RESEARCH ARTICLE • OPEN ACCESS

Latent Markov Modeling of HIV Vulnerability Transitions Among Adolescent Girls and Young Women in Zimbabwe

Fungai Hamilton Mudzengerere, Oliver Bodhlyera, and Henry Mwambi

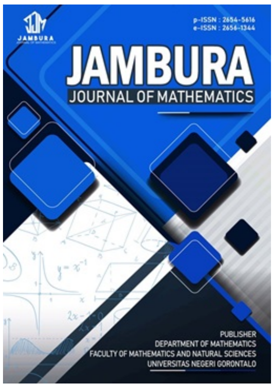
JAMBURA
JOURNAL OF MATHEMATICS

Volume 8, Issue 2, Pages 273–285, August 2026

Received 25 June 2026, Revised 10 August 2026, Accepted 17 August 2026, Published 30 August 2026

To Cite this Article : F. H. Mudzengerere, O. Bodhlyera, and H. Mwambi, "Latent Markov Modeling of HIV Vulnerability Transitions Among Adolescent Girls and Young Women in Zimbabwe", *Jambura J. Math*, vol. 8, no. 2, pp. 273–285, 2026, doi: *https://doi.org/10.37905/jjom.v8i2.40071*.

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS



Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
Journal Abbreviation	:	Jambura J. Math.
Frequency	:	Biannual (February and August)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/jjom
Online ISSN	:	2656-1344
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
Google Scholar ID	:	iWLjgaUAAAAJ
Email	:	info.jjom@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Mathematics, Volume 8, Issue 2, Pages 273–285, August 2026 https://doi.org/10.37905/jjom.v8i2.40071

Research Article

Check for updates

Latent Markov Modeling of HIV Vulnerability Transitions Among Adolescent Girls and Young Women in Zimbabwe

Fungai Hamilton Mudzengerere[1], **Oliver Bodhlyera**[1], **Henry Mwambi**[1,*]

[1] *School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa*

ARTICLE HISTORY

Received 25 June 2026
Revised 10 August 2026
Accepted 17 August 2026
Published 30 August 2026

KEYWORDS

Latent markov modeling
Longitudinal data
HIV
Vulnerability
Adolescent girls and young women

ABSTRACT. *This study uses applied Latent Markov modeling (LMM) to assess longitudinal transitions in HIV vulnerability among adolescent girls and young women (AGYW) enrolled in the Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) program across nine districts in Zimbabwe, from October 2023 to September 2024. Using routinely collected program data for 4,341 AGYW aged 10–19 years, the analysis identified distinct latent classes of HIV vulnerability and tracked their transitions over a 12-month period. For AGYW aged 10–14 years, four latent classes, that is, very high, high, medium, and low risk were identified with vulnerability factors of in-school dropout risk, orphanhood, alcohol use, and exposure to violence. The probability of maintaining low-risk status was high (0.704), while the probability of transitioning from very high to high risk was comparatively modest (0.252), suggesting only limited movement out of the highest-risk category among AGYW. Among the 15–19-year-old AGYW, vulnerability to HIV was mainly driven by sexual behavior, condom use, pregnancy history, and transactional sex, with the highest probability of remaining in the same class observed in high-risk (0.819) and low-risk (0.714) states. Covariates such as area of residence and source of income significantly influenced transitions, where AGYW with family or informal income were less likely to shift to higher-risk states. Overall, the results demonstrate transitions consistent with reductions in HIV risk levels among AGYW during the period of DREAMS participation, alongside the program's social, educational, and economic empowerment components. In this study, Latent Markov modelling (LMM) was applied as a robust analytical framework for monitoring longitudinal changes in vulnerability and evaluating program impact in low-resource settings.*

