ANALYZING ECOLOGICAL MOMENTARY DATA
USING GROWTH MIXTURE MODELING

by

Mariya P. Shiyko

A dissertation submitted to the Graduate Faculty in Educational Psychology
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2009

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

David Rindskopf, Ph.D.

Date | Chair of Examining Committee

Mary Kopala, Ph.D.

Date | Executive Officer

Yuelin Li, Ph.D

Jay Verkuilen, Ph.D.

Jamie Ostroff, Ph.D.

Keith A. Markus, Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

ANALYZING ECOLOGICAL MOMENTARY DATA USING GROWTH MIXTURE MODELING

by

Mariya P. Shiyko

Adviser: Professor David Rindskopf

Real-time data capture, also known as *ecological momentary assessment* (EMA), is a unique data collection technique, which records moment-to-moment changes in human behavior as they occur in real time and in naturalistic settings. EMA is typically collected by electronic devices that prompt study participants to report behaviors (e.g., smoking) in real time, thereby minimizing problems associated with retrospective recall and reactivity. EMA has been heralded as a promising research tool in education, psychology, and behavioral medicine. It provides the needed data to examine patterns of behaviors as well as their temporal characteristics.

G*rowth mixture modeling* (GMM) is a statistical solution to many challenges associated with analyzing intensive EMA data. GMM estimates individual developmental profiles and classifies them into latent homogenous groups based on similarities in trajectories. This dissertation is a secondary data analysis of daily smoking rate of 74 newly-diagnosed cancer patients, who were enrolled in a randomized smoking cessation clinical trial prior to their cancer-related surgery. Patients' daily smoking rate was recorded over an average period of two weeks, yielding 896 assessments in total. The exploratory data assessment demonstrated substantial differences in patterns of smoking reduction across individuals during the intervention period. The goal of the GMM analysis was threefold: 1) to identify distinct smoking cessation patterns in a sample of patients awaiting a cancer-related surgery, 2) to investigate whether differences in tapering profiles are associated with differential smoking abstinence at surgery, 3) to identify personal and situational characteristics that are associated with each of the smoking cessation approaches.

The final model identified three latent developmental classes, which included abrupt, gradual, and slow reducers, varying in their personal characteristics and smoking cessation rates. This model is contrasted with a single-class solution alternative. Challenges of model enumeration and model identification processes are discussed. While growth mixture modeling widens the spectrum of research questions that can be addressed, it also poses technical and conceptual challenges for future research.

Dedication

I would like to dedicate this dissertation to my husband, Evangelos Pappas, who selflessly held my hand throughout this exciting and challenging journey and supported me in countless ways. His admirable generosity, patience, wisdom, sense of humor, and love allowed me to pursue goals I would not have dared setting alone. Thank you for believing in me!

Acknowledgements

There are so many people who provided their support and guidance and whose wisdom and generosity helped me grow professionally and personally. First and foremost, I would like to extend my gratitude to David Rindskopf, my mentor for the last five years. His great intelligence and teaching talent never cease to inspire me. Thank you for supporting me in all endeavors I undertook. My special “thank you” goes to Yuelin Li, who had an abundance of patience when teaching me so many valuable skills. His creativity and curiosity are contagious; thank you for believing in me and supporting me in any possible way.

Thank you to Jamie Ostroff, who nurtured me and inspired me with her intelligence and wise leadership. This dissertation would not have been possible without her generosity. Thank you to Jay Verkuilen for being always available to give a friendly advice in times of personal and professional turmoil. Thank you to Keith Markus for his time and insightful comments. Thank you to Linda Muthén for her instant responses to all technical questions I had.

Thank you to my colleagues at the Department of Psychiatry and Behavioral Sciences of Memorial Sloan-Kettering Cancer Center, especially to Susan Holland, Emily Sachs, and Catherine Mosher. You were always generous with your time and your friendship is invaluable. Thank you to my fellow students at the Educational Psychology program at the Graduate Center, CUNY. You inspired me, entertained me, and made my life so much saner. Special thanks to Esther Carvalhaes and Kallen Tsikalas for your friendship; I am blessed to have you in my life.

Last, but not least, an enormous “thank you” to my parents, Tamara and Pyotr Shiyko. They instilled in me the desire for knowledge from an early age; and their never-ceasing love, support, and encouragement made me less fearful of pursuing my dreams. Thank you for being so closely involved in my life from miles away. Thank you to my little brother, Pavel Shiyko, who is growing into a wonderful man and a close friend. Thank you to my in-laws, Varvara and Konstantinos Pappas for being proud and supportive of me. Without all of you I would have been far away from where I am today!

Table of contents

List of tables

List of figures

Longitudinal Data: Introduction and Overview

Longitudinal studies are a source of rich data for researchers in the behavioral sciences. By observing the same individuals over time and recording their responses and behaviors, researchers are able to make inferences about temporal sequences of events and identify causal relationships between them. Methodological advances in longitudinal data management and analysis go beyond repeated measures analysis of variance and stimulate more complex study designs and research questions. Such methods are capable of handling difficulties associated with data collected repeatedly from the same individuals, such as:

- a nested structure of data with correlated observations taken from the same individual
- a combination of time-varying and time-invariant measures
- data missingness at some time points for some individuals
- differences in timing of data collection across individuals.

Instead of relying on just a few waves of data, modern longitudinal studies heavily invest in a number of observations they are able to record for each study participant (Stone, Shiffman, Atienza, & Nebeling, 2007). With remarkable improvements in data collection, storage, and analysis, researchers are able to replicate naturally occurring processes, sampling data in real time over an extended time period. The resulting data are commonly referred as *intensive longitudinal data* (ILD) which has its own nuances and challenges.

Walls and Schafer (2006, p.xiii) state that: “ILD arise in any situation where quantitative or qualitative characteristics of multiple individuals or study units are recorded at more then a handful of time points.” Every person may have their own schedule of observations that differ in frequency, duration, and spacing. Moreover, ILD does not simply mean a larger volume of repeated measures. It is the scientific motivation and unique research hypotheses associated with such data that need to be translated into a statistical model. In ILD, the questions usually pertain to processes: their initial development and evolution, interaction with environmental and personal factors, and uniqueness across individuals.

This dissertation focuses on one type of ILD, referred to as *ecological momentary assessment*, collected in the context of a smoking cessation trial. Driven by specific research questions aimed at identifying patterns of smoking cessation across individuals, we examine the feasibility of applying *growth mixture modeling* to the data (Muthén, 1997). To date, growth mixture modeling has been tested in longitudinal data with a few waves of observations, mostly aligned across individuals.

We recognize the potential for growth mixture modeling as a tool of finding common developmental patterns in a pool of complex and varying trajectories. Qualitative differences in development may be associated with certain personality characteristics, and linked to long-term outcomes such as smoking abstinence rates. Despite our data coming from the field of health psychology, the methodology can be extended to educational, psychological, or medical research – fields interested in monitoring changes in human behavior that do not follow a uniform pattern.

In the following sections we will describe the nature of ecological momentary assessments and the uniqueness of data they yield. We will provide details of the smoking intervention trial and identify research hypotheses we plan to answer with the analysis. We will present an overview of growth mixture modeling, focusing on a conceptual understanding through linking it with more widely used methodologies. Model building and selection processes will be outlined focusing on statistical as well as theoretical indicators of fit. Results will be summarized and compared across various types of models. The final model will be interpreted in graphical and statistical terms. In conclusion, we will address the issues of feasibility, practical and methodological implications and challenges, as well as limitations and implications of the current study.

Ecological Momentary Assessments

Description of Ecological Momentary Assessments Methodology

In the field of psychology, ILD are referred to as *ecological momentary assessments* (EMAs) – a frequent sampling of persons' experiences, thoughts, feelings, or actions, recorded at the time when they arise (Stone & Shiffman, 2002). In EMAs, the data collection process is moved to naturalistic settings, where samples of experiences are recorded instantaneously. Rather than relying on paper records, portable electronic devices prompt individuals to answer a set of questions, storing responses with a matching time stamp. The frequency and timing of data sampling is guided by study questions and the phenomenon of interest. For instance, looking at associations between mood and a type of consumed food would require a record of both every time a person eats, while an end of the day report of pain symptoms and quality of life assessments would be sufficient for looking at recovery trajectories for patients after surgery.

While the field of behavioral research is accustomed to relying on self-reports for data collection purposes, there is a never ceasing need to assure data quality. One of the advantages of EMA is that it reduces the recollection time to "right now", "at this moment", "since you woke up", or "today", therefore minimizing memory bias, which is a common problem when an event is separated from recall by days or weeks (Moskowitz & Young, 2006).

Another advantage of EMA pertains to the phenomenon known as reactivity, first discussed by Campbell and Stanley in their classic manuscript on research design (1963). Reactivity implies that study participants' behavior may alter simply because of their awareness of being watched. As EMA takes place in natural settings, the hope is that individuals are not as self-conscious when responding to computer assessments as they are in laboratory settings. By distancing a researcher from the data collection process, participants are more likely to provide accurate authentic responses with better generalizability (Smyth & Stone, 2003).

Finally, due to the nature of the methodology, assessments are taken frequently and over an extended time period, yielding a record of numerous observational points. After a temporal assembly of observations, momentary recordings allow reconstructing dynamic psycho-social and behavioral processes without intervening into the privacy of people's lives (Stone et al., 2007).

The number of studies that employ EMA methodology is rising. To mention a few examples, real-time data collection is used to monitor patients' diet (Smyth et al., 2001; Wegner et al., 2002), adherence to a medication treatment plan (Moskowitz & Young, 2006), study emotional components of eating disorders (Engel et al., 2007; Smyth et al., 2007), learn about withdrawal symptoms associated with smoking cessation process (McCarthy, Piasecki, Fiore, & Baker, 2006; Shiffman et al., 2006), relate daily emotional responses to teachers' burnout (Carson, 2007), and identify factors explaining differences in student performance and motivation based on school-related and after-school experiences (Martinez, 2004; Rathunde & Csikszentmihalyi, 2005).

Challenges of EMA Data Analysis

Acknowledging the advantages of EMA, the price for the fine-grain recording is the complexity of collected data. Besides the challenges of repeated observations outlined in the introduction, the data collection process yields tens, hundreds, or even thousands of observational points spread throughout a period of days, weeks, or months for every participant. As a result of individual differences in sampling schedules, such data may have little temporal overlap and various patterns of frequency across participants. Furthermore, due to the increased demand of data collection, missing observations are the norm rather than the exception. They occur more frequently as the intervention progresses and subjects get more accustomed to reporting their responses and see little value in answering the same questions repeatedly.

Some researchers see the quality of EMA data primarily as a strength and advise aggregating data across time points to fit it into standard methodological approaches (Hufford & Shiffman, 2002; Shiffman & Hufford, 2001; Shiffman, Stone, & Hufford, 2008). While an average of daily mood ratings for seven days would be more accurate than simply asking a participant at the end of the week: "how was your mood during the past week?", such summary is not informative in detecting daily variations in mood changes. In some instances, aggregated averages would be sufficient, while in others, individual recordings need to be analyzed. Therefore, the choice of statistical methodology should be driven by the research questions.

Multilevel Modeling as an Approach to EMA Data Analysis

To date, the most widely used statistical approach to analyzing momentary data is *multilevel modeling* (MLM), also known as hierarchical linear modeling or mixed effects modeling (Affleck, Zautra, Tennen, & Armeli, 1999; Raudenbush & Bryk, 2002; Schwartz & Stone, 1998). Commonly used in analysis of complex longitudinal data, MLM is suited to deal with challenges of unequal frequencies of measures across individuals, missing observations, and different sources of data (time-varying measures, and person and environment related characteristics). Assuming a primary interest in temporal trajectories, MLM provides an estimation of population-average time trend, allowing for random variations around the average for individuals. To explain some of the variability in growth curves around the mean, MLM modifies the curve based on significant personal characteristics, fine-tuning the average for specific subgroups.

In MLM, the growth curve fitting process is designed for a homogenous group of people who follow, in principle, the same developmental pattern with minor deviations due to individualities. For instance, one can hypothesize that socially deviant behavior gradually increases across teenage years, peaking at the age of 18 – 19 years old, and then decreases as a result of maturation, so that the developmental trajectory follows a quadratic shaped curve (example inspired by Muthén, 2002). It may also be natural to expect boys to exhibit more of such behavior than girls, therefore, having a higher starting point, a steeper increase, and a higher peak. Allowing gender-specific differences in development, the qualitative patterns of change, however, are kept similar. Both groups are expected to start low, increase, and gradually decrease over time, although specific starting points and growth rates may differ. In this example, three types of parameters define growth trajectories: the intercepts, the linear slopes and the quadratic slopes, which are assumed to share similar distributions across genders. Any explained differences in development are due to a specified personal characteristic, such as gender.

One may argue, however, that the average trajectory does not accurately describe behavioral patterns between individuals. There may be four groups with qualitatively different trajectories: (1) the one we described above with a low start, a gradual raise, and a subsequent decrease; (2) a group that starts high and gradually decreases over time; (3) a group that starts low and increases, without a subsequent reduction in deviant behavior, and (4) a group that stays high throughout all years of observation. Each of the defined trajectories is an MLM model in itself with its own growth parameters, a random variation of individuals around the average curves, and a number of personal characteristics explaining variations around averages. Models for each of the developmental patterns are based on a different subsample of individuals with most people falling into Groups 1 and 2, and very few being in the last two groups. The patterns can have different shapes and a different set of predictors. Finally, each of the groups may have different chances of committing a crime later in life – a distant indicator of qualitative differences between groups.

The MLM methodology, however, would not be able to provide model estimations for four types of behavioral trajectories identified above (if these behaviors can not be predicted from a set of observed covariates), as one of its assumptions is that individuals share the same underlying average path of developmental. GMM, on the other hand, would be able to detect the differences (Muthén, 2002; Muthén & Shedden, 1999).

Growth Mixture Modeling as an Approach to EMA Data Analysis

One of the greatest advantages of EMA data is its ability to capture developmental patterns that are reflecting human individuality. In analyzing such data, it is natural to search for common prototypes; however, a simple average may not capture the complexity and variety of human behaviors. This is why *growth mixture modeling*

(GMM) may be an advantageous approach both for summarizing data and preserving existing differences.

While building upon the advantages on MLM, GMM extends it by relaxing the assumption of population homogeneity in the process of searching for common developmental patterns (Muthén, 1997). Assuming that data follows several distinct trajectories (with the number specified by the analyst), GMM produces unique parameter estimates for each of them. Individuals who share the same developmental patterns are identified as members of the same latent class. The class is latent simply because there is no observed indicator of a class membership prior to data analysis. Similar in nature to other analytic techniques dealing with unobserved grouping variables such as latent class analysis or latent profile analysis (McCutcheon, 1987), GMM estimates posterior probabilities of class membership for every individual in the sample. In line with the previous example, a person who exhibits deviant behavior at earlier age but gradually improves may have a 90% chance of falling into the second class, but s/he still has a (very small) chance of belonging to any of the other three classes.

The process of determining the number of latent classes may be confirmatory or exploratory. A data analyst may simply confirm the number of behavioral patterns in situations where previous research explicitly justifies them. For a new study, however, a researcher may rely more heavily on statistical methods to guide her through the process of class identification. In both instances, the interpretation of classes is very important as developmental trajectories are intended to reflect existing phenomena rather than data peculiarities.

Besides determining the number and kinds of developmental patterns, it is also important to learn what personal characteristics may serve as predictors of class membership for the purposes of intervention. For instance, parental involvement in their children's lives may determine whether children act antisocially. By knowing that early demonstration of deviant behavior is common for children with low parental involvement, and that children who exhibit early deviant behavior are also more likely to continue with such behavior over the years, such children and families may be targeted for an intervention. In educational, psychological, or medical settings, early identification of at-risk individuals may preclude negative consequences by means of an early intervention. At the same time, resources may be preserved on those who do not require them.

To date, GMM was applied to longitudinal data with a limited number of repeated observations (Muthén, 2000, 2001a). While EMA may extend growth trajectories and make them more precise by providing more data for every individual, it also carries unique complications. This dissertation evaluates the feasibility of the GMM application to EMA. Data that are used in this study came from a smoking cessation trial carried out at the Department of Psychiatry and Behavioral Sciences at the Memorial Sloan Kettering Cancer Center. Inherent differences associated with the process of smoking cessation provide an opportunity to learn about patterns of quitting behavior, to determine if particular patterns lead to more successful outcomes, and to identify individual profiles associated with each quitting trajectory.

Smoking Cessation Data

Data Overview

The smoking cessation trial was designed to test the effect of the *scheduled reduced smoking* (SRS) intervention in a sample of newly diagnosed cancer patients awaiting surgery. SRS is a behavioral intervention, where the inter-cigarette intervals are gradually increased over the course of several days and where tobacco users are prompted to have a cigarette at designated times rather than at their convenience. The intervention mechanism aims to increase self-efficacy of quitting smoking, to allow practice of coping with smoking urges prior to a complete abstinence, and to break associations between environmental and behavioral stimuli of tobacco use (Cinciripini et al., 1995; Cinciripini, Wetter, & McClure, 1997).

EMAs were only collected in the intervention arm, where *personal digital assistants* (PDAs) delivered a smoking schedule and recorded cigarettes that were smoked, missed, postponed, or smoked off-schedule on a daily basis. Every study participant had an individualized tapering program, which reflected their wake time, smoking rates, and surgery dates. All patients were strongly encouraged to follow the protocol due to medical necessity to abstain from tobacco prior to surgery. However, many users altered the schedule as they decided to quit abruptly on their own, were unable to comply with too few allocated cigarettes, or extended their schedule by failing to quit on a designated day (Ostroff, Burkhalter, Li, Shiyko, & Holland, 2008).

As a result, every study participant provided data that can be summarized in a trajectory of smoked cigarettes over the intervention time period, averaging ten days and ranging from 1 to 30 days. At the time of surgery admission, all participants were screened for their smoking status, based on self-report and biochemical verification.

Upon enrollment, all patients were assessed on a number of personal descriptors (Ostroff & Burkhalter, 2007), including socio-demographic characteristics, tobacco use history and nicotine dependence [Fagerstrom Test for Nicotine Dependence (Fagerstrom, Heatherton, & Kozlowski, 1990; Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991)], medical information such as cancer site, anxiety and depression levels [the Hospital Anxiety and Depression Scale, HADS (Snaith & Zigmond, 1986)], as well as quitting self-efficacy [ten items from The Confidence Questionnaire Form (Baer, Holt, & Lichtenstein, 1986)]. The original quit date (OQD) was selected by every patient based on surgery timing as well as personal preferences. It reflects the length of the originally planned tapering period. Table 1 summarizes measures taken at baseline and used in this study.

For the purpose of this dissertation, only EMA data collected from 74 patients in the SRS arm are used for the analyses. The momentary records of smoked cigarettes are used to recreate smoking profiles for every study participant across the intervention period. With the GMM approach, the smoking trajectories are represented by rates of daily smoking, computed from the reconstructed smoking profiles; smoking status at surgery can serve as an indicator of quitting success rates; and baseline personal characteristics can assist in defining profiles of patients following a particular behavioral pattern of development during the intervention period.

Table 1

Study participants' characteristics

Participants' Characteristics		SRS N = 74	Data Transformation
Gender (female)		36 (49%)	
Age		56.1 (10.3)	Centered at 55.0
Ethnicity (non-White)		13 (18%)	
Education (above High School)		46 (62%)	
Employment Status (Employed)		41 (55%)	
Original Quit Date – OQD		9.8 (4.0)	Centered at 10.0
Cancer Site (thoracic vs. others)		20 (27%)	
Anxiety		6.2 (3.2)	
Depression		9.7 (3.9)	
Smoking Related Variables	Nicotine Dependence: Fagerstrom Score	4.8 (1.8)	Centered at 5.0
	Number of Years Smoking	34.5 (12.7)	Centered at 35.0
	Self-Efficacy to Quit Smoking by Surgery	51.7 (19.9)	Centered at 50.0
Distal Outcome	Smoking at Surgery Admission (quit)	36 (49%)	

Statistics are N and percentages for dichotomous variables, and means (with standard deviations in parentheses) for continuous variables.

