averaged effect of the ungoverned covariates X on the relative frequencies of the latent classes can be expressed as the averaged log odds ratio

$$\begin{aligned}\Delta_j(x) &= E_g\left\{log\frac{Pr(z_{ij} = 1|X = x, w, G)/Pr(z_{i1} = 1|X = x, w, G)}{Pr(z_{ij} = 1|X = x - 1, w, G)/Pr(z_{i1} = 1|X = x - 1, w, G)}\right\} \\ &= \{(1 - \phi)\alpha_c + \phi\beta_c\}^T x\end{aligned}$$

Similarly, the model averaged effect of the governed covariates W can be expressed as the following averaged log odds ratio

$$\begin{aligned}\delta_j(w) &= E_g\left\{log\frac{Pr(z_{ij} = 1|x, W = w, G)/Pr(z_{i1} = 1|x, W = w, G)}{Pr(z_{ij} = 1|x, W = w - 1, G)/Pr(z_{i1} = 1|x, W = w - 1, G)}\right\} \\ &= \phi\gamma_c^T w\end{aligned}$$

Hence, we can interpret the covariate effects easily by reporting the model-averaged effects $\Delta_j(x)$ and $\delta_j(w)$. In order to compute the model-averaged effects, we can obtain standard errors of $\psi = (\alpha^T, \beta^T, \gamma^T, \theta^T)^T$ and define

$$Q(\psi, y) = \sum_{i=1}^{n} Q_i(\psi, y_i) = \sum_{i=1}^{n}\sum_{j=1}^{C} \tau_{ij}(\psi, y_i)q_{ij}(\psi, y_i)$$

where column vector $q_{ij}(\psi, y_i)$ is

$$q_{ij}(\psi, y_i) = \begin{pmatrix} \frac{(1-\phi)p_{ju}(\alpha,x_i)}{p_j(\alpha,\beta,\gamma,\phi,x_i,w_i)} \frac{\partial \log p_{ju}(\alpha,x_i)}{\partial \alpha} \\ \frac{\phi p_{jg}(\alpha,\beta,x_i,w_i)}{p_j(\alpha,\beta,\gamma,\phi,x_i,w_i)} \frac{\partial \log p_{jg}(\beta,\gamma,x_i,w_i)}{\partial \beta} \\ \frac{\phi p_{jg}(\alpha,\beta,x_i,w_i)}{p_j(\alpha,\beta,\gamma,\phi,x_i,w_i)} \frac{\partial \log p_{jg}(\beta,\gamma,x_i,w_i)}{\partial \gamma} \\ \frac{\partial \log f_j(y_i;\theta)}{\partial \theta} \end{pmatrix}$$

It follows that the maximum likelihood estimator $\hat{\psi}$ is the root of Q. The empirical Fisher's information matrix can be expressed as

$$I = \sum_{i=1}^{n} Q_i(\psi, y_i) Q_i(\psi, y_i)^T$$

and an approximation of the variance-covariance matrix of $\hat{\psi}$ is $\text{avar}(\hat{\psi}) = I^{-1}$. From this, the standard errors of parameter estimates $\hat{\psi}$, as well as the model-averaged effects $\Delta_j(x)$ and $\delta_j(w)$, can be computed.

3.2.4 Analysis of Underlying Population

Combining the relative frequencies model and maximum likelihood estimation from previous sections, we can outline the analysis of underlying population using the activity governor value ϕ, where Z_i is an indicator of latent class membership of each subject. Z_i is derived by a parametric relationship with ungoverned covariate x_i and governed covariates x_i and w_i, and represented with parameter estimates α, β and γ.

This diagram emphasizes the necessity of using the activity governor values ϕ and $1-\phi$ to incorporate all relevant covariates in the form of governed and ungoverned covariates. The ungoverned component $(1-\phi)p_{ju}(\alpha, x_i)$ and the governed component $\phi p_{jg}(\beta, \gamma, x_i, w_i)$ provide relevant clinical interpretations of the latent classes from latent class-specific means of feature variables, $\hat{\theta}_{j1}$,..., $\hat{\theta}_{jM}$.

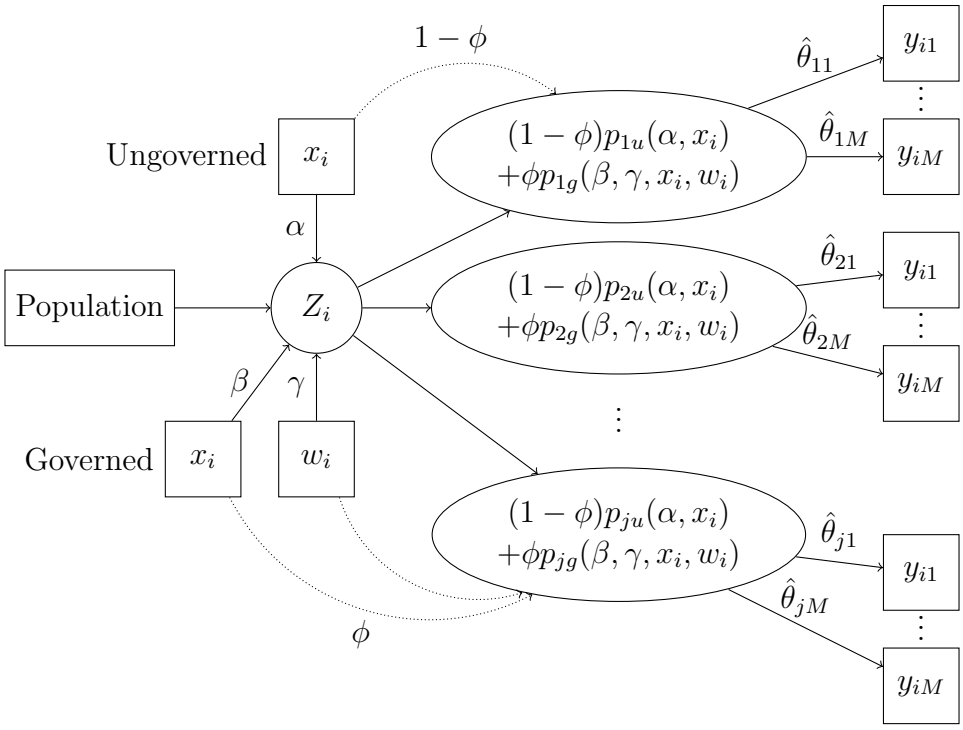

Figure 3.1: Latent Class Regression Model with Covariate Activity Governor

3.3 Results

3.3.1 Simulation Studies

Simulation studies were conducted to investigate scenarios where the presence of covariates might affect the conceptualization of latent classes, and gain insight into the practicality of incorporating the covariate activity governor.

3.3.1.1 Design A: Unstructured Covariates Independent of Manifest Variables

The first simulation study is based on a 2-class finite mixture model where covariates do not have a structure. A dataset was generated with one ungoverned covariate x_1 and 8 governed covariates $w_1, ..., w_8$. Let $x_1, w_1, ..., w_8 \overset{iid}{\sim} N(0, 1)$. Outcome variables were generated as a mixture of two populations for 500 subjects:

	Class 1 (70%)	Class 2 (30%)
Y_1	Bern(0.1)	Bern (0.25)
Y_2	Bern(0.6)	Bern(0.4)
Y_3	Pois(1)	Pois(3)
Y_4	N(5,1)	N(5,1)
Y_5	N(0,1)	N(-1.5,1)

Covariates x_i and w_i were unrelated to latent class memberships, so that the probability of an individual i belonging to the first class, given covariates x_i and w_i, is 0.7, or $Pr(z_{i1} = 1|x_i, w_i) = 0.7$.

We fitted latent class models with covariate activity governor ϕ =0.001, 0.25, 0.50, 0.75 and 0.999. The results showed that ungoverned covariate x_1 and governed covariates $w_1, ..., w_8$ were not statistically significant (Table 3.1). The optimal model is obtained when $\phi = 0.25$, supported by the minimum ICL-BIC value. However, the ICL-BIC values did not vary much between different models (Table 3.3). It is further observed that the latent class-specific means for different activity governor values remained similar (Table 3.2).

Table 3.1: Design A - Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor				
Estimate	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
Ungoverned Component						
Intercept	2	0.37 (2.92)	0.96 (0.20)	0.03 (0.13)	1.03 (0.12)	0.71 (0.10)
x_1	2	-0.35 (2.98)	0.13 (0.20)	0.29 (0.13)	-0.32 (0.12)	0.05 (0.10)
Governed Component						
Intercept	2	0.71 (0.10)	0.83 (0.12)	-2.77 (0.33)	-1.32 (0.60)	0.44 (3.89)
x_1	2	0.04 (0.10)	0.05 (0.11)	-0.96 (0.25)	5.25 (1.36)	0.28 (3.77)
w_1	2	0.12 (0.10)	0.18 (0.12)	-1.09 (0.25)	1.25 (0.56)	0.09 (3.92)
w_2	2	0.02 (0.10)	0.01 (0.12)	0 (0.22)	1.71 (0.61)	0.47 (4.10)
w_3	2	-0.10 (0.09)	-0.16 (0.11)	0.33 (0.21)	-6.56 (1.68)	-0.42 (3.99)
w_4	2	0.14 (0.09)	0.17 (0.11)	-0.50 (0.23)	3.77 (1.03)	-0.05 (3.57)
w_5	2	0.08 (0.09)	0.14 (0.11)	-0.75 (0.22)	0.90 (0.53)	0.74 (4.08)
w_6	2	-0.04 (0.09)	-0.06 (0.11)	0.29 (0.22)	-2.78 (0.77)	1.24 (4.51)
w_7	2	0.10 (0.10)	0.11 (0.12)	-0.08 (0.23)	2.68 (0.79)	1.51 (5.42)
w_8	2	-0.13 (0.09)	-0.18 (0.11)	0.14 (0.22)	-1.75 (0.63)	-0.49 (3.86)
Estimate		Activity Governor				
(Model-Averaged)	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
x_1	2	0.04 (0.09)	0.78 (0.10)	-0.34 (0.06)	1.07 (0.09)	0.05 (0.09)
w_1	2	0.12 (0.09)	0.13 (0.13)	-0.54 (0.10)	0.31 (0.18)	0 (53.72)
w_2	2	0.02 (0.10)	0 (0.14)	0 (0.10)	0.43 (0.20)	0 (60.98)
w_3	2	-0.10 (0.10)	-0.12 (0.13)	0.16 (0.09)	-1.64 (0.21)	0 (58.67)
w_4	2	0.14 (0.09)	0.13 (0.12)	-0.25 (0.09)	0.94 (0.19)	0 (55.10)
w_5	2	0.08 (0.09)	0.11 (0.12)	-0.38 (0.09)	0.23 (0.18)	0 (50.69)
w_6	2	-0.04 (0.09)	-0.05 (0.12)	0.15 (0.09)	-0.69 (0.20)	0 (53.30)
w_7	2	0.10 (0.11)	0.09 (0.15)	-0.04 (0.09)	0.67 (0.21)	0 (61.12)
w_8	2	-0.13 (0.09)	-0.14 (0.12)	0.07 (0.09)	-0.44 (0.19)	0 (59.01)

Table 3.2: Design A - Latent Class-Specific Means (and Standard Errors)

	Class	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Y_1	1	0.22 (0.42)	0.23 (0.42)	0.23 (0.42)	0.22 (0.42)	0.23 (0.42)
	2	0.09 (0.28)	0.09 (0.28)	0.08 (0.28)	0.09 (0.28)	0.09 (0.28)
Y_2	1	0.46 (0.50)	0.46 (0.50)	0.46 (0.50)	0.47 (0.50)	0.47 (0.50)
	2	0.57 (0.50)	0.56 (0.50)	0.56 (0.50)	0.56 (0.50)	0.56 (0.50)
Y_3	1	2.94 (1.71)	3.01 (1.73)	2.93 (1.74)	2.90 (1.73)	2.97 (1.70)
	2	1.06 (1.02)	1.11 (1.06)	1.09 (1.05)	1.07 (1.03)	1.06 (1.02)
Y_4	1	4.93 (1.07)	4.93 (1.07)	4.97 (1.01)	4.94 (1.06)	4.92 (1.08)
	2	4.97 (1.00)	4.96 (1.00)	4.92 (1.05)	4.96 (1.00)	4.97 (1.00)
Y_5	1	-1.47 (0.93)	-1.56 (0.90)	-1.54 (0.90)	-1.52 (0.91)	-1.47 (0.94)
	2	0.02 (0.98)	0 (0.97)	0.03 (0.96)	0.05 (0.95)	0 (0.98)

Table 3.3: Design A-Model Selection

	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Log-Likelihood	-1476.47	-1476.45	-1474.88	-1470.78	-1478.78
BIC	3114.52	3114.49	3111.34	3103.14	3119.13
Entropy	150.24	140.29	142.28	140.14	152.33
ICL-BIC	3414.99	3395.06	3395.91	3383.42	3423.80

3.3.1.2 Design B: Structured Covariates Independent of Manifest Variables

The second simulation study is updated to have a 2-class finite mixture model where covariates, in addition to an underlying structure, are independent of manifest variables. That is, an ungoverned covariate x_1 and governed covariates $w_1, ..., w_8$ were generated to be independent of the latent classes of manifest variables within a generated dataset. If we let $x_1 \sim Bern(0.5)$, governed covariates were generated as $w_1, ..., w_8 \overset{iid}{\sim}$ N(2, 1) for $x_1 = 1$ and $w_1, ..., w_8 \overset{iid}{\sim} N(-2, 1)$ otherwise. Outcome variables were generated as a mixture of two populations for 500 subjects,

	Class 1 (70%)	Class 2 (30%)
Y_1	Bern(0.1)	Bern (0.25)
Y_2	Bern(0.6)	Bern(0.4)
Y_3	Pois(1)	Pois(3)
Y_4	N(5,1)	N(5,1)
Y_5	N(0,1)	N(-1.5,1)

and covariates x_i and w_i were unrelated to latent class memberships, where the probability of an individual i belonging to the first class, given covariates x_i and w_i, is 0.7, or $Pr(z_{i1} = 1|x_i, w_i) = 0.7$.

We fitted latent class models with covariate activity governor ϕ =0.001, 0.25, 0.50, 0.75 and 0.999. Governed covariates tended to be statistically significant when fully governed (ϕ = 0.999), and became less statistically significant as activity governor decreased (Table 3.4). Changes in statistical significance did not happen until $\phi \leq 0.25$ for model-averaged values (Table 3.4, Model-Averaged values). With minimum ICL-BIC model selection criterion value of 3325.55, the 2-class solution when $\phi = 0.50$ was the best fitting model (Table 3.6). However, as in Design A, ICL-BIC values did not vary between models despite different

ϕ values. In addition, conceptualization of latent class solutions did not vary much across different models (Table 3.5).

Table 3.4: Design B - Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor				
Estimate	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
Ungoverned Component						
Intercept	2	-0.02 (2.84)	-0.21 (0.19)	0.12 (0.13)	-0.56 (0.11)	0.65 (0.09)
x_1	2	-0.04 (2.85)	0.70 (0.21)	-0.78 (0.15)	-0.42 (0.12)	0.14 (0.10)
Governed Component						
Intercept	2	-3.20 (0.44)	1.05 (0.13)	-6.07 (1.01)	-6.64 (1.72)	71.50 (878)
x_1	2	-6.07 (0.85)	-0.06 (0.12)	2.85 (0.57)	4.26 (1.19)	28.51 (343.3)
w_1	2	1.81 (0.30)	0.30 (0.12)	-1.43 (0.37)	-3.08 (0.86)	19.82 (260.7)
w_2	2	-0.20 (0.20)	0 (0.11)	-0.50 (0.32)	-1.55 (0.58)	84.63 (1029.5)
w_3	2	-2.48 (0.40)	0.28 (0.12)	-3.19 (0.62)	-3.54 (0.98)	56.13 (676.8)
w_4	2	0.67 (0.23)	0.18 (0.11)	0.94 (0.31)	1.33 (0.52)	-4.62 (98.61)
w_5	2	3.24 (0.45)	-0.09 (0.11)	-0.24 (0.29)	-0.14 (0.43)	-33.07 (393.6)
w_6	2	-4.49 (0.61)	-0.24 (0.12)	0.67 (0.33)	1.11 (0.53)	45.14 (538.2)
w_7	2	-0.05 (0.21)	-0.20 (0.11)	1.83 (0.42)	3.52 (1.01)	-168.4 (2050.5)
w_8	2	2.12 (0.35)	-0.17 (0.11)	1.18 (0.33)	1.11 (0.47)	16.98 (222.7)
Estimate		Activity Governor				
(Model-Averaged)	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
x_1	2	-6.07 (0.06)	0.93 (0.09)	1.03 (0.06)	0.75 (0.06)	0.17 (0.09)
w_1	2	1.80 (0.05)	0.22 (0.13)	-0.72 (0.09)	-0.77 (0.19)	0.02 (41.95)
w_2	2	-0.20 (0.05)	0 (0.12)	-0.25 (0.09)	-0.39 (0.19)	0.09 (36.33)
w_3	2	-2.48 (0.06)	0.21 (0.13)	-1.59 (0.09)	-0.89 (0.18)	0.06 (38.67)
w_4	2	0.67 (0.05)	0.13 (0.12)	0.47 (0.09)	0.33 (0.20)	0 (36.21)
w_5	2	3.24 (0.05)	-0.07 (0.14)	-0.12 (0.08)	-0.03 (0.17)	-0.03 (41.67)
w_6	2	-4.49 (0.06)	-0.20 (0.14)	0.33 (0.08)	0.28 (0.17)	0.05 (42.49)
w_7	2	-0.05 (0.05)	-0.15 (0.13)	0.92 (0.09)	0.88 (0.18)	-0.17 (42.42)
w_8	2	2.11 (0.05)	-0.12 (0.15)	0.59 (0.08)	0.28 (0.17)	0.02 (48.78)

Table 3.5: Design B - Latent Class-Specific Means (and Standard Errors)

	Class	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Y_1	1	0.12 (0.33)	0.08 (0.27)	0.08 (0.27)	0.08 (0.27)	0.08 (0.27)
	2	0.19 (0.39)	0.26 (0.44)	0.27 (0.44)	0.27 (0.44)	0.27 (0.44)
Y_2	1	0.57 (0.50)	0.56 (0.50)	0.56 (0.50)	0.56 (0.50)	0.56 (0.50)
	2	0.44 (0.50)	0.46 (0.50)	0.45 (0.50)	0.46 (0.50)	0.46 (0.50)
Y_3	1	1.85 (1.71)	1.11 (1.03)	1.13 (1.04)	1.13 (1.05)	1.12 (1.04)
	2	1.69 (1.55)	3.07 (1.86)	3.12 (1.86)	3.11 (1.86)	3.09 (1.86)
Y_4	1	4.98 (1.05)	5.06 (1.04)	5.05 (1.04)	5.06 (1.04)	5.06 (1.04)
	2	5.24 (0.95)	5.07 (0.99)	5.08 (1.00)	5.08 (1.00)	5.07 (1.00)
Y_5	1	-0.39 (1.30)	0.16 (0.93)	0.15 (0.94)	0.16 (0.93)	0.16 (0.93)
	2	-0.37 (1.02)	-1.40 (1.01)	-1.42 (1.01)	-1.45 (0.99)	-1.42 (1.01)

Table 3.6: Design B -Model Selection

	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Log-Likelihood	-1519.42	-1458.40	-1450.56	-1453.81	-1462.84
BIC	3200.42	3078.38	3062.70	3069.20	3087.26
Entropy	65.48	139.62	131.43	131.23	140.61
ICL-BIC	3331.37	3357.61	3325.55	3331.66	3368.48

3.3.1.3 Comorbidity Design

The comorbidity design mimics the mild cognitive impairment study, where the dataset contains not only covariates representing risk factors of disease, x_i, but also covariates representing strong risk factors of comorbidity, w_i.

