Trajectory Modelling Techniques Useful to Epidemiological Research

Supervised by:

Dr.Mehdi Moradinazar

Presented by:
Ramin Ghasemi(PhD candidate)
Anita Ahmadi(MSc student)

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Trajectory Modelling Techniques Useful to Epidemiological Research: A Comparative Narrative Review of Approaches

Hermine Lore
Nguena Nguefack[1]
M Gabrielle Pagé[2,3]
Joel Katz[4]
Manon Choinière[2,3]
Alain Vanasse[5,6]
Marc Dorais[7]
Oumar Mallé Samb[1]
Anaïs Lacasse[1]

This article was published in the following Dove Press journal:
Clinical Epidemiology

Clinical Epidemiology 2020:12 1205–1222

3.2 (Q1) : 2024 Impact Factor

CONTENTS

01

Why Model Trajectories?

Why Model Trajectories?

Oversimplification of Data

In many studies, measured health outcomes are averaged out and their evolution across the entire study sample or pre-specified observed subgroups is analyzed.

describing populations of individuals using averaged estimates amounts to oversimplifying the complex intra- and inter-individual variability of the real-life clinical context.

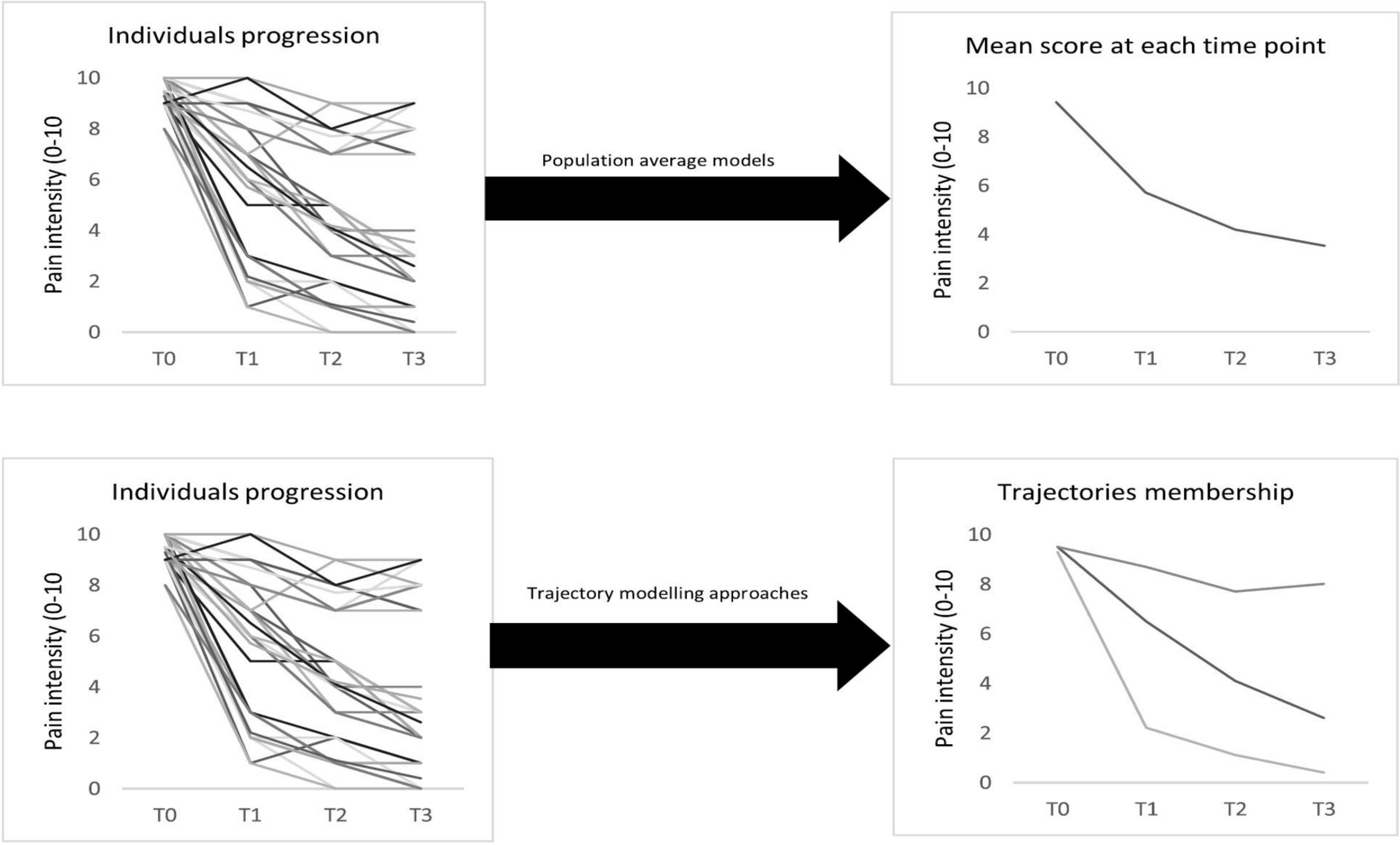
Importance of Subgroups

However, in most cases, unknown or unexpected subgroups of individuals share similar patterns of clinical symptoms, behaviours, or healthcare utilization.

Need for Precision

Such approaches focus on the relationships among individuals; their purpose is to classify individuals into distinct subgroups or classes based on personal response patterns.

Classification is done so that individuals within a given subgroup share greater similarities than individuals from separate subgroups.


Individuals progression
Pain intensity (0-10
10
8
6
4
2
0
T0
T1
T2
T3
Population average models
Mean score at each time point
Pain intensity (0-10
10
8
6
4
2
0
T0
T1
T2
T3
Individuals progression
Pain intensity (0-10
10
8
6
4
2
0
T0
T1
T2
T3
Trajectory modelling approaches
Trajectories membership
Pain intensity (0-10
10
8
6
4
2
0
T0
T1
T2
T3

Advantages of using terajectory modeling compared to measures based on sample means:

01

Allows the researcher to better characterize and understand intra- and inter-individual variability and patterns of health outcomes over time.

Such approaches can provide scientific evidence to optimize personalized healthcare focused on the needs of specific subpopulations.

02

It is useful in exploring heterogeneity of health profiles, to identify vulnerable populations who require better healthcare, and to identify trajectories leading to the best health outcomes.

02

Latent Class Framework

Latent Class Modelling Approaches:

Latent Classes

Latent class modelling are statistical models which include random variables that cannot be directly observed.Individuals are assigned to latent trajectory subgroups on the basis of their observed symptoms or behaviours.
Each subgroup is composed of individuals with relatively similar observations/scores on observed behaviours.

Temporal Patterns

Rather than evaluating individual time points or change between adjacent time points, longitudinal latent class modelling approaches identify subgroups of subjects who have a similar outcome pattern over the study period as a whole.

Four Latent Class Trajectory

Classification Based on Data Type:

longitudinal data:

- ✓ **GMM** **(Growth Mixture Modelling)**
- ✓ **GBTM** **(Group-Based Trajectory Modelling)**
- ✓ **LTA** **(Latent Transition Analysis)**

sectional data:

- ✓ **LCA** **(Latent Class Analysis)**

03

Growth Mixture Modelling

Growth Mixture Modelling

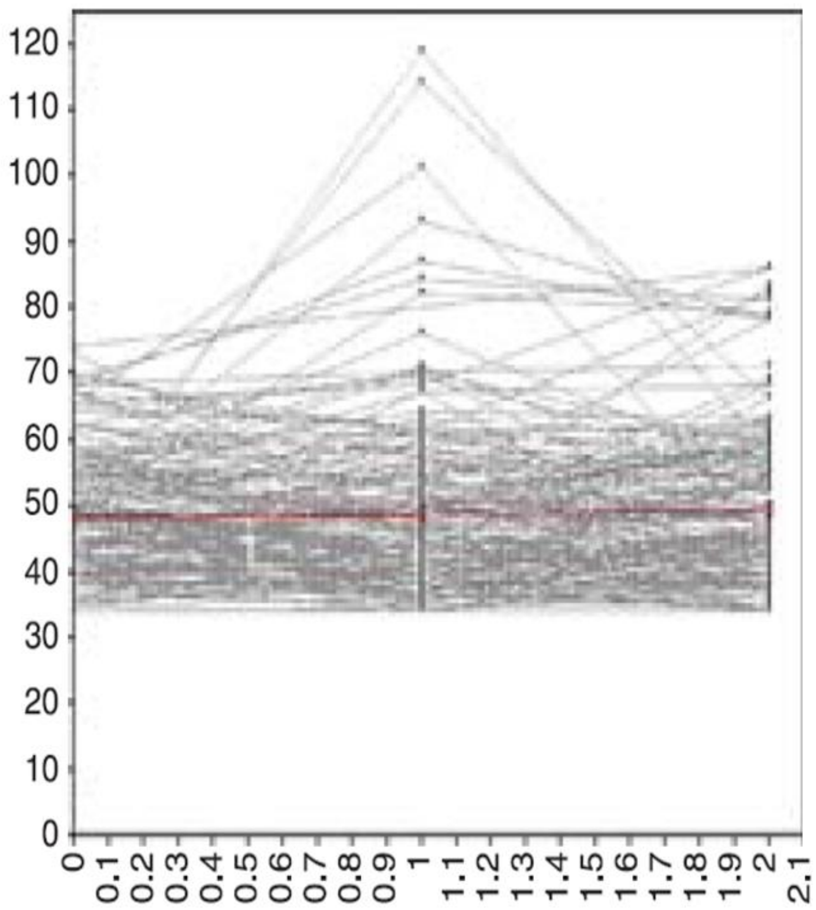
Finite Mixture Model

GMM is a finite mixture model. It assumes that in any given population, there exists a finite number of unobserved subpopulations or classes (latent classes) with similar behaviours or experiences.

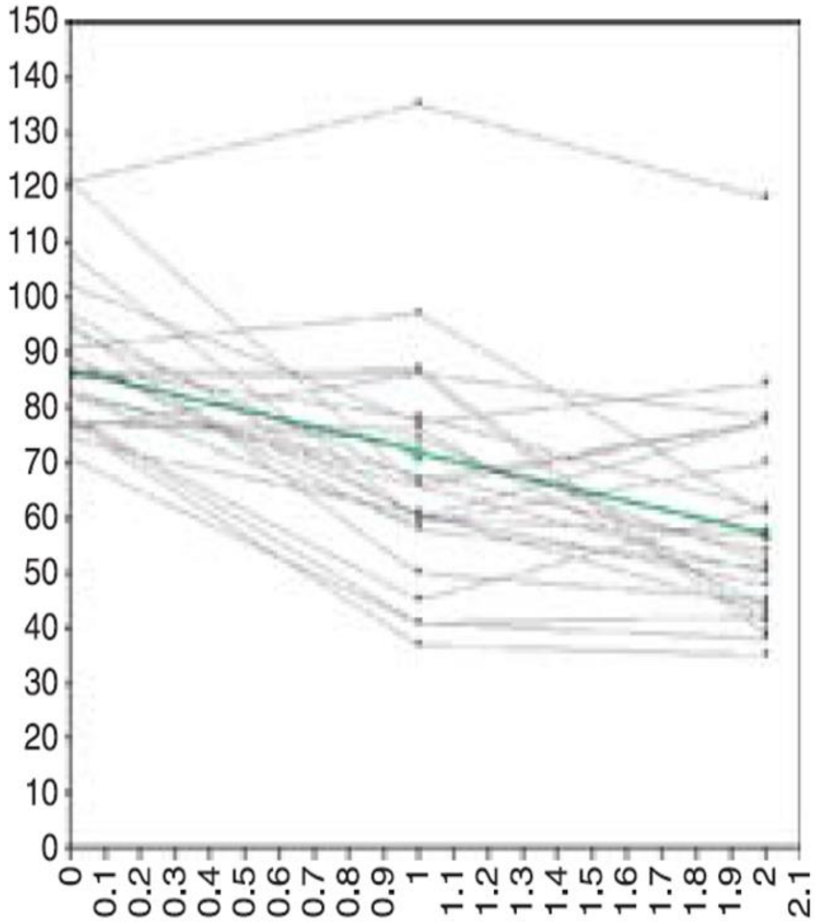
This stands in contrast to classic statistical models which assume that all individuals come from the same population with common population parameters.

It estimates an average growth curve for each class and allows for variations between individuals of the same class. This heterogeneity within classes is captured by introducing random effects in the model through which variances of the growth parameters can be estimated (intercepts and slopes).