Differences in Smoking Trajectories

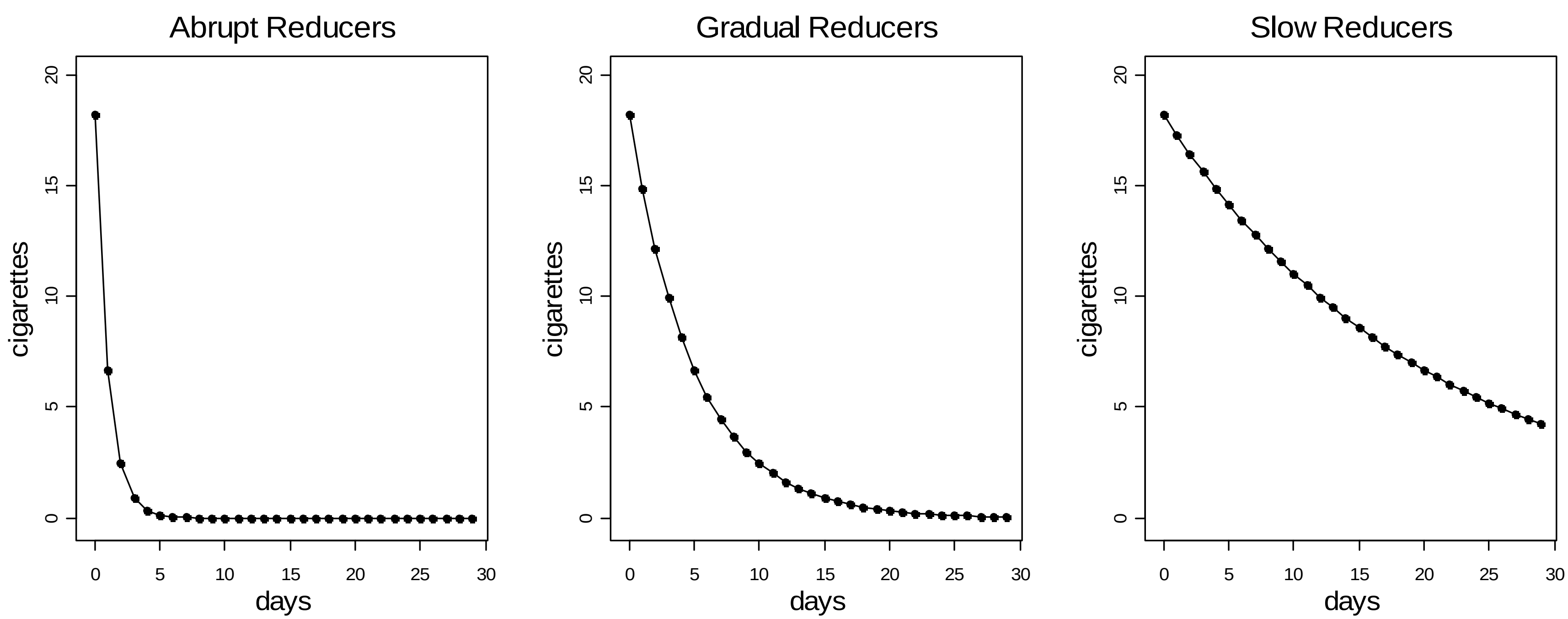
Due to the nature of the SRS program, patients were expected to reduce their daily rates of smoking over the intervention period, exhibiting a downward trajectory over time. Largely, the tapering schedule was influenced by how much time patients had before surgery and their initial smoking rates. Beyond that, patients had some flexibility in determining their target quit date based on personal preferences. From preliminary adherence analysis, only about half of the sample followed the originally set quit date, with the other half speeding or postponing it (Ostroff et al., 2008). As a result, there appears to be a great deal of variability in patients' quitting behavior, which may reflect qualitatively different patterns of smoking cessation. Moreover, some of the patterns may be associated with more successful outcomes than others.

We can hypothesize about quitting trajectories based on preliminary examination of a sample of patients and previous research that provides some insight into existing behavioral patterns in smoking cessation. The literature identifies two major quitting prototypes: the first one involves a gradual reduction in the number of cigarettes smoked on a daily basis prior to a complete abstinence, while the second involves an abrupt abstinence and is referred to as "cold turkey" (Cheong, Yong, & Borland, 2007). There are mixed findings on the success rates among the two methods (Cheong et al., 2007; Cinciripini et al., 1995; Gunther, Gritsch, & Meise, 1992; Marcos, Godas, & Corominas, 2004).

Due to the design of the SRS intervention, we would expect most participants to follow the gradual reduction behavior. However, since many patients initiated an early quit attempt or prolonged their cessation period, we presume that the average trajectory would not reflect the diversity of quitting patterns. Therefore, we are looking for two or more groups that may reflect differences in behavior more accurately and hypothesize about three plausible trajectories, summarized in Figure 1 as groups of:

- “gradual reducers”: individuals who follow the tapering program, starting at a high smoking rate and gradually reducing it;
- “abrupt reducers”: individuals who abruptly stop smoking early in the intervention, although the exact timing differs across individuals;
- “slow reducers”: individuals who seem to be struggling with the schedule and their trajectory of cigarette reduction is shallow or, possibly, flat. Their smoking rates stay relatively high throughout the intervention.

Figure 1. Trajectories of smoking behavior over the course of the SRS intervention for three hypothesized latent classes.



Conceptual Relationships between Study Measures

Once the latent groups with different cessation patterns are identified, it is important to learn whether selecting a particular quitting approach has an effect on the ultimate goal of the program – cigarette abstinence. For instance, "slow reducers" may have lower quitting rates by surgery day compared to "gradual" or 'abrupt" reducers. Therefore, quitting rates for each behavioral pattern will be estimated and compared.

Assuming that behavioral responses to intervention have an impact on abstinence rates, personal characteristics may provide an insight into the profiles of individuals who tend to exhibit a particular kind of response. Taking into account background variables, important personal characteristics will be identified to predict what latent pattern a particular person is more likely to follow. For instance, people with high baseline quitting self-efficacy and less addiction may prefer a "cold turkey" approach as they are sure in their ability to quit and are not strongly dependent on the nicotine, while patients with high addiction and lower confidence in their ability to quit on their own may rely mostly on the program. Establishing profiles of patients falling in each latent category would help to learn about intervention efficacy for various groups of people, to identify at-risk individuals that fail to quit and to learn about possible similarities in their behavior during the intervention period, and, finally, to tailor interventions based on individual needs of patients.

With the use of GMM analysis, we intend to answer the following research questions related to the smoking cessation data:

1. How many developmental trajectories are needed to describe major smoking cessation patterns of behavior in this group of patients following the SRS intervention?
2. If more than one pattern is identified, what are qualitative differences between them?
3. Are different cessation patterns related to abstinence rates at surgery?
4. What personal characteristics predict the likelihood of exhibiting a certain type of cessation behavior?

The next section offers details on the growth mixture modeling methodology, providing a conceptual as well as statistical overview.

Growth Mixture Modeling

Overview

Growth mixture modeling is a statistical approach to modeling longitudinal data. GMM integrates and extends a number of modeling frameworks such as latent class analysis, finite mixture modeling, and structural equation modeling (Muthén, 2001a, 2004; Muthén & Asparouhov, 2008a, 2008b). Referred by some authors as "second generation structural equation modeling," GMM represents a system of equations, connecting time-varying and time-constant, observed and latent variables to each other (Muthén, 2001b). For EMA type of data, however, it may be useful to distance ourselves from the conceptualization of the GMM as an extension of SEM in order to avoid restrictions related to the number of repeated observation as well as their balanced sampling across individuals (Willett & Sayer, 1994).

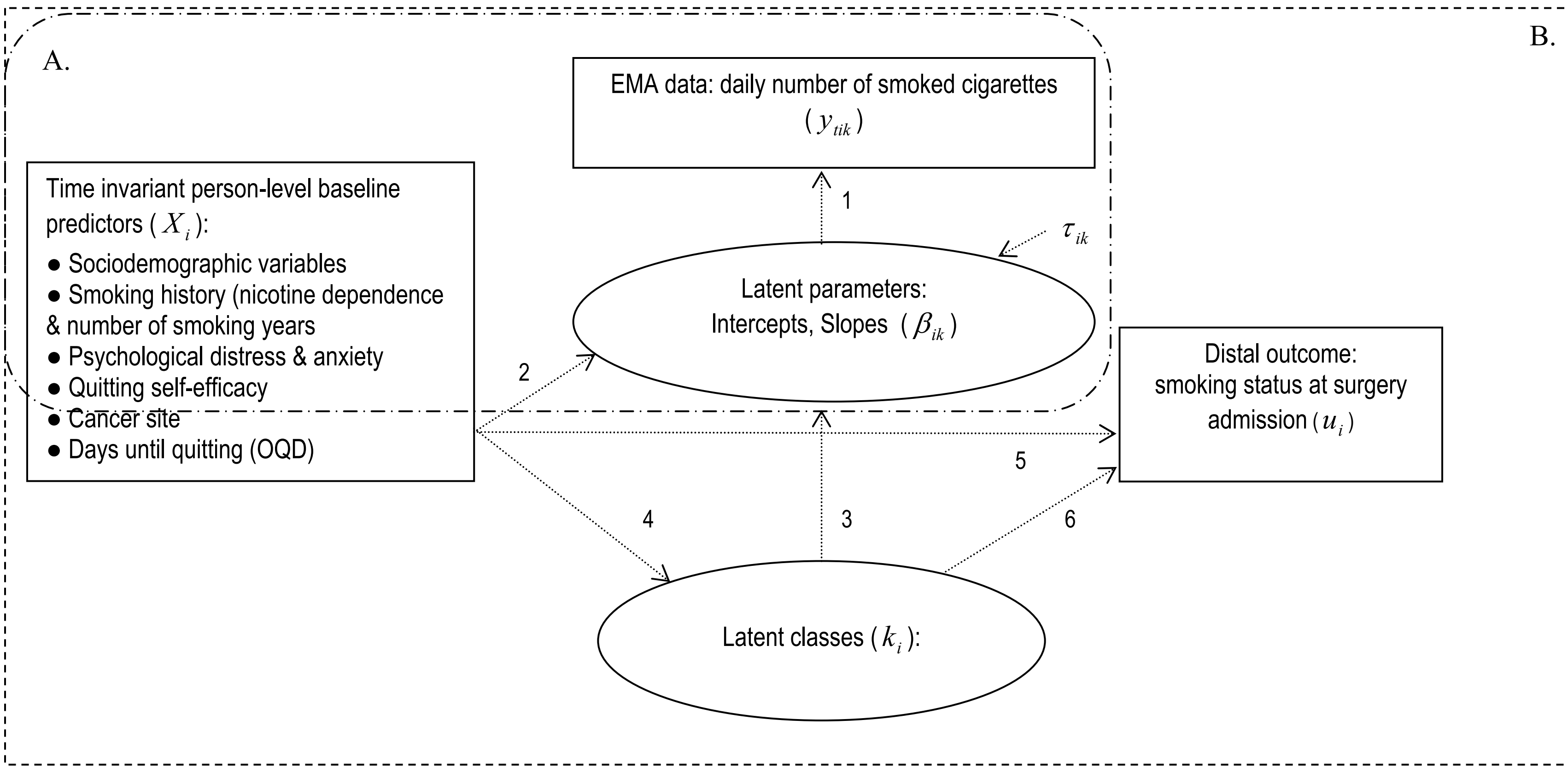
Another commonly used methodology for analyzing longitudinal data is multilevel modeling. It allows great flexibility in how the repeated assessments are modeled, but it is not fit to estimate differences in developmental trajectories based on classes that are not observed. However, relying on the MLM for the conceptualization of the longitudinal data structure would allow tailoring GMM to analysis of EMA data.

Figure 2 displays a conceptual map of a GMM for SRS data, where observed variables are represented as rectangles and latent variables as ovals (inspired by Muthén & Shedden, 1999). The centerpiece of the diagram, framed in part A of the figure, represents a model, where the outcome variable, the daily number of smoked cigarettes recorded for every individual in the study (y_{tik}), is estimated by growth trajectories, defined by intercept and slope parameters β_k . This is the part of the model where repeated assessments collected with large frequency need to be translated into growth trajectories. Arrow 3, connecting latent classes K to the model parameters β_k, as well as the subscript k indicate differences in growth parameters across latent classes. Both growth parameters are allowed to be random within each latent class due to variability in baseline smoking rates and daily cigarette reduction rates. This variability is captured by the variance parameter τ_k.

The time-stable personal characteristics X_i, measured at baseline, play a three-fold role in this model. First, they may account for some inter-personal variability τ_k in intercept and slope parameters β_k within each latent class (link 2). The flexibility of the GMM model allows differences in the effects of baseline covariates across latent classes. Second, X_i serve as predictors of latent class membership, helping to establish profiles of individuals who follow a particular behavioral pattern (link 4). Third, they can have a potential effect on the distal smoking outcome u_i, with patients of certain baseline characteristics being more likely to quit than others (link 5).

Finally, the probability of quitting at surgery is estimated for each latent class. Link 6 captures this relationship. Thus, the probability of quitting u_i reflects the conditional dependence on K. The following sections provide details about each part of the conceptual model outlined in Figure 2.

Figure 2. GMM model for smoking cessation data.



Multilevel Modeling and its Extension to GMM

MLM is a flexible approach to longitudinal data analysis, which takes into account complexities of repeated measures (Raudenbush & Bryk, 2002; Singer & Willett, 2003). MLM differentiates between two levels of data. Individual observations recorded at different time points represent the micro level. They are nested within individuals with their time-stable characteristics, corresponding to a macro level. By separating the variances into intra-person and inter-person components, MLM accounts for dependence of measures taken from the same individuals, allowing repeated observations to correlate with each other, while holding higher level units independent.

The main purpose of the MLM is to estimate developmental trajectories for every person in the study. In the process of maximum likelihood estimation, MLM produces two types of parameters, referred to as fixed and random effects. Fixed effects define an average developmental trajectory for the entire group of individuals, while random effects represent fluctuations around the averages.

In SRS data, with a single level of nesting where observations are nested within patients, the model can be expressed as a system of regression equations at two levels. The first level defines an individual trajectory for every person in the study by estimating intercept and growth slope parameters. These individual parameters, estimated at the micro level, appear as outcomes at the second level of the model, where they are regressed on sample-wide intercepts, adjusting for person-level covariates.

For the current study, the outcome y_{it} is a count, which we assume to be a Poisson-distributed variable, modeled by the probability density function:

$$P(y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad for\ y = 0, 1, 2, \ldots,$$

where λ is the mean (and standard deviation) parameter. The log of the daily smoking rate y_{it} is estimated by a natural logarithm of the expected value $\ln(\lambda_{ti})$ (Agresti, 2002). Incorporating the Poisson link function, the following MLM model (where $K = 1$) defines smoking developmental trajectories for patients in the SRS study (notation adopted from Raudenbush & Bryk, 2002, chap. 6 [1]):

$$\ln(\lambda_{ti}) = \pi_{0i} + \pi_{1i} * days_{ti}, \tag{1}$$

$$\pi_{0i} = \beta_{00} + \beta_{01} * X_i + r_{0i} \tag{2}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * X_i + r_{1i}$$

In this model, the natural logarithm of a smoking rate on day t for a person i is determined by two types of parameters: the intercept π_{0i} and the slope π_{1i}. The residual ε_{ti} is omitted in the first level of the model as the outcome $\ln(\lambda_{ti})$ represents an expected rate value of the Poisson distributed data. After accounting for non-normality in the outcome variable with the Poisson link function, the rest of the model follows a linear format. The variable *days* is centered at baseline, coded as '0', and the last day taking a value of $(\max(days_i) - 1)$. Due to differences in intervention length across individuals, the value for the last day in our data varies, with the maximum being 29 days.

In Equation 1, π_{0i} represents the rate of smoking for person i, measured on the logarithmic scale, at the launch of the study (i.e., $days = 0$). The slope π_{1i} represents the rate of daily change in smoking for a particular individual i. These parameters, estimated at the first level of the MLM, appear as outcomes at the second level, where they are estimated from the corresponding intercepts and person-level covariates. From Equation 2, the intercept π_{0i} is estimated from the group-based intercept β_{00} and person-level baseline predictors X_i. Allowing for inter-individual differences in initial smoking rates, the random intercept-related error term r_{0i} is added to the model. Similarly, the daily change in the rate of smoking, π_{1i}, is modeled from the overall intercept β_{10} and baseline person-level predictors X_i. A random intercept variable r_{1i} is included to reflect unaccounted differences in slope variations across individuals.

Often, the 2-level system of Equations is combined by substituting Equations (2) in (1), yielding:

$$\ln(\lambda_{it}) = \beta_{00} + \beta_{01} * X_i + \beta_{10} * days_{ti} + \beta_{11} * X_i * days_{ti} \qquad (3)$$

$$+ r_{0i} + r_{1i} * days_{ti}.$$

The combined model is more explicit about individual trajectories being estimated from the overall group parameters: $\beta_{00} + \beta_{01} * X_i$ for the intercept and $\beta_{10} * days_{ti} + \beta_{11} * X_i * days_{ti}$ for the slope – both representing fixed effects of the model. The random effects $r_{0i} + r_{1i} * days_{ti}$ reflect individual differences in those parameters. The distributions of all residuals in the model are assumed to be independent across individuals and normally distributed with a mean of 0 and estimated variances τ_{00} and τ_{11}:

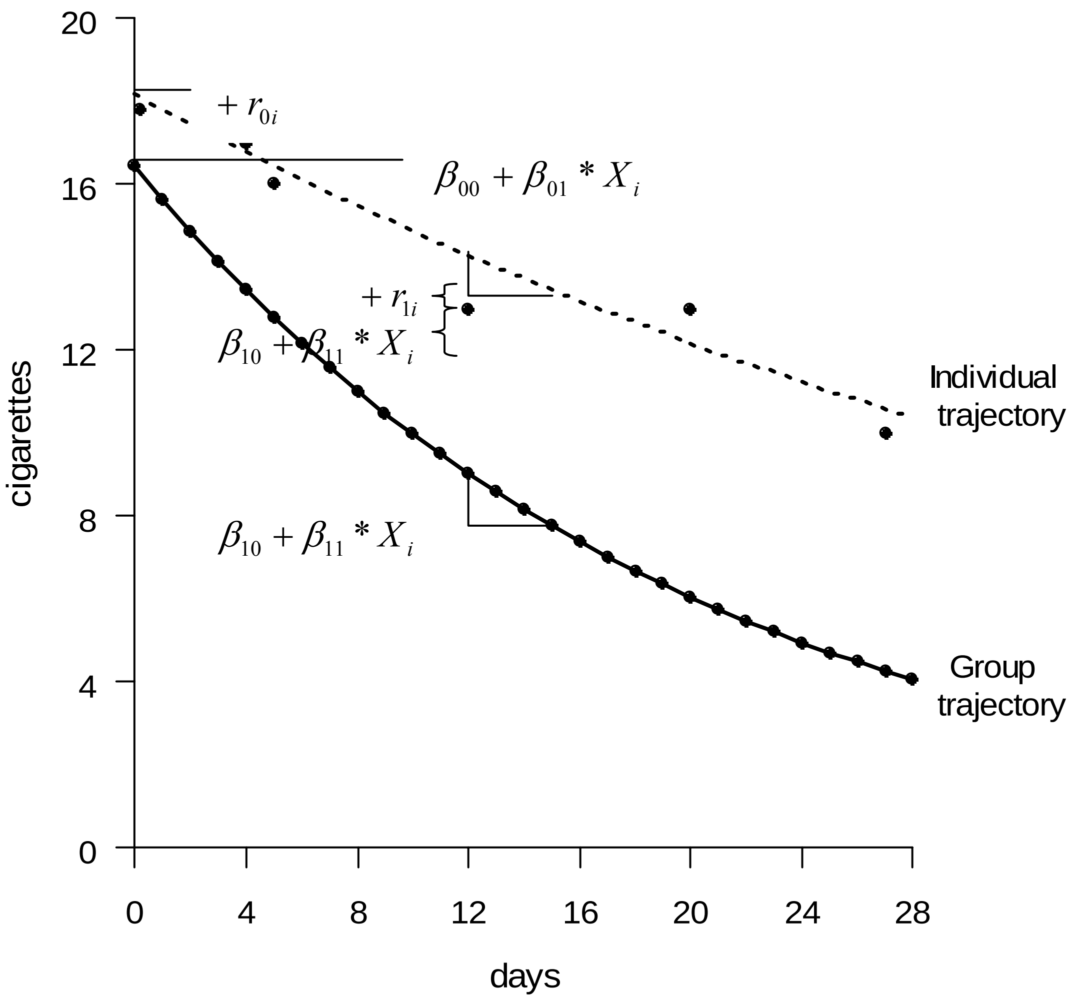
$$r_{0i} \sim N(0, \tau_{00}) \qquad (4)$$

$$r_{1i} \sim N(0, \tau_{11}).$$

The covariance τ_{01} between intercept and slope random terms is also estimated to allow for correlation between the initial rate of smoking and the magnitude of daily tapering.

A graphical representation of the model parameters is given in Figure 3, where a developmental trajectory from a hypothetical individual is plotted against a group-based trajectory estimated from data for a particular combination of independent variables X_i. The group trajectory is defined by fixed intercept and slope parameters presented in the Equation 3. Differences between the group and individual trajectories are captured in the intercept and slope random effects, r_{0i} and r_{1i}.

Figure 3. Graphical representation of the multilevel model.



In summary, MLM can be used to learn about the overall behavioral pattern of smoking cessation for patients in the study. Fixed effects provide an estimate of general development, taking into account patient baseline characteristics. Random effects are incorporated to reflect individual differences in trajectories.

We stress the importance of handling time in a flexible way as a big advantage of MLM for analyzing EMA, where staggered observations and observations of unequal length are common. In MLM, time is treated as a continuous random variable, taking various values across individuals. The univariate format of a data structure assumes a single column of data for the outcome vector y_{it}, accompanied by the variable day_{ti}, specifying timing of each observation (Table 2). This format of data structure is referred to as the *long format*, and it avoids matching observational points across individuals. In this approach, the time variable is frequently conceptualized as continuous. In comparison, the *wide or multivariate format*, summarized in Table 3, forces the observation alignment across individuals for every assessment time, yielding a large volume of missing data for study designs where individuals are observed frequently at different time points. In this format, time is commonly treated as a categorical variable. This data structure approach is widespread in the structural equation model analysis of longitudinal data.