Let $x_i, w_i \overset{iid}{\sim} N(0,1)$. For any given individual, let $Z_1 \sim Bern(p_1)$ be the indicator of the disease class, where $p_1 = \frac{exp(\eta_1)}{1+exp(\eta_1)}$ and $\eta_1 = x_i$. Similarly, let $Z_2 \sim Bern(p_2)$ be the indicator of comorbidity, where $p_2 = \frac{exp(\eta_2)}{1+exp(\eta_2)}$ and $\eta_2 = logit(0.1) + 3 * w_i$. One response variable, Y_1, was generated to be a manifestation of disease, while the other response variable, Y_2, was primarily a manifestation of comorbidity. Specifically, $Y_1 \sim N(-1.5, 1)$ if $Z_1 = 1$ and $Y_1 \sim N(0,1)$ otherwise. The distribution of Y_2 depended on both Z_1 and Z_2: $Y_2 \sim N(2,1)$ if $(Z_1, Z_2) = (1,1)$, $Y_2 \sim N(1.1,1)$ if $(Z_1, Z_2) = (0,1)$, $Y_2 \sim N(0.3,1)$ if $(Z_1, Z_2) = (1,0)$ and $Y_2 \sim N(-0.3,1)$ if $(Z_1, Z_2) = (0,0)$. If we were to marginalize over Z_2, then Y_2 has a mean of 0.75 in the disease group ($Z_1 = 1$) and 0.07 in the non-disease group ($Z_1 = 0$). Alternatively, if we were to marginalize over Z_1, then Y_2 has a mean of 1.55 in the comorbidity group ($Z_2 = 1$) and 0.00 in the group without comorbidity ($Z_2 = 0$).

We fitted latent class models with covariate activity governor ϕ =0.001, 0.10, 0.20, 0.50 and 0.999. The model-averaged effect of ungoverned covariates x_i was significant for $0.20 \leq \phi \leq 0.50$, whereas the governed covariate w_i was significant when $\phi \geq 0.50$ (Table 3.7). To interpret the latent classes, we looked for class-specific means in Table 3.8 that were at least 1.0 unit apart. At the extremes, a model in which w_i was almost fully inactive ($\phi = 0.001$) revealed the structure of the disease only, whereas a model in which w_i was almost fully active ($\phi = 0.999$) revealed the structure of comorbidity only. Values of ϕ between 0.10 and

0.25 yielded latent classes that separated the sample based on both disease and comorbidity. The ICL-BIC criterion selected the model with $\phi = 0.50$ (Table 3.9).

Table 3.7: Comorbidity Design - Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor				
Estimate	Class	ϕ=0.999	ϕ=0.50	ϕ=0.20	ϕ=0.10	ϕ=0.001
Ungoverned Component						
Intercept	2	2.35 (430.66)	-0.57 (0.14)	-0.71 (0.33)	-0.66 (0.34)	-0.16 (0.23)
x_1 (Risk Factor of Disease)	2	3.40 (481.77)	0.72 (0.15)	0.90 (0.29)	1.14 (0.28)	1.02 (0.19)
Governed Component						
Intercept	2	-1.13 (0.29)	-2.55 (0.46)	-1.69 (2.87)	-1.16 (4.99)	-1.49 (845.11)
x_1 (Risk Factor of Disease)	2	0.64 (0.23)	0.71 (0.78)	1.32 (2.73)	2.02 (6.92)	6.28 (2850.97)
w_1 (Risk Factor of Comorbidity)	2	1.51 (0.24)	5.17 (0.85)	6.07 (8.04)	6.08 (16.45)	13.90 (6022.45)
Estimate		Activity Governor				
(Model-Averaged)	Class	ϕ=0.999	ϕ=0.50	ϕ=0.20	ϕ=0.10	ϕ=0.001
Intercept	2	−1.13(0.39)	−1.64(0.61)	−0.91(0.53)	−0.71(0.50)	−0.16(0.85)
x_1	2	0.64(0.54)	0.72(0.29)	0.98(0.46)	1.23(0.64)	1.02(2.83)
w_1	2	1.51(0.24)	2.59(1.24)	1.21(1.61)	0.61(1.64)	0.01(6.02)

Table 3.8: Comorbidity Design - Latent Class-Specific Means (and Standard Errors)

	Class	$\phi = 0.999$	$\phi = 0.50$	$\phi = 0.20$	$\phi = 0.10$	$\phi = 0.001$
Y_1	1	-0.46 (1.24)	-0.43 (0.09)	-0.29 (1.20)	-0.09 (1.12)	0.12 (0.10)
	2	-1.20 (1.33)	-1.22 (0.12)	-1.37 (1.21)	-1.64 (1.01)	-1.58 (1.00)
Y_2	1	-0.22 (0.96)	-0.28 (0.07)	-0.26 (0.97)	-0.10 (1.09)	-0.10 (1.11)
	2	1.65 (0.89)	1.67 (0.07)	1.42 (1.02)	1.09 (1.20)	0.87 (1.26)

Table 3.9: Comorbidity Design-Model Selection

	$\phi = 0.999$	$\phi = 0.50$	$\phi = 0.20$	$\phi = 0.10$	$\phi = 0.001$
Log-Likelihood	-695.06	-691.99	-708.36	-713.35	-717.43
BIC	1470.91	1464.78	1497.52	1507.49	1515.65
Entropy	115.88	114.97	157.16	163.47	161.42
ICL-BIC	1702.67	1694.71	1811.84	1834.43	1838.48

3.3.1.4 Missingness Design

For each of 500 subjects, let covariates $x_1, w_1, w_2, w_3 \overset{iid}{\sim} N(0,1)$ where the three governed covariates each have an independent 10% probability of being missing. Moreover, assume that the latent classes based on manifest variable Y_1 are more widely separated when any governed covariate is missing than when w_1, w_2, w_3 do not have any missing values. Standard latent class analysis excludes observations with missing values of covariates, which may yield a misleading interpretation of the latent classes. Specifically, let the true latent class membership indicator $Z \sim Bern(p)$, where $p = \frac{exp(\eta)}{1+exp(\eta)}$ and $\eta = logit(0.7) + 2 * x_1$. Assume that the distribution of Y_1 depends on both Z and the missingness of covariates: $Y_1 \sim N(-0.4, 1)$ if $Z = 1$ and there are any missing values, $Y_1 \sim N(0.4, 1)$ if $Z = 1$ and there are no missing values, $Y_1 \sim N(1.9, 1)$ if $Z = 0$ and there are any missing values, and $Y_1 \sim N(1.1, 1)$ if $Z = 0$ and there are no missing values.

We fitted latent class models with covariate activity governor ϕ =0.001, 0.25, 0.50, 0.75 and 0.999. Results for selected values of ϕ are shown in Table 3.10 and Table 3.11. The model-averaged effect of ungoverned covariates x_i was significant for $\phi = 0.50$ and 0.20, whereas the governed covariate w_i was significant when $\phi \geq 0.50$ (Table 3.10). To interpret the latent classes, we looked for class-specific means in Table 3.11 that were at least 1.0 unit apart. At the extremes, a model in which w_i was almost fully inactive ($\phi = 0.001$) revealed the structure of the disease only, whereas a model in which w_i was almost fully active ($\phi = 0.999$) revealed the structure of comorbidity only. Values of ϕ between 0.10 and 0.25 yielded latent classes that detected the influence of both disease and comorbidity. The ICL-BIC criterion selected the model with $\phi = 0.50$.

Table 3.10: Missingness Design - Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor				
Estimate	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
Ungoverned Component						
Intercept	2	-2.18 (1099.02)	-1.02 (1.13)	0.59 (0.57)	0.54 (0.41)	0.65 (0.48)
x_1	2	13.13 (6237.23)	1.50 (0.98)	1.66 (0.53)	1.08 (0.26)	1.16 (0.29)
Governed Component						
Intercept	2	5.24 (2.80)	-4.08 (1.94)	6.02 (4.72)	3.81 (6.44)	2.12 (762.13)
x_1	2	0.20 (0.92)	-0.21 (0.65)	-1.12 (1.13)	0.75 (2.19)	0.45 (413.35)
w_1	2	-2.20 (1.52)	-1.78 (1.09)	-3.97 (3.28)	-6.27 (9.84)	-2.41 (798.06)
w_2	2	1.11 (1.28)	-0.38 (0.66)	1.40 (1.17)	0.70 (1.92)	0.68 (337.61)
w_3	2	0.87 (0.82)	-0.60 (0.67)	2.43 (2.18)	2.46 (3.74)	1.11 (444.20)
Estimate		Activity Governor				
(Model-Averaged)	Class	ϕ=0.999	ϕ=0.75	ϕ=0.50	ϕ=0.25	ϕ=0.001
Intercept	2	5.23 (2.93)	-3.32 (1.41)	3.31 (2.33)	1.35 (1.61)	0.65 (0.84)
x_1	2	0.22 (6.26)	0.22 (0.48)	0.27 (0.60)	1.00 (0.54)	1.16 (0.40)
w_1	2	-2.20 (1.52)	-1.33 (0.82)	-1.98 (1.64)	-1.57 (2.46)	0 (0.80)
w_2	2	1.11 (1.38)	-0.29 (0.50)	0.70 (0.58)	0.17 (0.48)	0 (0.34)
w_3	2	0.87 (0.82)	-0.45 (0.50)	1.21 (1.09)	0.61 (0.94)	0 (0.44)

Table 3.11: Missingness Design - Latent Class-Specific Means (and Standard Errors)

	Class	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Y_1	1	1.18 (1.58)	1.01 (1.10)	1.67 (1.06)	1.57 (1.03)	1.43 (1.12)
	2	0.59 (1.05)	-0.46 (0.81)	0.30 (1.01)	0.19 (0.98)	0.30 (1.05)

Table 3.12: Missingness Design-Model Selection

	$\phi = 0.999$	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$	$\phi = 0.001$
Log-Likelihood	-641.74	-646.44	-643.57	-648.22	-662.38
BIC	1359.47	1368.87	1363.12	1372.43	1400.75
Entropy	114.84	282.49	369.06	414.25	457.25
ICL-BIC	1589.15	1933.85	2101.25	2200.94	2315.25

3.3.2 MCI Dataset

3.3.2.1 Overview

In this section, we further analyzed the MCI dataset from the motivating example (Chapter 1, Section 1.2), introducing the covariate activity governor to control for the influence of 7 covariates related to vascular comorbidity. All 7 vascular covariates were used as governed covariates and only an intercept (i.e. no covariates) was used in the ungoverned component

of the model. By specifying different values of covariate activity governor, ϕ, changes in clinical interpretation of the 5 latent classes were detectable. Assuming that the 7-covariate model from the motivating example is equivalent to when the activity governor value is $\phi = 1$ (ICL-BIC=213773.64), and the 0-covariate model is equivalent to when the activity governor value is $\phi = 0$ (ICL-BIC=208670.62), we found that $\phi = 0.75$ was preferred according to both the objective ICL-BIC criterion and clinical judgment when we compared the results from the motivating example (Table 3.16).

3.3.2.2 Analysis of Underlying Population

Figure 3.2 outlines the application of activity governor value $\phi = 0.75$ in the UDS dataset. In this diagram, we assigned 7 vascular covariates to be governed covariates in order to derive latent class membership through Z_i and find the parameter estimate γ. $\phi = 0.75$ was also used to derive relative frequencies model $(1-\phi)p_{ju}(\alpha, x_i) + \phi p_{jg}(\gamma, x_i, w_i)$, which can control for the effect of vascular covariates on clinical interpretations of the latent classes derived from functional, neuropsychiatric and cognitive test scores. The clinical interpretations were based on latent class-specific means of functional, neuropsychiatric and cognitive tests and written as $\hat{\theta}_{j1}..., \hat{\theta}_{jM}$ for M features and $j = 1, ..., C$ classes.

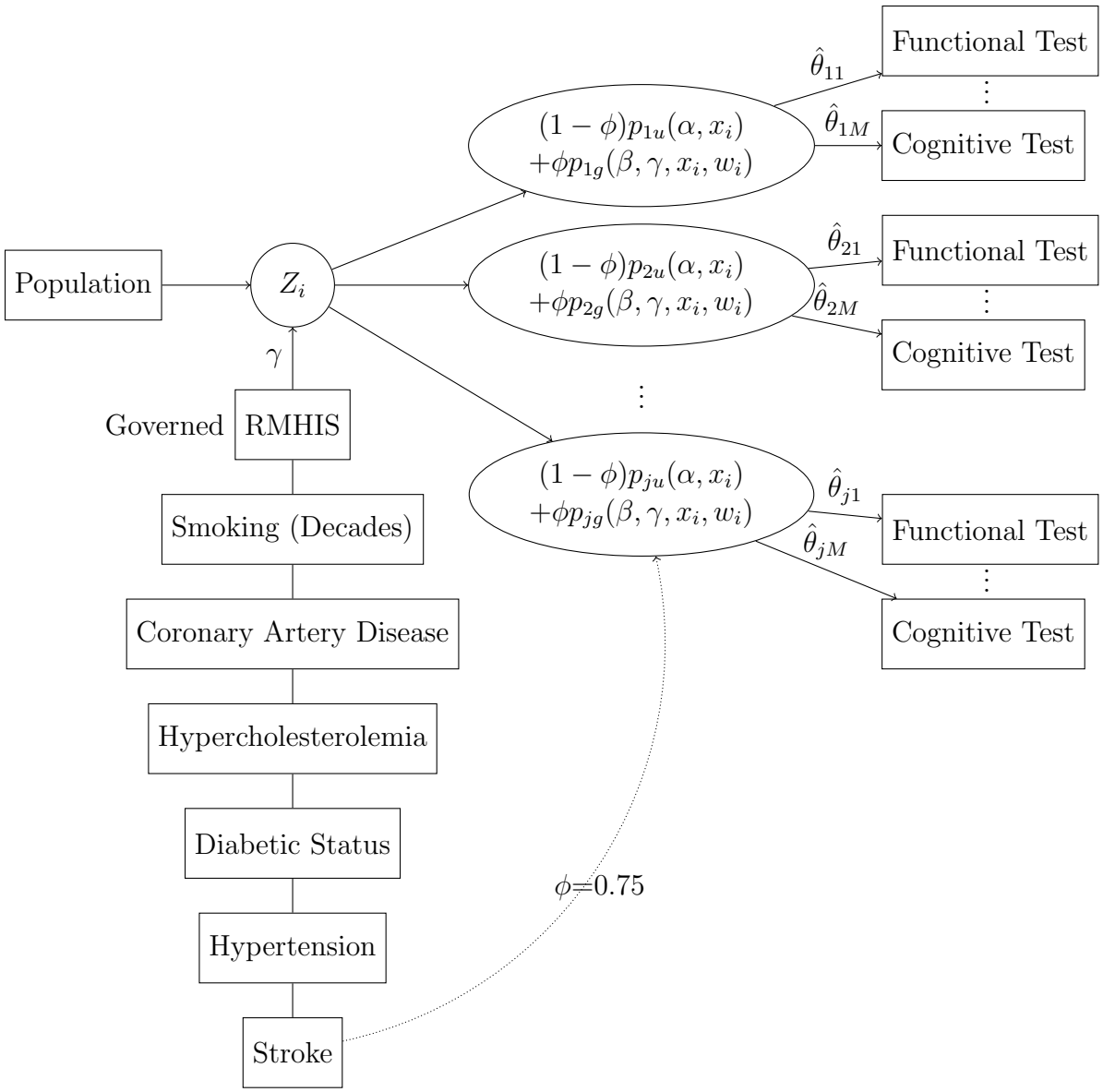

Figure 3.2: Application to MCI Data with Covariate Activity Governor

3.3.2.3 Analysis

The changes in statistical significance were prominent in parameter estimates as ϕ decreased (Table 3.13). When $\phi = 0.75$, the model-averaged effects of all covariates, except stroke, were significant (Table 3.14). As seen in Table 1.2, the resulting 5-class solution was quite similar to the 0-covariate solution when $\phi = 0.75$,

1. Non-amnestic with functional impairment and neuropsychiatric features
2. Mildly impaired
3. Functional impairment and neuropsychiatric features
4. Amnestic with functional impairment and neuropsychiatric features
5. Amnestic multi-domain with functional impairment and neuropsychiatric features

where with the exception that the first class, "non-amnestic with functional impairment and neuropsychiatric features", had an expanded phenotype that included cognitive impairment in not only executive function but also attention and language as well as neuropsychiatric features. Hence, this model successfully partially incorporated the information from vascular covariates, demonstrating that these covariates were risk factors affecting latent class frequencies and, importantly, also identifying the class characterized by non-amnestic cognitive impairment (as found in the solution with no covariates but missed in the model with 7 fully active covariates).

Using "non-amnestic with functional impairment and neuropsychiatric features" class as a reference group, we interpreted the exponentiated model-averaged log odds ratios of covariates when $\phi = 0.75$ (Table 3.14). The results indicated that mildly impaired participants were 0.15 times as likely to suffer from probable cerebrovascular disease, 0.73 times as likely to suffer from coronary artery disease and 0.55 times as likely to suffer from diabetes compared to the reference group (OR: $e^{-1.90} = 0.15$, $e^{-0.32} = 0.73$, $e^{-0.59} = 0.55$). Moreover, participants in the "functional impairment and neuropsychiatric features" class were 0.22

times as likely to suffer from probable cerebrovascular disease, 0.58 times as likely to suffer from diabetes but 1.79 times more likely to have high levels of cholesterol compared to the reference group (OR: $e^{-1.51} = 0.22$, $e^{-0.54} = 0.58$, $e^{0.58} = 1.79$). Participants in the "amnestic with functional impairment and neuropsychiatric features" were 0.20 times as likely to suffer from probable cerebrovascular disease, 0.91 times as likely to smoke, 0.49 times as likely to suffer from diabetes, 0.64 times as likely to suffer from hypertension but 1.77 times more likely to have high levels of cholesterol compared to the reference group (OR: $e^{-1.59} = 0.20$, $e^{-0.09} = 0.91$, $e^{-0.72} = 0.49$, $e^{-0.45} = 0.64$, $e^{0.57} = 1.77$). Finally, participants in the "amnestic multi-domain with functional impairment and neuropsychiatric features" class were 0.21 times as likely to suffer from probable cerebrovascular disease, 0.86 times as likely to smoke, 0.41 times as likely to suffer from diabetes but 1.52 times more likely to have high levels of cholesterol compared to the reference group (OR: $e^{-1.54} = 0.21$, $e^{-0.15} = 0.86$, $e^{-0.90} = 0.41$, $e^{0.42} = 1.52$).