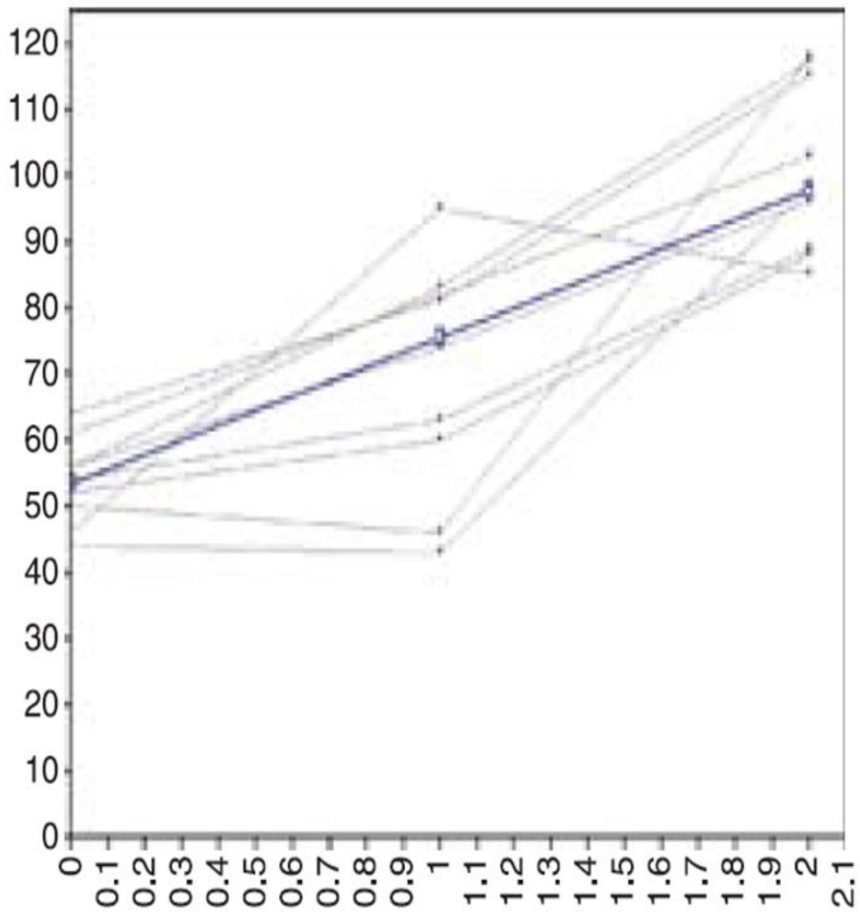
Growth Mixture Modelling



Class 1



Class 2



Class 3

Growth Mixture Modelling

The GMM is a model for longitudinal data that has been developed for the study of continuous data.

However, it was adapted to handle other types of data such as count data (with or without inflation at zero) and categorical data.

Four steps for conducting a GMM analysis:

Step 1: Definition of the Problem and Specification of the Number of trajectory Subgroups

The link between the research area and the method is formalized.

An appropriate analysis plan is developed. The expected number of latent classes and the shape of the curve for each class are hypothesized based on the researchers knowledge of the field and a descriptive analysis of the raw data.

Step 2: Model Specification

- During this step, a set of models can be specified and estimated. Researchers may make decisions about **growth parameters** (intercept, slope variance, and covariance) and the addition of covariates. Substantive theory and previous research should be used to guide these decisions, as much as possible.
- During these steps, researchers should decide if the **shape** of each trajectory over time should be linear, quadratic or cubic (intercept and slope parameters).
- They should also decide if growth factor **variances** should be specific to each class, if within-class growth factor covariances should be different from zero, and if outcome residual variances should be invariant with respect to class.

Step 3: Model Estimation

GMM can be estimated by maximum likelihood or by Bayesian methods.

Step 4: Model Selection and Interpretation

The objective of this step is to determine which of the models tested provides the best or most reasonable.

The goodness of fit of the various models should be compared using the Lo- Mendell-Rubin adjusted likelihood ratio test

and/or the parametric bootstrapped likelihood ratio test ($p<0.05$ indicates better fit)

and/or the Bayesian Information Criteria (BIC) (best models have smaller BIC).

Researchers should also take into account convergence, the ability of the model to provide well separated classes (**entropy** near 1.0), the proportion of the sample in each trajectory (more than 5% is recommended), average posterior probabilities (near 1.0), parsimony and the usefulness of the observed latent classes in practice.