A related advantage of MLM is its flexibility in handling missing data at the first level of the model (Raudenbush & Bryk, 2002). In Figure 3, an individual developmental trajectory was estimated from six data points. For a different person, with only three repeated observations, the trajectory would still be estimated based on the available data. Thus, an individual with fewer data points is not discarded from the analysis. Instead, MLM incorporates all available information to estimate parameters with a maximum likelihood algorithm, which has an underlying assumption of data *missing at random* (Allison, 2001; Little & Rubin, 1987). This approach maximizes the number of individuals for the analysis, thus improving the overall power. This quality is essential for the analysis of EMA, as missing data are very common due to a heightened burden of responses that occur multiple times throughout a day and, often, in situations when a person is not available.

Table 2

Long format of longitudinal data structure

Subject ID	Cigarettes per day (Y_{it})	Days (t)	Age (X_i)	Quit at Surgery (u_i)
01	20	0	62	0
01	18	1	62	0
01	17	2	62	0
01	12	4	62	0
01	40	0	55	0
01	36	1	55	0
01	36	2	55	0
01	25	7	55	0
…	…	…	…	…
74	15	1	54	1
74	3	4	54	1
74	0	6	54	1
74	0	7	54	1

Table 3

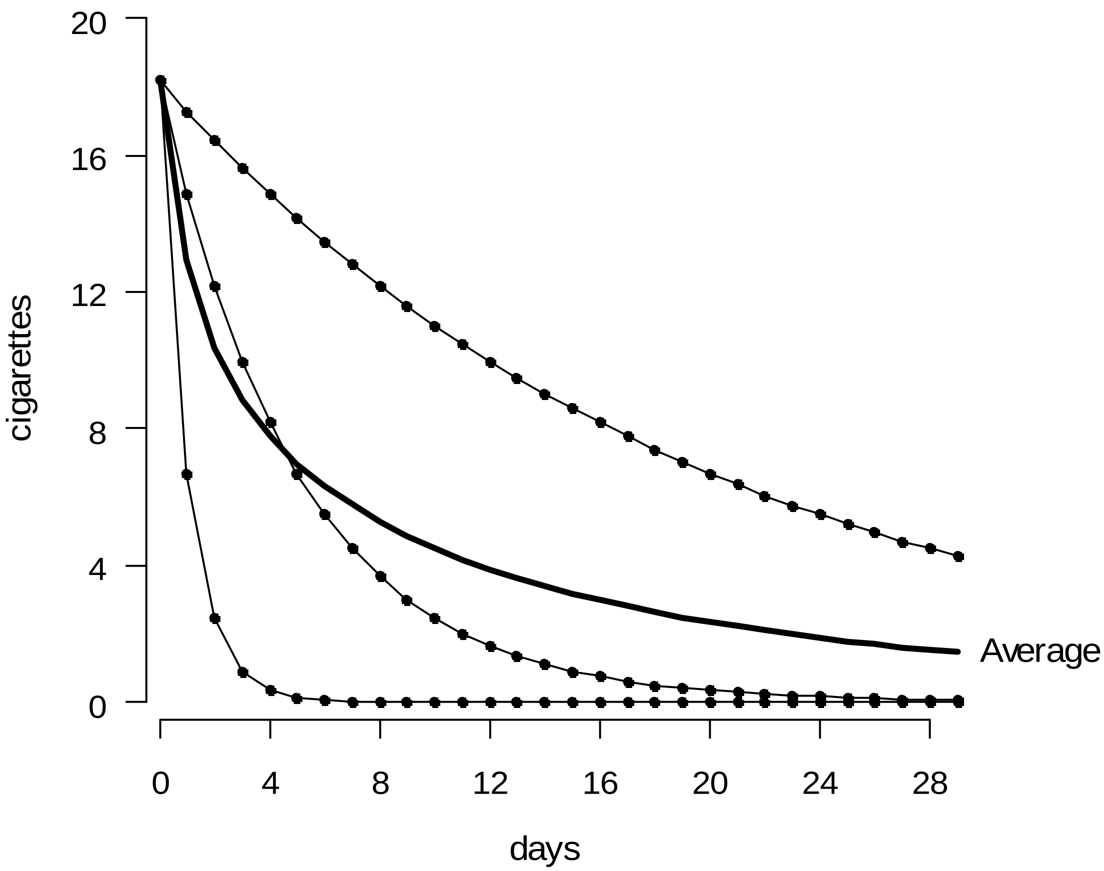
Wide format of longitudinal data structure

Subject ID	Cigarettes per day (Y_{it}) Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Age (X_i)	Quit at Surgery (u_i)
01	20	18	17	NA	12	NA	NA	NA	62	0
02	40	36	36	NA	NA	NA	NA	25	55	0
…	…	…	…	…	…	…	…	…	…	…
74	NA	15	NA	NA	3	NA	0	0	54	1

Extending the Multilevel Model to Incorporate Latent Classes

MLM advantages notwithstanding, its application is limited to data coming from a single population with common population parameters (Muthén, 2004). In Equations 1-3, we assumed that a mean-based growth curve represents a true population-wide trajectory of development. However, based on data exploration and literature review, we expect important differences in behavioral responses to the SRS intervention. While random effects of the MLM model take into account individual fluctuations around the average, they are reflective of sampling variability rather than qualitative differences in development. An assumption of a single underlying population trajectory precludes MLM from identifying distinct developmental patterns from a mixture, especially in situations where observed covariates can not predict differences in class trajectories perfectly. As demonstrated in Figure 4, the average trajectory is able to capture only a general developmental trend such as the overall decline in smoking over the course of the intervention. However, fine granulated behavioral differences are omitted. From the three hypothetical classes, previously presented in Figure 1, the average seems to capture behavior of gradual reducers, those, who follow the PDA regimen. However, non-standard behaviors of abrupt and slow reducers fall outside the prediction range of the average trajectory.

Figure 4. The average MLM estimated developmental trajectory versus three hypothetical latent developmental classes.



To overcome this limitation, GMM extends MLM by allowing model parameters to vary across latent classes K. It produces an individual set of parameters for each developmental trajectory. This part of the GMM can be summarized in the following series of equations:

$$\ln(\lambda_{itk}) = \pi_{0ik} + \pi_{1ik} * days_{ti} \tag{5}$$

$$\pi_{0ik} = \beta_{00k} + \beta_{01k} * X_i + r_{0ik} \tag{6}$$

$$\pi_{1ik} = \beta_{10k} + \beta_{11k} * X_i + r_{1ik}\,.^{2}$$

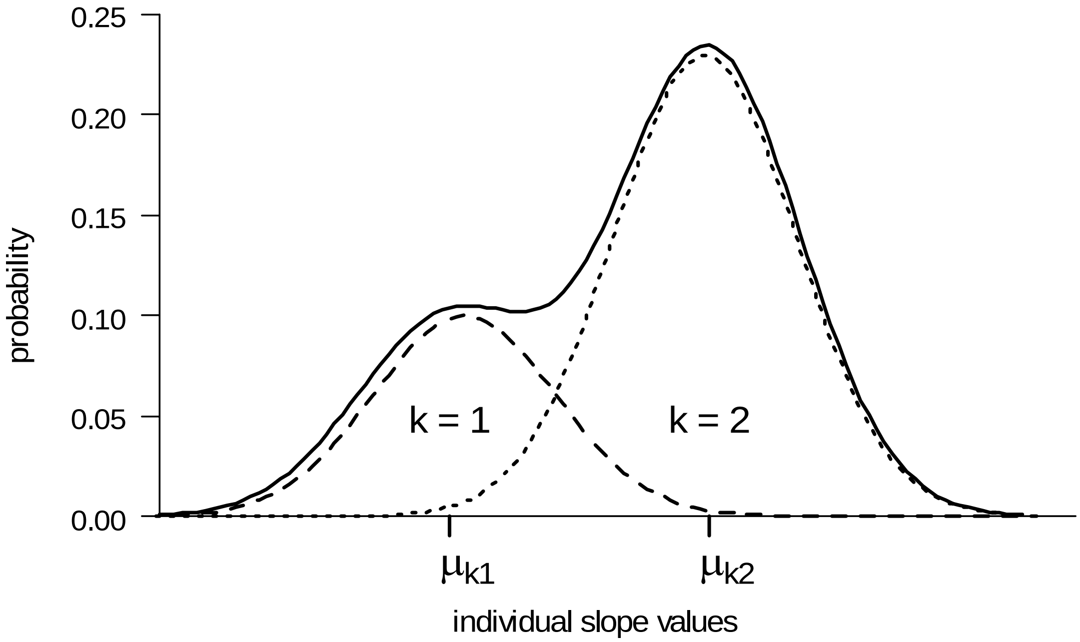
When Equations 5 and 6 are combined, the model simplifies to

$$\ln(\lambda_{itk}) = \beta_{00k} + \beta_{01k} * x_i + \beta_{10k} * days_{ti} + \beta_{11k} * x_i * days_{ti} + r_{0ik} + r_{1ik} * days_{ti} \tag{7}$$

In principle, Equations 7 and 3 are very similar. They are estimating a developmental trajectory defined by fixed intercept and slope parameters, with random variations reflected in residuals. The major difference, however, lies in adding the subscript k to all parameters, highlighting that model parameters are no longer common for all individuals in the sample. Rather, they are class specific.

Whereas in MLM all model parameters were assumed to have a single underlying distribution, in GMM, every parameter is represented by a mixture of distributions with their own means and variances. As a result, parameter estimates are conditional on the class they represent. Figure 5 demonstrates an example of a distribution mixture for a model parameter with two latent classes. For instance, if slopes do not share a single underlying normal distribution, their plot may deviate from a symmetric bell curve. The non-normality is created as a result of a mixture of two normal distributions with corresponding parameters μ_k and σ_k^2.

Figure 5. A hypothetical mixture of two normal distributions of slope parameters.



The concept of model parameter variability across classes is applied to fixed and random effects, and the following equations explicitly define this class variability:

$$\beta_{00} \sim N(\mu_{0k}, \sigma_{0k}^2) \quad (8)$$

$$\beta_{10} \sim N(\mu_{1k}, \sigma_{1k}^2)$$

$$(r_{0ik}, r_{1ik}) \sim BVN(0, \tau_k)$$

$$\tau_k = \begin{bmatrix} \tau_{00k} & \tau_{01k} \\ \tau_{10k} & \tau_{11k} \end{bmatrix}$$

Here, intercepts and slopes are assumed to come from a mixture of two or more normal distributions, defined by their corresponding mean and variance parameters. The distribution of random effects may also vary across classes.

Modifying MLM to allow differences across latent classes makes it possible to define the model in a flexible way. Specifically, the model parameters can be modified depending on data fit and model interpretation. For instance, a linear fit can be examined for one latent class and a quadratic development for a second class. In a similar fashion, random effects r_{0ik} and r_{1ik} may be present in one class but constrained to zero for others. Finally, the influence of person-level covariates X_i may differ across developmental classes, explaining the intercept variability in one class, but not the other. Due to model specifications at a class level, GMM is flexible in describing distinct developmental patters for a heterogeneous sample of individuals.

Relating Baseline Predictors, Latent Classes, and Distal Outcomes

Another important part of the GMM relates person-level baseline covariates X_i to latent classes K, defining profiles of patients representing each latent group. This relationship is statistically characterized by the multinomial logistic regression for unordered responses (Muthén, 2001b, p.297). In the case of three latent classes, selecting "gradual" quitters as a reference category, there are two equations modeling the likelihood of being a member of a particular class, expressed on the *logit* (logarithm of odds) scale:

$$\ln\left[\frac{P(K = abrupt \mid X_i)}{P(K = gradual \mid X_i)}\right] = \omega_{01} + \omega_{11} * X_i \qquad (9)$$

$$\ln\left[\frac{P(K = slow \mid X_i)}{P(K = gradual \mid X_i)}\right] = \omega_{02} + \omega_{12} * X_i$$

The likelihood of being in a particular class versus being in a reference class is modeled as a function of the intercept ω_{0k} and person-level baseline covariates X_i. In the process of model estimation, every individual has a set of non-zero posterior probabilities of being a member of each latent class. A person's most likely class is chosen based on the highest probability value.

Finally, the ultimate outcome of the study, smoking status at surgery u_i, measured on a binary scale: "quit" versus "smoking", is linked to baseline covariates X_i and latent classes K by means of a binary logistic regression:

$$\ln\left[\frac{P(u_1 = 1 \mid k_i, X_i)}{P(u_1 = 0 \mid k_i, X_i)}\right] = \nu_{0k} + \nu_{1k} * X_i \qquad (10)$$

Here, the log odds of quitting at surgery ν_{0k} is conditional on latent class membership as well as baseline personal characteristics, estimated by X_i. The slope parameter ν_{1k} can vary across classes, allowing some predictors to play a role for one latent group but not for others.

Software for GMM Analysis

In all parts of the GMM model defined above, latent classes play a central role in the estimation process. Information on latent classes is incorporated in defining developmental trajectories for each subgroup of individuals, in determining profiles of patients following certain behavioral patterns, and in looking at short- and long- term abstinence rates. Were the latent class variable observed, the estimation process would be straightforward. However, in a situation where groups are undefined or latent, the maximum likelihood estimation process is based on an iterative procedure known as the *expectation maximization* (EM) algorithm, developed by Dempster, Laird, and Rubin (1977).

To date, the only software available for estimating GMM, employing the expectation maximization algorithm is Mplus [3], developed by Bengt and Linda Muthén and Tihomir Asparouhov (Muthén & Muthén, 2007). The software is heavily influenced by the structural equation modeling methodology, but it is flexible enough to allow estimation of the longitudinal data through the MLM's perspective of nested data. This allows applying GMM to EMAs.

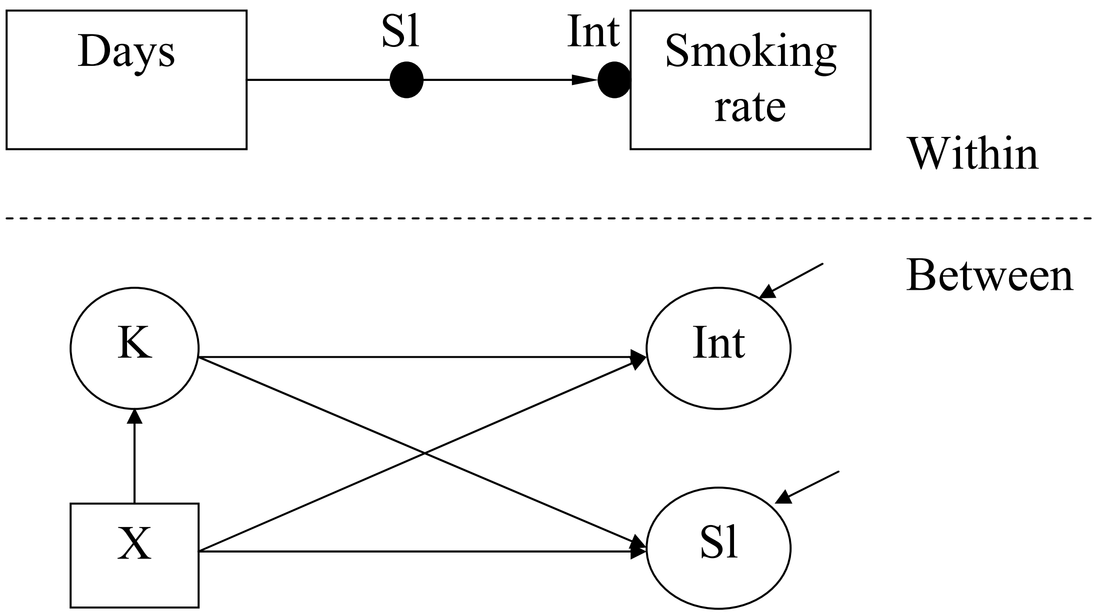
In Mplus, most longitudinal analyses are conceptualized as a single-level problem, where data are structured in a multivariate format (i.e. Table 3). A multilevel approach is not common for GMM; however, in special cases, like EMA, with large differences in length of recorded data across individuals and large volumes of missing data, the multi-level approach is preferred. Having too many observations stretched in a wide format, where columns represent measurement points, would place a huge demand on the model estimation process. In fact, during the exploratory stage of our analysis, this approach failed, and model estimation became possible only after eliminating data measured beyond the first 12 days of observations.

Figure 6 provides a conceptual overview of GMM in Mplus and it can be directly translated into syntax, examples of which are summarized in Appendices 2 and 3. Similar to Figure 2, observed variables $smoking\ \ rate$, $days$, and X are represented by rectangles, and latent continuous and categorical variables K, Int, and Sl are depicted as ovals. The model is separated into two large groups of within and between sections, separated by the horizontal dashed line. The within part of the model corresponds to Equation 5 of the GMM, and the between part is defined by Equation 6. To capture randomness of intercepts and slopes within latent classes, black circles Int and Sl are added to the arrow connecting the time variable $days$ to the outcome $smoking\ \ rate$. Additionally, these parameters are explicitly present as latent in the between model part, with arrows pointing towards them, capturing randomness. The values of Int and Sl depend on the latent person-level class K as well as baseline covariates X.

It is important to note the latent categorical variable K is a between-level variable, illustrating the formation of classes on the personal level. Since individuals are conceptualized to belong to the same latent class for the entire intervention period, the class-level variable is time-stable. It is possible, however, to conceptualize K as a within-level variable, where individuals move between classes across time (see e.g. Asparouhov & Muthén, 2008).

While Mplus was primarily used for growth mixture modeling analyses, data manipulations, graphical summaries, and descriptive statistics were constructed in the open-source statistical package R, version 2.8.2 (R Development Core Team, 2008). The HLM software, version 6.04 (Raudenbush, Bryk, & Congdon, 2007) was used to confirm the one-class GMM solution and to extract random intercepts and slopes for examination of model fit.

Figure 6. Mplus two-level conceptual diagram of GMM for smoking data.



Model Building and Selection Processes

Multilevel Modeling

A Stepwise Approach to Multilevel Model Building

The process of MLM model building can be performed in several stages. The fitting process starts off with the simplest longitudinal model, where the average trajectory of development is estimated for all individuals in the sample. The random intercept and slope parameters are constrained to zero in order to identify the overall developmental pattern:

$$\ln(\lambda_{it}) = \beta_{00} + \beta_{10} * days_{ti} \qquad (11)$$

In the subsequent steps, the intercept β_{00} and the slope β_{10} are modeled as random to allow for variation in baseline smoking rates and in daily reduction of smoking across individuals. This is reflected in the subscripts *i*, signifying individual trajectory parameters for each study participant, and random residuals r_{0i} and r_{1i}:

$$\ln(\lambda_{it}) = \beta_{00} + \beta_{10} * days_{ti} + r_{0i} + r_{1i} * days_{ti} \qquad (12)$$

Finally, a number of baseline predictors (for illustration purposes we use only one X_i), such as Fagerstrom and Self-Efficacy scores, among others (Table 1), are added to the model to explain the randomness of variability in intercept and slope parameters:

$$\ln(\lambda_{it}) = \beta_{00} + \beta_{01} * X_i + \beta_{10} * days_{ti} + \beta_{11} * X_i * days_{ti} + r_{0i} + r_{1i} * days_{ti} \qquad (13)$$

Different baseline predictors can characterize slope and intercept variability, such that the Fagerstrom scale can only be a predictor of the intercept, and the Self-Efficacy scale can only be a predictor of the slope.

Deciding on the Best Fitting MLM

Two models are considered nested if arriving to the second model requires freeing some of the restricted parameters of the first. For example, Models 11 and 12 are nested, since Model 12 was created by freeing random parameters r_{0i} and r_{1i}, which were fixed to zero in Model 11. Only nested models can be compared to each other with a Log Likelihood (LL) ratio test. Its null hypothesis (H_0) states that the simpler model is sufficient to describe the data, while the alternative hypothesis (H_a) is that additional parameters improve the overall model fit. The decision is made by taking the difference of two LL values from nested models and multiplying the difference by -2. The resulting product follows a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters in the two compared models. This product is compared to the critical value of the chi-squared distribution with an alpha level of .05. If the product is smaller than the critical value, the H_0 is retained, and a simpler model is used.

Individual model parameters are evaluated based on the traditional t-statistics. Those that are not significant are fixed at zero. Generally, if a significant model parameter is added to the model, the overall LL ratio test would also yield significant model improvement. Based on the principle of model parsimony, the final model should contain only parameters that significantly (or meaningfully) contribute to the overall model description.

Predicting Smoking Status at Surgery Admission from Smoking Trajectories Estimated by MLM

One of the MLM assumptions is that trajectories of smoking cessation behavior, defined by intercept and slope parameters, share the same underlying distributions. Some of their variation from overall means can be predicted from person-level descriptors, and the rest is random. Despite a single-class assumption, it is possible that the magnitude of baseline smoking and daily rates of cigarette reduction helps to differentiate between quitters and smokers at the end of the intervention period. To test this hypothesis, random intercept and slope parameters (if randomness is statistically justified) will serve as predictors of the smoking status at surgery in the logistic regression analysis, incorporated in the overall model fitting procedure in Mplus:

$$Logit\ (P(Quit_{\ i})) = \nu_0 + \nu_1 * Intercept_{\ i} + \nu_2 * Slope_{\ i} \qquad (14)$$

Significance of individual tests, the magnitude, and directionality of the parameters ν_1 and ν_2 would inform how the probability of quitting at surgery changes with changes in baseline smoking rates and daily reduction in smoking. The Mplus software allows estimation of the model defined in Equations 13 and 14 simultaneously (see Appendix 2 for syntax details).

The final model is explored for the overall fit to data by extracting individual intercept and slope parameter estimates and studying their distributions. Corresponding residuals are also investigated for normality with graphical and statistical summaries.