Table 3.13: Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor		
Covariate	Class	ϕ=0.75	ϕ=0.50	ϕ=0.25
Ungoverned Component				
Intercept	1	—	—	—
	2	-0.90 (1.45)	0.11 (0.80)	0.04 (0.21)
	3	-0.98 (1.57)	-0.64 (0.56)	0.11 (0.28)
	4	-3.77 (24.94)	0.09 (0.77)	0.18 (0.25)
	5	-2.90 (8.84)	-0.04 (0.74)	0.14 (0.27)
Governed Component				
Intercept	1	—	—	—
	2	1.41 (0.75)	0.30 (1.02)	0.25 (1.32)
	3	0.88 (0.96)	-1.40 (0.86)	1.00 (1.32)
	4	1.83 (0.58)	0.54 (0.86)	1.03 (1.23)
	5	1.18 (0.76)	0.99 (0.78)	1.37 (1.17)
RMHIS	1	—	—	—
	2	-2.54 (0.77)	-0.55 (0.52)	-0.52 (1.08)
	3	-2.01 (0.53)	1.56 (0.61)	-6.22 (41.92)
	4	-2.12 (0.49)	-6.05 (96.28)	-5.83 (31.65)
	5	-2.05 (0.49)	-2.31 (1.11)	-5.08 (8.97)
Smoking (Per Decade)	1	—	—	—
	2	-0.09 (0.06)	-0.27 (0.12)	0.17 (0.15)
	3	-0.02 (0.08)	-0.09 (0.09)	0 (0.13)
	4	-0.12 (0.06)	-0.10 (0.07)	0.08 (0.15)
	5	-0.20 (0.09)	-0.11 (0.07)	-0.18 (0.19)
Coronary Artery Disease	1	—	—	—
	2	-0.42 (0.21)	-0.16 (0.26)	0.47 (0.58)
	3	-0.20 (0.23)	-0.16 (0.34)	-0.10 (0.50)
	4	-0.34 (0.20)	-0.41 (0.24)	-0.17 (0.56)
	5	-0.28 (0.22)	-0.27 (0.22)	0.10 (0.52)
Hypercholesterolemia	1	—	—	—
	2	0.37 (0.26)	-0.03 (0.28)	1.03 (0.82)
	3	0.77 (0.27)	-0.72 (0.41)	0.20 (0.73)
	4	0.76 (0.26)	-0.28 (0.22)	1.33 (0.69)
	5	0.56 (0.26)	-0.40 (0.20)	0.88 (0.64)
Diabetic Status	1	—	—	—
	2	-0.78 (0.28)	-1.12 (0.24)	-0.31 (0.67)
	3	-0.71 (0.28)	0.83 (0.41)	-0.80 (0.65)
	4	-0.96 (0.26)	-0.46 (0.24)	-1.59 (0.67)
	5	-1.19 (0.26)	-0.85 (0.21)	-2.52 (0.62)
Hypertension	1	—	—	—
	2	-0.34 (0.29)	0.15 (0.53)	-1.57 (0.78)
	3	-0.57 (0.31)	0.42 (0.49)	0.16 (0.73)
	4	-0.60 (0.30)	0.03 (0.31)	-1.16 (0.87)
	5	-0.40 (0.37)	-0.41 (0.30)	-0.50 (1.38)
Stroke	1	—	—	—
	2	-0.17 (0.32)	0.23 (0.47)	-1.34 (0.91)
	3	-0.47 (0.40)	0.86 (0.45)	-0.54 (0.89)
	4	-0.41 (0.29)	-0.30 (0.55)	-1.53 (1.72)
	5	-0.23 (0.33)	-0.45 (0.52)	-0.33 (1.01)

Table 3.14: Model-Averaged Estimated Log Odds Ratios (and Standard Errors) of Governed and Ungoverned Covariates with Class 1 as a Reference

		Activity Governor		
Covariate	Class	ϕ=0.75	ϕ=0.50	ϕ=0.25
Intercept	1	—	—	—
	2	0.83 (0.42)	0.20 (0.27)	0.09 (0.20)
	3	0.41 (0.43)	-1.02 (0.33)	0.33 (0.17)
	4	0.43 (6.08)	0.32 (0.42)	0.40 (0.18)
	5	0.16 (1.85)	0.48 (0.26)	0.45 (0.17)
RMHIS	1	—	—	—
	2	-1.90 (0.58)	-0.28 (0.26)	-0.13 (0.27)
	3	-1.51 (0.40)	0.78 (0.31)	-1.56 (10.48)
	4	-1.59 (0.37)	-3.03 (48.14)	-1.46 (7.91)
	5	-1.54 (0.37)	-1.16 (0.55)	-1.27 (2.24)
Smoking (Per Decade)	1	—	—	—
	2	-0.07 (0.04)	-0.14 (0.06)	0.04 (0.04)
	3	-0.02 (0.06)	-0.04 (0.05)	0 (0.03)
	4	-0.09 (0.04)	-0.05 (0.04)	0.02 (0.04)
	5	-0.15 (0.07)	-0.06 (0.04)	-0.05 (0.05)
Coronary Artery Disease	1	—	—	—
	2	-0.32 (0.16)	-0.08 (0.13)	0.12 (0.14)
	3	-0.15 (0.17)	-0.08 (0.17)	-0.03 (0.12)
	4	-0.25 (0.15)	-0.21 (0.12)	-0.04 (0.14)
	5	-0.21 (0.16)	-0.13 (0.11)	0.03 (0.13)
Hypercholesterolemia	1	—	—	—
	2	0.28 (0.20)	-0.02 (0.14)	0.26 (0.20)
	3	0.58 (0.20)	-0.36 (0.20)	0.05 (0.18)
	4	0.57 (0.19)	-0.14 (0.11)	0.33 (0.17)
	5	0.42 (0.20)	0.20 (0.10)	0.22 (0.16)
Diabetic Status	1	—	—	—
	2	-0.59 (0.21)	-0.56 (0.12)	-0.08 (0.17)
	3	-0.54 (0.21)	0.42 (0.20)	-0.20 (0.16)
	4	-0.72 (0.19)	-0.23 (0.12)	-0.40 (0.17)
	5	-0.90 (0.19)	-0.43 (0.11)	-0.63 (0.15)
Hypertension	1	—	—	—
	2	-0.25 (0.22)	0.08 (0.26)	-0.39 (0.20)
	3	-0.43 (0.23)	0.21 (0.24)	0.04 (0.18)
	4	-0.45 (0.22)	0.02 (0.15)	-0.29 (0.22)
	5	-0.30 (0.28)	-0.21 (0.15)	-0.13 (0.34)
Stroke	1	—	—	—
	2	-0.12 (0.24)	0.11 (0.23)	-0.34 (0.23)
	3	-0.35 (0.30)	0.43 (0.22)	-0.13 (0.22)
	4	-0.31 (0.22)	-0.15 (0.28)	-0.38 (0.43)
	5	-0.17 (0.25)	-0.22 (0.26)	-0.08 (0.25)

Table 3.15: Class-Specific Means or Proportions of Functional, Neuropsychiatric, and Cognitive Features in Five-Class Model. Shown are the results for $\phi = 0.75, 0.50, 0.25$.Cognitive test scores were standardized by demographics so that a cognitive value of -1.5 indicates that an individual with MCI performed 1.5 standard deviations worse than a cognitively normal person of the same age, race, and years of education. Neuropsychologists typically regard cognitive standardized scores of -1.5 or worse as evidence of impairment in a specific cognitive domain.

Model Type	Test Type		Non-Amnestic With Functional Impairment And Neuropsychiatric Features (Relative Frequency=22%)	Mildly Impaired (Relative Frequency=22%)	Functional Impairment And Neuropsychiatric Features (Relative Frequency=17%)	Amnestic With Functional Impairment And Neuropsychiatric Features (Relative Frequency=26%)	Amnestic Multi-Domain With Functional Impairment And Neuropsychiatric Features (Relative Frequency=13%)
ϕ=0.75	Functional	No. of IADL impaired	3.53 (2.87)	0 (<0.01)	2.03 (2.00)	2.82 (2.50)	3.51(2.60)
(7 Governed	Neuropsychiatric	% with GDS $\geq$ 5	29.43 (0.46)%	9.89 (0.30)%	18.84 (0.39)%	15.91 (0.37)%	19.91 (0.40)%
Covariates)		No. of NPI-Q symptoms present	2.53 (2.29)	0 (<0.01)	2.05 (1.82)	2.19 (1.92)	2.16 (1.92)
	Cognitive	Global					
		MMSE	-2.21 (1.92)	-1.27 (1.78)	-0.51 (1.20)	-1.50 (1.69)	-3.05 (1.88)
		Logical Memory					
		Immediate	-1.04 (0.88)	-0.95 (1.11)	-0.05 (0.69)	-1.59 (0.68)	-2.54 (0.63)
		Delayed	-1.11 (0.89)	-0.97 (1.11)	-0.09 (0.71)	-1.81 (0.74)	-2.64 (0.47)
		Semantic Memory					
		Category Fluency	-1.23 (0.97)	-0.76 (0.94)	-0.48 (0.94)	-0.77 (0.86)	-1.60 (0.85)
		Attention					
		Trails A	2.34 (2.22)	0.52 (1.51)	0.10 (0.85)	-0.04 (0.69)	0.76 (1.03)*
		Digit Span Forward	-0.64 (1.04)	-0.27 (1.01)	-0.10 (0.99)	-0.07 (0.98)	-0.54 (1.07)
		Language					
		Boston Naming	-1.67 (2.09)	-1.07 (1.80)	-0.43 (1.22)	-0.44 (1.12)	-2.04 (2.36)
		Executive Function					
		Trails B	3.18 (1.74)	1.02 (1.68)	0.39 (1.03)	0.22 (0.75)	1.91 (1.80)*
		Digit Span Backward	-0.85 (0.89)	-0.44 (0.95)	-0.19 (0.95)	-0.28 (0.98)	-0.76 (0.84)
		Visuomotor					
		Digit Symbol	-1.70 (0.98)	-0.61 (1.03)	-0.40 (0.93)	-0.38 (0.90)	-1.17 (1.01)

Model Type	Test Type		Executive Function With Functional Impairment And Neuropsychiatric Features (Relative Frequency=21%)	Amnestic Multi-Domain With Functional Impairment And Neuropsychiatric Features (Relative Frequency=20%)	Non-Amnestic With Functional Impairment And Neuropsychiatric Features (Relative Frequency=9%)	Mildly Impaired (Relative Frequency=22%)	Functional Impairment And Neuropsychiatric Features (Relative Frequency=28%)
ϕ=0.50	Functional	No. of IADL impaired	3.13 (0.04)	3.69 (0.03)	3.48 (0.03)	0 (0.03)	2.10 (0.06)
(7 Governed	Neuropsychiatric	% with GDS $\geq$ 5	26.79 (0.08)%	21.43 (0.36)%	30.55 (0.32)%	9.84 (0.05)%	13.09 (0.04)%
Covariates)		No. of NPI-Q symptoms present	2.77 (0.03)	2.34 (0.05)	2.21 (0.05)	0 (0.08)	1.79 (0.09)
	Cognitive	Global					
		MMSE	-1.03 (0.07)	-2.94 (0.06)	-2.61 (0.03)	-1.25 (0.04)	-1.09 (0.03)
		Logical Memory					
		Immediate	-0.63 (0.03)	-2.32 (0.08)	-0.93 (0.07)	-0.95 (0.04)	-1.05 (0.02)
		Delayed	-0.70 (0.05)	-2.45 (0.04)	-0.99 (0.03)	-0.97 (0.03)	-1.21 (0.03)
		Semantic Memory					
		Category Fluency	-0.84 (0.07)	-1.50 (0.04)	-1.40 (0.06)	-0.75 (0.13)	-0.55 (0.10)
		Attention					
		Trails A	0.73 (0.77)	1.24 (0.72)	3.01 (0.05)	0.48 (0.06)	-0.23 (0.03)*
		Digit Span Forward	-0.36 (0.08)	-0.51 (0.11)	-0.76 (0.16)	-0.27 (0.11)	0.01 (0.11)
		Language					
		Boston Naming	-0.70 (0.02)	-1.82 (0.02)	-2.24 (0.03)	-1.04 (0.05)	-0.36 (0.03)
		Executive Function					
		Trails B	1.57 (0.03)	1.95 (0.03)	3.76 (0.05)	0.99 (0.06)	-0.02 (0.03)*
		Digit Span Backward	-0.57 (0.06)	-0.76 (0.06)	-0.96 (0.05)	-0.43 (0.05)	-0.10 (0.13)
		Visuomotor					
		Digit Symbol	-1.08 (0.06)	-1.21 (0.06)	-2.01 (0.09)	-0.60 (0.06)	-0.13 (0.05)

Model Type	Test Type		Amnestic Multi-Domain With Functional Impairment And Neuropsychiatric Features (Relative Frequency=18%)	Executive Function With Functional Impairment And Neuropsychiatric Features (Relative Frequency=18%)	Mildly Impaired (Relative Frequency=21%)	Mildly Impaired (Relative Frequency=21%)	Amnestic With Functional Impairment And Neuropsychiatric Features (Relative Frequency=22%)
ϕ=0.25	Functional	No. of IADL impaired	3.27 (0.04)	3.80 (0.03)	0.11 (0.04)	1.40 (0.08)	3.44 (0.05)
(7 Governed	Neuropsychiatric	% with GDS $\geq$ 5	27.96 (0.08)%	31.08 (0.10)%	8.52 (0.08)%	11.94 (0.07)%	17.66 (0.06)%
Covariates)		No. of NPI-Q symptoms present	2.14 (0.03)	3.23 (0.05)	0.18 (0.09)	1.39 (0.08)	2.19 (0.10)
	Cognitive	Global					
		MMSE	-2.88 (0.08)	-0.97 (0.05)	-1.28 (0.04)	-0.62 (0.04)	-2.52 (0.04)
		Logical Memory					
		Immediate	-1.50 (0.03)	-0.55 (0.04)	-0.93 (0.06)	-0.71 (0.03)	-2.18 (0.05)
		Delayed	-1.56 (0.06)	-0.64 (0.02)	-0.95 (0.03)	-0.81 (0.03)	-2.39 (0.03)
		Semantic Memory					
		Category Fluency	-1.51 (0.05)	-0.81 (0.03)	-0.81 (0.06)	-0.35 (0.06)	-1.25 (0.11)
		Attention					
		Trails A	2.98 (0.05)	0.71 (0.05)	0.50 (0.04)	-0.34 (0.05)	0.26 (0.03)*
		Digit Span Forward	-0.76 (0.04)	-0.30 (0.12)	-0.35 (0.02)	0.08 (0.01)	-0.29 (0.05)
		Language					
		Boston Naming	-2.42 (0.03)	-0.50 (0.05)	-1.13 (0.03)	-0.18 (0.02)	-1.21 (0.06)
		Executive Function					
		Trails B	3.69 (0.03)	1.45 (0.03)	1.18 (0.02)	-0.19 (0.02)	0.87 (0.03)*
		Digit Span Backward	-1.00 (0.06)	-0.49 (0.03)	-0.55 (0.03)	0.02 (0.03)	-0.52 (0.04)
		Visuomotor					
		Digit Symbol	-1.91 (0.06)	-1.06 (0.03)	-0.75 (0.08)	0.08 (0.08)	-0.75 (0.06)

* Higher scores on Trail A and Trail B indicate worse performance.

Table 3.16: MCI-Model Selection

	$\phi = 0.75$	$\phi = 0.50$	$\phi = 0.25$
Log-Likelihood	-33788.04	-34170.53	-52815.73
BIC	68977.62	69742.58	107032.99
Entropy	1360.32	1676.21	2021.19
ICL-BIC	71698.25	73095.01	111075.37

3.4 Discussion

In this chapter, we recognized a problem that arises with latent class regression models, where covariates affect both latent class frequencies (i.e., mixture probabilities) and conceptualization of the latent classes. We provided different simulation studies to explore scenarios where fully active covariates might affect the conceptualization of latent classes. We found that despite having a pronounced internal structure, covariates did not change the interpretation of the latent classes when they were unrelated to the patterns among the manifest variables (i.e., $\alpha = 0$, with the possible exception of intercepts, Section 3.2.1-Equation 2). When covariates had a structure related to the manifest variables, however, we found that fully active covariates can alter the conceptualization of latent classes.

The comorbidity design (Section 3.3.1) revealed that the activity governor can provide the flexibility to explore the structure of a combination of disease and comorbidity and find an optimal model that is based on clinical judgement. Hence, this simulation study successfully demonstrated that, when the relationship between covariates and manifest variables is more complicated than assumed by the standard latent class regression model, our covariate activity governor provides a flexible method to explore and customize the extent to which covariates alter the clinical interpretation of the latent classes. The missingness design (Section 3.3.1) demonstrated the possibility to explore the extent to which the structure of the population will change with different activity governor values with inclusion of covariates.

In the MCI dataset, we selected seven covariates related to vascular comorbidity to be governed, which were deemed important by an investigator. The results showed that the activity governor was able to reveal unique MCI subtypes that were unobservable when the covariates were fully active ($\phi = 1$). In further application of our method, investigators will have the option of including covariates suspected of influencing the clinical interpretation of the latent classes as candidates to be governed.

Our method primarily depends on the investigator's expertise to determine which covariates should be governed, and we should consider techniques to empirically select covariates

that should be governed. Moreover, the number of classes remain the same as values for the covariates activity governor ϕ varies, but one could allow the number of classes to vary as ϕ varies to explore how different number of components influence the scientific interpretation of the latent classes.

By providing straightforward clinical interpretation of the latent classes and an option to govern the activity of covariates, the activity governor will provide more flexibility than the standard model in investigating the effects of covariates.

3.5 Appendix

From section 3.2.3, we know

$$Pr(z_{ij} = 1|x, w, G = 0) = \frac{exp(x^T\alpha_j)}{\sum_{k=1}^{C} exp(x^T\alpha_k)}, j = 1, ..., C$$

and

$$Pr(z_{ij} = 1|x, w, G = 1) = \frac{exp(x^T\beta_j + w^T\gamma_j)}{\sum_{k=1}^{C} exp(x^T\beta_k + w^T\gamma_k)}, j = 1, ..., C.$$

Since $G \sim Bernoulli(\phi)$, where G is independent of covariates (X_i, W_i), we can write

$$Pr(z_{ij} = 1|x, w, G) = \left(\frac{exp(x^T\beta_j + w^T\gamma_j)}{\sum_{k=1}^{C} exp(x^T\beta_k + w^T\gamma_k)}\right)^G \left(\frac{exp(x^T\alpha_j)}{\sum_{k=1}^{C} exp(x^T\alpha_k)}\right)^{1-G}, j = 1, ..., C$$

where $\alpha_1 = 0$, $\beta_1 = 0$ and $\gamma_1 = 0$ for identifiability and $\alpha = (\alpha_1^T, ..., \alpha_C^T)^T$, $\beta = (\beta_1^T, ..., \beta_C^T)^T$ and $\gamma = (\gamma_1^T, ..., \gamma_C^T)^T$.

The model-averaged log odds ratio of the effect of the ungoverned covariates X from section 3.3 can be calculated as:

$$\begin{aligned}
\Delta_j(x) &= E_G\left\{log\frac{Pr(z_{ij} = 1|X = x, w, G)/Pr(z_{i1} = 1|X = x, w, G)}{Pr(z_{ij} = 1|X = x - 1, w, G)/Pr(z_{i1} = 1|X = x - 1, w, G)}\right\} \\
&= E_G\left\{log\frac{Pr(z_{ij} = 1|X = x, w, G)}{Pr(z_{i1} = 1|X = x, w, G)}\right\} - E_G\left\{log\frac{Pr(z_{ij} = 1|X = x - 1, w, G)}{Pr(z_{i1} = 1|X = x - 1, w, G)}\right\} \\
&= \phi(x^T\beta_j + w^T\gamma_j) + (1 - \phi)(x^T\alpha_j) - [\phi\{(x - 1)^T\beta_j + w^T\gamma_j\} + (1 - \phi)\{(x - 1)^T\alpha_j\}] \\
&= \{(1 - \phi)\alpha_j + \phi\beta_j\}^T
\end{aligned}$$

where

$$\frac{Pr(z_{ij} = 1|X = x, w, G)}{Pr(z_{i1} = 1|X = x, w, G)} = \{exp(x^T\beta_j + w^T\gamma_j)\}^G\{exp(x^T\alpha_j)\}^{1-G}$$

and

$$\frac{Pr(z_{ij}=1|X=x-1,w,G)}{Pr(z_{i1}=1|X=x-1,w,G)}=\{exp((x-1)^T\beta_j+w^T\gamma_j)\}^G\{exp((x-1)^T\alpha_j)\}^{1-G}$$

for $j=1,...,C$.

Similarly, the model averaged effect of the governed covariates W can be expressed as

$$\begin{aligned}\delta_j(w)&=E_G\left\{log\frac{Pr(z_{ij}=1|x,W=w,G)/Pr(z_{i1}=1|x,W=w,G)}{Pr(z_{ij}=1|x,W=w-1,G)/Pr(z_{i1}=1|x,W=w-1,G)}\right\}\\&=E_G\left\{log\frac{Pr(z_{ij}=1|x,W=w,G)}{Pr(z_{i1}=1|x,W=w,G)}\right\}-E_G\left\{log\frac{Pr(z_{ij}=1|x,W=w-1,G)}{Pr(z_{i1}=1|x,W=w-1,G)}\right\}\\&=\phi(x^T\beta_j+w^T\gamma_j)+(1-\phi)(x^T\alpha_j)-[\phi\{x^T\beta_j+(w-1)^T\gamma_j\}+(1-\phi)(x^T\alpha_j)]\\&=\phi\gamma_j^T\end{aligned}$$

where

$$\frac{Pr(z_{ij}=1|x,W=w-1,G)}{Pr(z_{i1}=1|x,W=w-1,G)}=\{exp(x^T\beta_j+(w-1)^T\gamma_j)\}^G\{exp(x^T\alpha_j)\}^{1-G}.$$

Chapter 4

Compound Latent Class Analysis

4.1 Overview

Standard latent class analysis (LCA) methods were used in the first aim to explore the relative frequencies of the latent classes resulting from covariates of the latent class model. However, these methods are limited to low-dimensional covariates. When covariates are high-dimensional and potentially correlated, standard methods are not feasible options. A two-step procedure could be potentially used, where, any standard method can reduce the dimension of the covariates, and the results of dimension reduction can be used as fixed inputs into standard LCA. Although it is intuitive to implement standard dimension reduction methods, the two-step procedure ignores uncertainties or "fuzziness" in dimension reduction that are propagated in the LCA.

Instead, we propose an alternative method, namely compound LCA, that introduces a second set of latent classes that are formulated based on the observed high-dimensional covariate patterns, and that has advantage of fully incorporating the fuzziness of the dimension reduction in the LCA.