1. Introduction

Tracking longitudinal transitions of categorical variables observed on specific individuals or groups is critical in determining the impact of an intervention as pointed out by Melendez-Torres et al. [1]. However, the categorical latent variables pose challenges in tracking transitions by virtue of their complexity in nature, hence the need for use of powerful approaches like Latent Markov modelling (LMM) for tracking longitudinal changes in variables. Kim et al. [2] defined Latent Markov modelling as a statistical tool for tracking sequential dynamic probabilistic transitioning of categorical variables over time. The technique is an extension of Latent Class Analysis (LCA) where subgroups are created for individuals based on observed categorical variables and then tracked for longitudinal transition between the classes over time [3]. Latent Markov modelling has been widely used in education, psychological, health, engineering, and environment studies among others. For example, Melendez-Torres et al and Bonell et al. [1, 4] researched on longitudinal tracking of prevention of bullying in a school set-up and the technique proved useful in measuring transitioning from high risk to lower risk classes. Noor et al. [5] researched on drug and substance use among men who have sex with men (MSM) in the human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) epidemic and obtained the probabilities of transitions of different participating groups with a history of drug use. The technique can also be applied in tracking longitudinal transitions across of HIV risk classes among adolescent girls and young women (AGYW) enrolled in the Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) program. The DREAMS program was introduced to reduce the burden of HIV in low resource settings as pointed out by Wambiya et al. [6] and was being implemented in sub-Saharan African countries including Zimbabwe. The DREAMS program was offered through a layering approach where AGYW form social asset building clubs (SABC) for sharing information, running savings clubs and participating in income generation activities, and participants were referred for clinical, social and post violence support as outlined by United States Agency for International Development (USAID) [7].

Latent Markov models describe changes in latent classes of categorical response variables through a mixture of Markov processes which assume that the future probability of being in a latent class after a transition period is determined by the current state [8, 9]. The technique accounts for time-changing repeated measures and the effect of observable covariates on the response variables through estimating the probability of the transition.

The transition probability of being in a particular latent

*Corresponding Author.

Email : *fmudzengerere@gmail.com* (F. H. Mudzengerere)
Homepage : http://ejurnal.ung.ac.id/index.php/jjom/index / E-ISSN : 2656-1344

state (or class) at time t, that is, X_t given the previous state (class) X_{t-1} having set of parameters w and a set of covariates z_{it} is given by:

$$P\left(X_t \mid X_{t-1}, w, z_{it}\right),$$

where X_t is the value of the variable of interest at time t. X_{t-1} is the value of the variable of interest at the previous time step with a Markovian dependency. w is a set of covariates affecting the transition. z_{it} is an subject-specific observed covariate for individual i at time t.

Tracking of longitudinal data through LMM uses repeated measures over time and this is critical in evaluating the effectiveness of an intervention through comparing latent class probabilities at different times. Noor et al. [5] suggested that in conducting LMM, multinomial logistic regression is used to model the log-odds of the probability of an event occurring, denoted by:

$$\text{logit}\left(p_i\right) = \log\left(\frac{p_i}{p_R}\right),$$

where p_i is the probability of event i, and p_R is the probability of the reference or base category, which serves as the pivot for the multinomial extension across the latent classes $\left(\frac{p_i}{p_R}\right)$ is the odds of the event. $\log\left(\frac{p_i}{p_R}\right)$ is the final transformed value which is useful in regression modeling because it allows fitting linear models on a transformed scale for latent classes compared at different time intervals.

Kam et al. [10] highlighted the advantage of using LMM over other categorical trajectory methods like latent growth models and simple Markov chains. Thus, LMM permits a more accurate assessment of the nature of sample-level changes in distinct latent classes over time. Saengtrakul et al. and Zyberaj et al., [11, 12] suggest that LMM answers some of the questions about transition of latent variables which include the distribution of individuals within classes, class prevalence, probability of transition across classes and the factors associated with these transitions. In a study assessing the stability of latent classes among men who have sex with men and drug use, Card et al. [13] applied Latent Markov Modelling (LMM), leveraging its key advantage over static latent class analysis with its ability to model transitions in class membership over time rather than treating group membership as fixed.