Enumeration and Specification of a Growth Mixture Model

The process of building a growth mixture model is complex and can be subdivided into two procedures: model enumeration and model specification. Model enumeration refers to the process of selecting the correct number of latent classes (Nylund, Asparouhov, & Muthén, 2007), which is essential for understanding distinctions in developmental trajectories. Model specification refers to defining model parameters and may entail specification of fixed and random effects; establishing relationship between baseline predictors and growth terms, latent classes, and distal outcomes; and constraining or freeing parameter variability across latent classes. Both procedures are iterative and repeated multiple times until the best fitting model is selected. Both are essential for identifying the most appropriate model.

Model Enumeration

Unless GMM serves a confirmatory purpose, the number of latent classes is rarely predefined. In the process of model enumeration, a researcher would gradually compare models with k = 1, 2, 3, etc number of latent classes until a more complex model does not improve the overall model fit. Seemingly straightforward, this procedure carries a lot of complications (Muthén, 2007; Nylund et al., 2007).

Potential Pitfalls in Estimating the Number of Latent Classes

Bauer & Curran (Bauer & Curran, 2003) raised the issue of overestimating the number of classes when repeated observations follow a non-normal distribution but substantially constitute a single population. In cases of non-normally distributed data, latent classes may emerge artificially, breaking down continuous heterogeneity into smaller homogeneous subgroups. Plausible solutions were offered by Rindskopf (2003) and Muthén (2003). With the Bayesian approach, one would need to generate data from the best fitting model and compare it to the original data; thus, testing posterior versus a priori distributions (Rindskopf, 2003). If the 'a priori' data was not sampled from a multiclass distribution estimated by the model, the fit would be poor. On the opposite, when a multiclass solution is correct, the fit would be close. Currently, this approach is not incorporated in statistical packages available for running GMM analysis. For classical methodologists, Muthén offered an alternative approach, testing multivariate skewness and kurtosis of an estimated model against data (Muthén, 2003, 2004). While the test of multivariate skewness and kurtosis is implemented in the current version of the MPlus software, it is not available for Poisson distributed data (Muthén & Muthén, 2007).

Substantive Model Checking

Another important way of checking model fit is by relying on the substantive theory (Muthén, 2003). While little theoretical background may support or refute $k > 1$ classes, auxiliary information in forms of antecedents, concurrent events, or consequences may shed light on the appropriate model. Antecedents are covariates measured prior to repeated observations. They help to predict class membership, defining profiles of individuals following a particular developmental trend. Not only they are useful in the process of class enumeration, they are also important substantively, allowing early identification of individuals who are likely to follow a certain developmental path. It is recommended that covariates are included early on in the model selection process, as their presence or absence may change the final selection of the number of latent classes (Asparouhov & Muthén, 2008).

Concurrent events are time-varying covariates, which may have different effects across latent classes, when added as covariates or parallel processes. Similarly to antecedents, concurrent events help to differentiate substantively between different developmental profiles. While this may be an important research question in the future, the current study does not incorporate concurrent information.

Finally, consequences, or distal outcomes, are essential for learning whether particular groups of people experience different distal events. Practically, early detection of "at risk" individuals may lead to early interventions. In model selection process, an identification of "at risk" class may justify its existence due to practical importance. Moreover, differences in outcomes for individuals that follow different developmental pattern strengthen the importance of learning about these patterns.

Model Fit Indices

While relying on substantive proof of existence of multiple classes is important, there is a number of statistical tests and fit indices aiding the model enumeration process. Unlike some multilevel models, GMM with different number of classes are not nested; thus, violating the assumption for the LL ratio test. The difference in LL between the models with $k=1$ and $k=2$ does not follow a chi-squared distribution and the alternative hypothesis of better fit of a model with the higher number of classes cannot be tested (Muthén & Asparouhov, 2008a). Therefore, other model fit indices, such as BIC (Bayesian Information Criterion), ABIC (Adjusted Bayesian Information Criterion), LMRT (Lo-Mendell-Rubin test), and BLRT (Bootstrapped Likelihood Ratio test), are used for the purpose of model comparison.

Bayesian Information Criterion (BIC)

The most commonly used index is BIC (Schwarz, 1978), which is computed from LL of the estimated model, correcting for the number of free model parameters *r* and a sample size *N*:

$$BIC = -2 * LL + r * \log(N) .$$

To select a model with the best fitting number of classes, one would look for the smallest BIC value. BIC penalizes complex models with larger number of latent classes as they require more parameters to be estimated and compromise parsimony. By making a sample size adjustment, BIC allows comparing models estimated from unequal samples by design or due to data missingness.

Adjusted Bayesian Information Criterion (ABIC)

A closely related model fit index is ABIC (Sclove, 1987), which modifies the BIC formula by substituting *N* with *(N+2)/24*:

$$ABIC = -2 * LL + r * \log\left(\frac{N+2}{24}\right).$$

BIC is a consistent statistic, meaning, its ability to identify the correct number of classes increases as the sample size increases (Haughton, 1988). ABIC should perform better in situations where the sample size is small, although, there are no guidelines about the cut-off point between small and sufficient sample sizes.

Three Monte-Carlo simulation studies compared performance of information-based indices in samples ranging from 50 to 2000 individuals with a continuous outcome measured at four repeated observations (Nylund et al., 2007; Tofighi & Enders, 2008; Tolvanen, 2007). Varying the proportion of individuals in each latent class and a degree of class separation, researchers concluded that both BIC and ABIC demonstrate good sensitivity. In the Nylund et al study with the smallest sample of 200 individuals, BIC identified the correct number of latent classes 84% of the time (Nylund et al., 2007). The authors warned, however, that BIC's performance worsens as the sample size decreases. In the same study, ABIC performed worse than the BIC in smaller samples, correctly enumerating only 66% of the time. Tolvanen's work demonstrated a good performance of the BIC index in samples as small as 50, with misclassification errors ranging from 5 to 7% (Tolvanen, 2007). ABIC, on the other hand, was reported to have a very high error rate (70% and higher), performing much better in samples with N > 500. Tofighy & Enders recommended ABIC as the best performing fit index, with sensitivity ranging from 80 to 100% in a sample of 1000 individuals (Tofighi & Enders, 2008). BIC's performance, on the other hand, was very inconsistent, ranging from almost 0% to 100%. Overall, it appears that BIC is the most suitable fit index for studies with small samples, while the performance of ABIC is less consistent.

Lo-Mendell-Rubin Test (LMRT)

A modified likelihood ratio test was developed by Lo, Mendell, and Rubin and is referred to as the Lo-Mendell-Rubin test (LMRT) (Lo, Mendell, & Rubin, 2001). Unlike the traditional likelihood ratio test, which requires a chi-square distribution of differences between two compared models, LMRT approximates the distribution of LL differences. With H_0 being that the number of classes equals to *k-1*, it compares the improvement in the model fit from adding an additional class (*K*=*k*). With no significant improvement, the model with the *k-1* number of classes is retained.

Simulation studies demonstrate good sensitivity of LMRT for GMM data, ranging from 81% for a sample size of 200 individuals to 90% for a sample of 500 (Nylund et al., 2007). In Tolvanen's work, LMRT identified the correct number of latent classes in over 95% of cases with a sample size of 50 (Tolvanen, 2007), which is an indicator of good performance.

Bootstrapped Likelihood Ratio Test (BLRT)

Another way to obtain a p-value when comparing two neighboring models is through repeated sampling. One of the most recent techniques for testing non-nested models is the bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000). It draws repeated samples, which constitute a distribution, under the null assumption that the *k-1* class model fits the data well. The p-value of the real difference in LL between *k* and *k-1* models is obtained from the generated distribution. If the difference is not unlikely ($p > .05$), the null hypothesis is retained, if it is rare ($p < .05$), the *k* class model is considered to fit data better. While BLRT was only recently incorporated in the MPlus software, it appears promising in identifying the correct number of classes and having high accuracy (above 90%, Nylund et al., 2007; Tolvanen, 2007).

Besides the aforementioned model fit indices, there are a number of alternative fit statistics that are used for GMM model selection process (e.g. Akaike's information criterion, adjusted Akaike's information criterion). However, their accuracy of detecting the correct number of classes is much lower than the accuracy of the four indices described above. Therefore, for the analyses, we will primarily rely on a combination of four statistical indices of model fit described above as well as substantive checking of the number of latent classes.

Model Specification

Specifying a model in the process of selecting a correct number of latent classes can be a very laborious process, as there are multiple ways to alter the model by changing its parameters. Correct model specification can also affect the number of classes selected. Usually, one would start with the simplest model with fixed growth parameters and no covariates, and progress towards a fuller model, including possible random variability within classes, covariates, and distal outcomes. While the process of model specification is very dynamic and it is rarely performed in a linear fashion, several major stages that are involved in this procedure can be spelled out.

The simplest model would involve mean developmental trajectories across latent classes:

$$\ln(\lambda_{itk}) = \beta_{00k} + \beta_{10k} * days_{ti}, \tag{15}$$

so that individuals that are estimated to be members of the same developmental class share common baseline rates of smoking as well as daily reduction rates. This simplest approach to model specification is often referred to as latent growth curve analysis (Duncan, Duncan, Strycker, Li, & Alpert, 1999; Muthén, 2004) with no random variability at any level of the model.

The next step in the model modification process involves the possibility of freeing intercept or slope variances:

$$\ln(\lambda_{itk}) = \beta_{00k} + \beta_{10k} * days_{ti} + r_{0ik} + r_{1ik} * days_{ti} \qquad (16)$$

This strategy may be questionable in a small sample of data (in this study, N = 74) and may lead to technical difficulties in model convergence (Jung & Wickrama, 2008). However, some random effects are worth exploring. For example, a good approach would be to allow for random variability in intercept or slope parameters only for some latent classes (Asparouhov & Muthén, 2008). Conceptually, for this study, it may be justifiable to allow for randomness in intercepts to account for differences in baseline smoking rates across individuals who belong to the same latent class (provided that freeing intercepts benefits the overall model fit). However, slopes may stay fixed and different in magnitude across classes, to be able to capture mean patterns of smoking cessation.

Further on, intercept or slope predictors in forms of baseline person-level covariates can be added to the model to estimate trajectories more precisely for people with specific baseline profiles:

$$\ln(\lambda_{itk}) = \beta_{00k} + \beta_{01k} * X_i + \beta_{10k} * days_{ti} + \beta_{11k} * days_{ti} * X_i + r_{0ik} + r_{1ik} days_{ti} \qquad (17)$$

Based on this model, an intercept and a slope can be specified separately for men and women (if gender is used as a covariate) in each developmental class. Additionally, class membership can also be predicted from person-level characteristics, so that placement of a person in a latent class is estimated based on longitudinal data as well as baseline predictors (Equation 9).

Inclusion of covariates in Equation 9 is reported to be essential for correct identification of the number of latent classes, for computing class proportions and class membership (Muthén, 2004). There were contradictory findings from a simulation study (Tofighi & Enders, 2008), however, which reported that inclusion of covariates actually increased the probability of incorrect class extraction. Nevertheless, the authors recommend taking this finding cautiously as it appeared that the inclusion of covariates had a different effect across various sample sizes: specifically, being more beneficial for smaller sample data. Inclusion of covariates may complicate the model as more parameters are being estimated, but it may also benefit the model selection process. There are many related decisions that a researcher has to make when including covariates. Some of them pertain to the number of covariates and whether or not they have similar effects across latent classes. These are conceptual as well as model fitting decisions. Statistically, significance of covariates is checked through individual t-tests. For the current study, the alpha level of .1 was chosen due to the exploratory nature of the analyses and the small sample size.

Finally, adding a distal outcome, smoking status at surgery, is an important step in the model building process, since the ultimate goal of defining latent cessation trajectories is to be able to differentiate between successful quitters and those who were not able to quit. After adding a distal dichotomous variable, each class gets an estimation of the probability of smoking at surgery, which can be compared across classes. Baseline personal descriptors can also be used to predict the ultimate outcome of the study (Equation 10).

The process of model specification is essential for correct class extraction. It is achieved through comparison of model fit indices described above for *K=1, 2, 3,* 4, …, in conjunction with conceptual justification.

Other Model Fitting Methods

There are a number of additional ways to assess the fit of the model. They include classification tables (Boscardin, Muthén, Francis, & Baker, 2008; Wang & Bodner, 2007) and graphical summaries (Boscardin et al., 2008; Muthén & Asparouhov, 2008a). Classification tables involve summary of average probability of falling in each latent class for individuals with an assigned class membership, decided based on the likeliest category. High values on the diagonal and small off-diagonals inform that people who were estimated to belong to a particular category are very likely members of that category. Graphical summaries include comparisons of observed and estimated mean curves as well as estimated mean curves and actual developmental trajectories. When many model assessment approaches point to the same solution, the selection of the correct model appears to be justified.

Data Manipulations and Structure

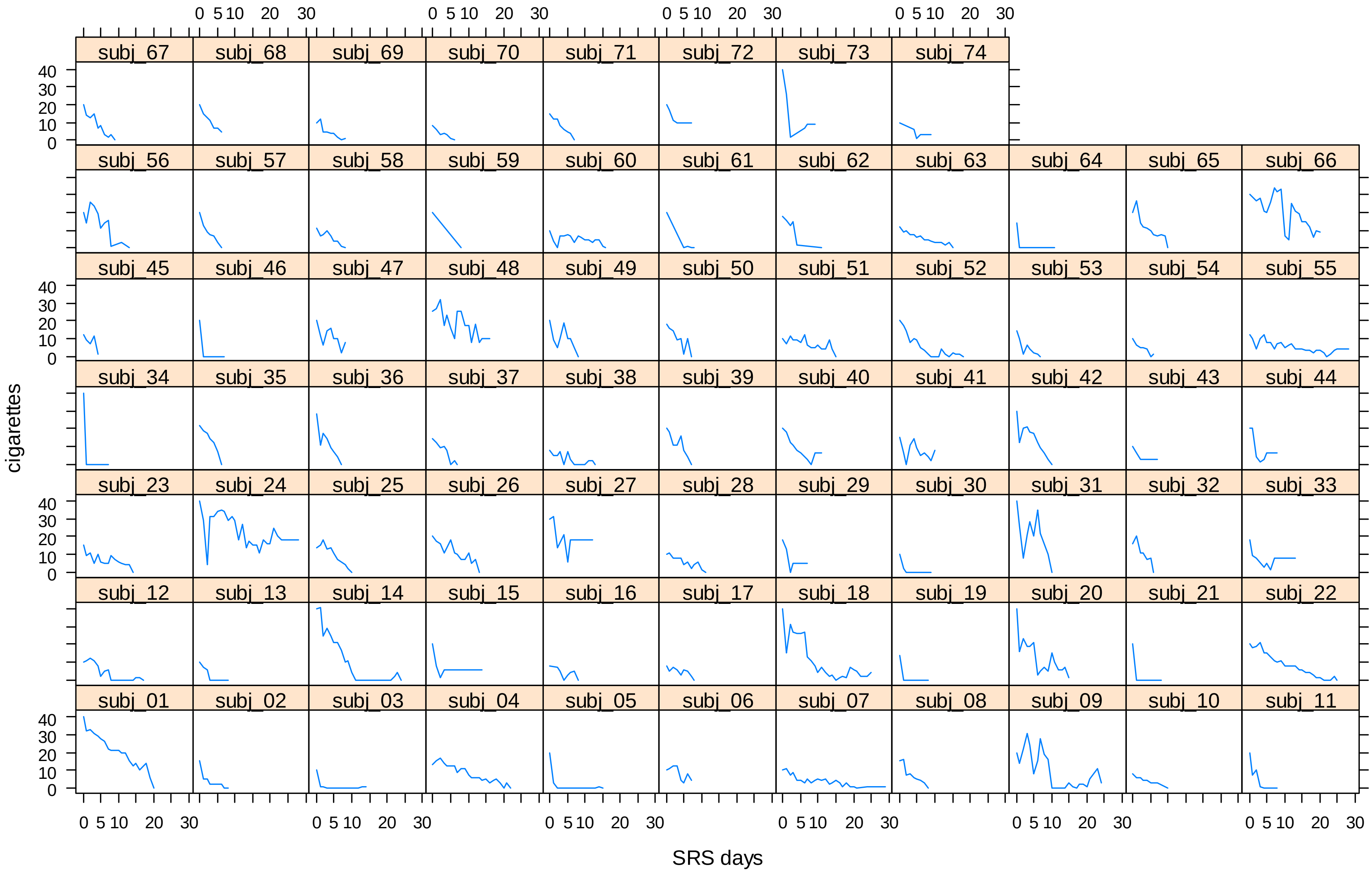
Frequency of Daily Smoking (Number of Cigarettes per Day)

Daily rates of smoking were recorded in real time for all individuals in the study. All participants were instructed to smoke on schedule, when there was a prompt from a PDA (related to the nature of the intervention). Off-schedule cigarettes were also recorded by individuals either in real time or at the end of each day. As a result, all scheduled and unscheduled cigarettes were added to yield a daily total for the entire intervention period until a pre-determined quit date. On average, patients were scheduled to stay with the program for 10 days (SD ≈ 4). Thirty eight patients (52% of the sample) adhered to the original schedule, 15 (20%) shortened their originally chosen quit date by an average of 5 days (SD = 3.9), and 21 (28%) extended it by an average of 8.5 days (SD = 7.0). On average, patients provided about 10 days of valid data (SD = 7.0), with 739 total daily assessments. Three (4%) patients provided data only for one day of the intervention, and six (8%) patients for two days.

At the end of the intervention, when patients were admitted to surgery, each was questioned on whether or not they were smoking (CO_2 breath samples were collected to verify this information and was used as the outcome indicator) and their average number of cigarettes per day in the past week; if patients quit, they were also asked about their last smoking date. This information was used to extrapolate data for individuals who had very few assessment points (fewer than determined by the difference between the starting date and the OQD). For patients who reported quitting on a particular day, a frequency of zero was added on that and subsequent days until the OQD. For patients who reported smoking, an average number of weekly cigarettes was used instead. As a result of these data additions, the total number of observations increased to 896, with a mean of 12 days per person (SD = 6.1) and the minimum data frequency being two days for one participant only.

Individual smoking trajectories for 74 patients in the study after the aforementioned data manipulation are summarized in Figure 7. The SRS day of zero corresponds to subjects' baseline rate of smoking, and subsequent days demonstrate how participants' smoking fluctuated over the course of the intervention.

Figure 7. Individual smoking trajectories (N = 74).



Covariates

All study covariates are summarized in Table 1. Scale variables such as Fagerstrom scores, number of years of smoking, quitting self-efficacy, and days until the originally scheduled quit date (OQD) were centered about the distributions' midpoints. Dichotomous variables were dummy-coded. Basic descriptive statistics for baseline covariates are presented in Table 1 in forms of frequencies, means and standard deviations.

Data structure

Similarly to traditional MLM analyses, data were structured in a long format (as described in Table 2), such that every individual was represented by several rows of data, with the number of rows being equal to the number of days in the study. Only days with available cigarette counts were represented in the data matrix. Individuals had different rows of data and days that were not always overlapping.

This approach to data analysis is not common for GMM, where data are represented in the wide format with the outcomes aligned across time points. Users of this methodology emphasize the strength of this approach due to flexibility of estimating developmental patterns, which do not need to follow either a linear or a quadratic curve, since the development can be estimated at each time point. However, for data that has large volumes of repeated observations (which are common in EMA) that are not aligned across individuals and their frequency varying substantially from person to person [e.g. the shortest outcome vector is of length 2 (days) and the longest is of length 29 (days)], the traditional conceptualization of the model is disadvantageous. A lot of missing data as well as large column dimensionality of the outcome matrix (74 x 29) place a big strain on model estimation in the Mplus software. Moreover, in most cases of social sciences data, linear or quadratic slope parameters suffice to describe developmental patterns and are easier to generalize.

Results

Results of the MLM analyses are presented first to share findings that would have been obtained with the traditionally used statistical approach for EMA data. After assessing the fit of the MLM model, results of the multi-class solutions are outlined. The final GMM model is interpreted in detail.

Single-Class Multilevel Model

Results of the stepwise model fitting procedure are summarized in the first part of Table 4, with nested multilevel models listed in order from the simplest to the fullest. Applying the LL ratio test to neighboring models, freeing the slope variance, τ_{11}, significantly improved the overall model fit ($\chi^2_{(df = 2)}$ = -2*(-1571.5 + 1076.5) = 990, p < .001). The only significant intercept predictor, Fagerstrom, chosen through repetitive model selection, improved the overall model fit ($\chi^2_{(df = 1)}$ = -2*(-1076.5 +1050.0) = 53, p < .001). Two covariates, Number of Years Smoking and Self-Efficacy, were selected as significant predictors of the slope ($\chi^2_{(df = 2)}$ = -2*(-1050.0 + 1007.4) = 85.2, p < .001).

Parameter estimates of the final multilevel model are summarized in the second half of Table 4, yielding the following prediction Equation of smoking trajectories for study participants:

$$\ln(\lambda_{it}) = 2.839 + .167 * FagerstromC_i - .323 * time_{ti} - .01 * NumYRsmoC_i * time_{ti} \\ - .006 * SEC_i * time_{ti}$$

with the covariance matrix for random effects

$$\tau = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \begin{bmatrix} .116 & .001 \\ .001 & .100 \end{bmatrix},$$

where the variances of τ_{00} and τ_{11} are significantly different from zero (p < .001).