4.2 Methods

4.2.1 Relative Frequencies Model

For subject i ($i \in \{1, ..., n\}$), let x_i be the observed covariate vector and let y_i be the observed response vector. Let $a \in \{1, ..., A\}$ denote the latent classes of covariates, and let $b \in \{1, ..., B\}$ denote the latent classes of responses. Assume that A and B, the numbers of latent classes, are known. Then standard LCA is based on following finite mixture model of the responses:

$$f(y_i|x_i) = \sum_b \pi(b|x_i) f(y_i|b, x_i) = \sum_b \pi(b|x_i) f(y_i|b)$$

where we assume $f(y_i|b, x_i) = f(y_i|b)$ and $\pi(b|x_i)$ is a parametric model of the direct effect of the covariates on the relative frequencies of the latent classes of the responses. When x_i is

high dimensional and possibly correlated, we propose replacing the relative frequency model $\pi(b|x_i)$ with one of lower dimension, $\pi(b|a)$, based on latent classes of covariates.

When we assume conditional independence between the covariates and responses given the latent classes

$$f(x_i, y_i|a, b) = f(x_i|a)f(y_i|b)$$

then under compound LCA, the finite mixture model is given by

$$\begin{aligned} e^{\ell_i} = f(x_i, y_i; \alpha, \beta) &= \sum_a \sum_b \pi(a, b) f(x_i, y_i|a, b; \alpha, \beta) \\ &= \sum_a \left\{ \pi(a) f(x_i|a; \alpha) \sum_b \pi(b|a) f(y_i|b; \beta) \right\} \end{aligned}$$

with log-likelihood $\ell = \sum_i \ell_i$.

4.2.2 Maximum Likelihood Estimation

Using the EM algorithm, estimation can be carried out by iterating between estimation of relative frequencies of the two sets of latent classes using EM algorithm:

$$\hat{\pi}(a) = n^{-1} \sum_i \psi_{ia}$$

$$\hat{\pi}(b|a) = n^{-1} \hat{\pi}(a)^{-1} \sum_i \tau_{iab}$$

and solving the weighted score equations for the parameters α and β of the conditional distributions:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \sum_i \sum_a \psi_{ia} \frac{\partial log f(x_i|a; \alpha)}{\partial \alpha} = 0 \\ \frac{\partial \ell}{\partial \beta} &= \sum_i \sum_a \sum_b \tau_{iab} \frac{\partial log f(y_i|b; \beta)}{\partial \beta} = 0. \end{aligned}$$

Then posterior probabilities of membership in the two sets of latent classes are simultaneously updated:

$$\tau_{iab} = Pr(a, b|x_i, y_i; \alpha, \beta) = \frac{\pi(a)f(x_i|a;\alpha)\pi(b|a)f(y_i|b;\beta)}{\sum_{a'}\{\pi(a')f(x_i|a';\alpha)\sum_{b'}\pi(b'|a')f(y_i|b';\beta)\}}$$
$$\psi_{ia} = Pr(a|x_i, y_i; \alpha, \beta) = \sum_b \tau_{iab}. \tag{5}$$

We assume conditional independence of the covariates, so that $f(x_i|a;\alpha) = \prod_k f(x_{ik}|a;\alpha)$ with distinct means α_{ak}. Additionally, we assume conditional independence of the responses, so that $f(y_i|b;\beta) = \prod_j f(y_{ij}|b;\beta)$ with distinct means β_{bj}. Then we use the following EM algorithm to solve for the posterior probabilities of membership and latent class specific means:

1. Initialize values of τ_{iab}, ψ_{ia}, A and B.

2. Update the latent class specific means at each iteration of the EM algorithm, where

$$\hat{\alpha}_{ak} = \frac{\sum_i \psi_{ia} x_{ik}}{\sum_i \psi_{ia}}$$
$$\hat{\beta}_{bj} = \frac{\sum_i \sum_a \tau_{iab} y_{ij}}{\sum_i \sum_a \tau_{iab}}$$

3. Update the posterior probabilities of membership in the two sets of latent classes, τ_{iab} and ψ_{ia}, using Equation (5).

4. Repeat steps 2-3 until convergence.

4.2.3 Information Matrix

The empirical Fisher's information matrix can be computed to find the standard errors of parameter estimates. We can compute a column vector of weighted score equations for

$\zeta = (\pi(a)^T, \pi(b|a)^T, \alpha^T, \beta^T)^T$, defined by $Q(\zeta, x_i, y_i)$, where

$$Q(\zeta, x_i, y_i) = \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix}$$

and the solution to $Q(\zeta, x_i, y_i)$ is the maximum likelihood estimator $\hat{\zeta}$. Then the empirical Fisher's information matrix can be written as

$$\begin{aligned} I &= \sum_i Q(\zeta, x_i, y_i) Q(\zeta, x_i, y_i)^T \\ &= \sum_i \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix} \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix}^T \end{aligned}$$

where

$$\begin{aligned} \frac{\partial \ell}{\partial \pi(a)} &= \sum_{i=1}^{n} \frac{\psi_{ia}}{\pi(a)}, \quad a = 1, ..., A \\ \frac{\partial \ell}{\partial \pi(b|a)} &= \sum_{i=1}^{n} \sum_{a=1}^{A} \frac{\tau_{iab}}{\pi(b|a)}, \quad b = 1, ..., B \\ \frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^{n} \psi_{ia} \frac{\partial \log f(x_i|a; \alpha)}{\partial \alpha}, \quad a = 1, ..., A \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^{n} \sum_{a=1}^{A} \tau_{iab} \frac{\partial \log f(y_i|b; \beta)}{\partial \beta}, \quad b = 1, ..., B. \end{aligned}$$

We can derive the standard errors of parameter estimates $\hat{\zeta}$ by computing an estimator of the asymptotic covariance matrix $\text{avar}(\hat{\zeta}) = I^{-1}$.

4.2.4 Model Selection Criterion

McLachlan and Peel (2000) recommend using the ICL-BIC model selection criterion to find number of components C for latent class analysis, which includes penalty terms for parsimony and complexity of the model. The ICL-BIC model section criterion can be defined as the following objective function

$$-2\log L(\hat{\zeta}) + 2EN(\hat{\tau}) + d\log n$$

that is minimized, where $\log L(\hat{\zeta})$ is the log-likelihood function evaluated at parameter estimates $\hat{\zeta}$, d is the number of parameters in the model and n is the number of subjects. $EN(\hat{\tau})$ is the entropy of the fuzzy classification matrix $((\tau_{ij}))$, defined by

$$EN(\tau) = -\sum_{i=1}^{n}\sum_{j=1}^{C}\tau_{ij}\log\tau_{ij}$$

for C components of i subject.

We can expand the objective function to estimate the numbers of latent classes, A and B, by following the guidelines of McLachlan and Peel (2000) to include both a BIC term for lack of parsimony and an entropy penalty for fuzziness of latent class membership probabilities

$$-2\log L(\hat{\zeta}) + 2EN(\hat{\tau}) + p\log n$$

where

$$EN(\tau) = -\sum_{i=1}^{n}\sum_{a=1}^{A}\sum_{b=1}^{B}\tau_{iab}\log\tau_{iab}$$

and the number of parameters is given by

$$p = \dim(\alpha) + \dim(\beta) + (A-1) + A(B-1).$$

Entropy R^2 is another model selection criterion that is often used in practice. Let entropy

be the entropy of fuzzy classification matrix $C = ((\tau_{ij}))$, A be the number of classes of covariates, B be the number of classes of feature variables and n be the sample size. Then we can derive the following equation

$$\begin{aligned} R^2 &= 1 - \frac{\text{Entropy}}{\text{Highest Possible Entropy}} \\ &= 1 - \frac{\text{Entropy}}{-2\sum_i \sum_a \sum_b \frac{1}{AB} \log \frac{1}{AB}} \\ &= 1 - \frac{\text{Entropy}}{2 * n \log(AB)} \end{aligned}$$

where ideally, $R^2 \geq 0.80$.

4.2.5 Analysis of Underlying Subpopulations

Figure 4.1 visualizes derivation of methods in compound LCA. From the latent variable Z_i, we can derive the prevalence of classes of covariates, $\pi(a)$, from $a = 1, ..., A$. From covariates $x_{i1}..., x_{iK}$, we can derive latent class specific means $\hat{\alpha}_{a1}.., \hat{\alpha}_{aK}$ for $a = 1, ..., A$.

To conduct dimension reduction in one step, we replace the relative frequency model $\pi(b|x_i)$ with lower dimension $\pi(b|a)$. Then for a fixed value of a, we can find the prevalence of feature class given covariate class, $\pi(b|a)$, ranging from $b = 1, .., B$. From feature variables $y_{i1}..., y_{iJ}$, we can derive latent class specific means $\hat{\beta}_{b1}.., \hat{\beta}_{bJ}$ for $b = 1, ..., B$.

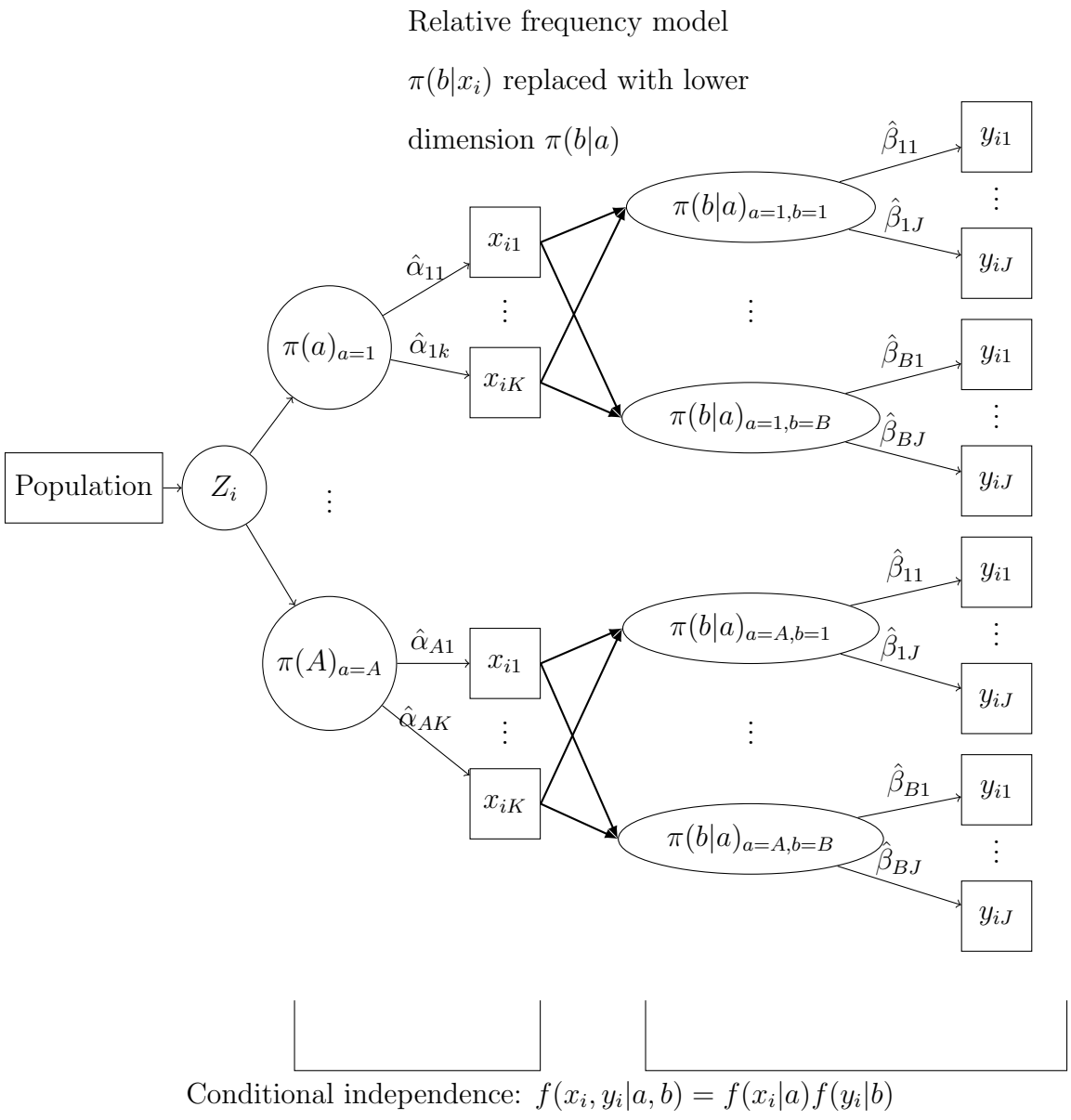

Figure 4.1: Compound LCA - Analysis of Underlying Subpopulation

4.3 Results

A simulation study was conducted to evaluate the performance of our model with high-dimensional and possibly correlated covariates.

4.3.1 Simulation Studies

Let a be the classes of covariates and b be the classes of feature variables. A sample with a size of 600 was randomly generated from 2 classes of 6 covariates with the following distribution,

Table 4.1: Data Generation - Covariates

	Class 1	Class 2
x_1	$N(0,1)$	$N(-1,1)$
x_2	$N(0.5,1)$	$N(-0.5,1)$
x_3	Pois(1)	Pois(3)
x_4	Pois(2.5)	Pois(4)
x_5	Bern(0.3)	Bern(0.45)
x_6	Bern(0.5)	Bern(0.25)

where class 1 has a prevalence of 40% and class 2 has a prevalence of 60%. We then generated 2 classes of 4 feature variables with the following distribution where if $a = 1$, we generate the following 4 feature variables

Table 4.2: Data Generation - Feature Variables ($a = 1$)

	Class 1	Class 2
y_1	$N(2,1)$	$N(-1,1)$
y_2	$N(0.5,1)$	$N(-2,1)$
y_3	Pois(4)	Pois(2)
y_4	Bern(0.4)	Bern(0.6)

where class 1 has a prevalence of 2/3% and class 2 has a prevalence of 1/3%. When $a = 2$, we again generate 4 feature variables

Table 4.3: Data Generation - Feature Variables ($a = 2$)

	Class 1	Class 2
y_1	$N(2,1)$	$N(-1,1)$
y_2	$N(0.5,1)$	$N(-2,1)$
y_3	Pois(4)	Pois(2)
y_4	Bern(0.4)	Bern(0.6)

where class 1 has a prevalence of 1/3% and class 2 has a prevalence of 2/3%.

Then we define estimators of $\hat{\pi}(a)$, $\hat{\pi}(b|a)$, $\hat{\alpha}_{ak}$ and $\hat{\beta}_{bj}$ as $I^{(1)}, I^{(2)}, I^{(3)}$ and $I^{(4)}$ and conduct a simulation study, where we

1. Generate independent draws of $x_1, x_2, x_3, x_4, x_5, x_6$ and y_1, y_2, y_3, y_4
2. Compute estimators $I^{(1)}, I^{(2)}, I^{(3)}$ and $I^{(4)}$
3. Repeat n times and obtain $I_1^{(1)}, ..., I_n^{(1)}$, $I_1^{(2)}, ..., I_n^{(2)}$, $I_1^{(3)}, ..., I_n^{(3)}$ and $I_1^{(4)}, ..., I_n^{(4)}$
4. For different estimators, compute the bias and standard errors where

$$\text{Bias}_{\pi(a)} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(1)} - \hat{\pi}(a)), \quad \text{Standard Error}_{\pi(a)} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(1)} - \hat{\pi}(a))^2$$

$$\text{Bias}_{\pi(b|a)} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(2)} - \hat{\pi}(b|a)), \quad \text{Standard Error}_{\pi(b|a)} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(2)} - \hat{\pi}(b|a))^2$$

$$\text{Bias}_{\hat{\alpha}_{ak}} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(3)} - \hat{\alpha}_{ak}), \quad \text{Standard Error}_{\hat{\alpha}_{ak}} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(3)} - \hat{\alpha}_{ak})^2$$

$$\text{Bias}_{\hat{\beta}_{bj}} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(4)} - \hat{\beta}_{bj}), \quad \text{Standard Error}_{\hat{\beta}_{bj}} = \frac{1}{n}\sum_{i=1}^{n}(I_i^{(4)} - \hat{\beta}_{bj})^2$$

where n is the number of replicates, 500 in this study.

4.3.1.1 Simulation Results: Sample Size=600

Using a sample size of 600, we computed the bias and standard error of estimators of $\hat{\pi}(a)$, $\hat{\pi}(b|a)$, $\hat{\alpha}_{ak}$ and $\hat{\beta}_{bj}$. We additionally computed average of analytical standard error, which is derived from the observed Fisher's information matrix. The estimators had small bias

and standard error values, although average of analytical standard errors tended to be larger than standard errors.

The standard error values of estimates of $\hat{\alpha}_{15}, \hat{\alpha}_{16}, \hat{\alpha}_{25}$ and $\hat{\alpha}_{26}$ were much smaller than average of analytical standard errors (Table 4.6), while the bias and standard error values of estimate of $\hat{\beta}_{24}$ were higher than that of other feature variables (Table 4.7).

Table 4.4: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\pi}(a)$

a	Bias	Standard Error	Average of Analytical Standard Error
1	0.0043	0.0013	0.0289
2	-0.0043	0.0013	0.0417

Table 4.5: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\pi}(b|a)$

b	Bias	Standard Error	Average of Analytical Standard Error
1	0.0004	0.0025	0.0313
2	-0.0004	0.0025	0.0358

Table 4.6: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\alpha}_{ak}$

$\hat{\alpha}_{ak}$	a	Bias	Standard Error	Average of Analytical Standard Error
$\hat{\alpha}_{11}$	1	-0.0062	0.0068	0.0323
$\hat{\alpha}_{12}$	1	0.0039	0.0076	0.032
$\hat{\alpha}_{13}$	1	0.0069	0.0086	0.0372
$\hat{\alpha}_{14}$	1	-0.0078	0.0158	0.0207
$\hat{\alpha}_{15}$	1	0.0022	0.0013	0.0673
$\hat{\alpha}_{16}$	1	0.0012	0.0015	0.0617
$\hat{\alpha}_{21}$	2	0.0019	0.004	0.0253
$\hat{\alpha}_{22}$	2	-0.009	0.0041	0.0255
$\hat{\alpha}_{23}$	2	0.0067	0.0143	0.0154
$\hat{\alpha}_{24}$	2	0.0092	0.0145	0.0123
$\hat{\alpha}_{25}$	2	0.0014	0.0009	0.0479
$\hat{\alpha}_{26}$	2	-0.0008	0.0008	0.0581

Table 4.7: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\beta}_{bj}$

$\hat{\beta}_{bj}$	b	Bias	Standard Error	Average of Analytical Standard Error
$\hat{\beta}_{11}$	1	0.0001	0.0044	0.0256
$\hat{\beta}_{12}$	1	0.00009	0.0042	0.0254
$\hat{\beta}_{13}$	1	0.0038	0.0143	0.0125
$\hat{\beta}_{14}$	1	-0.0003	0.0009	0.0502
$\hat{\beta}_{21}$	2	-0.001	0.0037	0.0238
$\hat{\beta}_{22}$	2	-0.0003	0.0036	0.0239
$\hat{\beta}_{23}$	2	0.0053	0.0075	0.0165
$\hat{\beta}_{24}$	2	-0.1018	0.0112	0.0463

4.3.1.2 Simulation Results: Sample Size=2000

Using a sample size of 2000, we computed the bias, standard error and average of analytical standard error of estimators of $\hat{\pi}(a)$, $\hat{\pi}(b|a)$, $\hat{\alpha}_{ak}$ and $\hat{\beta}_{bj}$.

For a larger sample size, the estimators had even smaller values of bias, standard error, and average of analytical standard error (Table 4.8-4.11). The standard error values of estimates of $\hat{\alpha}_{15}, \hat{\alpha}_{16}, \hat{\alpha}_{25}$ and $\hat{\alpha}_{26}$ were still smaller than average of analytical standard errors (Table 4.10), and the bias and standard error values of estimate of $\hat{\beta}_{24}$ were higher than that of other feature variables (Table 4.11).