Lo- Mendell-Rubin adjusted

	2-class vs 3-class	3-class vs 4-class	4-class vs 5-class
Value	1088.997***	392.121**	204.070
P value	0.000	0.001	0.051

*p<.05, **p<.01, ***p<.001.

Growth Mixture Modelling

Advantages

- ✓ -Handling missing data
- ✓ -Allowing for correlated residuals
- ✓ -Latent class modelling
- ✓ -Estimates a mean growth curve
- ✓ -Identification of individual heterogeneity

- ✓ -Complexity of interpreting results

Limitations

04

Group-Based Trajectories Modelling

Group-Based Trajectories Modelling

Like GMM, GBTM—also known as Latent Class Growth Analysis (LCGA)—is a finite mixture model.

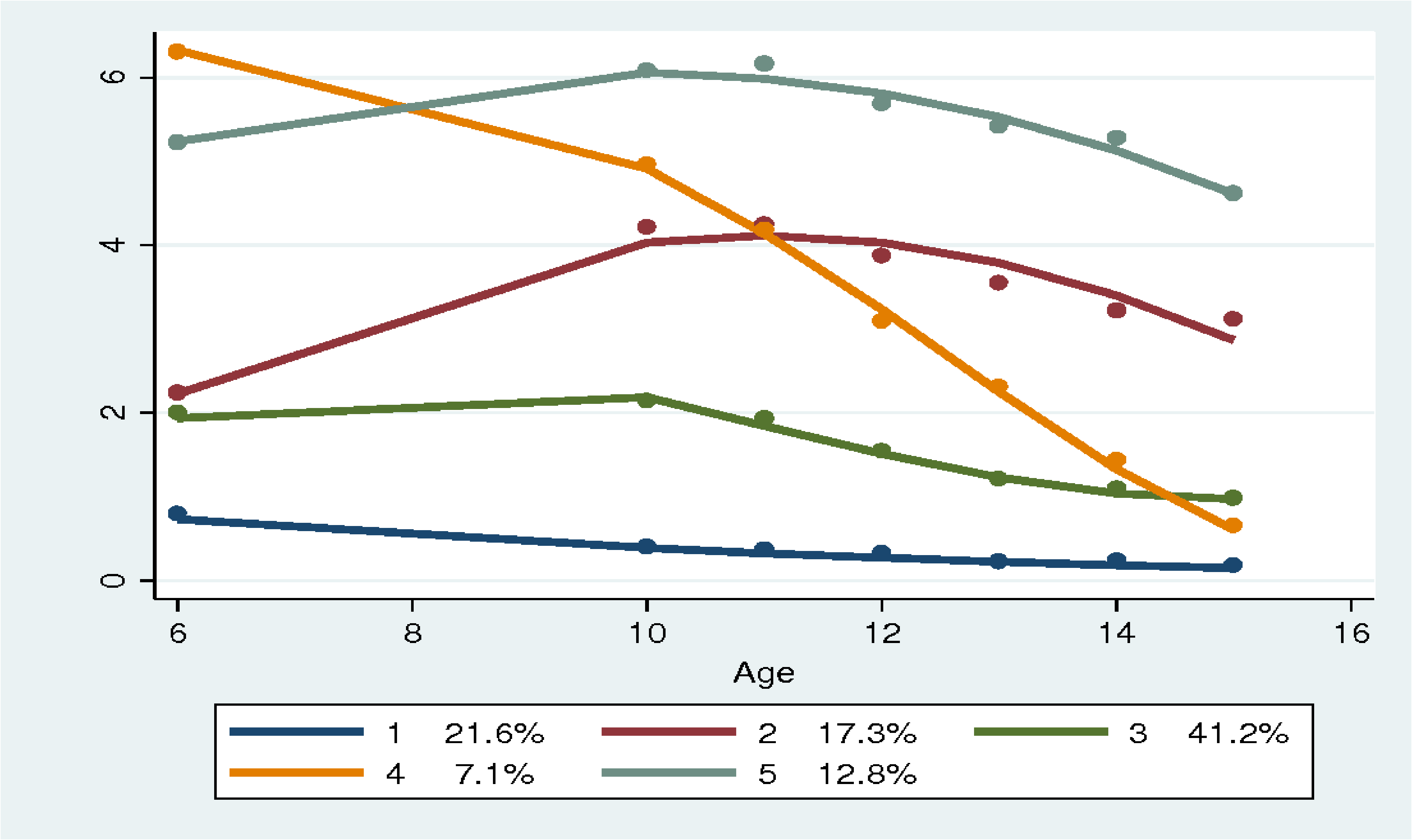
While GMM estimates the withinclass variance, GBTM assumes that there is no variation between individuals in the same class (no within-class variance on the growth factors). Indeed, GBTM is a simplified version of GMM.