Based on this summary, the baseline smoking rate, expressed on the logarithmic scale, is estimated to be 2.839 (17 cigarettes per day) for an individual with average nicotine dependence (Fagerstrom = 5). Every unit increase on the Fagerstrom scale is predicted to change the original baseline rate by the amount of .167 (or increase it by about 18%). The slope value - .323 is the average decline in the number of cigarettes per day on the logarithmic scale for a patient with 35 years of smoking history (*NumYRsmoC*) and baseline self-efficacy of 50 (*SEC*) [or a decrease of about 28% of the original smoking rate on a daily basis]. The graphical summary of the average smoking trajectory as well as the effect of nicotine dependence on the initial smoking rates are presented in Figures 8a and 8b. Higher initial nicotine dependence, measured by the Fagerstrom scale, shifts the intercept upwards, whereas weaker dependence results in fewer baseline cigarettes (Figure 8b demonstrates changes in the intercept for a standard unit change in the Fagerstrom scale). Higher self-efficacy and longer smoking history correspond to steeper cessation patterns: -.006 and -.010, respectively, for a unit change in the covariates. Figures 8c and 8d demonstrate the average decline in smoking over time as well as the effect of a standard unit change in self-efficacy and smoking history on tapering.

Using model-estimated raw intercept and slope predictors, computed for every individual in the sample, we explored whether the distal outcome can be predicted from the baseline smoking and the magnitude of daily reduction in smoking rates. The last two rows of Table 4 summarize results of the logistic regression incorporated in the overall

MLM model. They demonstrate that neither intercepts nor slopes of individual smoking trajectories are useful predictors of the outcome.

Table 4

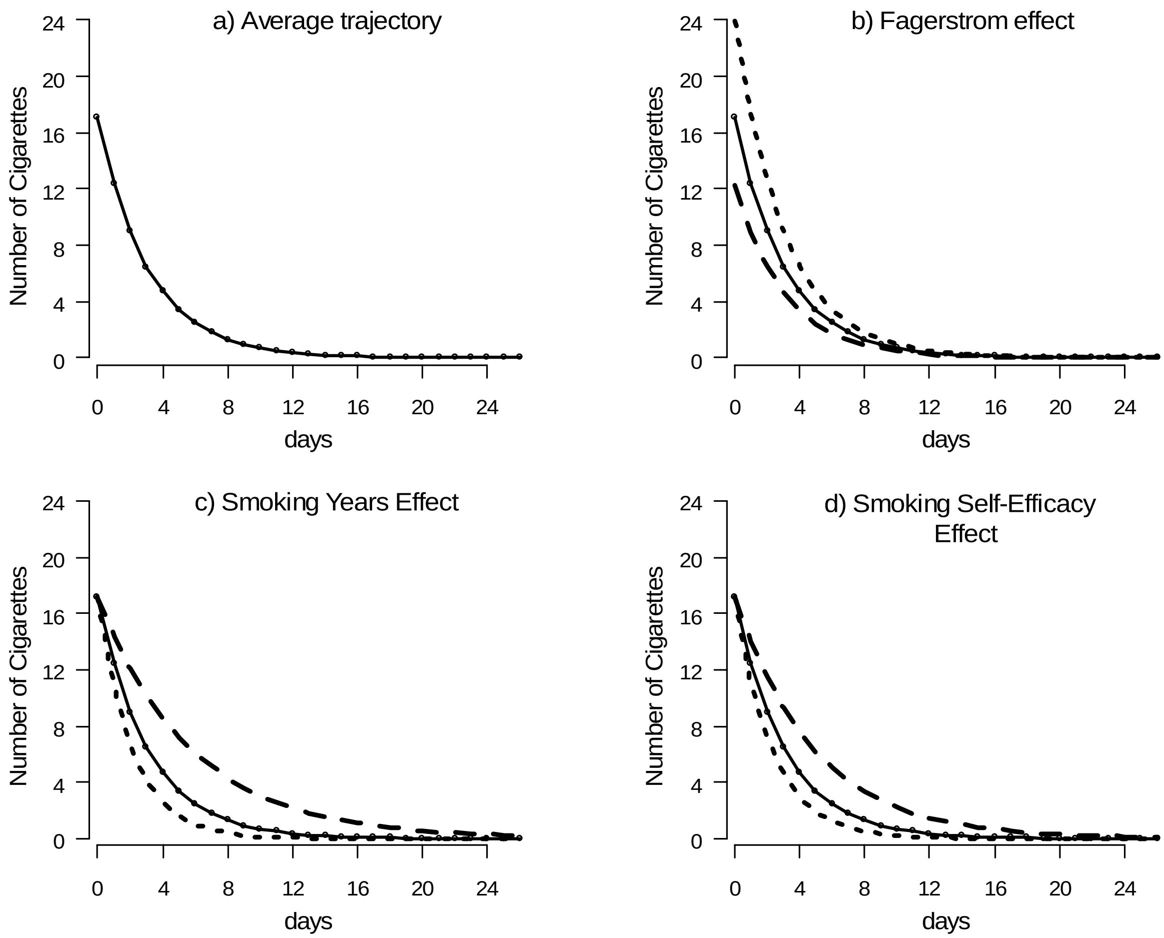
Results of multilevel Poisson models

Model description	Number of parameters	Log Likelihood
Fixed intercept and slope	2	
Random intercept and fixed slope	3	-1571.5
Random intercept and slope	5	-1076.5
Random intercept with covariates and random slope	6	-1050.0
Random intercept and slope with covariates (F)	8	-1007.4

Model parameters of the final model (F)

Parameter	Parameter Estimate (logarithm/ logit scale)	95%CI (logarithm/ logit scale)	t statistic	p value	Exp (parameter estimate)
Intercept, β_{00}	2.839	2.747, 2.931	61.76	<.001	17.099
Fagerstrom$_C$, β_{01}	0.167	.113, .221	6.133	<.001	1.182
Slope, β_{10}	-0.323	-.402, -.244	-8.157	<.001	0.724
Years Smoking, β_{11}	-0.010	-.010, -.002	-2.968	.005	0.990
Self-Efficacy$_C$, β_{12}	-0.006	-.010, -.020	-3.052	.004	0.994
SD r_{0i}	0.341			<.001	
SD r_{1i}	0.317			<.001	
Intercept $_i$ ν_1	0.692	-2.30, 3.68	.460	.645	1.998
Slope $_i$ ν_2	-3.169	-8.61, 2.34	-1.144	.253	.042

Figure 8. Predicted smoking trajectories for a single-class multilevel model with covariates.



Solid lines represent average trajectories for *X*s centered at group means; dotted lines represent trajectories for $+1 * SD$ change in *X*s; dashed lines represent trajectories for a $-1 * SD$ change in *X*s.

Individual estimates of intercept and slope parameters as well as their raw and empirical Bayes (EB) residuals were extracted to investigate the fit of the final MLM model. The distributions of raw (skewness = -.23, SE = .29; kurtosis = .33, SE = .57; Figure 9) and EB intercept residuals (skewness = .15, SE = .29; kurtosis = -.47, SE = .57) resembled a normal curve, thus, satisfying the assumption of residual normality: $r_{0i} \sim N(0, \tau_{00})$. Both actual and EB slope residuals, however, did not follow a normal pattern, which was evident from the graphical summary, as well as skewness and kurtosis statistics (skewness = -2.28, SE = .29; kurtosis = 5.52, SE = .57 and skewness = -1.27, SE = .29; kurtosis = 2.23, SE = .57, respectively). Based on the first graph of raw slope residuals in Figure 10, there appears to be two clusters of data points: one around the interval of -.5 and +.5 and a second around -2 to -1, on the logarithmic scale. This violates the assumption of $r_{1i} \sim N(0, \tau_{11})$, thus, suggesting that the one-class MLM solution does not fit the data. Comparable clustering was observed in a distribution of EB residuals.

Similarly, the distribution of predicted intercept parameters was normal around the center (skewness = .27, SE = .29; kurtosis = -.57, SE = .57; Figure 9), satisfying the assumption of normality $\beta_{00} \sim N(\mu, \sigma^2)$. The distribution of slopes, however, was highly skewed (p <.001), with the large proportion of slope values clustering between the value of zero and negative .5, and the left tail extending to negative 1.6 (Figure 10), thus, violating the assumption of normality $\beta_{11} \sim N(\mu, \sigma^2)$. Interestingly, there was a large positive association between the slope values and residuals, with Pearson product-moment correlation of .92 (p < .001), such that individuals with the steepest estimated slopes also had large residuals (last graph in Figure 10).

From these observations, it appears that there are, at least, two groups of individuals: those with a relatively gradual reduction in smoking and those with a sharp decrease. The MLM model's assumption of a homogenous population does not appear to fit the data well. Average parameter estimates do not capture the steepness of slopes for a subgroup of people. Large residuals for the same individuals point toward discrepancy between the group average and the actual developmental trajectories. Moreover, the outcome *smoking status* remains unexplained from the modeled cessation behavior during the intervention.

The next section summarizes results of models with cessation trajectories varying across latent classes of individuals. We demonstrate the search process for the best fitting model as well as interpret the final results.

Figure 9. Diagnostic graphs for intercept parameters from a single-class multilevel model.

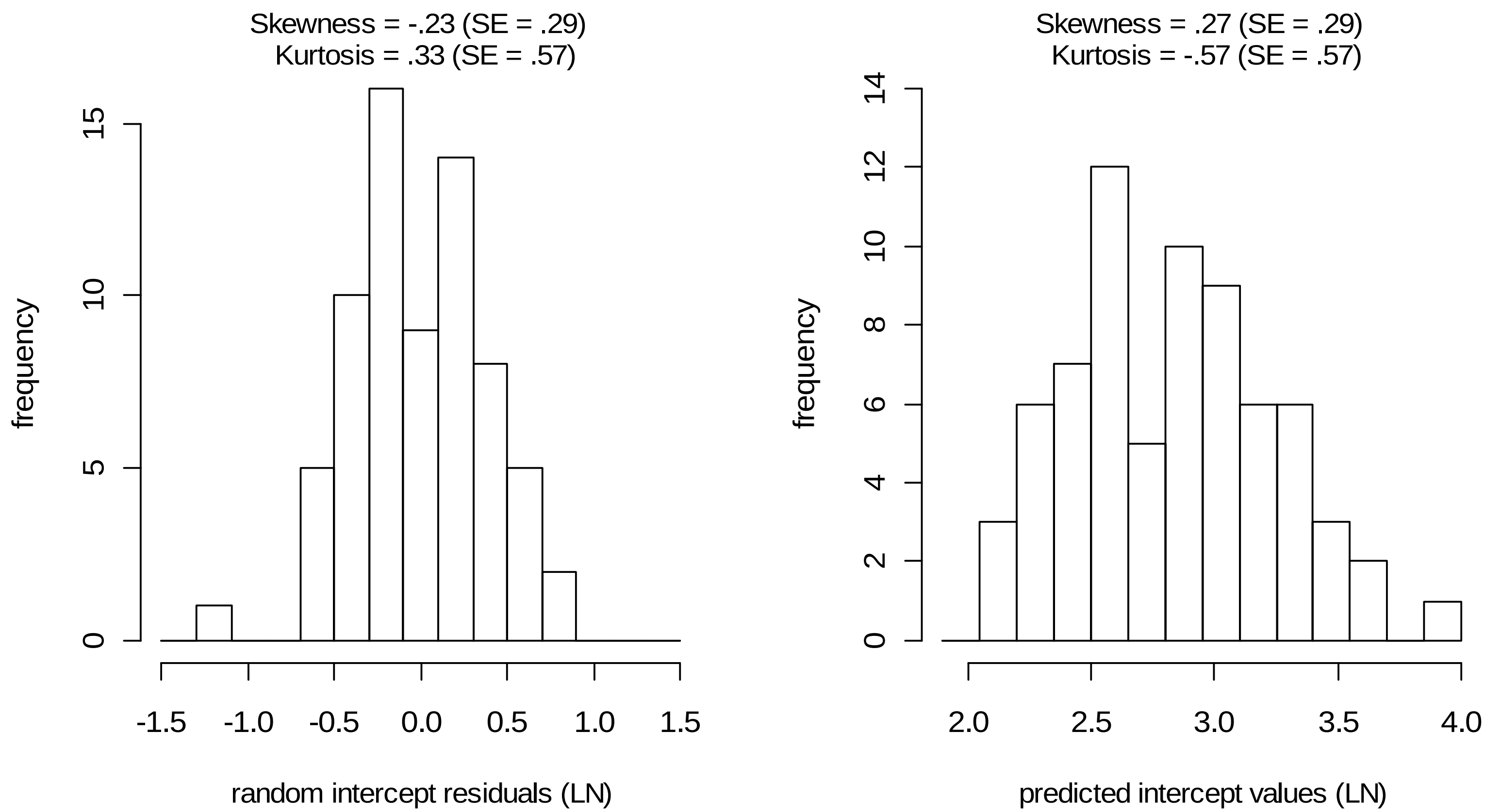
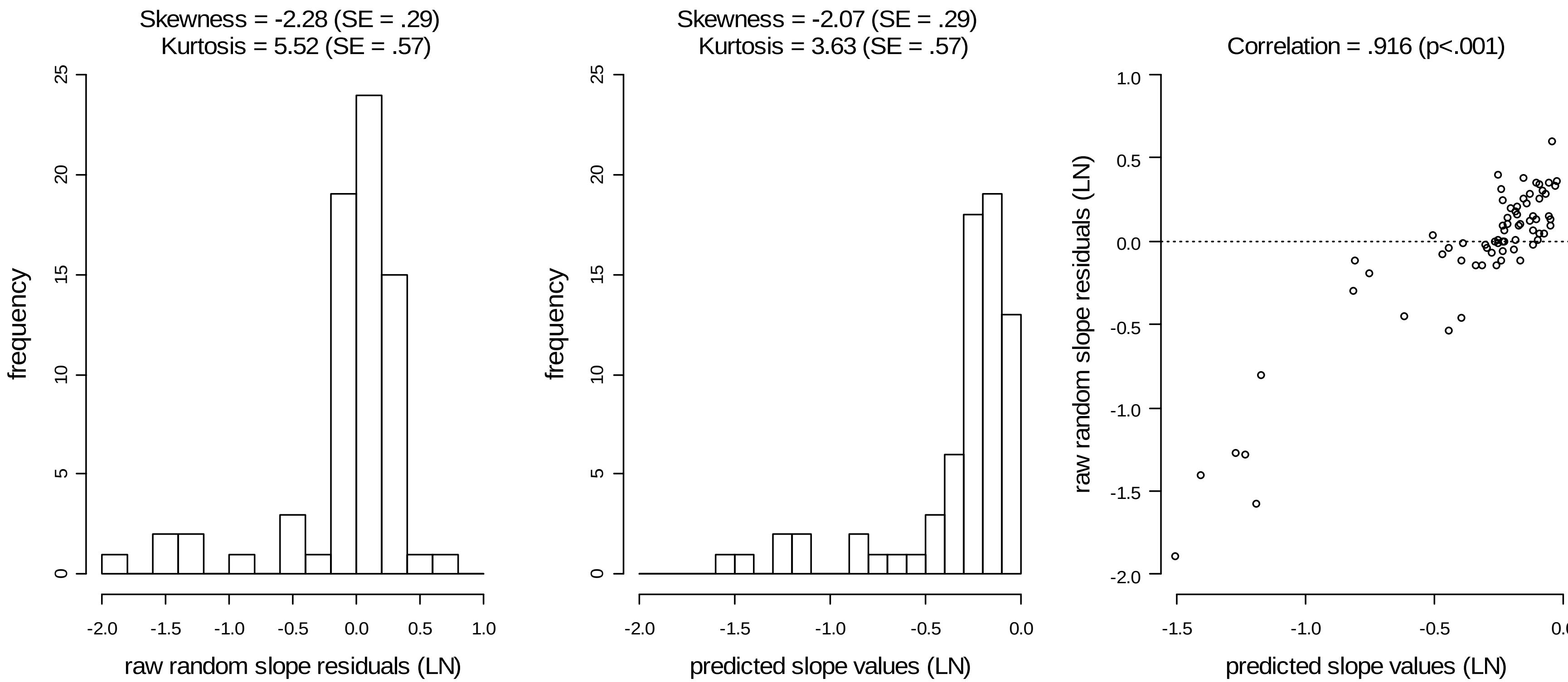


Figure 10. Diagnostic graphs for slope parameters from a single-class multilevel model.



Growth Mixture Models

It is important to note that all analyses summarized below were obtained from data structured in the long format as described in Table 2. Exploratory analyses with data in the wide format failed and forced us to eliminate any observations beyond the twelfth day.

Fixed Growth Parameters

Initially, models with intercept and slope residual parameters constrained to zero ($\tau_{00k} = 0$ and $\tau_{11k} = 0$), were fitted sequentially for classes *K* equals to 1 through 5. BIC and ABIC indices were compared across neighboring models; both statistics are summarized in Table 5. There was a continuous reduction in BIC and ABIC, with additional classes adding a significant improvement to the overall model fit. Similarly, BLRT suggested improvement in the overall model fit when additional classes were added. LMRT, however, did not confirm these findings. In fact, it suggested that *K=1* class fits the data as well as the model with *K=2* classes (p = .564). However, three classes were identified as fitting better than two (p = .015), with the fourth and fifth classes adding no significant information to the model fit (p = .237 and p = .161, respectively).

Values of the slope parameters in each latent class were another indicator of the model meaningfulness. For *K=5*, slopes in the last two classes were very comparable in magnitude: $\beta_{104} = -.118\ (SE = .019)$ and $\beta_{105} = -.133\ (SE = .012)$, with the only differences being in intercepts: $\beta_{004} = 3.330\ (SE = .12)$ and $\beta_{005} = 2.642\ (SE = .08)$, respectively. Similarity in slopes reflects no differences in smoking trajectories for these two classes, but rather different baseline smoking rates, which can be captured by allowing variability in the intercept parameters such that $\tau_{00k} \neq 0$ rather than complicating a model with an additional latent group.

Based on these results, five latent classes appear to be redundant for estimating latent smoking trajectories, and freeing intercept residual variance has potential for improving the model. The next section summarizes model comparisons for $K = 1$ through *4* with random intercept parameters τ_{00k}, varying across classes.

Table 5

Model fit indices for fixed effects growth mixture models

	Number of latent classes				
	1	2	3	4	5
BIC	9462.9	7022.7	5880.1	5494.6	5277.0
ABIC	9456.5	7006.8	5854.7	5459.6	5232.5
Number of parameters	2	5	8	11	14
LMRT (p value)		.564	.015	.237	.161
BLRT (p value)		< .001	< .001	< .001	< .001

Note: BIC and ABIC evaluate each model, while LMRT and BLRT compare adjacent models

Random Intercepts Varying Across Latent Class and Fixed Slopes

Table 6 summarizes results of growth mixture models with $K = 1$ through 4. Overall, adding random intercepts τ_{00k} had a positive effect on the model fit. The improvement is especially visible when models with the same number of latent classes are compared to each other on the basis of BIC and ABIC reduction (Figure 11). While both BIC and ABIC keep getting smaller as more latent classes are added, technical difficulties with replicating the best log likelihood, described below, suggest a misfit.

During the model estimation process, several sets of random starts (a minimum of 10) are selected for the EM algorithm. The best log likelihoods are recorded to assure that the final solution represents a global rather than a local maximum. That is evident from replicating the best likelihood at least two times from different random starts. When the best log likelihood is not replicated, parameter estimates can not be trusted. That is especially true when parameters from two best but not identical likelihoods are different in magnitude.

In the model with $K = 4$, the best log likelihood was not replicated. In an attempt to solve this problem, the number of random starts was increased to 1000 (following the guidelines outlined in Muthén & Muthén, 2007, pp. 379 – 381) with no success. After examining the model parameters, the variance of the intercept in one of the classes was not significantly different from zero (p = .119) and equal $.113\ (SE = .072)$. To simplify the model, this variance was constrained to zero. Repeated estimation of the modified model did not lead to replication of the best log likelihood; thus, all parameter estimations are questionable as they could reflect a local maximum solution.

LMRT did not yield consistent results, demonstrating a better fit of 4 latent classes over 3 latent classes (p = .019), while also indicating that 3 classes show no improvement in model fit over a 2-class solution (p = .142). BLRT, on the other hand, consistently favored models with a larger number of latent classes. However, for the last model with three random and one fixed intercept parameters, the likelihood values were not replicated in all bootstrap draws, thus, biasing the p-value. Finally, the slope parameters for classes *1* and *3* in the final four-class solution model were close in magnitude: $\beta_{101} = -.034\ (SE = .01)$ and $\beta_{103} = -.096\ (SE = .01)$, with larger differences in intercepts: $\beta_{001} = 3.30\ (SE = .08),\ \tau_{001} = .031$ and $\beta_{003} = 2.64\ (SE = .13),\ \tau_{001} = .163$, respectively.