Table 4.8: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\pi}(a)$

a	Bias	Standard Error	Average of Analytical Standard Error
1	-0.0002	0.0004	0.0085
2	0.0002	0.0004	0.0126

Table 4.9: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\pi}(b|a)$

b	Bias	Standard Error	Average of Analytical Standard Error
1	0.0009	0.0008	0.0093
2	-0.0009	0.0008	0.0107

Table 4.10: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\alpha}_{ak}$

$\hat{\alpha}_{ak}$	a	Bias	Standard Error	Average of Analytical Standard Error
$\hat{\alpha}_{11}$	1	0.0036	0.0022	0.01
$\hat{\alpha}_{12}$	1	0.0023	0.0023	0.0099
$\hat{\alpha}_{13}$	1	-0.0024	0.0026	0.0112
$\hat{\alpha}_{14}$	1	-0.0023	0.0053	0.0062
$\hat{\alpha}_{15}$	1	-0.0002	0.0004	0.0204
$\hat{\alpha}_{16}$	1	0.0008	0.0004	0.0185
$\hat{\alpha}_{21}$	2	0.0003	0.0012	0.0075
$\hat{\alpha}_{22}$	2	0.0013	0.0013	0.0075
$\hat{\alpha}_{23}$	2	0.0016	0.004	0.0046
$\hat{\alpha}_{24}$	2	-0.002	0.0047	0.0037
$\hat{\alpha}_{25}$	2	0.0001	0.0003	0.014
$\hat{\alpha}_{26}$	2	-0.0002	0.0002	0.0172

Table 4.11: Bias, Empirical Standard Error and Average of Analytical Standard Error of $\hat{\beta}_{bj}$

$\hat{\beta}_{bj}$	b	Bias	Standard Error	Average of Analytical Standard Error
$\hat{\beta}_{11}$	1	-0.0001	0.0013	0.0077
$\hat{\beta}_{12}$	1	-0.0025	0.0011	0.0076
$\hat{\beta}_{13}$	1	0.0019	0.0044	0.0038
$\hat{\beta}_{14}$	1	-0.0007	0.0003	0.0152
$\hat{\beta}_{21}$	2	-0.0018	0.001	0.0071
$\hat{\beta}_{22}$	2	0.0002	0.001	0.0071
$\hat{\beta}_{23}$	2	-0.0022	0.0019	0.005
$\hat{\beta}_{24}$	2	-0.1005	0.0103	0.0139

4.3.2 MCI Dataset

4.3.2.1 Study Sample

Individuals from the Uniform Data Set (UDS) of the National Alzheimer's Coordinating Center (NACC), which is a longitudinal study that includes patients who have dementia, mild cognitive impairment and who are cognitively normal (National Alzheimer's Coordinating Center, 2021b), were included in the analysis. Genetic data was available for a limited NACC participants, where more than 75% of patients had records of APOE $\varepsilon 4$ allele (National Alzheimer's Coordinating Center, 2021a). We focused on a sample of 6034 participants as of June 2015 freeze date and included standardized evaluations of functional abilities, neuropsychiatric symptoms and assessments of cognitions from the motivating example as manifest variables (Section 1.2). We included vascular risk factors, demographic characteristics, and APOE $\varepsilon 4$ carrier status as covariates.

4.3.2.2 Vascular Risk Factors

The Rosen Modification of Hachinski Ischemic Score (RMHIS) was used to assess cerebrovascular disease status of participants, which is a scale modified from Hachinski Ischemic Score to include 8 features that would increase the accuracy of diagnosis of multi infarct dementia (MID), a vascular disorder (Rosen et al., 1980). Additional risk factors of cerebrovascular disease such as diabetic status, hypercholesterolemia and hypertension were included. Both coronary vascular disease and cardiac dysrhythmia were dichotomized to indicate diagnosis

of corresponding disease for each participant.

4.3.2.3 Demographic Characteristics

Age was measured in decades at baseline, centered at age 70. Race was dichotomized to indicate whether a participant is African-American or not. Education was measured in years. Gender was additionally included as a covariate.

4.3.2.4 APOE

APOE $\varepsilon 4$ carrier status was measured by whether a person is APOE $\varepsilon 4$ positive or negative. APOE data was missing for about 20% of participants.

4.3.2.5 Analysis

We analyzed the UDS data by applying the methods derived from compound LCA. Different models were fit with the number of classes of covariates ranging from 1 to 4 classes and the the number of classes of feature variables ranging from 2 to 4 classes (Table 4.12-Table 4.15). The ICL-BIC model selection criterion (ICL-BIC=42385.99, Table 4.15), entropy R^2 value (R^2 =0.84, Table 4.15) and analysis of clinical interpretation of the latent classes indicated that the model with 3 classes of covariates ($A = 1, 2, 3$) and 3 classes of feature variables ($B = 1, 2, 3$) is the best fitting model for the dataset. Using this solution, the latent class solution of covariates revealed that the dataset consists of (Table 4.12):

1. $A = 1$: African-American patients, older, more females, with high risk of cholesterol, diabetes and hypertension

2. $A = 2$: Non-African American patients, younger, with low vascular risk and high prevalence of APOE

3. $A = 3$: Non-African American patients, older, more males, high vascular risk and relatively high prevalence of APOE.

The latent class solution of features revealed that different types of patients from the covariate space can be diagnosed into following subtypes of MCI (Table 4.16):

1. $B = 1$: Mildly impaired
2. $B = 2$: Amnestic multi-domain with functional impairment and neuropsychiatric features
3. $B = 3$: Amnestic with functional impairment and neuropsychiatric features.

Combining the solutions of covariates and feature variables, we can interpret relative frequencies of latent classes by focusing on the prevalence of feature class given covariate class (Table 4.13). Among patients in the first covariate class, or who are African-American patients, older, more females, with high risk of cholesterol, diabetes and hypertension, the prevalence of feature classes show: the probability of being diagnosed with mild impairment is 0.24, the probability of being diagnosed with amnestic multi-domain with functional impairment and neuropsychiatric features is 0.76 and the probability of being diagnosed with amnestic with functional impairment and neuropsychiatric features is 0.

The prevalence of feature classes with respect to the next two covariate classes can be interpreted in a similar fashion. Among patients in the second covariate class, or patients who are non-African American patients, younger, with low vascular risk and high prevalence of APOE, the prevalence of feature classes indicate: the probability of being diagnosed with mild impairment is 0.32, the probability of being diagnosed with amnestic multi-domain with functional impairment and neuropsychiatric features is 0.07 and the probability of being diagnosed with amnestic with functional impairment and neuropsychiatric features is 0.61. Among patients in the third covariate class, or patients who are non-African American patients, older, more males, high vascular risk and relatively high prevalence of APOE, the prevalence of feature classes reveal: the probability of being diagnosed with mild impairment is 0.26, the probability of being diagnosed with amnestic multi-domain with functional impairment and neuropsychiatric features is 0.05 and the probability of being diagnosed with

amnestic with functional impairment and neuropsychiatric features is 0.69.

Even though we assumed conditional independence between the covariates and responses given the latent classes, there exists some evidence that feature classes may be nested within covariate classes. Using the posterior probabilities of membership in the two sets of latent classes derived from 3 classes of covariates ($A = 1, 2, 3$) and 3 classes of feature variables ($B = 1, 2, 3$) by using compound LCA, we applied a modal method on both covariate and feature classes. Then we incorporated the results from using a modal method into linear and logistic regression models, which indicated that the feature classes are nested within the covariate classes (Table 4.17).

Table 4.12: Latent Class-Specific Means (and Standard Errors) of Covariates

	Class	1 Class of Covariates, 2 Classes of Feature Variables	1 Class of Covariates, 3 Classes of Feature Variables	1 Class of Covariates, 4 Classes of Feature Variables	2 Classes of Covariates, 2 Classes of Feature Variables	2 Classes of Covariates, 3 Classes of Feature Variables	2 Classes of Covariates, 4 Classes of Feature Variables
Hachinski Ischaemia Score	1	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.04 (0.02)	0.09 (0.01)	0.04 (0.02)
	2	—	—	—	0.09 (0.01)	0.04 (0.02)	0.09 (0.02)
Hypercholesterolemia	1	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.36 (0.01)	0.77 (0.02)	0.34 (0.01)
	2	—	—	—	0.76 (0.01)	0.35 (0.01)	0.78 (0.02)
Diabetic Status	1	0.15 (0.01)	0.15 (0.01)	0.15 (0.01)	0.03 (0.03)	0.28 (0.01)	0.03 (0.03)
	2	—	—	—	0.28 (0.01)	0.03 (0.03)	0.28 (0.01)
Hypertension	1	0.56 (<0.01)	0.56 (0.01)	0.56 (0.01)	0.31 (0.01)	0.83 (0.02)	0.31 (0.01)
	2	—	—	—	0.84 (0.02)	0.31 (0.01)	0.83 (0.02)
Education	1	15 (0.01)	15 (0.01)	15 (0.01)	16 (0.01)	15 (0.02)	16 (0.02)
	2	—	—	—	14 (0.02)	16 (0.01)	15 (0.02)
Age	1	0.40 (<0.01)	0.40 (<0.01)	0.40 (<0.01)	0.24 (<0.01)	0.57 (0.01)	0.24 (0.01)
	2	—	—	—	0.57 (<0.01)	0.24 (0.01)	0.57 (0.01)
Gender	1	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.54 (0.01)	0.46 (0.01)	0.54 (0.01)
	2	—	—	—	0.46 (0.01)	0.54 (0.01)	0.46 (0.01)
Race	1	0.16 (0.01)	0.16 (0.01)	0.16 (0.01)	0.09 (0.02)	0.22 (0.01)	0.10 (0.02)
	2	—	—	—	0.23 (0.01)	0.09 (0.02)	0.22 (0.01)
Cardiac Dysrhythmia	1	0.10 (0.01)	0.10 (0.01)	0.10 (0.01)	0.05 (0.02)	0.15 (0.01)	0.05 (0.02)
	2	—	—	—	0.15 (0.01)	0.05 (0.02)	0.15 (0.01)
Coronary Vascular Disease	1	0.14 (0.01)	0.14 (0.01)	0.14 (0.01)	0.02 (0.05)	0.29 (0.01)	0.01 (0.01)
	2	—	—	—	0.28 (0.01)	0.02 (0.07)	0.29 (0.01)
APOE	1	0.41 (<0.01)	0.41 (0.01)	0.41 (0.01)	0.43 (0.01)	0.40 (0.01)	0.43 (0.01)
	2	—	—	—	0.40 (0.01)	0.43 (0.01)	0.40 (0.01)

	Class	3 Classes of Covariates, 2 Classes of Feature Variables	3 Classes of Covariates, 3 Classes of Feature Variables	3 Classes of Covariates, 4 Classes of Feature Variables	4 Classes of Covariates, 2 Classes of Feature Variables	4 Classes of Covariates, 3 Classes of Feature Variables	4 Classes of Covariates, 4 Classes of Feature Variables
Hachinski Ischaemia Score	1	0.10 (0.02)	0.08 (0.03)	0.04 (0.02)	0.04 (0.02)	0.03 (1.53)	0.06 (0.03)
	2	0.08 (0.03)	0.04 (0.02)	0.07 (0.03)	0.20 (0.02)	0.05 (0.03)	0.10 (0.11)
	3	0.04 (0.02)	0.10 (0.03)	0.10 (0.03)	0.09 (0.04)	0.07 (0.03)	0.07 (0.04)
	4	—	—	—	0.08 (0.03)	0.12 (0.03)	0.04 (0.03)
Hypercholesterolemia	1	0.82 (0.02)	0.64 (0.02)	0.34 (0.03)	0.04 (0.01)	0.52 (0.02)	0.25 (0.03)
	2	0.65 (0.01)	0.35 (0.01)	0.67 (0.02)	0.84 (0.03)	0.29 (0.03)	0.88 (0.02)
	3	0.35 (0.01)	0.82 (0.03)	0.81 (0.01)	0.34 (0.03)	0.71 (0.02)	0.67 (0.02)
	4	—	—	—	0.74 (0.02)	0.82 (0.04)	0.38 (0.02)
Diabetic Status	1	0.21 (0.01)	0.31 (0.02)	0.03 (0.02)	0.04 (0.03)	0.08 (0.03)	0.02 (0.02)
	2	0.33 (0.02)	0.03 (0.04)	0.34 (0.02)	0.20 (0.02)	0.02 (0.02)	0.22 (0.04)
	3	0.04 (0.03)	0.20 (0.02)	0.19 (0.04)	0.09 (0.05)	0.37 (0.02)	0.35 (0.10)
	4	—	—	—	0.41 (0.02)	0.19 (0.12)	0.04 (0.02)
Hypertension	1	0.76 (0.02)	0.85 (0.03)	0.29 (0.02)	0.29 (0.01)	0.37 (0.02)	0.50 (0.02)
	2	0.87 (0.03)	0.29 (0.01)	0.89 (0.04)	0.76 (0.02)	0.30 (0.02)	0.78 (0.02)
	3	0.30 (0.01)	0.76 (0.02)	0.75 (0.01)	0.62 (0.03)	0.91 (0.04)	0.89 (0.02)
	4	—	—	—	0.95 (0.07)	0.77 (0.03)	0.26 (0.05)
Education	1	16 (0.02)	13 (0.03)	16 (0.02)	16 (0.01)	18 (0.02)	15 (0.03)
	2	13 (0.03)	16 (0.02)	13 (0.03)	16 (0.02)	15 (0.02)	16 (0.02)
	3	16 (0.01)	16 (0.03)	16 (0.02)	13 (0.08)	13 (0.03)	13 (0.03)
	4	—	—	—	13 (0.03)	16 (0.04)	16 (0.03)
Age	1	0.68 (0.01)	0.47 (0.01)	0.23 (0.01)	0.12 (0.01)	0.25 (0.01)	1.34 (0.01)
	2	0.44 (0.01)	0.21 (0.01)	0.42 (0.01)	0.68 (0.01)	0.26 (0.01)	0.61 (0.01)
	3	0.22 (¡0.01)	0.67 (0.01)	0.68 (0.01)	1.08 (0.01)	0.38 (0.01)	0.34 (0.01)
	4	—	—	—	0.27 (0.01)	0.77 (0.01)	-0.02 (0.01)
Gender	1	0.23 (0.02)	0.72 (0.02)	0.55 (0.02)	0.54 (0.01	0.35 (0.02)	0.55 (0.03)
	2	0.72 (0.02)	0.54 (0.01)	0.72 (0.02)	0.21 (0.03)	0.64 (0.03)	0.23 (0.02)
	3	0.55 (0.01)	0.21 (0.02)	0.22 (0.01)	0.65 (0.03)	0.71 (0.02)	0.74 (0.02)
	4	—	—	—	0.72 (0.03)	0.22 (0.02)	0.53 (0.02)
Race	1	0.02 (0.10)	0.47 (0.02)	0.07 (0.10)	0.07 (0.02)	0.01 (0.05)	0.05 (0.16)
	2	0.49 (0.02)	0.06 (0.02)	0.49 (0.02)	0.02 (0.12)	0.12 (0.18)	0.02 (0.11)
	3	0.07 (0.02)	0.01 (0.14)	0.01 (0.03)	0.20 (0.03)	0.48 (0.02)	0.53 (0.03)
	4	—	—	—	0.55 (0.02)	0.02 (0.03)	0.07 (0.02)
Cardiac Dysrhythmia	1	0.21 (0.02)	0.08 (0.03)	0.05 (0.02)	0.04 (0.03)	0.06 (0.05)	0.19 (0.02)
	2	0.07 (0.03)	0.05 (0.03)	0.07 (0.03)	0.21 (0.02)	0.05 (0.02)	0.20 (0.05)
	3	0.05 (0.02)	0.21 (0.02)	0.21 (0.03)	0.16 (0.04)	0.07 (0.03)	0.06 (0.04)
	4	—	—	—	0.06(0.04)	0.24(0.04)	0.03(0.04)
Coronary Vascular Disease	1	0.42 (0.01)	0.12 (0.03)	0 (0.02)	0 (0.19)	0.02 (0.17)	0.08 (0.03)
	2	0.12 (0.03)	0 (0.35)	0.13 (0.03)	0.44 (0.02)	0 (0.02)	0.42 (0.02)
	3	0 (0.22)	0.42 (0.02)	0.41 (0.33)	0.06 (0.06)	0.13 (0.03)	0.12 (0.74)
	4	—	—	—	0.15 (0.03)	0.50 (0.47)	0 (0.03)
APOE	1	0.40 (0.02)	0.39 (0.02)	0.42 (0.02)	0.46 (0.01)	0.53 (0.02)	0.16 (0.02)
	2	0.40 (0.02)	0.43 (0.01)	0.41 (0.02)	0.40 (0.02)	0.36 (0.02)	0.43 (0.02)
	3	0.43 (0.01)	0.41 (0.02)	0.41 (0.01)	0.17 (0.05)	0.42 (0.02)	0.42 (0.02)
	4	—	—	—	0.47 (0.02)	0.37 (0.03)	0.48 (0.02)

Table 4.13: Latent Class-Specific Means (and Standard Errors) of Feature Variables

		Class	1 Class of Covariates, 2 Classes of Feature Variables	1 Class of Covariates, 3 Classes of Feature Variables	1 Class of Covariates, 4 Classes of Feature Variables	2 Classes of Covariates, 2 Classes of Feature Variables	2 Classes of Covariates, 3 Classes of Feature Variables	2 Classes of Covariates, 4 Classes of Feature Variables
Functional	No. of IADL impaired	1	1.58 (0.01)	0.19 (¡0.01)	1.32 (0.02)	1.59 (0.01)	0.19 (<0.01)	0 (0.01)
		2	3.18 (0.01)	3.08 (0.17)	2.94(0.01)	3.17 (0.01)	3.27 (0.03)	3.65 (0.04)
		3	—	3.23 (0.07)	2.94 (0.02)	—	3.03 (0.06)	3.65 (0.04)
		4	—	—	2.94 (0.01)	—	—	3.60 (0.02)
Neuropsychiatric	% with GDS ≥ 5	1	13.44 (0.01)%	9.31 (0.02)%	10.92 (0.02)%	13.49 (0.01)%	9.32 (0.02)%	13.80 (0.01)%
		2	24.64 (0.01)%	22.72 (0.04)%	23.38 (0.02)%	24.62 (0.01)%	22.32 (0.02)%	21.35 (0.02)%
		3	—	22.26 (0.02)%	23.38 (0.02)%	—	22.55 (0.03)%	21.34 (0.04)%
		4	—	—	23.38 (0.01)%	—	—	21.24 (0.02)%
	No. of NPI-Q symptoms present	1	1.39 (0.01)	0.54 (0.01)	0 (0.02)	1.40 (0.01)	0.54 (0.01)	0.87 (0.01)
		2	2.19 (0.01)	2.16 (0.05)	2.86 (0.01)	2.18 (0.01)	2.32 (0.02)	2.27 (0.03)
		3	—	2.30 (0.02)	2.86 (0.01)	—	2.14 (0.04)	2.27 (0.03)
		4	—	—	2.86 (0.01)	—	—	2.27 (0.02)
Cognitive	Global							
	MMSE	1	-0.96 (<0.01)	-0.82 (0.01)	-1.55 (0.01)	-0.97 (0.01)	-0.83 (0.01)	-1.28 (0.01)
		2	-2.42 (0.01)	-2.19 (0.04)	-1.69 (0.01)	-2.42 (0.01)	-2.87 (0.02)	-1.83 (0.02)
		3	—	-1.87 (0.02)	-1.69 (0.01)	—	-3.27 (0.03)	-1.84 (0.03)
		4	—	—	-1.69 (0.01)	—	—	-1.86 (0.02)
	Logical Memory							
	Immediate	1	-0.94 (0.01)	-0.85 (0.01)	-1.08 (0.01)	-0.95 (0.01)	-0.85 (0.01)	-1.01 (0.01)
		2	-1.48 (0.01)	-1.48 (0.04)	-1.26 (0.02)	-1.47 (0.01)	-2.30 (0.02)	-1.30 (0.02)
		3	—	-1.27 (0.02)	-1.26 (0.01)	—	-2.39 (0.03)	-1.29 (0.02)
		4	—	—	-1.26 (0.01)	—	—	-1.28 (0.01)
	Delayed	1	-1.05 (0.01)	-0.89 (0.01)	-1.16 (0.01)	-1.06 (0.01)	-0.89 (0.01)	-1.03 (0.01)
		2	-1.56 (0.01)	-1.58 (0.04)	-1.36 (0.01)	-1.55 (0.01)	-1.43 (0.02)	-1.44 (0.02)
		3	—	-1.39 (0.02)	-1.36 (<0.01)	—	-1.49 (0.03)	-1.44 (0.02)
		4	—	—	-1.36 (0.01)	—	—	-1.41 (0.01)
	Semantic Memory							
	Category Fluency	1	-0.63 (<0.01)	-0.59 (<0.01)	-0.84 (0.01)	-0.63 (<0.01)	-0.59 (<0.01)	-0.77 (¡0.01)
		2	-1.28 (<0.01)	-1.19 (0.02)	-0.98 (0.01)	-1.28 (¡0.01)	-1.03 (0.01)	-1.02 (0.01)
		3	—	-1.01 (0.01)	-0.98 (0.01)	—	-1.13 (0.02)	-1.02 (0.01)
		4	—	—	-0.98 (0.01)	—	—	-0.99 (0.01)
	Attention							
	Trails A	1	-0.03 (<0.01)	0 (<0.01)	0.63 (0.02)	-0.03 (<0.01)	0 (0.01)	0.57 (<0.01)*
		2	1.61 (0.01)	1.12 (0.05)	0.78 (0.02)	1.62 (0.01)	0.94 (0.03)	0.81 (0.03)
		3	—	0.98 (0.02)	0.78 (0.01)	—	1.21 (0.04)	0.81 (0.04)
		4	—	—	0.78 (0.01)	—	—	0.87 (0.02)
	Digit Span Forward	1	-0.08 (<0.01)	-0.14 (0.01)	-0.28 (0.01)	-0.08 (<0.01)	-0.14 (0.01)	-0.27 (0.01)
		2	-0.57 (<0.01)	-0.45 (0.02)	-0.32 (0.01)	-0.57 (<0.01)	-0.34 (0.01)	-0.33 (0.01)
		3	—	-0.34 (0.01)	-0.32 (0.01)	—	-0.45 (0.02)	-0.33 (0.01)
		4	—	—	-0.32 (0.01)	—	—	-0.33 (0.01)
	Language							
	Boston Naming	1	-0.45 (<0.01)	-0.56 (0.01)	-1.10 (0.01)	-0.46 (<0.01)	-0.57 (0.01)	-1.05 (0.01)
		2	-1.75 (0.01)	-1.57 (0.03)	-1.02 (0.02)	-1.76 (0.01)	-1.09 (0.01)	-1.04 (0.03)
		3	—	-1.11 (0.01)	-1.02 (0.01)	—	-1.56 (0.03)	-1.05 (0.03)
		4	—	—	-1.02 (0.01)	—	—	-1.06 (0.02)
	Executive Function							
	Trails B	1	0.23 (<0.01)	0.29 (0.01)	1.14 (0.01)	0.23 (<0.01)	0.29 (0.01)	1.07 (0.01)*
		2	2.52 (0.01)	1.84 (0.04)	1.33 (0.01)	2.54 (0.01)	1.52 (0.02)	1.35 (0.03)
		3	—	1.59 (0.02)	1.33 (0.01)	—	1.95 (0.03)	1.36 (0.02)
		4	—	—	1.33 (0.01)	—	—	1.42 (0.02)
	Digit Span Backward	1	-0.22 (<0.01)	-0.28 (0.01)	-0.45 (0.01)	-0.22 (<0.01)	-0.28 (0.01)	-0.46 (0.01)
		2	-0.79 (<0.01)	-0.65 (0.02)	-0.50 (0.01)	-0.80 (<0.01)	-0.52 (0.01)	-0.50 (0.01)
		3	—	-0.53 (0.01)	-0.50 (0.01)	—	-0.66 (0.02)	-0.50 (0.01)
		4	—	—	-0.50 (0.01)	—	—	-0.51 (0.01)
	Visuomotor							
	Digit Symbol	1	-0.32 (<0.01)	-0.34 (0.01)	-0.68 (0.01)	-0.32 (<0.01)	-0.34 (0.01)	-0.66 (0.01)
		2	-1.41 (<0.01)	-1.12 (0.03)	-0.90 (0.01)	-1.41 (0.01)	-0.95 (0.01)	-0.90 (0.01)
		3	—	-0.97 (0.01)	-0.90 (0.01)	—	-1.15 (0.02)	-0.91 (0.01)
		4	—	—	-0.90 (0.01)	—	—	-0.93 (0.01)