Group-Based Trajectories Modelling

For example, in the case of the aforementioned three pain intensity trajectory subgroups(no improvement, gradual improvement, rapid improvement), GBTM assumes that in each class, individuals have the same pain intensity evolution. The proportion of the population belonging to each of these subgroups is then estimated.

- The model also determines, for each individual, the probability of belonging to one subgroup or another (posterior group probability).

- As in GMM, Parameters are estimated by maximizing likelihood. Covariates (that vary or not over time) can also be included in the model.



0
2
4
6
6
8
10
12
14
16
Age
1 21.6%
2 17.3%
3 41.2%
4 7.1%
5 12.8%

Four steps for conducting a GBTM analysis:

Step 1: Definition of the Problem and Specification of the Number of Trajectory Subgroups.

Same as previously described for GMM.

Step 2: Model Specification

- ❑ It is suggested to first test a one-group model, then gradually adjust the maximum logical number of subgroups.
- ❑ This maximum logical number of subgroups should be greater than the expected number of subgroups.
- ❑ If the quadratic component of this model is not significant, the model for a linear trajectory is to be run to determine the BIC value of this model.
- ❑ If the quadratic component of the model for a trajectory is significant, the analysis of the quadratic model for two trajectories is performed

Step 3: Model Estimation

Same as previously described for GMM.

Step 4: Model Selection and Interpretation

1) preference for a useful and parsimonious model that fits the data well

2) close correspondence between the estimated probability of each subgroup and the proportion of individuals classified in such subgroup according to the rule of attribution of the maximum probability of belonging

3) average posterior probabilities of subgroup membership greater than or equal to 0.7 for each subgroup

4) sufficient number of individuals in each subgroup (more than 5%)

5) reasonably narrow confidence intervals

6) difference of BICs between two models with different numbers of trajectory subgroups

Group-Based Trajectories Modelling

Advantages

GBTM is a simpler specification of the GMM, and both have the same advantages regarding handling missing data and allowing for correlated residuals.

Unlike GMM, GBTM estimates fewer parameters and can thus run faster with fewer errors.

GBTM supposes that all individuals in a trajectory class have the same behaviour, whereas GMM allows for within-class variation). This means that, when GBTM is used, researchers can discuss differences between subgroups, but not differences within Subgroups(.