Failure to replicate results for the 4-class model as well as absence of meaningful differences in the slope parameters made us cautious about considering a four-class solution. Further results focus on models with *K=1* through *3.*

Table 6

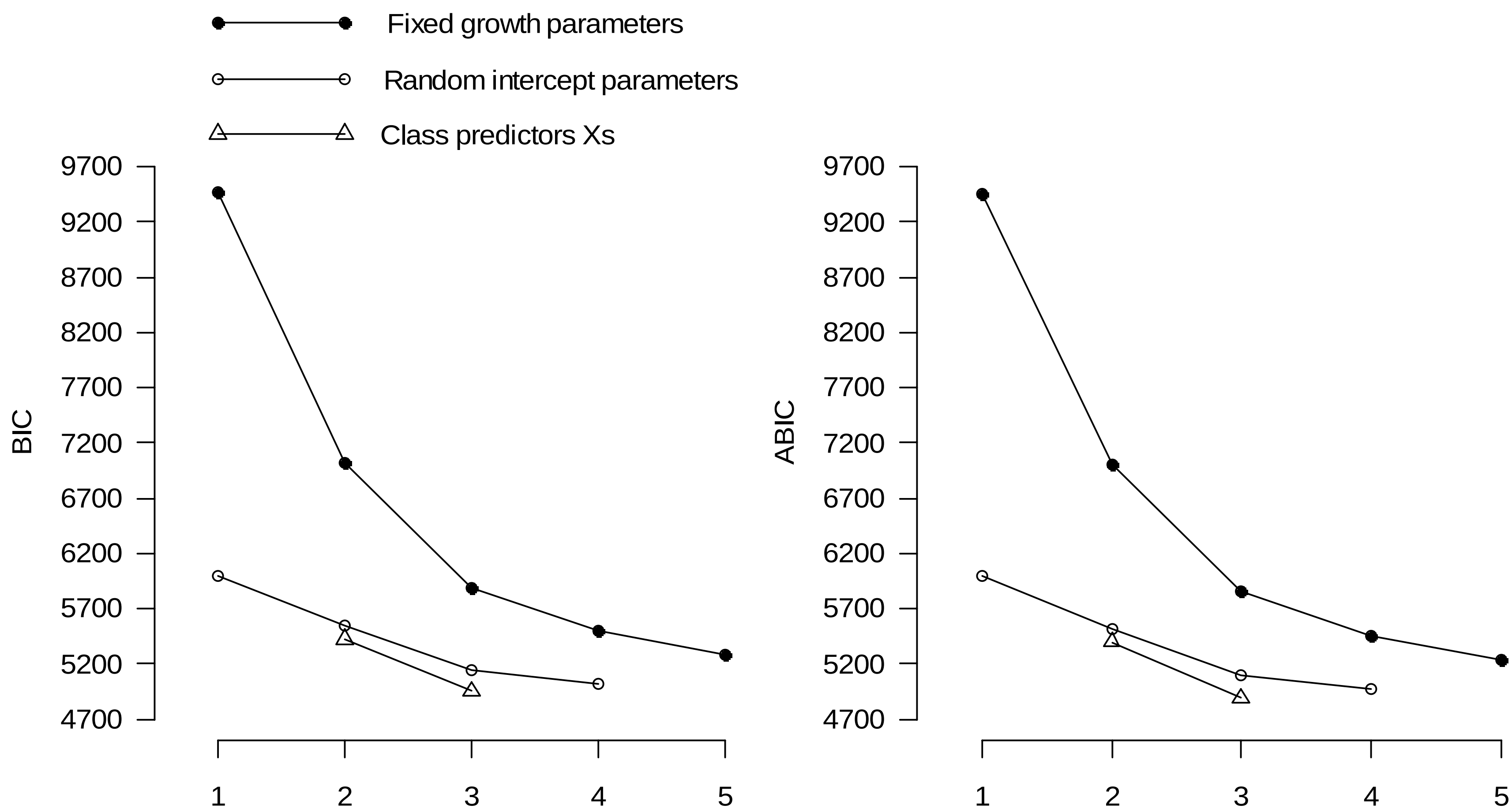
Model fit indices for random intercepts and fixed slopes growth mixture models

	Number of latent classes				
	1	2	3	4	4†
BIC	6002.4	5541.5	5135.7	5025.5	5035.7
ABIC	5992.9	5519.3	5100.8	4977.9	4991.3
Number of parameters	3	7	11	15	14
LMRT (p value)		0.032	0.142	0.019	0.021
BLRT (p value)		< .001	< .001	< .001	< .001

Note: BIC and ABIC evaluate each model, while LMRT and BLRT compare adjacent models

† In class 4, the random intercept parameter is fixed to 0: $\tau_{004} = 0$.

Figure 11. Comparisons of growth mixture models with different number of latent classes *K* based on the BIC and ABIC indices.



Incorporating Person-Level Predictors of Random Intercepts

Fagerstrom scores were found to be a useful predictor of baseline smoking rate in a single-class model described earlier. To test whether Fagerstrom explains variability in random intercepts across individuals within latent classes, this person-level covariate (X_i) was added to the model. Log likelihoods were not replicated for $K = 2$ or $K = 3$. Modifying the model with Fagerstrom scores serving as a predictor only for some latent classes did not solve the problem of the local maxima solution. Thus, results of these analyses are omitted.

Freeing Slope Parameters

Exploratory analyses were performed to investigate whether allowing random variation in the slope parameters τ_{11k} would improve the model fit. Multiple estimation problems as well as non-replication of the best log likelihood were experienced in this process. Due to the small sample size, the decision was made to constrain all slope variances to zero. Conceptually, we were looking for groups of individuals who followed distinct smoking cessation patterns, but who were also alike within each group. Thus, constraining the slope variances to zero did not impede the estimation process of classifying individuals into distinct latent classes.

Incorporating Person-Level Predictors of Latent Classes

Baseline covariates outlined in Table 1 were added as predictors of latent classes as defined by Equation 9. Non-significant predictors were eliminated from the model step by step to leave only covariates that significantly predicted the latent class membership.

For a two-class model, quitting self-efficacy and the originally scheduled length of the intervention period (OQD) were found to be significant predictors of latent classes ($p = .012$ and $p = .091$, respectively). Compared to the model with no covariates, both the BIC and the ABIC reduced significantly (Table 7 and Figure 11), suggesting a better fit of the model with covariates. Additionally, LMRT and BLRT statistics demonstrated a significant improvement over the model with a single-class solution.

For a three-class model, four baseline covariates were useful differentiators between latent classes of individuals. Similarly to the two-class model, quitting self-efficacy and the length of the originally planned tapering period (OQD) were identified as useful covariates. Additionally, age and the length of smoking, measured in years, were selected as useful predictors. Based on the BIC and the ABIC fit indices, this three-class model carried a substantial improvement over two classes. The LMRT confirmed the improved fit of the model with three latent classes. The p-value of the BLRT test, however, was based on a very small number of converged solutions and, therefore, was biased. Since the p-value of the BLRT depends heavily on the number of draws, non-estimable solutions bias the statistic. The Mplus recommendations are to increase the number of random starts to overcome this problem (Muthén & Muthén, 2007). Increasing the starts to 2000 did not resolve the difficulty.

Table 7

Model fit indices for random intercepts and fixed slopes growth mixture models with person-level predictors of latent classes (X_i) and distal outcomes (u_i)

	Number of latent classes			
	2	2 $u^{\dagger}$	3	3 $u^{\dagger}$
BIC	5426.5	5383.4	4950.5	5061.7
ABIC	5397.9	5348.5	4890.2	4991.9
Number of parameters	9	11	19	22
LMRT (p value)	0.032	0.044	0.044	0.042
BLRT (p value) ├	< .001	< .001	‡	‡

Note: BIC and ABIC evaluate each model, while LMRT and BLRT compare adjacent models

$u^{\dagger}$ is a GMM model with a distal outcome

├ To estimate BLRT, the OPTSEED value was set to the seed with the best LL (Muthén & Muthén, 2007, p. 500)

‡ 48 out of 51 bootstrap draws did not converge, the p-value is biased

Adding a Distal Outcome

Finally, a distal outcome u_i was added to study the probability of quitting smoking at surgery for each latent class (Equation 10). Based on the BIC, the ABIC, and the LMRT statistics, the model with a three-class solution and a distal outcome fit the data noticeably better than the two-class model with a distal outcome (Table 7). Both the BIC and the ABIC indices decreased when a third class was added. The significant p-value of the LMRT demonstrates that the three-class model fits the data better than the two-class model. The non-significant value of the BLRT should not be interpreted as it is based only on three successful bootstrap draws.

Additionally, when a distal outcome was added to the two-class solution, the magnitude of growth parameters changed substantially. For instance, the slopes of two latent classes were -1.338 (SE = .286) and -0.093 (SE = .018) on the logarithmic scale in the model without the distal outcome. They changed to -0.248 (SE = .016) and -0.070 (SE = .015), respectively, when the outcome was added. Such changes demonstrate a non-stable solution, which is easily affected by the addition of a variable that should not have an effect on defining latent classes. On the other hand, parameters of the three-class solution model stayed stable once the distal outcome was added.

During this process of model building and comparisons, we relied on a combination of statistical and conceptual indicators of model fit. At the end, it appears that the model with a three-class solution, where intercepts are random and differing across latent classes, slopes are fixed and varying across latent classes, baseline covariates predicting a class membership, and a distal outcome estimated for each latent class, fits the data the best. The detailed summary of this model is presented in the following section.

The Three-Class Final Growth Mixture Model

Latent Developmental Trajectories

The parameters of the final growth mixture model are summarized in Table 8. Based on the results, the baseline smoking rate was comparable across three latent classes, ranging from 2.73 to 2.806 on the logarithmic scale (or 15.33 to 16.54 on the cigarettes per day scale), thus, indicating that the initial smoking rate did not have a large impact on identifying the class membership. However, the variability in baseline smoking differed across classes, with the first class having a more homogenous sample of individuals ($\tau_{001} = .121$) compared to the other two classes ($\tau_{002} = .237$ and $\tau_{003} = .281$). The variability in all classes was substantially different from zero (p < .02). Figure 12 provides a graphical summary of latent classes, where the initial rate of smoking is, clearly, comparable.

Qualitative differences in smoking trajectories are captured by slope parameters, with the three latent classes exhibiting different reduction rates in smoking over time. The first latent class has the steepest slope ($\beta_{101} = -1.391$), indicating an immediate drop in smoking within a few days of the intervention. By days 3-4, study participants who are likely members of this latent class appear to reduce to nearly zero cigarettes per day (Figure 12). This class is relatively small, with about 13% of individuals following this behavioral pattern.

The second latent class is composed of individuals who followed the prescribed schedule and exhibited a gradual reduction in smoking on a daily basis: $\beta_{102} = -.224$. Participants in this class tended to reduce gradually, taking up to three weeks to reach a near zero mark on a daily smoking rate. This appears to be the most prevalent class with more than half of individuals (56%) falling in this behavioral category.

Finally, the third group of participants had a shallow negative slope ($\beta_{103} = -.07$) that was significantly different from zero in spite of its small magnitude ($p < .001$). Patients who were estimated to follow this behavioral pattern appeared to reduce their smoking very slowly and never came close to a zero point during the course of the intervention. About a third of the sample (31%) was estimated to be likely members of this latent class.

Model Fit Indicators

In the process of model estimation, every study participant had a non-zero probability of falling within each latent class. Posterior probabilities were computed for every individual, with the highest probability determining the most likely class assignment. To evaluate the classification quality, average posterior probabilities of class membership were computed for individuals falling in each latent class (Table 9). For all three latent classes, the average posterior probability of belonging to a particular class was very high, indicating a clear class separation, with all individuals having a high posterior probability only for one latent class.

Another indicator of model fit is a comparison between predicted and actual mean smoking trajectories. The actual trajectories were constructed based on daily smoking averages for individuals estimated to be members of a particular latent class. Figure 12 demonstrates a very close fit between the estimated and actual trajectories for classes 1 and 2, and a looser fit for class 3. For the latter class, the average seems to capture a general shallow downward developmental trend, but there appears to be a lot of variability among class members.

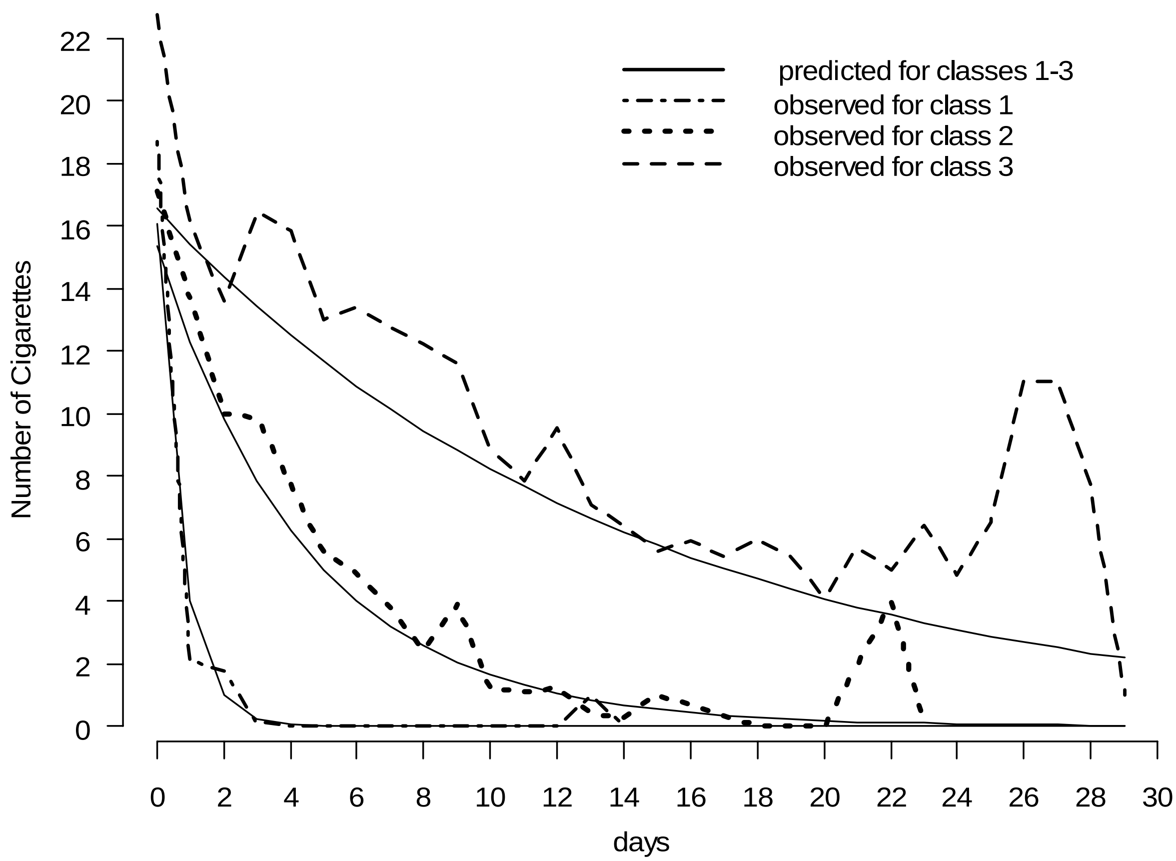
Table 8

Parameter estimates of the final three class growth mixture model

Parameters	Parameter Estimate	Standard Error	p-value	Exp (parameter estimate)
Intercept-related parameters				
β_{001}	2.777	0.145	<.001	16.071
τ_{001}	0.121	0.051	.018	
β_{002}	2.730	0.073	<.001	15.333
τ_{002}	0.237	0.069	.001	
β_{003}	2.806	0.124	<.001	16.544
τ_{003}	0.281	0.062	<.001	
Slope parameters				
β_{101}	-1.391	0.321	<.001	.249
β_{102}	-0.224	0.010	<.001	.799
β_{103}	-0.070	0.014	<.001	.932
Quitting at surgery				
v_{01}	1.252	0.802	.118	3.497
v_{02}	0.093	0.332	.780	1.097
v_{03}	0.968	0.483	.045	2.633
Baseline covariates as predictors of class membership				
$\omega_{01}{}^{\dagger}$	12.048	4.858	.013	170757.6
$\omega_{11} * SEc$	0.088	0.031	.004	1.092
$\omega_{21} * OQD$	-0.926	0.739	.210	0.396
$\omega_{31} * age.c$	-0.785	0.423	.064	0.456
$\omega_{41} * YrSmoking$	1.348	0.532	.011	3.850
$\omega_{03}{}^{\dagger}$	-0.316	0.609	.604	0.729
$\omega_{13} * SEc$	-0.046	0.029	.118	0.955
$\omega_{23} * OQD$	0.987	0.270	<.001	2.683
$\omega_{33} * age.c$	0.018	0.044	.680	1.018
$\omega_{43} * YrSmoking$	0.070	0.048	.147	1.073

† second class serves as a reference group

Figure 12. Predicted and observed average developmental trajectories for three latent classes estimated as part of the final GMM model.



Note: As the number of days increases, the averages are based on fewer observations.

Table 9

Classification of average posterior probabilities for three latent classes based on the final growth mixture model

	Average posterior probabilities		
Assigned classes based on $p \geq .5$	1	2	3
1	.999	.001	0
2	0	.993	.007
3	0	.016	.984

The comparison between predicted trajectories and actual observed smoking patterns yielded a similar conclusion (Figure 13). The first and second classes appeared to have captured well the developmental trends with modest variability around the averages. The third class, however, demonstrated a lot of variability, both in the intercept and slope parameters. While the majority of individuals in this class appeared to reduce in comparison to their baseline smoking rate, their daily behavior emerged as somewhat erratic with multiple peaks and falls during the course of the intervention. Freeing the slope parameter τ_{113} for this latent class led to problems with model estimation and could not be assessed.

Smoking Status at Surgery Admission

The distal outcome smoking status as surgery admission was used as an indicator of the quality as well as practical usefulness of the model. Within each latent class none of the baseline characteristics were identified as useful predictors of the smoking status. Thus, the probability of smoking at surgery is only conditional on one's latent class: $P(Quit \mid k_i)$. Specifically, for someone in latent class 1, the log odds of quitting $v_0 \mid K = 1$, defined in Equation 10, was estimated to be 1.252 (SE = .802). This translates into a probability of .78, which is very high.

In comparison, for someone in the second latent class, the log odds of quitting $v_0 \mid K = 2$ was estimated to be .093 (SE = .332). The probability of quitting in this class was about .52. Finally, the log odds of quitting $v_0 \mid K = 3$ was estimated to be -.968 (SE = .483). For this latent class, the probability of quitting was estimated to be .28, which is significantly lower than in the other two classes.

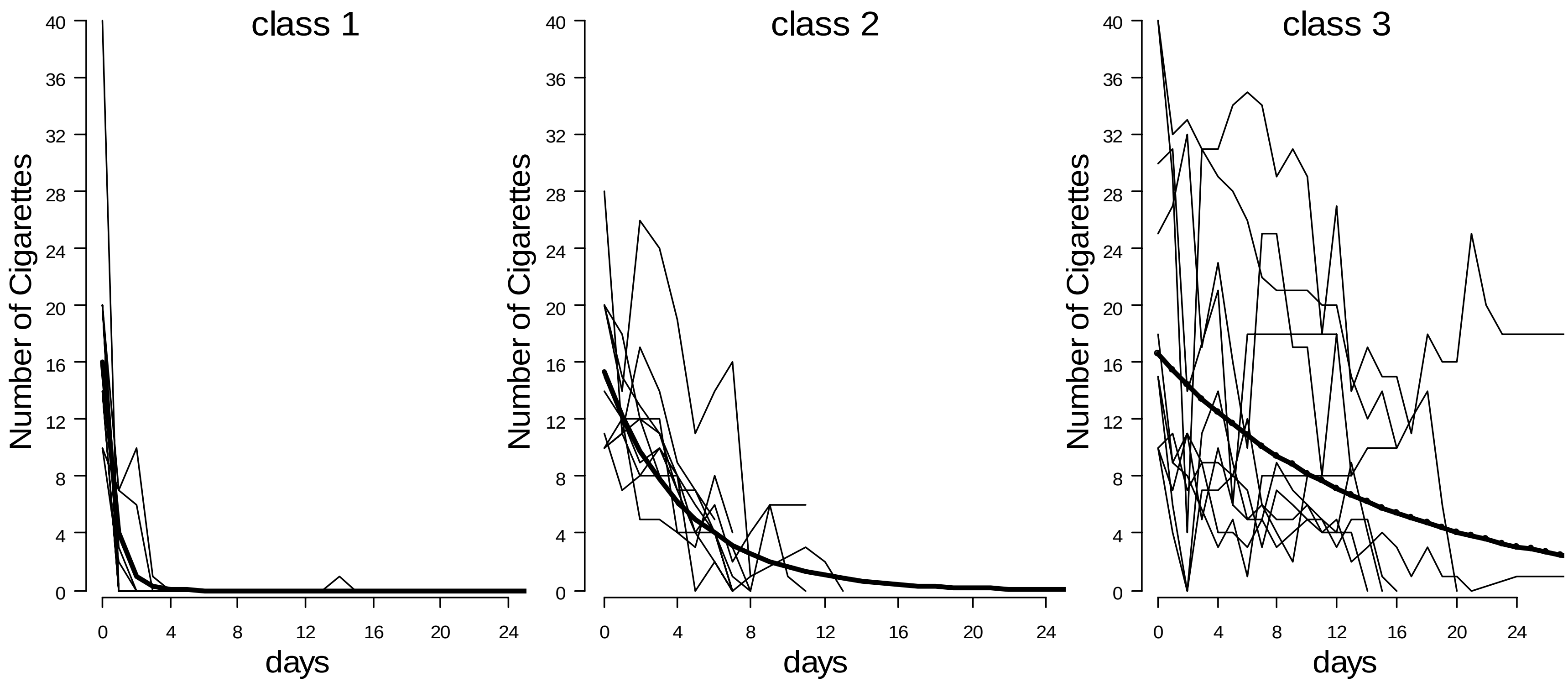
Individual Profiles of Patients in Each Latent Class

Results of the multinomial unordered logistic regression, defined by Equation 18, are summarized in the lower half of Table 8. Four baseline covariates: quitting self-efficacy, length of the originally scheduled intervention period (OQD), age, and the number of smoking years were identified as useful predictors of class membership. After comparing predictors across latent classes, three patient profiles emerged.