Longitudinal tracking of latent classes is therefore valuable for identifying behavioral patterns and how they change over time. Zych et al. and Weinberger et al. [14, 15] used LMM to examine transitions among drug users, accounting for frequency of use, drug type, changes in community acceptability, and mental health status. Okuda et al. [16] applied Latent Transition Analysis (LTA) which is a LMM-based approach to track how children's balance-skill classes changed from childhood to adolescence in a British birth cohort. This shows LMM's key strength in longitudinal transition modeling rather than treating birthweight's effect as a fixed association which is common in cross-sectional analysis. Ni et al. [17], Hultgren et al. [18] and Liao et al. [19] show LMM's core advantage of modeling how subgroups change over time across categorical and binary outcomes. Di Mari et al. [20] use the same framework to guard against mistaking measurement drift for real change. Nguena Nguefack et al. [21] and Zyberaj et al. [12] noted the advantage of LMM where each individual occupies one mutually exclusive class per time point, with transition probabilities estimated directly, thus capturing both cross-sectional and longitudinal structure at once. Thus, LMM's key strength lies in its capacity for longitudinal tracking of the same individuals across time to reveal not just where subgroups stand at any one point, but how and when they move between classes. The impact of the DREAMS program has been documented through both quantitative and qualitative approaches, where a dynamic view, demonstrated across health, behavioral, and economic applications captures change for cross-sectional snapshots which miss entirely modeling this impact using LMM. Thus, this study develops latent Markov models that capture both cross-sectional and longitudinal dependence between latent classes.

The choice of software for fitting latent transition models also matters for capturing this complexity. In a study modeling longitudinal transitions using school bullying data, Ryoo et al. [22] used Mplus, a standalone latent-variable modeling program with dedicated procedures for latent class and latent transition analysis; however, Mplus has been noted to fall short when handling multivariate categorical outcomes, motivating the use of alternative packages such as LMest, which additionally allows for missing responses [8]. Muthén and Asparouhov [23] and Vogelsmeier et al. [24] similarly used LMest for assessing transitions of classes because of its capacity to run large sets of models and efficiently produce publication-formatted output. Noor et al. [5] adopted a comparable technique in modeling latent transition analysis for substance use among men who have sex with men, producing model coefficients that showed association with covariates. Turner [3] applied multinomial regression to yield transition probabilities describing the joint distribution of latent classes, while Nylund-Gibson et al. [25] highlighted that the inclusion of covariates is critical for identifying variables that predict transitions between latent classes. In a study assessing transitions in mathematics performance, Saengtrakul et al. [11] used covariates to model transitions, a technique also adopted in the present analysis. For model selection, Perra [26] used BAIC to identify the best-fitting model for categorical and continuous data assumed to be normally distributed.

2. Methods

This section outlines the study sample, the process of selecting the best-fitting model, and the diagnostic checks used to evaluate a model of Latent Markov transitions between vulnerability-to-HIV classes among AGYW. Using routinely collected program data, we conducted a longitudinal cohort analysis of AGYW enrolled in the Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) program in Zimbabwe, with assessments at enrolment (T1) and again after 12 months (T2). AGYW were linked across the two time points using their unique program identifiers, and the transition analysis was restricted to individuals with matched records at both assessments.

2.1. Data variable measurements

This analysis drew on routine program data for AGYW aged 10–19 years who were at risk of HIV acquisition and enrolled in the DREAMS program. Records were extracted from nine

DREAMS-supported districts in Zimbabwe: Bulawayo, Gweru, Mazowe, Beitbridge, Bulilima, Gwanda, Insiza, Mangwe, and Matobo, where standardized screening and enrolment tools were applied consistently across all sites. Predictor variables were selected to capture the socio-demographic and economic factors likely to shape AGYW's risk profiles. Source of income served as a key predictor, categorized as formal employment, informal work, support from a family member, support from a husband or boyfriend, or engagement in transactional sex. Location was included as urban, rural, or peri-urban, capturing potential differences in service access, social norms, and risk environment. Marital status that is single, married, or divorced/separated was also included to account for how partnership status shapes vulnerability and behavioral outcomes. Together, these variables were used to examine their associations with AGYW's risk behaviors, health outcomes, and social vulnerabilities.