Limitations

05

Latent Transition Analysis

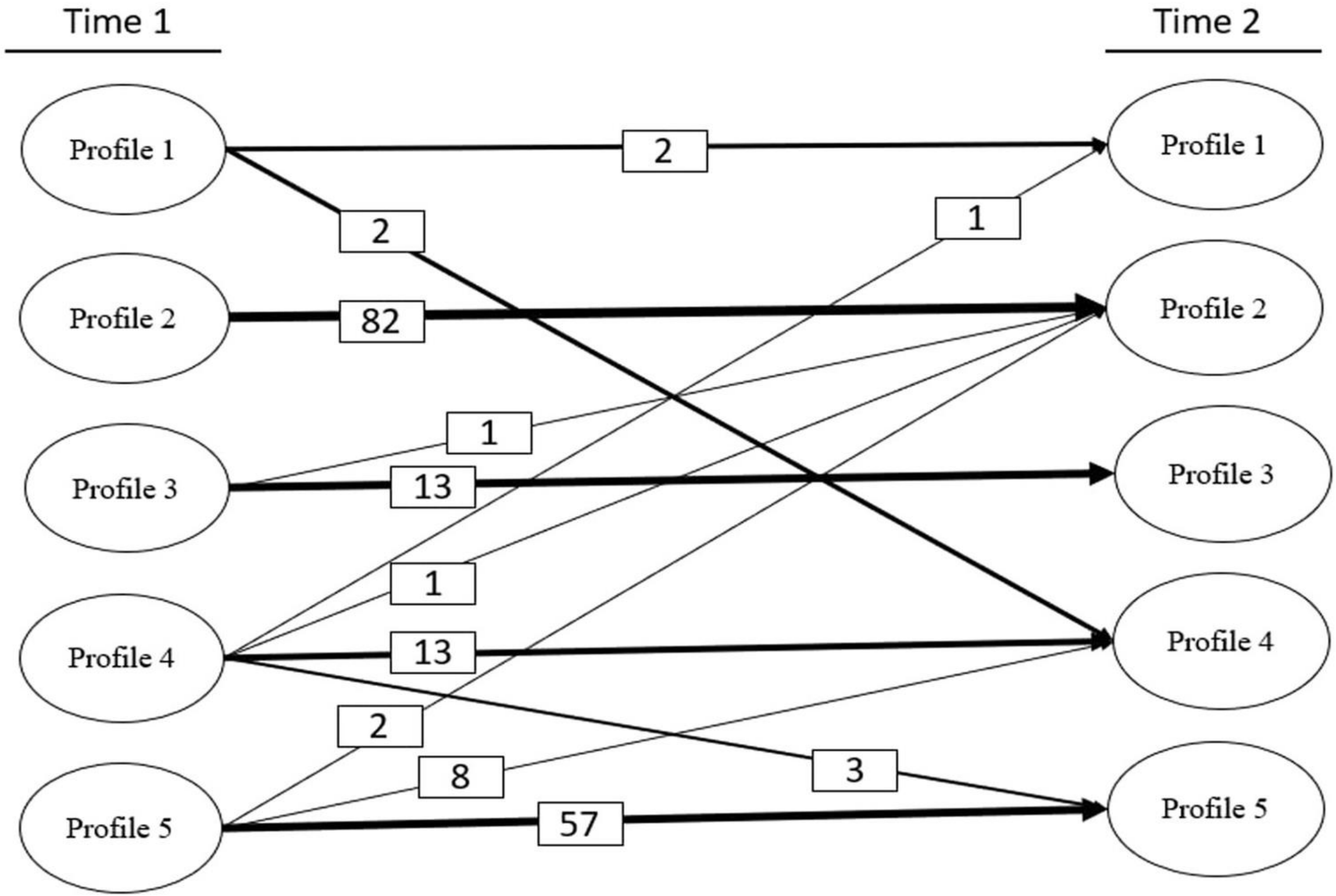
Latent Transition Analysis (LTA):