Individuals who were identified as members of the ‘abrupt reducers’ class (class 1) were more likely to have a higher than average quitting self-efficacy, a shorter than 10 days OQD (although their baseline smoking rates were comparable to the other two classes), more than 35 years of smoking history, and younger than 55 years old, when compared to the class of “gradual reducers”. In contrast, “slow reducers” (class 3), were very similar to the second class of “gradual reducers” with the exception of having a longer originally scheduled quit date (> 10 days).

Taking into account baseline covariates helps to predict one’s latent class status prior to observing the actual smoking behavior. The class of “abrupt reducers” appears to be especially distinct from the other two classes.

Figure 13. Predicted developmental trajectories for three latent classes and actual smoking trajectories for a random sample of 10 patients selected in each latent class.



Discussion

The method of collecting person-reported data by means of electronic diaries was initially introduced more than 20 years ago (Shiffman, 2008; Shiffman et al., 2008). Since then, it found applications in many areas of behavioral research, including the fields of personality, clinical, health, developmental, and industrial-organizational psychology (Rodriguez & Audrain-McGovern, 2004; Rodriguez, Moss, & Audrain-McGovern, 2005; Stanger, 2006). Multiple studies addressed the issue of methodological advantages and challenges associated with this methodology (see Shiffman, Stone, & Hufford, 2008, for a comprehensive review). At the same time, the statistical techniques available for the analysis of such data are limited (Walls & Schafer, 2006).

One of the strengths of EMA involves the ability to capture personal experiences, behaviors, and psychological and physical responses in real life over a certain time period. As a result, researchers are presented with fine-grain records of longitudinal data, which allow them to study natural history and sequences of events, contextual associations, and individual differences (Shiffman et al., 2008). Traditional statistical approaches require aggregation of these data (Hufford & Shiffman, 2002), thus, eliminating the advantages of the detailed recording. More advanced methods such as MLM (Hedeker, Mermelstein, & Demirtas, 2008; Schwartz & Stone, 1998; Walls, Jung, & Schwartz, 2006) and generalized estimation equations (Schafer, 2006) account for some complexities of EMA, but assume a homogenous developmental pattern across individuals. With traditional methods, the inter-personal diversity is captured by random variability rather than structural differences in parameter estimates.

The purpose of this dissertation was to demonstrate the application of GMM methodology to EMA data collected as part of the smoking cessation trial. Relaxing the assumption of homogeneity allows for flexible modeling of developmental trajectories as well as for preservation of information collected at various time points. The potential for GMM applications with EMA data is very broad, thus, it was important for us to assess the feasibility of this method as well as to address its practical and methodological implications.

Practical Findings

Addictive behaviors such as smoking and drinking are topics of great concern for researchers. Due to large variability among individuals in the way these behaviors originate and develop, there is great interest in longitudinal profiles that capture behavioral progression, identify characteristics of at-risk individuals, and link developmental patterns with long term outcomes. Several studies in the area of smoking research have addressed adolescent developmental profiles and their effects on smoking in adulthood by means of GMM (Colder et al., 2001; Maggi, 2008; Maggi, Hertzman, & Vaillancourt, 2007; Rodriguez & Audrain-McGovern, 2004; Rodriguez et al., 2005). Different smoking patterns were identified, and individuals with the highest risk of adulthood smoking were singled out. Their work addressed longitudinal data on the macro-level, where observations were recorded on an annual or biennial basis. In the current study, we concentrated on the issue of smoking cessation on a micro-level, where individuals were monitored during their cessation period on a daily basis.

Analyses of the smoking cessation data in the current study was heavily guided by preliminary exploration of smoking trajectories as well as literature summaries about natural patterns of quitting smoking. Guided by a priori hypotheses of multiple classes, we initially explored the fit of the single-class multilevel model. MLM captured some variability in intercept and slope parameters by incorporating random effects, and estimated an average cessation pattern for the overall group of individuals (Figure 8). Based on this model, all patients tapered their smoking, although with some variability in the daily reduction rates. This model mostly reflects the hypothetical behavior of what should have happened in the trial if everybody followed the prescribed schedule, with variability in slopes capturing some individual differences.

More importantly, we attempted to answer the question of whether a particular behavior during the intervention as well as personal characteristics assessed at baseline were associated with long-term smoking outcomes. In our sample, only 50% of patients were able to quit smoking by their hospital admission date. Results of the MLM analysis in combination with logistic regression yielded no significant predictors (Table 4). The utility of that model is questionable. Relying on the MLM results, we can discriminate between individuals with different baseline smoking rates based on their Fagerstrom scores, and between daily smoking reduction rates based on the number of smoking years as well as quitting self-efficacy. At the same time, the main study outcome remains unexplained. In this case, had we not hypothesized about different latent classes of development, the results of the MLM analyses would be somewhat unsatisfactory.

However, examination of the actual smoking trajectories that were captured with real-time data pointed to a group of individuals who deviated from the schedule set by the

SRS program. From those, some participants preferred an immediate smoking reduction while others continuously postponed their quitting date. This non-compliance could have contributed to differences in long-term outcomes.

The final GMM model was selected based on statistical indicators of model fit as well as the interpretability of the results and their practical usefulness. For GMM, the theoretical groundwork is essential as statistical parsimony does not always result in interpretable findings. Therefore, practical matters and theory should dominate over the statistical fit. Because of various difficulties in the model selection process, some researchers recommend using GMM methodology only for confirmatory analysis (Wang & Bodner, 2007). In our case, we were guided by statistical improvements in the model fit; however, the final model was also selected based on practical considerations.

Three latent classes of individuals were identified by GMM. The largest class included slightly over 50% of study participants, whose developmental trajectories resembled the schedule set by the program. The quitting rate for this group was estimated to be .5. In fact, the development of this group resembled a solution proposed by a single-class MLM.

The most interesting findings, however, came from the other two classes, whose behavior is a deviation from the expected. A relatively small class of “rapid reducers”, comprising 13% of the overall sample, was characterized by a very steep slope. The behavioral pattern of participants in this group resembled a behavior known as “cold turkey” or sudden abstinence (Cheong et al., 2007). Some findings identify this method as the most successful among self-quitters (Cheong et al., 2007), while others see no advantages (Law & Tang, 1995). For the participants of this study, the “cold turkey” method was very advantageous with the probability of quitting close to .8. In comparison, about 30% of individuals constituted a third class, characterized by a very slow reduction in smoking in comparison to the normative group. Their chance of quitting at hospital admission was as low as .28.

By extracting two additional classes, we were able to identify patients who were the most and the least successful. Both of the groups deviated from the prescribed intervention, but the direction of the deviation as well as the final outcome of their behavior were very different. The model was able to identify personal characteristics of patients falling within each latent class, thus, making it possible to predict a certain behavior prior to observing it. In comparison to the average class, cold turkey quitters were younger, with a shorter time period between the study enrollment and surgery, they had longer smoking history and reported high baseline quitting self-efficacy. In contrast, slow reducers had longer time between study enrollment and surgery. One can speculate that with no time pressure, patients were consistently postponing their quitting date and did not commit to a predefined smoking schedule.

Identifying behavioral smoking patterns with personal profiles and quitting success rates has important practical implications for assessing the efficacy of the intervention and for indentifying individuals who are at risk of poor outcomes. It would be beneficial to replicate the findings by incorporating a more distal smoking status (e.g. 6 month follow-up) to study if behavioral differences in the course of the intervention also have long-lasting effects.

Feasibility

To our knowledge, this is the first study that applied GMM to EMA type of data. One of the major goals for this dissertation was to demonstrate how the analysis can be conceptualized and carried out. Up to date, most longitudinal studies using GMM as their analytic approach relied on 4 to 8 waves of repeated observations for estimation of developmental trajectories. With EMA, the number of repeated assessments often exceeds 20 or 30 (Collins, 2006; Hufford & Shiffman, 2002), placing a great demand on the computational capacity of the program.

Practically speaking, the GMM approach is carried out with a wide data structure, where repeated observations are recorded in a wide data matrix with the number of columns representing the number of repeated assessments. During the model estimation procedure, numerical integration is applied at every step of the iterative EM algorithm. As the number of observational points increases, so does the number of dimensions of integration, placing a high computational load on the estimation process (Wang & Bodner, 2007). There is scarce information about the maximum number of observations that the Mplus software can handle. From the exploratory stage of analysis for the smoking cessation data, 12 days were the maximum.

While longitudinal data are commonly structured in a wide format, internal capacities of the Mplus program allow for the long-way arrangement. The long format is primarily used in twin studies and studies that address issues of patients nested within hospitals or students nested within schools (Muthén & Muthén, 2007). With these nested designs, the correlation between the first-level units is accounted for. In a similar fashion, longitudinal data can be conceptualized as a two-level problem with observations being nested within individuals (Raudenbush & Bryk, 2002). While this approach of modeling longitudinal data may constrain possibilities for developmental shapes, linear and quadratic patterns are easily implemented and would account for many developmental processes in social sciences.

While taking advantage of the MLM theoretical approach to account for long vectors of repeated assessments and the flexibility of handling time, other GMM strengths can still be incorporated. For instance, it is possible to regress latent parameters on each other (Equation 8 in Appendix 1), which may be of importance when several developmental processes are modeled simultaneously or when there is a need to control for one latent parameter while estimating another. Antecedents and distal outcomes are also easily included.

Some authors view GMM as an extension of structural equation modeling and referred to it as a "second generation SEM" (e.g. Muthén, 2001b). When working with the EMA data, it may be beneficial to move away from this conceptualization of GMM. When defining the model, we tried to build bridges between MLM and GMM methods, emphasizing the hierarchical structure of EMA. This relationship should not be overestimated, however, as the MLM approach does not capture the richness of the growth mixture methodology.

The Mplus program can be used to model EMA from the hierarchical perspective and there are no alternative user-friendly statistical packages that allow GMM analysis and would account for complexities of EMA data [3]. Examples of annotated syntax are summarized in Appendices 2 and 3 and can serve as guides for future analyses. One should keep in mind, however, the limitations of Mplus when estimating longitudinal models from the multi-level perspective. Specifically, all graphical summaries in forms of mean latent class trajectories, mean trajectories in comparison to actual trajectories, and mixtures for model parameters, are disabled when data are structured in the long way. Posterior summaries of class proportions and frequencies are misleading as they are computed for level one observations rather than for individuals falling within each latent class. Based on the estimate from Mplus for the final model, the prevalence of the three latent classes would be 10%, 43%, and 47%, respectively. Correct percentages were 13%, 56%, and 31%, which were computed from saved class probabilities data. In Mplus, the long format of data structure is primarily used for survey data, where individuals are nested within institutions or geographic areas. In those instances, counting individuals as level one units does not pose a problem. In comparison, in longitudinal designs, level one units represent individual observations and should not be counted when the class assignment is carried out on the personal level. While these are not major drawbacks, they slow down the model building process as well as make the interpretation of the results more laborious.

Methodological Issues

Model Selection Process

One of the most daunting tasks in performing the GMM analysis is selecting an appropriate model that accurately builds on the theoretical applications as well as data realities. The iterative process of model building is very time-consuming due to multiple combinations of model parameters that have to be compared across different numbers of latent classes. Major steps outlined in the enumeration and specification sections of this dissertation draw upon general recommendations from GMM and MLM literature. Real-life technical difficulties, however, such as the time of model estimation (which ranged between several minutes for simpler models with $K = 1$ or with fixed growth parameters to more than a day for complex models with multiple latent classes and random growth parameters) as well as non-replicability of the best LL or untrustworthy results for the BLRT force a very careful examination and comparison across various fit statistics as well as great reliance on theory.

Small Sample Size and Class Enumeration

GMM emerges from the field of asymptotic statistics, which in practice means that its results are trustworthy with large sample sizes, usually exceeding 1000 level two observations (Tolvanen, 2007). In practice, many empirical studies have much smaller samples but they still rely on the large-sample statistical methods. A small sample size has very practical implications for selecting the appropriate number of latent classes. While the number of observations for each individual is much larger with EMA than what researchers would normally encounter in traditional longitudinal designs, the latent class classification, largely depends on the number of level two units.

In general, classes with a small representation of individuals may not be extracted or may have unstable parameter estimates (Wang & Bodner, 2007). In the current study with a sample size of 74 individuals, class prevalence was estimated to be 13%, 56%, and 31%. It would be problematic to increase the number of latent classes, as even with 3 classes parameters for the smallest class were estimated based on a sample of 10 individuals.

Model Identification

Another drawback of small samples relates to model identification. The number of free parameters that are estimated for each latent class has to be considered very carefully. While GMM's flexibility allows for differences in all model parameters across classes as well as randomness in intercepts and slopes within a class, a larger number of free parameters places a great demand on program convergence and result quality.

In our example, we were able to test explicitly whether the addition of random intercepts improved the model fit. Based on statistical indicators BIC and ABIC, allowing variability in intercepts significantly improved the model. A different approach was taken for the slope parameters. Due to estimation difficulties, it was problematic to assess whether there was a need for freeing slope variances. Graphical summaries of data (Figure 13) demonstrated that the first two classes of abrupt and gradual reducers appeared to develop in a homogenous fashion, with average slope parameters accurately capturing their developmental trajectories. Conversely, the class of slow reducers appeared to have more variability in daily tapering rates. Therefore, the alternative tested model involved freed slope parameters in the third class. Thus, graphical summaries can be very useful in exploring the model fit when statistical comparisons are problematic.

Model Fit Indices

Two simulation studies addressed the issue of small samples and the sensitivity of model fit indices for enumeration purposes. Both studies simulated four waves of data with sample sizes of 50 (Tolvanen, 2007) and 200 (Nylund et al., 2007) individuals. BIC was recommended as the most accurate fit index for small sample data, with ABIC being more reliable for larger samples. During the course of model comparisons in the current study, BIC and ABIC exhibited a very high degree of agreement and we were able to rely on both indices in the model selection process.

LMRT did not exhibit consistent performance during the class enumeration process. Depending on the model specification, this statistic was either favoring or rejecting the model with a larger number of latent classes. One would expect that in a model with only mean trajectories (Table 5), the number of classes needed to capture variability in the data would be higher. Based on the LMRT statistic, that was not the case, and a one-class model with fixed intercept and slope parameters fit better than a two-class model. Conversely, in a more complex model with random intercepts and baseline covariates, a larger number of latent classes was preferable (Table 7). Although previous simulation studies indicated that LMRT identified the correct number of latent classes in 95% of cases with a sample size as small as $N = 50$, results of the current analyses demonstrate somewhat inconsistent performance of this statistical index. Future simulation studies may be warranted to clarify the role of this indicator for GMM model selection in EMA data.

The BLRT was the most questionable index. During the initial process of model building, p-values for this statistic were very small, with probabilities falling below .001. In comparison to the LMRT, these values were very small. As a re-sampling technique, BLRT may be problematic with small samples, underestimating the variability in parameters (Chernick, 1999). Conversely, with more complex models, summarized in Table 7, the initial settings for BLRT did not produce trustworthy results as a large number of bootstrap draws did not converge. After modifying the settings (Muthén &

Muthén, 2007), the p-value for the final model still could not be accurately estimated because of a high proportion of non-converged draws. While we made an attempt to evaluate models with different number of latent classes based on the BLRT statistic, we did not rely on it heavily. Although two simulation studies found BLRT reliable (Nylund et al., 2007; Tolvanen, 2007), further studies need to address the issue of accuracy for this model fit index.

A combination of statistical indicators of model fit is used in the process of GMM model selection. In instances when all of them point to a significant improvement in model fit, one can safely assume that a more complex model fits the data better. More controversial situations arise when there is disagreement among fit indicators. In that case, theory should guide the process of model selection and future research should address the issues of sensitivity and specificity of statistical indices.

Incorporating a Distal Outcome

This section raises the issue of how a distal outcome is incorporated in the GMM model. Specifically, we are concerned with the way probabilities of quitting are computed for individuals in each latent class. As a result of model estimation, participants are assigned to latent classes based on the highest probability of being in a particular latent class. Once the assignment is complete, the probability of quitting is estimated by the proportion of successful quitters falling in each latent class. The value of the parameter estimate does not reflect classification uncertainty and is computed as if class memberships were a non-probabilistic entity. While in our example we observed very high mean probability values of being in a specific latent class, this may vary across studies. It would be a problem if a person with a 52% chance of being in class 1 is treated as if he/she is a definite member of the class and this person's smoking status contributes to the overall estimation of the quitting rate without taking into account membership uncertainty.

Limitations of the Current Study

While contributing to the literature of EMA data analyses, this study has a number of limitations that need to be explicitly noted. First and foremost, our sample size was limited to 74 individuals. Despite the fact that the number of observations is very large in EMA (N = 896), level two sample size is a primary indicator of power in these types of models. There are no clear guidelines about the magnitude of the sample sufficient for GMM. SEM recommendations are to have 5 participants for every estimated model parameter (Bentler & Chih-Ping, 1987). With 22 parameters estimated in the final model, the sample size may be lacking 35 subjects; however, this is a very crude way of estimating the required sample size, which assumes $K = 1$ class.

When determining a sample size, there are several factors that have to be considered. First, as the number of latent classes increases, the overall number of participants should also increase to allow for a reasonable number of people within each latent class. In cases when particular classes are rare (which is common for serious addictions or deviant behaviors), the sample size should be large enough to obtain a reasonable representation from the smallest latent class (Wang & Bodner, 2007). In our study, the smallest class constituted 13% of the sample. While it may have been a relatively small size, the differences in parameter estimates were very pronounced. This is another important criterion to keep in mind. When classes are not well separated, the estimation becomes more difficult and larger samples are needed. More importantly, one needs to be concerned with the theoretical justification for several latent classes. Random parameters from a single-class solution may well capture small developmental differences and a multiple-class approach can be avoided.

Another limitation of the current study relates to the accuracy of data recording. Although EMA are precise in capturing daily events, the importance of PDA adherence is heightened when the total number of smoked cigarettes is considered. While study participants were given multiple opportunities to update their smoking records to account for missed or off-schedule cigarettes, there is a possibility that some events were not captured. We believe our EMA data reflects general smoking trajectories accurately, but some inconsistencies are possible.

Finally, inclusion of antecedents, concurrent events, and consequences is very important for an accurate class enumeration. In the current study, we were limited to antecedents in the form of baseline covariates, and consequences in the form of the hospital admission smoking outcome, while concurrent events were not considered. Baseline covariates played an important role in defining profiles of individuals falling within each latent class. The distal outcome was useful in examining the impact that each behavioral pattern had on the smoking status. This smoking condition was assessed at hospital admission, which, for most people was a few days after their originally scheduled quit date. A more distal measure would have been a stronger indicator of latent class differences. It would also allow examining the long-term impact of cessation behavior.

Implications for Future Research

EMA are increasingly used as a data collection tool in social sciences, capturing personal experiences in real-time. Such data are rich, but very complex methodologically. As an increasing number of scientists recognize that traditionally used statistical methods limit research questions that can be answered, there will be a greater demand for more flexible statistical tools. Growth Mixture Modeling is a methodology with great potential, designed for working with longitudinal data but it has never been applied to Intensive Longitudinal Data.

In this study, we demonstrated how data with long vectors of repeated assessments can be incorporated in an existing statistical package for GMM analysis. The available software may not be ideal for the analysis of EMA-type of data, but, with effort, it allows answering complex questions that may have important practical implications. EMA allows examination of micro-behavior on a daily basis across individuals, while GMM allows studying the qualitative differences in these behaviors, identifying profiles of individuals exhibiting similar developmental patterns, and determining their long-lasting effects.

This work opens exciting opportunities but it also poses multiple methodological questions that need to be addressed. To mention a few, there is a need to know more about the sample size requirements, performance of various model fit indices, effects of missing data on parameter estimations, and incorporation of time-varying covariates measured at different time points from the main outcome data. It is also hard to underestimate the importance of the theoretical work that should precede the data analysis. Reliance on the theoretical justification of class selection is very important. One should consider statistical, graphical, and theoretical indicators of model fit in combination.

There are other complex parametric and non-parametric methods available for analyzing ILD. Most of them find applications in engineering and medicine, and take into account a cyclical nature of processes as well as dynamic interrelationships between variables. While they are still to be applied and disseminated in social sciences, GMM has been extensively used and can greatly benefit the analysis and interpretation of EMA data.

Notes

[1] SEM notation for the GMM is outlined in the Appendix 1.

[2] Latent parameters in forms of intercepts and slopes can also be added as predictors to the second level of the model. An example of such extended equation is given in Appendix 1 in Equation 8.

[3] Another software package, WinBUGS (Bayesian inference Using Gibbs Sampling, 2008), allows estimating mixture models from the Bayesian perspective employing Markov chain Monte Carlo (MCMC) methods.