This analysis relied on response variables reflecting the risk behaviors, health outcomes, and social vulnerabilities of AGYW enrolled in the DREAMS program, informed by indicators identified in Wambiya et al. and USAID [6, 7]. School-related factors were current enrollment status (in-school versus out-of-school) and risk of dropping out were included as markers of educational engagement and associated vulnerability. Household and social circumstances were captured through orphanhood status, which signals heightened vulnerability and diminished social support. Sexual behavior and risk were assessed using binary ("Yes"/"No") indicators: ever having had sex (ages 10–14), involvement in transactional sex, multiple sexual partnerships (ages 15–19), and inconsistent or absent condom use (ages 15–19) which were critical for evaluating HIV risk. Health-related outcomes comprised a history of sexually transmitted infections and a history of pregnancy. Violence exposure was measured through experience of any violence (ages 10–14) and experience of sexual violence (ages 15–19), while substance use was captured via alcohol use (ages 10–14) and alcohol misuse (ages 15–19), each contributing to a broader picture of behavioral risk. Taken together, these variables formed the core indicators used to assess vulnerability and inform targeted programming within DREAMS.

2.2. Latent Markov Model

Although screening for eligibility into the DREAMS program captures a static picture of HIV vulnerability among AGYW, Latent Markov modeling enables longitudinal tracking of unobserved typologies within a population, capturing how class membership shifts over time [25]. Ryoo et al. [22] outline a five-step framework for fitting an LMM which are assessing the data, testing for longitudinal invariance, defining latent statuses, testing for transitions between latent groups, and finally fitting and evaluating the model. Within this procedure, Nguena Nguefack et al. [21] emphasize the importance of clearly defining the research problem and specifying the initial latent groups at time $t = 1$, noting that the number of latent classes should be guided by the hypothesis under investigation and by relevant considerations from the field of study. Saengtrakul et al. [11] highlight that model specification is a critical stage, as it determines whether item-response probabilities are treated as time-invariant and whether covariates are incorporated into the model. Estimation typically precedes model fitting and is conducted using the Expectation-Maximization algorithm, though Bayesian approaches using Markov Chain Monte Carlo methods offer an alternative, as demonstrated by Li et al. [27]. Model selection and interpretation in LMM are commonly guided by the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC), with the model yielding the lowest BIC and AIC values generally considered the best fit [28].

The incidence of a latent class transitioning over time is modeled using Markov chains, a type of stochastic process that describes a sequence of events in which the probability of the current event depends on the preceding event, as described by Bartolucci et al. [8]. Suppose we have n latent statuses estimated based on the dataset for adolescent girls and young women. The latent classes are derived from M categorical response items for vulnerabilities to HIV measured at each time t for T time points. The set of the M outcomes for individual i over the T time points can be presented as:

$$\begin{aligned}\mathbf{Y}_i = (&Y_{i11}, Y_{i12}, \ldots, Y_{i1M}, Y_{i21}, Y_{i22}, \ldots, Y_{i2M}, \ldots,\\ &Y_{iT1}, Y_{iT2}, \ldots, Y_{iTM})\end{aligned}$$