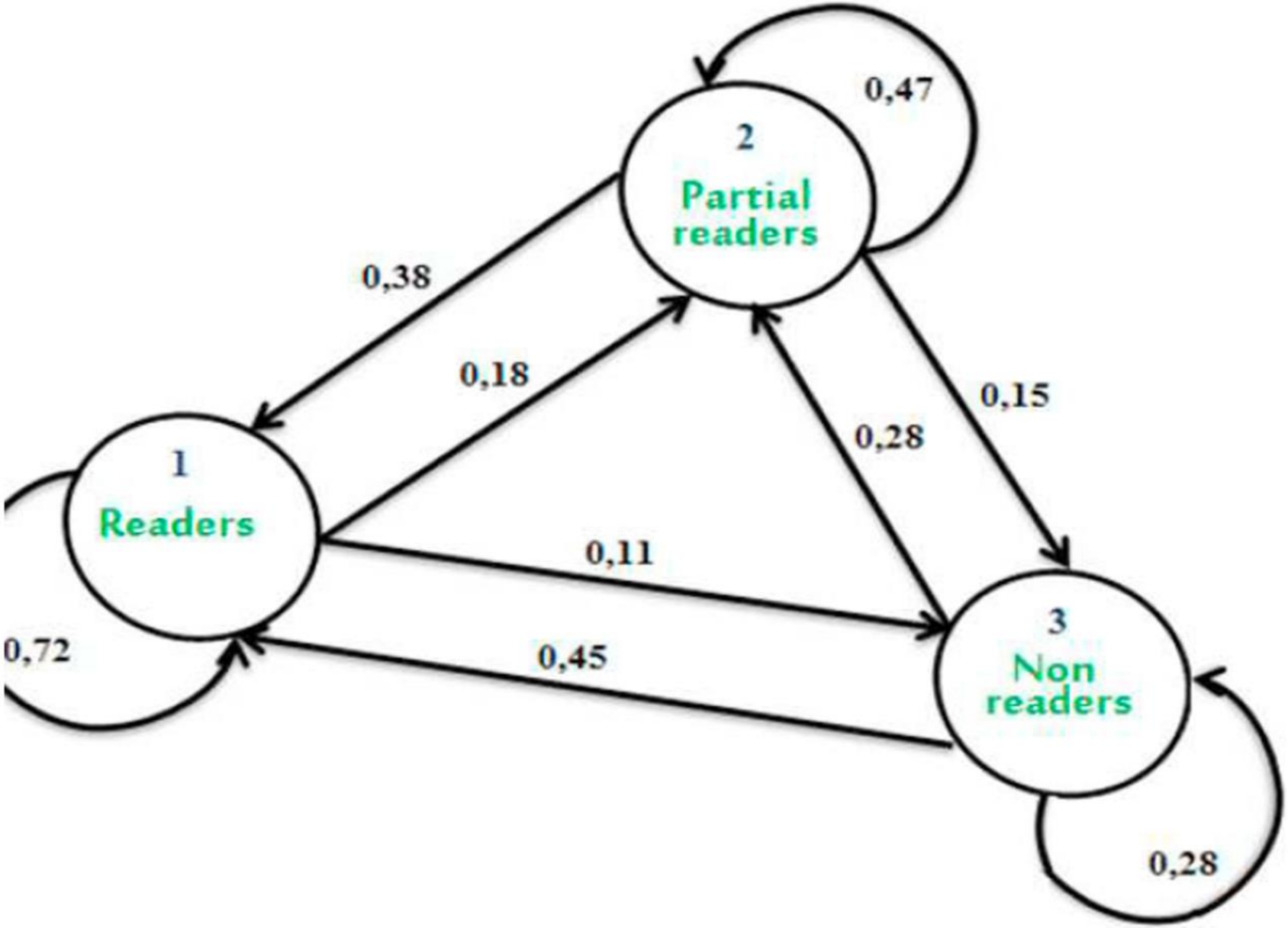
The model assumes that individuals can change their class membership over time.