Appendix 1

Mixed-Effects and Growth Mixture Models: SEM Notation

Mixed-Effects Model

Let y_{it} be a vector of outcomes, the number of cigarettes smoked per day, measured for all individuals *i* across all study days *t*, with the dimensions *n x 1*, where *n* is the total number of observations across all individuals. The frequency of the outcome varies across individuals due to differences in length of recorded observations. The outcome is a count, Poisson-distributed variable, modeled by the density function:

$$P(y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad for \; y = 0, 1, 2, \ldots,$$

where λ is the mean and standard deviation parameter. The outcome is defined such that y_{it} = ln (λ_{it}), where ln (λ_{it}) is the natural logarithm of the daily smoking rate . After incorporating the outcome link function, estimation of developmental trajectories for a single-class model is expressed in the following way (Muthén & Asparouhov, 2008a):

$$\ln(\lambda_{it}) = \eta_{0i} + \eta_{1i} t_{it} \tag{1}$$

$$\eta_{0i} = \alpha_0 + \gamma_0 x_i + \zeta_{0i} \tag{2}$$

$$\eta_{1i} = \alpha_1 + \gamma_1 x_i + \zeta_{1i}, \tag{3}$$

where η_{0i} and η_{1i} are random intercept and slope parameters estimated for each individual in the study. The expected value for η_{0i} is α_0, a baseline smoking rate for $x_i = 0$, varying with person-level predictors $\gamma_0 x_i$. The expected value for η_{1i} is α_1, a daily reduction in smoking for $x_i = 0$, varying with person-level predictors $\gamma_1 x_i$.

ζ_{0i} and ζ_{1i} are intercept and slope residuals that are normally distributed with the mean of 0 and variances Ψ_0 and Ψ_1, respectively. For the Poisson-distributed data, the first level residuals, ε_{it}, are omitted from the model as the outcome variable is an expected rate. While the outlined model in Equations 1 through 3 uses different symbolic representation of parameters, it is identical to the model defined by Raudenbush & Bryk (2002), with Equation 1 being the first level, and Equations 2 and 3 being the second-level model.

The above model can be expressed in general terms using structural equation modeling matrix notation, which is used in presenting GMM (Muthén & Asparouhov, 2008a; Muthén et al., 2002):

$$LN(\lambda_{it}) = \boldsymbol{\Lambda}\,\boldsymbol{\eta}_i \tag{4}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X}_i + \boldsymbol{\zeta}_i \tag{5}$$

In Equation 4, $\boldsymbol{\Lambda}$ is a n x p matrix, where p is the number of growth defining parameters (in this case, p = 2, as only intercept and slope define smoking trajectories):

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \cdots & \cdots \\ \cdots & \cdots \\ 1 & n \end{bmatrix}$$

The first column of this matrix contains intercepts, and the second column includes days of observations. Different people would have a different number of days listed in the second column of this matrix, skipping days for which data are not available.

$\boldsymbol{\eta}_i$ is a p x 1 vector of random intercept and slope parameters: $\boldsymbol{\eta}_i = \begin{bmatrix} \eta_{0i} \\ \eta_{1i} \end{bmatrix}$. In the Equation 4, $\boldsymbol{\eta}_i$ were defined as random, but they can also be fixed to $\boldsymbol{\eta} = \begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix}$ in the case when all individuals in a sample are assumed to follow the same developmental path. $\boldsymbol{\alpha}$ is a *p x 1* vector of expected values of $\boldsymbol{\eta}_i$ for $x_i = 0$: $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$. $\boldsymbol{\Gamma}$ is a *p x q* matrix, where *q* is the number of person-level predictors. With only one predictor, $\boldsymbol{\Gamma}$ is defined as $\begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix}$. $\boldsymbol{X}_i$ is a *q x p* matrix, and, if there is only one predictor, $\boldsymbol{X}_i = \begin{bmatrix} x1_i & x1_i \end{bmatrix}$. Finally, ζ_i is a *p x 1* vector of random effects: $\zeta_i = \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix}$.

Equations 4 and 5 define the mixed-effects model in a traditional way. In structural equation modeling language, Equation 4 would constitute the measurement part of the model, as the outcome variable LN (λ_{it}) defines continuous intercept and slope latent parameters $\boldsymbol{\eta}_i$, measured without error. The structural part is defined by Equation 5, where latent parameters are influenced by observed covariates $\boldsymbol{X}_i$. The structural part can be extended to incorporate a possible link between $\boldsymbol{\eta}_i$ parameters, so that

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{B}\,\boldsymbol{\eta}_i + \boldsymbol{\Gamma X}_i + \boldsymbol{\zeta}_i\ , \tag{6}$$

where $\boldsymbol{B}$ is a *p x p* matrix of structural parameters.

Growth Mixture Model

The model defined in Equations 4 and 6 can be easily extended to several latent classes. In growth mixture modeling, the outcome variable LN (λ_{kit}) is the rate of smoking for an individual i coming from a particular latent class k on day t. K is the latent class variable for each individual i, ranging from 1 to k unobserved classes. The general form of the model becomes

$$LN(\lambda_{itk}) = \boldsymbol{\Lambda}_k \boldsymbol{\eta}_{ik} \tag{7}$$

$$\boldsymbol{\eta}_{ik} = \boldsymbol{\alpha}_k + \boldsymbol{B}_k \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k \boldsymbol{X}_i + \boldsymbol{\zeta}_{ik} \tag{8}$$

$$\boldsymbol{\zeta}_{ik} \sim N(0, \boldsymbol{\Psi}_k), \tag{9}$$

where each parameter is class-specific, yielding unique estimates for each latent class K.

The model for the person-level categorical latent class variable K is a multinomial logistic regression for unordered responses:

$$LN\left[\frac{P(K = k \mid X_i)}{P(K = reference \mid X_i)}\right] = \boldsymbol{b}_{0k} + \mathbf{X}_i * \boldsymbol{b}_{1k} \tag{10}$$

The natural logarithm of the probability of being in a particular class versus being in a reference class is estimated from a class-specific average $\boldsymbol{b}_{0k}$, adjusted for baseline predictors $\boldsymbol{b}_{1k}$*$\mathbf{X}_i$. The outcome set of non-zero probabilities for every individual i is a n x k matrix, $\boldsymbol{b}_{0k}$ is a n x k matrix of model intercepts, $\mathbf{X}_i$ is a n x q matrix of covariates, and, $\boldsymbol{b}_{1k}$ is a q x k matrix of class-specific slopes.

Equations 7 through 10 define a general growth mixture model, which can be extended to incorporate connections between observed or latent variables to each other or to constrain some parameters. For example, random residuals $\boldsymbol{\zeta}_{ik}$ in Equation 8 can be constrained to zero in all or some latent classes to model developmental trajectories with fixed intercept or slope parameters. If all $\boldsymbol{\zeta}_{ik}$ are non-varying, the growth mixture model is referred to as latent growth curve analysis (Duncan et al., 1999; Muthén, 2004).

The model can also be extended to incorporate a distal outcome u_i. In the current study, u_i is a binary outcome of smoking status (quit or smoke), which is modeled with a logistic regression:

$$\ln\left[\frac{P(u_1 = quit \mid k_i, x_i)}{P(u_1 = smoke \mid k_i, x_i)}\right] = \boldsymbol{K}_i * \boldsymbol{v}_{0k} + \boldsymbol{X}_i * \boldsymbol{v}_{1k} \tag{11}$$

Based on this equation, the natural logarithm of the probability of quitting at surgery versus not quitting is modeled as a function of the overall class-specific quitting rate, adjusted by person-level covariates $\boldsymbol{X}_i$. The *n x k* matrix $\boldsymbol{K}_i$ contains indicators of class membership for each individual based on the highest posterior probability computed in Equation 10, such that

$$\boldsymbol{K}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ \dots & \dots & \dots \\ n & n & n \end{bmatrix}.$$

$\boldsymbol{v}_0$ is a *k x 1* vector of intercepts for latent classes *K*. $\boldsymbol{X}_i$ is a *n x q* matrix of covariates, and $\boldsymbol{v}_{1k}$ is a *q x k* matrix of class-specific slopes.

Appendix 2

Mplus Syntax for the Final 1 Class Model with a Distal Outcome

Syntax	Comments
DATA:	
FILE is data_long_modified.dat;	Specify the data file
DEFINE:	
NumYRsmo = NumYRsmo - 35;	Modify variables (centering)
VARIABLE:	
NAMES ARE SubjID SRSday allSmo …;	Names are assigned to all variables in the data file
MISSING = ALL (999);	Specify missing values
USEVARIABLE = SubjID SRSday allSmo NumYRsmo SEMeanC QuitSurg FagerC;	Specify variables for the current analysis
CLUSTER = SubjID;	Define the clustering variable for the two-level model
CATEGORICAL = QuitSurg;	Define the categorical distal outcome
COUNT = allSmo;	Define the Poisson distributed count outcome
WITHIN = SRSday;	Define the level 1 time variable
BETWEEN = NumYRsmo SEMeanC FagerC;	Define the level 2 covariates
ANALYSIS:	
TYPE = TWOLEVEL RANDOM;	Specify the two-level model with random effects
ALGORITHM = INTEGRATION;	Specify that numerical integration is used to obtain maximum likelihood estimates
CHOLESKY = OFF;	Turning off Cholesky optimization method for numerical integration
STARTS = 20 5;	Specify the number of EM random starts
MODEL:	
%WITHIN%	The level 1 MLM model
sl | allSmo ON SRSday;	The outcome allSmo (the number of daily cigarettes) is regressed ON the time variable SRSday. The slope sl is random, denoted by the sign |.
%BETWEEN%	The level 2 MLM model
f0 BY allSmo;	The latent intercept factor f0 is introduced, such that f0 is measured by the intercept allSmo. This is only needed for a single-class MLM, where level 2 intercepts and slopes are used to predict a distal outcome
allSmo@0;	The variance of the intercept allSmo is constrained to zero, while the variance of

	the intercept factor f0 is freely estimated. This is done to avoid two intercepts in the model
f0 ON FagerC;	The intercept f0 is regressed on the covariate FagerC
f1 BY sl;	The slope factor f1 is measured by the slope parameter sl
sl@0	The variance of the original slope parameter is constrained to zero
f1 ON SEMeanC NumYRsmo;	The slope parameter f1 is regressed on the covariates
f0 WITH f1@0;	The covariance of intercept and slope parameters is constrained to zero
QuitSurg ON f0 f1;	The distal outcome QuitSurg is regressed on intercept and slope parameters
OUTPUT:	
SAMPSTAT	Request sample descriptive statistics
PATTERNS;	Request a summary of missing data patterns

Appendix 3

Mplus Syntax for the Final 3 Class Model

Syntax	Comments
DATA: FILE is data_long_modified.dat;	
DEFINE:	
OQD = OQD - 10;	
NumYRsmo = NumYRsmo - 35;	
age = age - 55;	
VARIABLE:	
NAMES ARE SubjID SRSday allSmo … ;	
MISSING = ALL (999);	
USEVARIABLE = SubjID SRSday allSmo SEMeanC OQD age QuitSurg NumYRsmo;	
CLASSES = cb (3);	Specify the number of latent classes (3)
CLUSTER = SubjID;	
COUNT = allSmo;	
CATEGORICAL = QuitSurg;	
WITHIN = SRSday;	
BETWEEN = cb SEMeanC OQD age QuitSurg NumYRsmo;	The latent class variable cb is specified as a between-level factor, since classes are defined on the individual rather than observational level
ANALYSIS:	
TYPE = TWOLEVEL RANDOM MIXTURE;	Specify the two-level mixture model with random effects
ALGORITHM = INTEGRATION;	
CHOLESKY = OFF;	
STARTS = 1500 10;	
STITERATIONS = 20;	The maximum number of iterations in the initial EM stage is increased from 10 to 20 for a more thorough investigation of results
LRTSTARTS = 20 10 200 30;	Specify the number of random starts for the BLRT test
MODEL:	
%WITHIN%	The level 1 MLM model
%OVERALL%	The overall model
allSmo ON SRSday;	The outcome allSmo is regressed on the time variable SRSday. This model has a fixed slope parameter.
%cb#2% allSmo ON SRSday; %cb#3% allSmo ON SRSday;	Request a separate estimation of the model parameters (intercept and slope) for classes 2 and 3

%BETWEEN%	
%OVERALL%	The level 2 MLM model
allSmo;	The variance of the intercept is estimated
CB on SEMeanC OQD age NumYRsmo;	The class membership is predicted from covariates
%cb#2%	Request a separate estimation of the
allSmo;	intercept variances for classes 2 and 3
%cb#3%	
allSmo;	
OUTPUT:	
SAMPSTAT	
TECH7	Request sample statistics for each latent class
TECH11	Request the LMRT test of model fit
TECH14;	Request the BLRT test of model fit
SAVEDATA:	Saving data produced by the analysis
FILE is 3clRCBdistal4pr_LRTSTARTS.dat;	Assign a name to the saved file
SAVE = cprob;	Save posterior probabilities for each latent class for each individual

Bibliography

Affleck, G., Zautra, A., Tennen, H., & Armeli, S. (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*(5), 746-754.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Allison, P. (2001). *Missing data*. Thousand Oaks, CA: Sage Publication.

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27-51). Charlotte, NC: Information Age Publishing, Inc.

Baer, J. S., Holt, C. S., & Lichtenstein, E. (1986). Self-efficacy and smoking reexamined: Construct validity and clinical utility. *Journal of Consulting and Clinical Psychology, 54*(6), 846-852.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338-363.

Bayesian inference Using Gibbs Sampling (2008). Cambridge: Medical Research Council Biostatistics Unit. Available from www.mrc-bsu.cam.ac.uk/bugs.

Bentler, P. M., & Chih-Ping, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*(1), 78.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*(1), 192-208.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton Mifflin Company.

Carson, R. L. (2007). Exploring the episodic nature of teachers' emotions as it relates to teacher burnout. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 67, from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-99005-023&site=ehost-live

Cheong, Y., Yong, H.-H., & Borland, R. (2007). Does how you quit affect success? A comparison between abrupt and gradual methods using data from the International Tobacco Control Policy Evaluation Study. *Nicotine and Tobacco Research, 9*(8), 801-810.

Chernick, M. R. (1999). *Bootstrap methods: A practitioner' guide*. New York, NY: John Wiley and Sons Inc.

Cinciripini, P. M., Lapitsky, L., Seay, S., Wallfisch, A., Kitchens, K., & Van Vunakis, H. (1995). The effects of smoking schedules on cessation outcome: Can we improve on common methods of gradual and abrupt nicotine withdrawal? *Journal of Consulting and Clinical Psychology, 63*(3), 388-399.

Cinciripini, P. M., Wetter, D. W., & McClure, J. B. (1997). Scheduled reduced smoking: Effects on smoking abstinence and potential mechanisms of action. *Addictive Behaviors, 22*(6), 759-767.

Colder, C. R., Mehta, P., Balanda, K., Campbell, R. T., Mayhew, K., Stanton, W. R., et al. (2001). Identifying trajectories of adolescent smoking: An application of latent growth mixture modeling. *Health Psychology, 20*(2), 127-135.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57*(1), 505-528.

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

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Engel, S. G., Boseck, J. J., Crosby, R. D., Wonderlich, S. A., Mitchell, J. E., Smyth, J., et al. (2007). The relationship of momentary anger and impulsivity to bulimic behavior. *Behaviour Research and Therapy, 45*(3), 437-447.

Fagerstrom, K. O., Heatherton, T. F., & Kozlowski, L. T. (1990). Nicotine addiction and its assessment. *Ear, Nose, and Throat Journal, 69*(11), 763-765.

Gunther, V., Gritsch, S., & Meise, U. (1992). Smoking cessation: Gradual or sudden stopping? *Drug and Alcohol Dependence, 29*(3), 231-236.

Haughton, D. M. A. (1988). On the choice of a model to fit data from an exponential family. *Annals of Statistics, 16*(1), 342-355

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K.-O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119-1127.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64*(2), 627-634.

Hufford, M. R., & Shiffman, S. (2002). Methodological issues affecting the value of patient-reported outcomes data. *Expert Review of Pharmacoeconomics Outcomes, 2*(2), 89-98.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302–317.

Law, M., & Tang, J. L. (1995). An analysis of the effectiveness of interventions intended to help people stop smoking. *Archives of Internal Medicine, 155*(18), 1933.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika 88*(3), 767-778.

Maggi, S. (2008). Changes in smoking behaviors from late childhood to adolescence: 4 years later. *Drug and Alcohol Dependence, 94*, 251–253.

Maggi, S., Hertzman, C., & Vaillancourt, T. (2007). Changes in smoking behaviors from late childhood to adolescence: Insights from the Canadian National Longitudinal Survey of Children and Youth. *Health Psychology, 26*(2), 232-240.

Marcos, T., Godas, T., & Corominas, J. (2004). Nicotine replacement therapy versus gradual smoking withdrawal in smoking cessation. *Medicina Clinica, 123*(4), 127-130.

Martinez, S. (2004). *Patterns of disengagement of Mexican American adolescents in high school*. Paper presented at the Annual meeting of American Sociological Association.

McCarthy, D. E., Piasecki, T. M., Fiore, M. C., & Baker, T. B. (2006). Life before and after quitting smoking: An electronic diary study. *Journal of Abnormal Psychology, 115*(3), 454-466.

McCutcheon, A. L. (1987). *Latent Class Analysis*. Beverly Hills, CA: Sage Publications.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry & Neuroscience, 31*(4), 13-20.

Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data *Sociological Methodology, 27*, 453-480.

Muthén, B. (2000). Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking ages 18-37. In J. S. Rose, L. Chassin, C. C. Presson & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods and new questions* (pp. 113-140). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2001a). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Muthén, B. (2001b). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 291-322). Washington, DC: American Psychological Association.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81-117.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 369-377.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. (2007). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1-24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, B., & Asparouhov, T. (2008a). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143-165). Boca Raton: Chapman & Hall/CRC Press.

Muthén, B., & Asparouhov, T. (2008b). Multilevel regression mixture analysis. Muthén & Muthén, Retrieved from http://www.statmodel.com/papers.shtml.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*(4), 459-475.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463-469.

Muthén, L. K., & Muthén, B. (2007). *Mplus User's Guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

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

Ostroff, J., & Burkhalter, J. E. (2007). *Presurgical smoking cessation intervention for cancer patients: The resolve study* (No. IRB Protocol No. 02-094A(14)). New York, NY: Memorial Sloan Kettering Cancer Center.

Ostroff, J., Burkhalter, J. E., Li, Y., Shiyko, M. P., & Holland, S. (2008). *Tobacco treatment adherence in newly diagnosed cancer patients*. Paper presented at the annual meeting of the Society for Research on Nicotine and Tobacco.

R Development Core Team (2008). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org.

Rathunde, K., & Csikszentmihalyi, M. (2005). Middle school students' motivation and quality of experience: A comparison of Montessori and traditional school environments. *American Journal of Education, 111*(3), 341-371.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2007). Hierarchical linear and nonlinear modeling (Version 6.04).

Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 364-368.

Rodriguez, D., & Audrain-McGovern, J. (2004). Team sport participation and smoking: Analysis with general growth mixture modeling. *Journal of Pediatric Psychology, 29*(4), 299-308.

Rodriguez, D., Moss, H. B., & Audrain-McGovern, J. (2005). Developmental heterogeneity in adolescent depressive symptoms: Associations with smoking behavior. *Psychosomatic Medicine, 67*(2), 200-210.

Schafer, J. L. (2006). Marginal modeling of intensive longitudinal data by generalized estimating equations. In T. A. Walls & J. L. Schafer (Eds.), *Models of intensive longitudinal data* (pp. 38-62). New York, NY: Oxford University Press, Inc.

Schwartz, J. E., & Stone, A. A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology, 17*(1), 6-16.

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

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343.

Shiffman, S. (2008). Two decades of change for PROs. *Applied Clinical Trials, 17*(1), 66-66.

Shiffman, S., & Hufford, M. R. (2001). Ecological momentary assessment. *Applied Clinical Trials, 10*(3), 42.

Shiffman, S., Patten, C., Gwaltney, C., Paty, J., Gnys, M., Kassel, J., et al. (2006). Natural history of nicotine withdrawal. *Addiction, 101*(12), 1822-1832.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Smyth, J., Wonderlich, S., Crosby, R., Miltenberger, R., Mitchell, J., & Rorty, M. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders, 30*(1), 83-95.

Smyth, J. M., & Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies, 4*(1), 35-52.

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., et al. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology, 75*(4), 629-638.

Snaith, R. P., & Zigmond, A. S. (1986). The hospital anxiety and depression scale. *British Medical Journal (Clinical Research Ed.), 292*(6516), 344-344.

Stanger, C. (2006). Latent growth mixture models: An important new tool for developmental researchers. *Infant and Child Development, 15*(6), 635-637.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*(3), 236.

Stone, A. A., Shiffman, S., Atienza, A. A., & Nebeling, L. (2007). Historical roots and rationale of ecological momentary assessment (EMA). In A. A. Stone, S. Shiffman, A. A. Atienza & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 3-10). New York, NY: Oxford University Press, Inc.

Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent* *variable mixture models* (pp. 317-341). Charlotte, NC: Information Age Publishing, Inc.

Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study.* Unpublished doctoral dissertation, University of Jyvaskyla, Finland.

Walls, T. A., Jung, H., & Schwartz, J. E. (2006). Multilevel models for intensive longitudinal data. In T. A. Walls & J. L. Schafer (Eds.), *Intensive longitudinal data* (pp. 3-37). New York, NY: Oxford University Press, Inc.

Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for Intensive Longitudinal Data*. New York, NY: Oxford University Press.

Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods, 10*(4), 635-656.

Wegner, K. E., Smyth, J. M., Crosby, R. D., Wittrock, D., Wonderlich, S. A., & Mitchell, J. E. (2002). An evaluation of the relationship between mood and binge eating in the natural environment using ecological momentary assessment. *International Journal of Eating Disorders, 32*(3), 352-361.

Willett, J. B., & Sayer, A. G. (1994). Using covariance and structure analysis to detect correlates and predictors of individual change. *Psychological Bulletin, 116*(2), 363.