		Class	3 Classes of Covariates, 2 Classes of Feature Variables	3 Classes of Covariates, 3 Classes of Feature Variables	3 Classes of Covariates, 4 Classes of Feature Variables	4 Classes of Covariates, 2 Classes of Feature Variables	4 Classes of Covariates, 3 Classes of Feature Variables	4 Classes of Covariates, 4 Classes of Feature Variables
Functional	No. of IADL impaired	1	1.63 (0.01)	0.19 (<0.01)	0 (0.01)	1.64 (0.01)	0 (0.01)	0 (<0.01)
		2	3.14 (0.02)	2.85 (0.03)	3.66 (0.04)	3.13 (0.02)	3.61 (0.03)	3.56 (0.05)
		3	—	3.33 (0.02)	3.57 (0.03)	—	3.66 (0.02)	3.63 (0.04)
		4	—	—	3.66 (0.04)	—	—	3.66 (0.04)
Neuropsychiatric	% with GDS ≥ 5	1	13.53 (0.01)%	9.37 (0.02)%	13.80 (0.01)%	13.60 (0.01)%	13.80 (0.02)%	13.80 (0.03)%
		2	24.72 (0.01)%	24.71 (0.02)%	20.58 (0.03)%	24.62 (0.01)%	22.62 (0.02)%	22.84 (0.04)%
		3	—	21.00 (0.01)%	23.80 (0.02)%	—	20.68 (0.02)%	21.55 (0.03)%
		4	—	—	20.91 (0.03)%	—	—	21.00 (0.04)%
	No. of NPI-Q symptoms present	1	1.42 (0.01)	0.55 (0.01)	0.87 (0.01)	1.43 (0.01)	0.87 (0.01)	0.87 (0.01)
		2	2.17 (0.01)	2.01 (0.02)	2.29 (0.02)	2.16 (0.01)	2.25 (0.02)	2.23 (0.04)
		3	—	2.36 (0.01)	2.21 (0.02)	—	2.28 (0.02)	2.28 (0.03)
		4	—	—	2.28 (0.03)	—	—	2.28 (0.04)
Cognitive	Global							
	MMSE	1	-0.98 (0.01)	-0.80 (0.01)	-1.28 (0.01)	-0.98 (0.01)	-1.28 (0.01)	-1.28 (0.01)
		2	-2.42 (0.01)	-2.04 (0.02)	-1.83 (0.02)	-2.42 (0.01)	-1.88 (0.02)	-1.88 (0.03)
		3	—	-1.93 (0.01)	-1.90 (0.01)	—	-1.82 (0.01)	-1.84 (0.03)
		4	—	—	-1.83 (0.03)	—	—	-1.83 (0.03)
	Logical Memory							
	Immediate	1	-0.96 (0.01)	-0.84 (0.01)	-1.01 (0.01)	-0.97 (0.01)	-1.01 (0.01)	-1.01 (0.01)
		2	-1.46 (0.01)	-1.08 (0.02)	-1.33 (0.02)	-1.46 (0.01)	-1.23 (0.01)	-1.21 (0.03)
		3	—	-1.46 (0.01)	-1.17 (0.01)	—	-1.32 (0.01)	-1.28 (0.02)
		4	—	—	-1.31 (0.02)	—	—	-1.31 (0.03)
	Delayed	1	-1.07 (0.01)	-0.88 (0.01)	-1.03 (0.01)	-1.08 (0.01)	-1.03 (0.01)	-1.03 (0.02)
		2	-1.54 (0.01)	-1.18 (0.02)	-1.47 (0.02)	-1.54(0.01)	-1.36 (0.01)	-1.34 (0.03)
		3	—	-1.58 (0.01)	-1.30 (0.01)	—	-1.47 (0.01)	-1.42 (0.02)
		4	—	—	-1.45 (0.02)	—	—	-1.45 (0.03)
	Semantic Memory							
	Category Fluency	1	-0.64 (<0.01)	-0.58 (0.01)	-0.77 (<0.01)	-0.64 (<0.01)	-0.77 (<0.01)	-0.77 (0.01)
		2	-1.28 (<0.01)	-0.89 (0.01)	-1.03 (0.01)	-1.27 (0.01)	-0.94 (0.01)	-0.93 (0.02)
		3	—	-1.15 (0.01)	-0.88 (0.01)	—	-1.05 (0.01)	-1.01 (0.01)
		4	—	—	-1.04 (0.01)	—	—	-1.03 (0.02)
	Attention							
	Trails A	1	-0.02 (<0.01)	-0.01 (0.01)	0.57 (<0.01)	-0.03 (<0.01)	0.57 (<0.01)	0.57 (0.01)*
		2	1.64 (0.01)	1.35 (0.02)	0.78 (0.03)	1.64 (0.01)	0.90 (0.02)	0.96 (0.04)
		3	—	0.86 (0.01)	0.99 (0.02)	—	0.78 (0.02)	0.83 (0.02)
		4	—	—	0.79 (0.04)	—	—	0.79 (0.03)
	Digit Span Forward	1	-0.08 (<0.01)	-0.14 (0.01)	-0.27 (0.01)	-0.09 (<0.01)	-0.27 (0.01)	-0.27 (0.01)
		2	-0.57 (<0.01)	-0.38 (0.01)	-0.33 (0.01)	-0.57 (0.01)	-0.32 (0.01)	-0.33 (0.02)
		3	—	-0.37 (0.01)	-0.33 (0.01)	—	-0.33 (0.01)	-0.33 (0.01)
		4	—	—	-0.33 (0.02)	—	—	-0.33 (0.02)
	Language							
	Boston Naming	1	-0.45 (0.01)	-0.55 (0.01)	-1.05 (0.01)	-0.45 (0.01)	-1.05 (0.01)	-1.05 (0.01)
		2	-1.78 (0.01)	-1.75 (0.02)	-0.98 (0.03)	-1.78 (0.01)	-1.15 (0.02)	-1.22 (0.03)
		3	—	-0.99 (0.01)	-1.30 (0.02)	—	-1.00 (0.02)	-1.06 (0.03)
		4	—	—	-1.00 (0.03)	—	—	-1.02 (0.03)
	Executive Function							
	Trails B	1	0.22 (<0.01)	0.26 (0.01)	1.07 (0.01)	0.22 (<0.01)	1.07 (0.01)	1.07 (0.01)*
		2	2.57 (0.01)	2.14 (0.02)	1.31 (0.02)	2.58 (0.01)	1.47 (0.02)	1.53 (0.04)
		3	—	1.43 (0.01)	1.59 (0.02)	—	1.32 (0.02)	1.38 (0.03)
		4	—	—	1.33 (0.03)	—	—	1.34 (0.03)
	Digit Span Backward	1	-0.23 (<0.01)	-0.27 (0.01)	-0.46 (0.01)	-0.23 (<0.01)	-0.46 (0.01)	-0.46 (0.01)
		2	-0.80 (<0.01)	-0.56 (0.01)	-0.51 (0.01)	-0.79 (0.01)	-0.49 (0.01)	-0.51 (0.02)
		3	—	-0.57 (0.01)	-0.49 (0.01)	—	-0.50 (0.01)	-0.51 (0.01)
		4	—	—	-0.50 (0.01)	—	—	-0.50 (0.02)
	Visuomotor							
	Digit Symbol	1	-0.33 (<0.01)	-0.33 (0.01)	-0.66 (0.01)	-0.33 (<0.01)	-0.66 (0.01)	-0.66 (0.01)
		2	-1.42 (0.01)	-1.09 (0.01)	-0.90 (0.01)	-1.42 (0.01)	-0.93 (0.01)	-0.95 (0.02)
		3	—	-0.97 (0.01)	-0.94 (0.01)	—	-1.47 (0.01)	-0.92 (0.02)
		4	—	—	-0.90 (0.01)	—	—	-0.90 (0.02)

* Higher scores on Trail A and Trail B indicate worse performance.

Table 4.14: Relative Frequencies of Latent Classes

	a	b	$\pi(a)$	$\pi(b\|a)$
1 Class of Covariates,	1	1	—	0.54
2 Classes of Feature Variables	1	2	—	0.46
1 Class of Covariates,	1	1	—	0.29
3 Classes of Feature Variables	1	2	—	0.22
	1	3	—	0.49
1 Class of Covariates,	1	1	—	0.39
4 Classes of Feature Variables	1	2	—	0.14
	1	3	—	0.09
	1	4	—	0.38
2 Classes of Covariates,	1	1	0.52	0.61
2 Classes of Feature Variables	1	2	0.52	0.39
	2	1	0.48	0.46
	2	2	0.48	0.54
2 Classes of Covariates,	1	1	0.48	0.25
3 Classes of Feature Variables	1	2	0.48	0.40
	1	3	0.48	0.35
	2	1	0.52	0.33
	2	2	0.52	0.53
	2	3	0.52	0.14
2 Classes of Covariates,	1	1	0.52	0.38
4 Classes of Feature Variables	1	2	0.52	0.43
	1	3	0.52	0.13
	1	4	0.52	0.06
	2	1	0.48	0.35
	2	2	0.48	0.35
	2	3	0.48	0.12
	2	4	0.48	0.18
3 Classes of Covariates,	1	1	0.27	0.58
2 Classes of Feature Variables	1	2	0.27	0.42
	2	1	0.24	0.36
	2	2	0.24	0.63
	3	1	0.49	0.62
	3	2	0.49	0.38
3 Classes of Covariates,	1	1	0.26	0.24
3 Classes of Feature Variables	1	2	0.26	0.76
	1	3	0.26	0
	2	1	0.48	0.32
	2	2	0.48	0.07
	2	3	0.48	0.61
	3	1	0.26	0.26
	3	2	0.26	0.05
	3	3	0.26	0.69
3 Classes of Covariates,	1	1	0.48	0.37
4 Classes of Feature Variables	1	2	0.48	0.13
	1	3	0.48	0.04
	1	4	0.48	0.46
	2	1	0.25	0.43
	2	2	0.25	0.04
	2	3	0.25	0.30
	2	4	0.25	0.23
	3	1	0.27	0.29
	3	2	0.27	0.17
	3	3	0.27	0.03
	3	4	0.27	0.51
4 Classes of Covariates,	1	1	0.46	0.66
2 Classes of Feature Variables	1	2	0.46	0.34
	2	1	0.25	0.58
	2	2	0.25	0.42
	3	1	0.12	0.16
	3	2	0.12	0.84
	4	1	0.17	0.43
	5	2	0.17	0.57
4 Classes of Covariates,	1	1	0.21	0.33
3 Classes of Feature Variables	1	2	0.21	0.06
	1	3	0.21	0.61
	2	1	0.34	0.39
	2	2	0.34	0.18
	2	3	0.34	0.43
	3	1	0.23	0.42
	3	2	0.23	0.39
	3	3	0.23	0.19
	4	1	0.22	0.29
	4	2	0.22	0.21
	4	3	0.22	0.50
4 Classes of Covariates,	1	1	0.13	0.35
4 Classes of Feature Variables	1	2	0.13	0.03
	1	3	0.13	0.11
	1	4	0.13	0.51
	2	1	0.25	0.28
	2	2	0.25	0.07
	2	3	0.25	0.17
	2	4	0.25	0.48
	3	1	0.22	0.45
	3	2	0.22	0.17
	3	3	0.22	0.16
	3	4	0.22	0.22
	4	1	0.40	0.37
	4	2	0.40	0.03
	4	3	0.40	0.14
	4	4	0.40	0.46

Table 4.15: Model Selection

	1 Class of Covariates, 2 Classes of Feature Variables	1 Class of Covariates, 3 Classes of Feature Variables	1 Class of Covariates, 4 Classes of Feature Variables
Log-Likelihood	-77803.83	-18153.48	48327.67
BIC	156034.21	36951.15	-95793.53
Entropy	905.90	2973.94	3424.30
Entropy R^2	0.89	0.78	0.80
ICL-BIC	157846.02	42897.03	-88944.94
	2 Classes of Covariates, 2 Classes of Feature Variables	**2 Classes of Covariates, 3 Classes of Feature Variables**	**2 Classes of Covariates, 4 Classes of Feature Variables**
Log-Likelihood	-76781.58	-17070.85	66481.02
BIC	154424.96	35438.77	-131229.71
Entropy	2964.99	4977.68	5528.99
Entropy R^2	0.82	0.77	0.78
ICL-BIC	160354.94	45394.13	-120171.73
	3 Classes of Covariates, 2 Classes of Feature Variables	**3 Classes of Covariates, 3 Classes of Feature Variables**	**3 Classes of Covariates, 4 Classes of Feature Variables**
Log-Likelihood	-75904.07	-15965.70	66942.91
BIC	153105.22	33881.35	-131282.97
Entropy	3778.82	4252.32	5890.31
Entropy R^2	0.83	0.84	0.80
ICL-BIC	160664.87	42385.99	-119502.35
	4 Classes of Covariates, 2 Classes of Feature Variables	**4 Classes of Covariates, 3 Classes of Feature Variables**	**4 Classes of Covariates, 4 Classes of Feature Variables**
Log-Likelihood	-75442.85	67104.46	67102.22
BIC	152618.03	-131606.08	-130731.07
Entropy	4330.61	6080.12	6705.27
Entropy R^2	0.83	0.80	0.80
ICL-BIC	161279.26	-119445.84	-117320.54

Table 4.16: Interpretation of Relative Frequencies of Latent Classes for 3 Classes of Covariates and 3 Classes of Feature Variables using Compound LCA

Covariate Class (a=1)	Feature Class (b=1, 2, 3)	Prevalence of Covariate Class ($\pi(a)$)	Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)
African American patients, older, more females, with high risk of cholesterol, diabetes and hypertension	Mild Impaired	0.26	0.24
	AMN Multi + FX[1] + NP[2]	0.26	0.76
	AMN + FX + NP	0.26	0
Covariate Class (a=2)	**Feature Class (b=1, 2, 3)**	**Prevalence of Covariate Class ($\pi(a)$)**	**Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)**
Non-African American patients, younger, low vascular risk and high prevalence of APOE	Mild Impaired	0.48	0.32
	AMN Multi + FX + NP	0.48	0.07
	AMN + FX + NP	0.48	0.61
Covariate Class (a=3)	**Feature Class (b=1, 2, 3)**	**Prevalence of Covariate Class ($\pi(a)$)**	**Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)**
Non-African American patients, older, more males, high vascular risk and relatively high prevalence of APOE	Mild Impaired	0.26	0.26
	AMN Multi + FX + NP	0.26	0.05
	AMN + FX + NP	0.26	0.69

[1] FX: Functional Impairment [2] NP: Neuropsychiatric Features

Table 4.17: Logistic and Linear Regression Models

	Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Logistic Regression Model					
GDS	Intercept	1.25	0.06	416.24	<0.0001
	Class 2 (Covariates)	-0.05	0.05	1.04	0.31
	Class 3 (Covariates)	-0.15	0.06	6.61	0.01
	Class 2 (Feature Variables)	0.37	0.05	46.93	<0.0001
	Class 3 (Feature Variables)	0.33	0.04	71.15	<0.0001
	Source	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
Linear Regression Model					
IADL	Class 2 (Covariates)	85.27	85.27	14.66	0.0001
	Class 3 (Covariates)	65.19	65.19	11.21	0.0008
	Class 2 (Feature Variables)	937.64	937.64	161.18	<0.0001
	Class 3 (Feature Variables)	5407.34	5407.34	929.52	<0.0001
NPI	Class 2 (Covariates)	15.29	15.29	3.98	0.05
	Class 3 (Covariates)	24.39	24.39	6.35	0.01
	Class 2 (Feature Variables)	246.45	245.45	64.21	<0.0001
	Class 3 (Feature Variables)	1189.75	1189.75	309.96	<0.0001
MMSE	Class 2 (Covariates)	53.06	53.06	18.32	<0.0001
	Class 3 (Covariates)	24.84	24.84	8.57	0.0034
	Class 2 (Feature Variables)	371.29	371.29	128.17	<0.0001
	Class 3 (Feature Variables)	3804.48	3804.48	1313.36	<0.0001
Logical Memory - Immediate	Class 2 (Covariates)	0.15	0.15	0.16	0.69
	Class 3 (Covariates)	0.51	0.51	0.56	0.45
	Class 2 (Feature Variables)	42.83	42.83	47.29	<0.0001
	Class 3 (Feature Variables)	1345.53	1345.53	1485.67	<0.0001
Logical Memory - Delayed	Class 2 (Covariates)	0.16	0.16	0.17	0.68
	Class 3 (Covariates)	0.01	0.01	0.01	0.91
	Class 2 (Feature Variables)	72.80	72.80	76.64	<0.0001
	Class 3 (Feature Variables)	1378.97	1378.97	1451.82	<0.0001
Category Fluency	Class 2 (Covariates)	29.35	29.35	37.36	<0.0001
	Class 3 (Covariates)	25.37	25.37	32.29	<0.0001
	Class 2 (Feature Variables)	62.95	62.95	80.14	<0.0001
	Class 3 (Feature Variables)	935.93	935.93	1191.46	<0.0001
Trails A	Class 2 (Covariates)	27.23	27.23	12.90	0.0003
	Class 3 (Covariates)	11.87	11.87	5.62	0.02
	Class 2 (Feature Variables)	1634.06	1634.06	774.11	<0.0001
	Class 3 (Feature Variables)	2203.95	2203.95	1044.09	<0.0001
Digit Span Forward	Class 2 (Covariates)	1.90	1.90	1.86	0.17
	Class 3 (Covariates)	3.00	3.00	2.94	0.09
	Class 2 (Feature Variables)	131.65	131.65	128.97	<0.0001
	Class 3 (Feature Variables)	289.53	289.53	283.63	<0.0001
Boston Naming	Class 2 (Covariates)	164.86	164.86	57.25	<0.0001
	Class 3 (Covariates)	260.52	260.52	90.46	<0.0001
	Class 2 (Feature Variables)	591.72	591.72	205.46	<0.0001
	Class 3 (Feature Variables)	1808.38	1808.38	627.92	<0.0001
Trails B	Class 2 (Covariates)	29.61	29.61	15.08	0.0001
	Class 3 (Covariates)	19.32	19.32	9.84	0.002
	Class 2 (Feature Variables)	3492.09	3492.09	1778.77	<0.0001
	Class 3 (Feature Variables)	4219.84	4219.84	2149.47	<0.0001
Digit Span Backward	Class 2 (Covariates)	2.76	2.76	3.24	0.07
	Class 3 (Covariates)	5.86	5.86	6.88	0.009
	Class 2 (Feature Variables)	153.52	153.52	180.22	<0.0001
	Class 3 (Feature Variables)	410.07	410.07	481.39	<0.0001
Digit Symbol	Class 2 (Covariates)	0.94	0.94	1.02	0.31
	Class 3 (Covariates)	5.92	5.92	6.42	0.01
	Class 2 (Feature Variables)	663.68	663.68	720.07	<0.0001
	Class 3 (Feature Variables)	1200.52	1200.52	1302.53	<0.0001

4.4 Discussion

In this chapter, we first conducted a simulation study to investigate the performance of our estimators of $\hat{\pi}(a)$, $\hat{\pi}(b|a)$, $\hat{\alpha}_{ak}$ and $\hat{\beta}_{bj}$. We examined the bias, standard error and average of analytical standard error using sample sizes of 600 and 2000. These results showed that our estimators perform well when covariates are high-dimensional and correlated, and average of analytical standard errors decreased and tended to be similar to standard error values when the sample size was larger. Our results also reflected the difficulty of computing estimators based on Bernoulli distribution, where a bigger sample size did not necessarily improve the difference in standard errors and average of analytical standard errors for these estimators.