For example, in the case of three pain intensity subgroups (mild/moderate/severe), LTA allows for individuals to switch from the severe subgroup at one time point to the mild or moderate subgroup at the next time point, and so on.

LTA is used to analyze changes in multiple categorical variables over time (eg, yes/no, mild/moderate/severe).

LTA uses observed data from a set of categorical variables to define a latent variable for each time point.


Time 1
Time 2
Profile 1
Profile 2
Profile 3
Profile 4
Profile 5
Profile 1
Profile 2
Profile 3
Profile 4
Profile 5
2
2
1
82
1
13
1
13
2
8
3
57



2
Partial readers
0,47
0,38
0,18
0,15
0,28
1
Readers
0,11
0,72
0,45
3
Non readers
0,28

The primary objective of this approach is to study the probability of transition of an individual from one class at one time point to another class at the next time point.

Like GMM and GBTM, covariates can be added to LTA models. However, LTA requires that the number of classes be chosen before adding covariates principally to avoid a potential change in class number with and without covariates.

The model estimates the following parameters:

1

The latent status membership probabilities at Time 1

2

The proportion of the population in each latent class at each time point (latent status has been defined as a subgroup/class in which individuals' memberships can change over time)

3

The conditional probabilities of making a transition from one latent status to another over time (eg, probability of being in latent status L2 at Time t given a latent status L1 at Time t-1)

4

The item-response probabilities conditional on latent status membership (analogous to posterior group probabilities).

Four steps for conducting a LTA analysis:

Step 1: Definition of the Problem and Specification of the Number of trajectory Subgroups.

The choice of the number of latent classes is based on the result of a hypothesis test.

Step 2: Model Specification

During this step, researchers make a decision about the time invariance of item-response probabilities, the measurement invariance for transition probabilities (in order to achieve model identification and to facilitate the interpretation of classes prevalence), and the addition of covariates.

Step 3: Model Estimation

During this step, the estimation method should be chosen before fitting the models. LTA models can be estimated by maximum likelihood using the expectation maximization algorithm.
They can also be estimated with Bayesian methods using Markov chain Monte Carlo algorithms.

Step 4: Model Selection and Interpretation

In the LTA, the AIC and BIC are used to select the best model, favoring the one for which these two values are smaller in absolute value.

Advantages

-Comparing different subgroups

-Useful to model a change over time

-Large sample sizes

-Number of time points

Limitations

YOUR LOGO

THANK YOU