We aimed to explore different subtypes of MCI using compound LCA and its extension. With compound LCA, we were able to conduct latent class analysis for varying numbers of classes of covariates and response variables despite having high dimensional and possibly correlated covariates in the MCI dataset. Table 4.15 revealed different ICL-BIC model selection criterion values and entropy R^2 values for different models. Even though some models had negative ICL-BIC values, indicating a potential solution, entropy R^2 values were useful in narrowing the models down, ultimately leading us to select a model with 3 classes of covariates and 3 classes of feature variables when we also considered the relevance of clinical interpretation of our model.

The results from Table 4.16 revealed heterogeneity of MCI subgroups for high-dimensional and correlated covariates. For instance, as indicated by high risk of cholesterol, diabetes and hypertension, the patients in the first covariate class most likely have vascular dementia, where these conditions disproportionally affect African Americans. Vascular dementia tends to be represented by a non-amnestic feature such as executive dysfunction, and this characteristic was highlighted by an amnestic multi-domain feature class in our results, which includes both amnestic and non-amnestic features (Table 4.16, prevalence: 0.76). In addition, patients in the second and third covariate class most likely have Alzheimer's disease, as shown by latent class solution of covariates indicating the high prevalence of APOE $\varepsilon 4$

allele and additionally supported by the research on APOE $\varepsilon 4$ allele. This identification is further confirmed by the high prevalence of a pure amnestic class in the second and third covariate classes (Table 4.16, prevalence: 0.61, 0.69), as Alzheimer's disease is predominant among patients diagnosed with a pure amnestic MCI subtype.

However, the F-tests revealed that the classes of covariates are mostly statistically significant for each feature variable, implying that feature classes may be nested within covariate classes. Even though our results using compound LCA provided results that were consistent with underlying etiologies of MCI subgroups in literature, it is necessary to update the assumptions of conditional independence between the covariates and responses given the latent classes to reflect nesting between the classes of feature variables and classes of covariates. Moreover, these results highlight the need to consider developing a statistical method that would help us determine when a dataset has a nested structure.

Chapter 5

Extension of Compound Latent Class Analysis

5.1 Overview

In this chapter, we explore an extension of compound LCA that can handle nested classes of feature variables within classes of covariates when covariates are high dimensional and potentially correlated. We outline the updated relative frequencies model, maximum likelihood estimation and latent class specific means of covariates and feature variables. We provide a likelihood ratio test that will compare between compound LCA and its extension to find the best fitting model.

5.2 Methods

5.2.1 Relative Frequencies Model

We assume conditional independence between the covariates and responses given the latent classes, where the classes of responses are nested within the covariate space

$$f(x_i, y_i|a, b) = f(x_i|a)f(y_i|b, a)$$

Then under compound LCA, the finite mixture model is given by

$$\begin{aligned} e^{\ell_i} = f(x_i, y_i; \alpha, \beta) &= \sum_a \sum_b \pi(a, b) f(x_i, y_i|a, b; \alpha, \beta) \\ &= \sum_a \left\{ \pi(a) f(x_i|a; \alpha) \sum_b \pi(b|a) f(y_i|b, a; \beta, \alpha) \right\} \end{aligned}$$

with log-likelihood $\ell = \sum_i \ell_i$.

5.2.2 Maximum Likelihood Estimation

Using the EM algorithm, estimation can be carried out by iterating between estimation of relative frequencies of the two sets of latent classes using EM algorithm:

$$\hat{\pi}(a) = n^{-1} \sum_i \psi_{ia}$$
$$\hat{\pi}(b|a) = n^{-1} \hat{\pi}(a)^{-1} \sum_i \tau_{iab}$$

and solving for the following weighted score equations for the parameters α and β of the conditional distributions when the classes of responses are nested within the covariate space:

$$\frac{\partial \ell}{\partial \alpha} = \sum_i \sum_a \psi_{ia} \frac{\partial log f(x_i|a;\alpha)}{\partial \alpha} = 0$$
$$\frac{\partial \ell}{\partial \beta} = \sum_i \sum_a \sum_b \tau_{iab} \frac{\partial log f(y_i|b,a;\beta,\alpha)}{\partial \beta} = 0.$$

Then posterior probabilities of membership in the two sets of latent classes are simultaneously updated:

$$\tau_{iab} = Pr(a,b|x_i,y_i;\alpha,\beta) = \frac{\pi(a) f(x_i|a;\alpha) \pi(b|a) f(y_i|b,a;\beta,\alpha)}{\sum_{a'} \{\pi(a') f(x_i|a';\alpha) \sum_{b'} \pi(b'|a') f(y_i|b',a';\beta,\alpha)\}}$$
$$\psi_{ia} = Pr(a|x_i,y_i;\alpha,\beta) = \sum_b \tau_{iab}. \tag{6}$$

We assume conditional independence of the covariates, so that $f(x_i|a;\alpha) = \prod_k f(x_{ik}|a;\alpha)$ with distinct means α_{ak}. We also assume conditional independence of the responses, so that $f(y_i|b,a;\beta,\alpha) = \prod_j f(y_{ij}|b,a;\beta,\alpha)$ with distinct means β_{bj}. Then we use the following EM algorithm to solve for the posterior probabilities of membership and latent class specific means:

1. Initialize values of τ_{iab}, ψ_{ia}, A and B.

2. Update the latent class specific means at each iteration of the EM algorithm. Then we

update the parameter estimates as

$$\hat{\alpha}_{ak} = \frac{\sum_i \psi_{ia} x_{ik}}{\sum_i \psi_{ia}}$$
$$\hat{\beta}_{bj} = \frac{\sum_i \tau_{iab} y_{ij}}{\sum_i \tau_{iab}}$$

3. Update the posterior probabilities of membership in the two sets of latent classes, τ_{iab} and ψ_{ia}, using Equation 6.

4. Repeat steps 2-3 until convergence.

5.2.3 Information Matrix

The empirical Fisher's information matrix can be computed to find the standard errors of parameter estimates. We can compute a column vector of weighted score equations for $\zeta = (\pi(a)^T, \pi(b|a)^T, \alpha^T, \beta^T)^T$, defined by $Q(\zeta, x_i, y_i)$, where

$$Q(\zeta, x_i, y_i) = \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix}$$

and the solution to $Q(\zeta, x_i, y_i)$ is the maximum likelihood estimator $\hat{\zeta}$. Then the empirical Fisher's information matrix can be written as

$$
\begin{aligned}
I &= \sum_i Q(\zeta, x_i, y_i) Q(\zeta, x_i, y_i)^T \\
&= \sum_i \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix} \begin{pmatrix} \frac{\partial \ell}{\partial \pi(a)} \\ \frac{\partial \ell}{\partial \pi(b|a)} \\ \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix}^T
\end{aligned}
$$

where we can define the components of column vector $Q(\zeta, x_i, y_i)$ as

$$
\begin{aligned}
\frac{\partial \ell}{\partial \pi(a)} &= \sum_{i=1}^{n} \frac{\psi_{ia}}{\pi(a)}, \quad a = 1, ..., A \\
\frac{\partial \ell}{\partial \pi(b|a)} &= \sum_{i=1}^{n} \sum_{a=1}^{A} \frac{\tau_{iab}}{\pi(b|a)}, \quad b = 1, ..., B \\
\frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^{n} \psi_{ia} \frac{\partial \log f(x_i|a; \alpha)}{\partial \alpha}, \quad a = 1, ..., A \\
\frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^{n} \tau_{iab} \frac{\partial \log f(y_i|b, a; \beta, \alpha)}{\partial \beta}, \quad b = 1, ..., A * B.
\end{aligned}
$$

and the last component is changed to reflect the nested structure. We can derive the standard errors of parameter estimates $\hat{\zeta}$ by computing an estimator of the asymptotic covariance matrix $\text{avar}(\hat{\zeta}) = I^{-1}$.

5.2.4 Likelihood Ratio Test

In order to compare two different methods and find the best method for different scenarios, we propose using a likelihood ratio test. Define the null hypothesis to be

$$
H_0 : \beta_{1b} = \beta_{2b} = ... = \beta_{Ab}, \quad B = 1, ..., B
$$

which assumes that the feature means are the same regardless of given covariate classes. Define the alternative hypothesis to be

$$H_A : \text{there exists at least some inequalities for some } b.$$

The we can conduct a likelihood ratio test

$$\begin{aligned} T_L &= -2 * \frac{L_{compound}}{L_{compoundext}} \\ &= -2\{\ell(\beta|_b) - \ell(\beta|_{b,a})\} \sim \chi^2_{AB-B} \end{aligned}$$

where L_* denotes the likelihood of the respective model derived from compound LCA or its extension. If our test statistic T_L is statistically significant, we can reject the null hypothesis and conclude that our extension of compound LCA produces the best fitting model.

5.2.5 Analysis of Underlying Subpopulations

Figure 5.1 is an updated version of Figure 4.1, where red arrows indicate our assumption that feature classes are nested within covariate classes. The conditional independence is updated to be $f(x_i, y_i|a, b) = f(x_i|a)f(y_i|b, a)$.

Derivation of relative frequencies of the two sets of latent classes, $\pi(a)$ and $\pi(b|a)$, remains the same. However, latent class-specific means of J features for $b = 1, ..., B$, $\hat{\beta}_{b1}, ..., \hat{\beta}_{bJ}$, are updated to reflect the change in assumptions, where b classes of feature variables now include a combination of covariate classes and feature classes to be $b = 1, ..., A * B$.

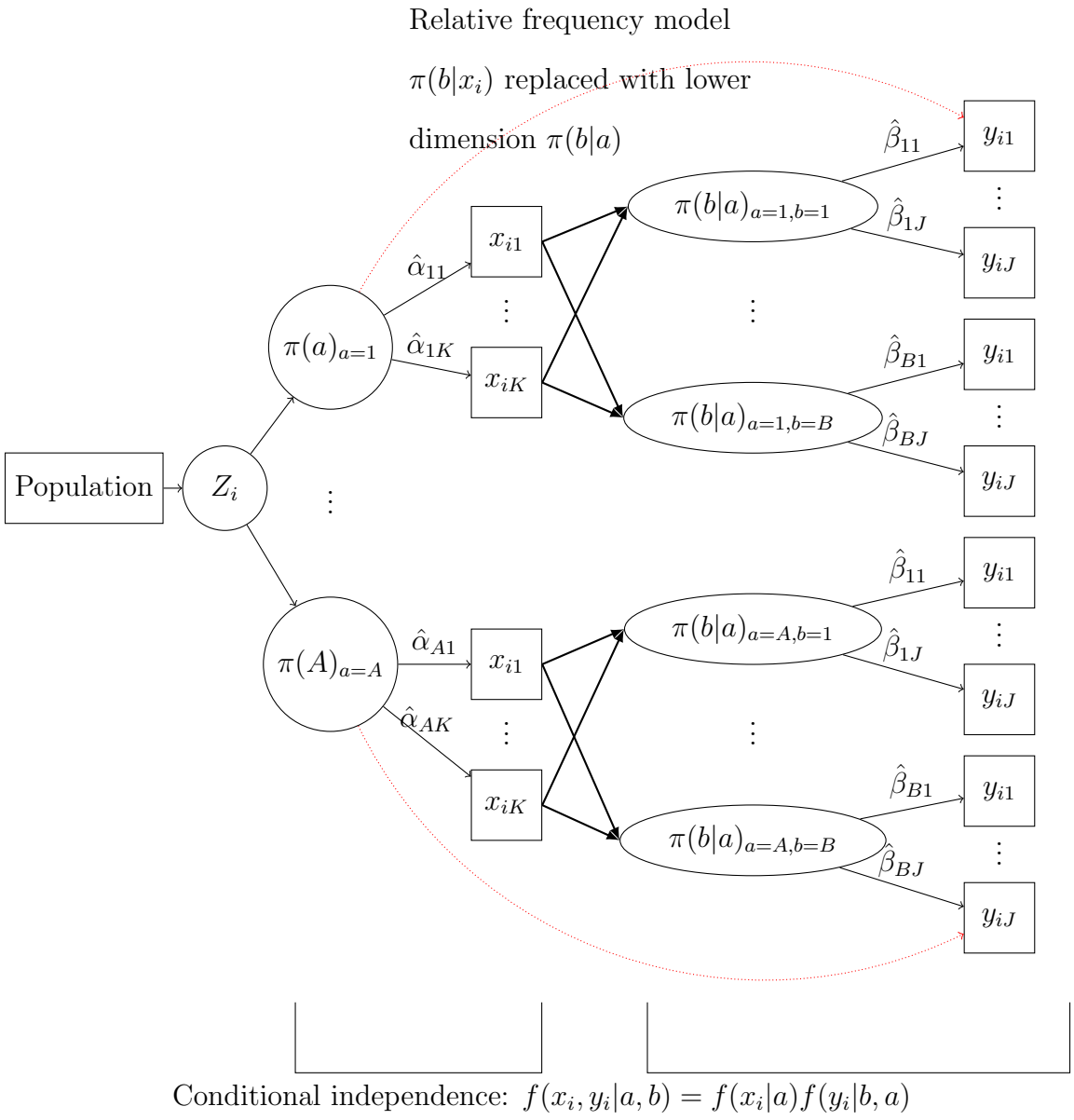

Figure 5.1: Extension of Compound LCA - Analysis of Underlying Subpopulation

5.3 Results

5.3.1 MCI Dataset

5.3.1.1 Overview

In order to compare the results from using our extension of compound LCA with previous results of compound LCA (Chapter 4), we again used 3 classes of covariates ($A = 1, 2, 3$) and 3 classes of feature variables ($B = 1, 2, 3$) to apply updated methods of compound LCA.

5.3.1.2 Analysis

We first interpret the latent class-specific means of covariates, which revealed a different group of patients in the dataset (Table 5.1):

1. $A = 1$: African American and non-African American patients with intermediate vascular risk
2. $A = 2$: African American patients with high vascular risk
3. $A = 3$: Non-African American patients with low vascular risk and high APOE.

The latent class solution of feature variables provided a unique clinical interpretation when we assumed that there is a nested structure between feature classes and covariate classes in the MCI dataset (Table 5.2):

1. $A = 1, B = 1$: Mildly impaired
2. $A = 1, B = 2$: Mild functional impairment and neuropsychiatric symptoms
3. $A = 1, B = 3$: Mild functional impairment and neuropsychiatric symptoms
4. $A = 2, B = 1$: Amnestic multi-domain with functional impairment
5. $A = 2, B = 2$: Amnestic multi-domain with functional impairment and neuropsychiatric symptoms

6. $A = 2, B = 3$: Functional impairment and neuropsychiatric symptoms with executive dysfunction

7. $A = 3, B = 1$: Amnestic multi-domain with functional impairment and neuropsychiatric symptoms

8. $A = 3, B = 2$: Non-Amnestic with functional impairment and neuropsychiatric symptoms

9. $A = 3, B = 3$: Amnestic with functional impairment and neuropsychiatric symptoms.

Combining the latent class definitions of both covariate and feature space, we can interpret the latent class frequencies when we use an extension of compound LCA (Table 5.3). Among patients who are in the first covariate class, or African American and non-African American patients with intermediate vascular risk, the prevalence of feature classes can be interpreted as: the probability of being mildly impaired is 0.60, the probability of having mild functional impairment and neuropsychiatric features are 0.26 and 0.14. Among patients in the second covariate class, or African American patients with high vascular risk, the prevalence of feature classes can be interpreted as: the probability of being diagnosed with amnestic multi-domain with functional impairment is 0.52, the probability of being diagnosed with amnestic multi-domain with functional impairment and neuropsychiatric features is 0.03, and the probability of being diagnosed with executive dysfunction with functional impairment and neuropsychiatric features is 0.45. Among patients in the third covariate class, or non-African American patients with low vascular risk and high APOE, the prevalence of feature classes can be interpreted as: the probability of being diagnosed with amnestic multi-domain with functional impairment and neuropsychiatric features is 0.20, the probability of being diagnosed with non-amnestic with functional impairment and neuropsychiatric features is 0.20, and the probability of being diagnosed with amnestic with functional impairment and neuropsychiatric features is 0.60.

We then used the likelihood ratio test to compare between the two methods and determine the best fitting model for our dataset. With the log-likelihood values of $\ell(\beta|_b) = -15965.70$ for compound LCA and $\ell(\beta|_b) = -10168.37$ for its extension, we computed $T_L = -2\{\ell(\beta|_b) - \ell(\beta|_b)\} = 11594.66$ which has a Chi-square distribution with 6 degrees of freedom. Using the test statistic, we conclude that with p-value <0.005, our extension of compound LCA provides the best fitting model.

Table 5.1: Latent Class-Specific Means (and Standard Errors) of Covariates

	Class	3 Classes of Covariates, 3 Classes of Feature Variables
Hachinski Ischaemia Score	1	0.04 (0.04)
	2	0.13 (0.05)
	3	0.08 (0.03)
Hypercholesterolemia	1	0.54 (0.02)
	2	0.65 (0.02)
	3	0.49 (0.02)
Diabetic Status	1	0.14 (0.02)
	2	0.28 (0.03)
	3	0.08 (0.03)
Hypertension	1	0.55 (0.02)
	2	0.72 (0.02)
	3	0.47 (0.02)
Education	1	15 (0.02)
	2	14 (0.04)
	3	16 (0.03)
Age	1	0.36 (0.01)
	2	0.50 (0.01)
	3	0.41 (0.01)
Gender	1	0.52 (0.02)
	2	0.51 (0.03)
	3	0.45 (0.02)
Race	1	0.17 (0.03)
	2	0.25 (0.02)
	3	0.04 (0.04)
Cardiac Dysrhythmia	1	0.09 (0.03)
	2	0.14 (0.03)
	3	0.09 (0.03)
Coronary Vascular Disease	1	0.14 (0.02)
	2	0.22 (0.03)
	3	0.10 (0.03)
APOE	1	0.38 (0.02)
	2	0.37 (0.02)
	3	0.52 (0.02)

Table 5.2: Latent Class-Specific Means (and Standard Errors) of Feature Variables

		Class	3 Classes of Covariates, 3 Classes of Feature Variables
Functional	No. of IADL impaired	1	0 (0.02)
		2	2.70 (0.02)
		3	2.38 (0.04)
		4	2.91 (0.01)
		5	3.51 (0.05)
		6	5.45 (0.09)
		7	4.55 (0.05)
		8	3.99 (0.06)
		9	4.00 (0.04)
Neuropsychiatric	% with GDS ≥ 5	1	13.90 (0.03)%
		2	14.84 (0.06)%
		3	17.34 (0.04)%
		4	18.42 (0.06)%
		5	38.88 (0.06)%
		6	36.72 (0.08)%
		7	30.09 (0.04)%
		8	31.18 (0.09)%
		9	16.04 (0.03)%
	No. of NPI-Q symptoms present	1	0.88 (0.02)
		2	1.85 (0.02)
		3	1.79 (0.03)
		4	1.31 (0.01)
		5	2.02 (0.03)
		6	4.08 (0.08)
		7	2.96 (0.04)
		8	2.28 (0.05)
		9	2.21 (0.03)
Cognitive	Global		
	MMSE	1	-1.30 (0.02)
		2	-1.15 (0.03)
		3	-0.23 (0.03)
		4	-2.49 (0.02)
		5	-3.62 (0.04)
		6	-1.66 (0.04)
		7	-3.31(0.03)
		8	-2.06 (0.05)
		9	-2.38 (0.02)
	Logical Memory		
	Immediate	1	-1.07 (0.01)
		2	-1.13 (0.03)
		3	0.37 (0.06)
		4	-1.40 (0.02)
		5	-1.76 (0.06)
		6	-0.74 (0.06)
		7	-2.15 (0.02)
		8	-0.72 (0.04)
		9	-2.31 (0.09)
	Delayed	1	-1.09 (0.02)
		2	-1.33 (0.03)
		3	0.34 (0.02)
		4	-1.48 (0.03)
		5	-1.69 (0.08)
		6	-0.84 (0.02)
		7	-2.21 (0.05)
		8	-0.77 (0.04)
		9	-2.56 (0.03)
	Semantic Memory		
	Category Fluency	1	-0.80 (0.03)
		2	-0.63 (0.02)
		3	-0.32 (0.03)
		4	-1.21 (0.08)
		5	-1.31 (0.02)
		6	-0.88 (0.04)
		7	-1.97 (0.03)
		8	-1.25 (0.03)
		9	-1.24 (0.05)
Cognitive	Attention		
	Trails A	1	0.52 (0.03)
		2	-0.08 (0.01)
		3	0 (0.03)
		4	1.85 (0.04)
		5	6.43 (0.02)
		6	0.98 (0.06)
		7	2.16 (0.02)
		8	2.64 (0.02)
		9	0.33 (0.04)*
	Digit Span Forward	1	-0.28 (0.04)
		2	-0.04 (0.02)
		3	-0.03 (0.02)
		4	-0.63 (0.08)
		5	-0.86 (0.01)
		6	-0.38 (0.05)
		7	-0.92 (0.03)
		8	-0.62 (0.05)
		9	-0.22 (0.06)
	Language		
	Boston Naming	1	-1.06 (0.03)
		2	-0.43 (0.03)
		3	-0.29 (0.01)
		4	-2.16 (0.04)
		5	-3.02 (0.04)
		6	-0.61 (0.02)
		7	-2.34 (0.05)
		8	-1.36 (0.02)
		9	-1.03 (0.02)
	Executive Function		
	Trails B	1	1.08 (0.06)
		2	0.17 (0.04)
		3	0.17 (0.02)
		4	2.82 (0.05)
		5	4.59 (0.08)
		6	1.89 (0.01)
		7	3.08 (0.09)
		8	3.45 (0.03)
		9	0.82 (0.05)*
	Digit Span Backward	1	-0.48 (0.04)
		2	-0.16 (0.02)
		3	-0.02 (0.02)
		4	-0.82 (0.09)
		5	-1.10 (0.01)
		6	-0.62 (0.03)
		7	-1.13 (0.04)
		8	-0.85 (0.04)
		9	-0.45 (0.05)
	Visuomotor		
	Digit Symbol	1	-0.68 (0.02)
		2	-0.27 (0.04)
		3	-0.25 (0.02)
		4	-1.47 (0.05)
		5	-1.72 (0.09)
		6	-1.28 (0.01)
		7	-1.89 (0.06)
		8	-1.87 (0.04)
		9	-0.69 (0.04)

* Higher scores on Trail A and Trail B indicate worse performance.

Table 5.3: Interpretation of Relative Frequencies of Latent Classes - Extension of Compound LCA

Covariate Class (a=1)	Feature Class (b=1, 2, 3)	Prevalence of Covariate Class ($\pi(a)$)	Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)
African American and non-African American patients with intermediate vascular risk	Mild Impaired	0.58	0.60
	Mild FX[1]+ NP[2]	0.58	0.26
	Mild FX + NP	0.58	0.14
Covariate Class (a=2)	**Feature Class (b=1, 2, 3)**	**Prevalence of Covariate Class ($\pi(a)$)**	**Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)**
African American patients with high vascular risk	AMN Multi+ FX	0.18	0.52
	AMN Multi + FX + NP	0.18	0.03
	FX + NP + Exec Dysfunction	0.18	0.45
Covariate Class (a=3)	**Feature Class (b=1, 2, 3)**	**Prevalence of Covariate Class ($\pi(a)$)**	**Prevalence of Feature Class Given Covariate Class ($\pi(b\|a)$)**
Non-African American patients with low vascular risk and high APOE	AMN Multi + FX + NP	0.24	0.20
	Non-amnestic + FX + NP	0.24	0.20
	AMN + FX + NP	0.24	0.60

[1] FX: Functional Impairment [2] NP: Neuropsychiatric Features

5.4 Discussion

In this chapter, we extended the methods of compound LCA from Chapter 4 to handle a nested model when the F-Tests indicated that the classes of feature variables are nested within the classes of covariates in the MCI dataset. The results showed that when the assumptions were broadened to include such a scenario, we were able to gain a better insight into heterogeneity of MCI subgroups.

For instance, the second covariate class revealed a unique solution of feature classes, where African American patients with vascular dementia were spread out between amnestic multi-domain and executive dysfunction features (Table 5.3, prevalence: 0.52, 0.03, 0.45), as opposed to a concentration of patients in an amnestic domain feature from compound LCA (Table 4.16, prevalence: 0.76). The results from the third covariate class were similar to those from compound LCA, where patients most likely have Alzheimer's disease, as indicated by the high prevalence of APOE $\varepsilon 4$ allele and diagnosis of a pure amnestic MCI subtype (Table 5.3, prevalence: 0.60).

Moreover, clinical interpretations revealed important features that were not available when we used compound LCA. Latent class solutions of feature variables in compound LCA were limited to 3 MCI subtypes–mildly impaired, amnestic multi-domain with functional impairment and neuropsychiatric features and amnestic with functional impairment and neuropsychiatric features–which were not sufficient in revealing non-amnestic features such as executive dysfunction and instead provided an empty class of a pure amnestic feature (Table 4.16, prevalence: 0). However, extended methods of compound LCA were able to expand MCI subtypes into 9 different categories, which included non-amnestic and executive dysfunction features.

Both compound LCA and its extension provided results that were consistent with underlying etiologies of MCI subgroups in literature, although compound LCA provided limited perspective in heterogeneity of MCI. Results from Chapter 4 and Chapter 5 reflect the importance of investigating and comprehending the relationship between the classes of covariates

and the classes of feature variables. When we finally applied accurate assumptions for the MCI dataset, the latent class solutions and relative frequencies of latent classes provided straightforward interpretations that can better explore heterogeneity of MCI from a clinical point of view. The likelihood ratio test confirmed our assumptions and proved to be a potential tool in the future for investigators to confirm the underlying structure of a dataset. Both methods will be able to provide clinicians with an opportunity to continue to use LCA with high-dimensional and correlated covariates, i.e., 20-30 covariates, without compromising the underlying probabilistic structure of those covariates.

Chapter 6

Future Research

6.1 Summary

This dissertation research can be divided into two broad approaches for handling risk factors within the latent class framework. The first approach is focused on exploring scenarios when the latent class regression model (Bandeen-Roche et al., 1997) fails to provide latent class definitions that are deemed plausible by clinical judgment, and introducing the activity governor that allows investigators to incorporate all clinically relevant covariates into the model to simultaneously control for the effect of covariates on latent class definitions. The second approach is focused on incorporating high-dimensional and possibly correlated covariates in LCA by using compound LCA, which can incorporate 20-30 covariates while preserving the underlying probabilistic structure. Additionally, an extension of compound LCA is introduced to handle a nested structure where classes of feature variables are nested within classes of covariates. Both approaches will be practical in exploring heterogeneity of MCI using LCA.

6.2 Future Research

In future research investigating heterogeneity of MCI structure, we can consider including biomarker data from imaging, proteomics and genetic sources to help better understand etiologically relevant MCI subtypes. However, biomarker data tends to be high dimensional, and we need to consider different directions for two approaches of this dissertation research.

In the application of latent class regression models, we can consider using the idea of penalized regression, a popular choice for high-dimensional data analysis. Penalized regression can be used to find parameter estimates with a penalty for complexity, which includes a tuning parameter λ. Types of penalized regression models include ridge regression model, LASSO (least absolute shrinkage and selection operator) (Tibshirani, 1996) and elastic net. Leoutsakos et al. (2011) developed penalized regression methods for latent class regression models by using the ridge and LASSO penalties on covariates related to development of

dementia. Additionally, these methods assume that the latent variables are discrete, which can be useful for exploring heterogeneity of discrete MCI subtypes. Leoutsakos et al. (2011) recommend using methods derived by Houseman et al. (2007), which is based on penalized item response theory models for latent variables on a continuous scale.

In future research, penalized regression methods within the Bayesian framework can be used in latent class regression models. For instance, Tibshirani (1996) noted that estimates obtained by LASSO can be derived as the Bayes posterior mode under independent double-exponential priors. Park and Casella (2008) introduced the concept of the Bayesian LASSO, where the Gibbs sampler is used on a conditional Laplace prior and a hyperprior is assigned to the tuning parameter λ. The Bayesian LASSO can be adapted to apply penalized regression methods on latent class regression models.

In the application of latent class analysis, we can consider developing a hierarchical model. Hierarchical model is a type of cluster analysis where the goal is to discover a natural grouping of variables without using a method of model-based clustering (Johnson and Wichern, 2007). The idea of incorporating a model-based hierarchical model within a latent class framework has been explored in literature. For instance, Wang et al. (2020) introduced a conditional independence model with hierarchical priors to construct a posterior distribution using Just Another Gibbs Sampler (JAGS). In future research, we can similarly apply the concept of model-based hierarchical models as a generalization of our nested compound latent class model, which would enable us to handle an even larger number of covariates in LCA.

Bibliography

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L. and Rathouz, P. J. (1997), 'Latent variable regression for multiple discrete outcomes', American Statistical Association **92**(440), 1375–1386.

Bernath, M. M., Bhattacharyya, S., Nho, K., Barupal, D. K., Fiehn, O., Baillie, R., Risacher, S. L., Arnold, M., Jacobson, T., Trojanowski, J. Q., Shaw, L. M., Weiner, M. W., Doraiswamy, P. M., Kaddurah-Daouk, R. and Saykin, A. J. (2020), 'Serum triglycerides in alzheimer disease: Relation to neuroimaging and csf biomarkers', Neurology **94**(20), e2088–e2098.

Biernacki, C., Celeux, G. and Govaert, G. (1999), 'An improvement of the nec criterion for assessing the number of clusters in a mixture model', Pattern Recognition Letters **20**, 267–272.

Biernacki, C. and Govaert, G. (1997), 'Using the classification likelihood to choose the number of clusters', Computing Science and Statistics **29**, 451–457.

Burke, S. L., Cadet, T. and Maddux, M. (2018), 'Chronic health illnesses as predictors of mild cognitive impairment among african american older adults', Journal of the National Medical Association **110**(4), 314–325.

Celeux, G. and Soromenho, G. (1996), 'An entropy criterion for assessing the number of clusters in a mixture model', Journal of Classification **13**, 195–212.

David Wechsler (2008), 'Wechsler Adult Intelligence Scale–Fourth Edition', https://doi.org/10.1037/t15169-000.

Diaz-Mardomingo, M. C., Garcia-Herranz, S., Rodriquez-Fernandez, R., Venero, C. and Peraita, H. (2017), 'Problems in classifying mild cognitive impairment (mci): One or multiple syndromes?', Brain Sciences **7**(9), 111.

Efron, B. and Hinkley, D. V. (1978), 'Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information', Biometrika **65**, 457–482.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975), 'A practical method for grading the cognitive state of patients for the clinician', Journal of Psychiatric Research **12**, 189–198.

Gavett, B. E., Gurnani, A. S., Saurman, J. L., Chapman, K. R., Steinberg, E. G., Martin, B., Chaisson, C. E., Mez, J., Tripodis, Y. and Stern, R. A. (2016), 'Practice effects on story memory and list learning tests in the neuropsychological assessment of older adults', PLOS One **11**(10), e0164492.

Hanfelt, J. J., Peng, L., Goldstein, F. C. and Lah, J. J. (2018), 'Latent classes of mild cognitive impairment are associated with clinical outcomes and neuropathology: Analysis of data from the national alzheimer's coordinating centert', Neurobiology of Disease **117**, 62–71.

Hanfelt, J. J., Wuu, J., Sollinger, A. B., Greenaway, M. C., Lah, J. J., Levey, A. I. and Goldstein, F. C. (2011), 'An exploration of subgroups of mild cognitive impairment based on cognitive, neuropsychiatric and functional features: Analysis of data from the national alzheimer's coordinating center', American Association for Geriatric Psychiatry **19**(11), 940–950.

Hathaway, R. J. (1986), 'Another interpretation of the em algorithm for mixture distributions', Statistics and Probability Letters **4**(2), 53–56.

Houseman, E. A., Marsit, C., Karagas, M. and Ryan, L. M. (2007), 'Penalized item response theory models: application to epigenetic alterations in bladder cancer', Biometrics **63**(4), 1269–1277.

Ito, K., Hutmacher, M. M. and Corrigan, B. W. (2012), 'Modeling of functional assessment questionnaire (faq) as continuous bounded data from the adni database', Journal of Pharmacokinetics and Pharmacodynamics **39**, 601–618.

Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., Dubois, B., Eriksdotter, M., Ewers, M., Graessel, E., Kramberger, M. G., Law, E., Mecocci, P., Molinuevo, J. L., Nygård, L., Olde-Rikkert, M. G., Orgogozo, J.-M., Pasquier, F., Peres, K., Salmon, E., Sikkes, S. A., Sobow, T., Spiegel, R., Tsolaki, M., Winblad, B. and Frölich, L. (2015), 'Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review', Alzheimer's Research and Therapy **7**(1), 17.

Johnson, R. A. and Wichern, D. W. (2007), Applied Multivariate Statistical Analysis, Pearson Prentice Hall.

Kaplan, E., Goodglass, H. and Weintraub, S. (1983), Boston Naming Test-Second Edition, Philadelphia : Lea & Febiger.

Kaufer, D. I., Cummings, J. L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., Lopez, O. L. and DeKosky, S. T. (2000), 'Validation of the npi-q, a brief clinical form of the neuropsychiatric inventory', The Journal of Neuropsychiatry and Clinical Neurosciences **12**(2), 233–239.

Lazarsfeld, P. F. and Henry, N. W. (1968), Latent Structure Analysis, Houghton Mifflin Company.

Leoutsakos, J.-M. S., Bandeen-Roche, K., Garrett-Mayer, E. and Zandid, P. P. (2011), 'Incorporating scientific knowledge into phenotype development: Penalized latent class regression', Statistics in Medicine **30**(7), 784–798.

Leroux, B. G. (1992), 'Consistent estimation of a mixing distribution', The Annals of Statistics **20**(3), 1350–1360.

Li, J.-Q., Tan, L., Wang, H.-F., Tan, M.-S., Tan, L., Xu, W., Zhao, Q.-F., Wang, J., Jiang, T. and Yu, J.-T. (2016), 'Risk factors for predicting progression from milld cognitive impairment to alzheimer's disease: a systematic review and meta-analysis of cohort studies', Journal of Neurology, Neurosurgery, and Psychiatry **87**(5), 476–484.

Louis, T. A. (1982), 'Finding the observed information matrix when using the em algorithm', Journal of the Royal Statistical Society. Series B (Methodological) **44**, 226–233.

Marshall, G. A., Zoller, A. S., Lorius, N., Amariglio, R. E., Locascio, J. J., Johnson, K. A., Sperling, R. A. and Rentz, D. M. (2015), 'Functional activities questionnaire items that best discriminate and predict progression from clinically normal to mild cognitive impairment', Current Alzheimer Research **12**(5), 493–502.

McLachlan, G. and Peel, D. (2000), Finite Mixture Models, Wiley.

Musa, G., Henríquez, F., noz Neira, C. M., Delgado, C., Lillo, P. and Slachevsky, A. (2017), 'Utility of the neuropsychiatric inventory questionnaire (npi-q) in the assessment of a sample of patients with alzheimer's disease in chile', Dementia & Neuropsychologia **11**(2), 129–136.

National Alzheimer's Coordinating Center (2021a), 'Biomarker & Imaging Data Set', `https://naccdata.org/data-collection/forms-documentation/biomarker-imaging`. Online; accessed 24-May-2021.

National Alzheimer's Coordinating Center (2021b), 'Uniform Data Set (UDS)', `https://naccdata.org/data-collection/forms-documentation/uds-3`. Online; accessed 23-May-2021.

Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S. A., Rodriguez, G. and Pagani, M. (2008), 'Principal component analysis of fdg pet in amnestic mci', European Journal of Nuclear Medicine and Molecular Imaging **35**(12), 2091–2202.

Park, T. and Casella, G. (2008), 'The bayesian lasso', Journal of the American Statistical Association **103**(482), 681–686.

Petersen, R. C. (2011), 'Mild cognitive impairment', The New England Journal of Medicine **364**(23), 2227–2234.

Petersen, R. C. (2016), 'Mild cognitive impairment', Continuum (Minneap Minn) **22**(2), 404–418.

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Change, J. M. and Filos, S. (1982), 'Measurement of functional activities in older adults in the community', Journal of Gerontology **37**, 323–329.

Richardson, J. T. (2007), 'Measures of short-term memory: A historical review', Cortex **43**(5), 635–650.

Rosen, W. G. (1980), 'Verbal fluency in aging and dementia', Journal of Clinical Neuropsychology **2**(2), 135–146.

Rosen, W. G., Terry, R. D., Fuld, P. A., Katzman, R. and Peck, A. (1980), 'Pathological verification of ischemic score in differentiation of dementias', Annals of Neurology **7**, 486–488.

Sanford, A. M. (2017), 'Mild cognitive impairment', Clinics in Geriatric Medicine **33**(3), 325–337.

Schwarz, G. (1978), 'Estimating the dimension of a model', Annals of Statistics **6**(2), 461–464.

Statistical Innovations Inc. (2016), 'Latent gold', https://www.statisticalinnovations.com/latent-gold-5-1/.

Tangalos, E. G. and Petersen, R. C. (2018), 'Mild cognitive impairment in geriatrics', Clinics in geriatric medicine **34**, 563–589.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', Journal of the Royal Statistical Society: Series B (Methodological) **58**(1), 267–288.

Tombaugh, T. N. (2004), 'Trail making test a and b: Normative data stratified by age and education', Archives of Clinical Neuropsychology **19**(2), 203–214.

Wang, C., Lin, X. and Nelson, K. P. (2020), 'Bayesian hierarchical latent class models for estimating diagnostic accuracy', Statistical Methods in Medical Research **29**(4), 1112–1128.

Yesavage, J. A. and Sheikh, J. I. (1986), 'Geriatric depression scale (gds): Recent evidence and development of a shorter version.', Clinical Gerontologist: The Journal of Aging and Mental Health **5**, 165–173.