Thus, $\mathbf{Y}_i$ denotes the vulnerabilities of an individual AGYW i, for all longitudinal tracking times $t = 1, \ldots, T$, and items $m = 1, \ldots, M$ where an AGYW response $\mathbf{Y}_{itm}$ may take on the values 1 or 0 corresponding to a Yes/No response on each vulnerability indicator. All vulnerability items are therefore binary outcomes for examples the 10–14-year-olds include in-school dropout risk (Yes/No), orphanhood (Yes/No), alcohol use (Yes/No), and experience of any violence (Yes/No). For the 15–19-year-olds, the variables include multiple sexual partners (Yes/No), no or irregular condom use (Yes/No), transactional sex (Yes/No), and history of pregnancy (Yes/No). Let S_{ti}, for $i = 1, 2, \ldots, n_s$ be individual i's latent status membership at Time t, and S_{Ti}, be individual i's latent status membership at time T, the endline reassessment point in this study. In general, status is defined at every observation time $t = 1, 2, \ldots, T$, so latent-status membership can be tracked at each occasion; in the DREAMS data used here, $T = 1, 2, \ldots, T$ months and reassessment was conducted at that endpoint to check the AGYW's vulnerability to HIV. Let $y = k$ be the indicator function which equals 1 if $y = k$ and 0 otherwise, and let G_i represent AGYW i's membership of a latent class which is vulnerability to HIV at enrolment which then changes due to transitions and let X_i be the value of the covariate **X** the value of **X** may be related to the probabilities of membership in the latent statuses (δs) as well as the transition probabilities (τs). The LTA models show changes in class membership itself over time for example moving from "high vulnerability" to "low vulnerability." The model identifies distinct subgroups whose outcome patterns cluster together statistically, based on response probabilities and trajectory shapes. Latent vulnerability classes capture unobserved heterogeneity in HIV risk among AGYW by clustering longitudinal outcome patterns. Baseline classes are identified using Latent Class Analysis, with item-response probabilities estimated via maximum likelihood. Latent Transition Analysis extends this by modelling probabilistic movement between classes over time, incorporating transition matrices (τ). Covariates are included through logistic regression to predict both initial class membership (δ) and transitions, linking social and structural factors to risk trajectories. This framework rigorously quantifies dynamic vulnerability while accounting for uncertainty in class assignment. Thus, the latent transition model is given by the full expression follows Bartolucci et al. [8] and Nguena Nguefack et al. [21], who provide the derivation of the joint response probability from the initial-state, transition, and measurement sub models. The model is given by:

$$P\left(Y_i = y \mid X_i = x,\ G_i = g\right) = \sum_{s_1=1}^{n_s} \sum_{s_2=1}^{n_s} \sum_{s_3=1}^{n_s} \delta_{s_1|g}(x) \cdot$$
$$\tau_{s_2|s_1,g}(x) \cdot \tau_{s_3|s_2,g}(x) \times \prod_{m=1}^{M} \prod_{k=1}^{r_m} \prod_{t=1}^{T} q_{mk}^{I(y_{mit}=k)}, \quad (1)$$

where:

- $\mathbf{Y}_i$– vector of observed outcomes for person i. The future latent group is linked to Y_i through the measurement sub model ϱ, so the next-period state depends on the observed responses only via the class-conditional response probabilities.
- X_i– vector of covariates or background characteristics.
- G_i– latent class the person belongs to at the preceding time point (i.e. at $t-1$); transitions are then modelled from the class at $t-1$ to the class at t.
- S_t– Hidden (latent) state at time t.
- $\delta_{s_1|g}(x)$– Probability of starting in state s_1, given latent class g and covariate vector x; here $s_1 \in \{1, \ldots, n_s\}$, $g \in \{1, \ldots, G\}$ indexes the population subgroup, and $x = (x_1, \ldots, x_p)^T$ is the covariate vector for the individual.
- $\tau_{s_t|s_{t-1},g}(x)$– Probability of moving from one state to another over time, given group g.
- $\varrho_{mk|s_t,g}(x)$– Probability of observing outcome m in category k when in state s_t. The subscript g indexes the population subgroup, so ϱ, δ and τ are all group-specific and g governs both the initial-state and transition probabilities.
- $I(y_m = k)$– Indicator showing which category was actually observed.
- n_s– Number of possible hidden states. (Note: the term $I(y_m = k)$ appearing as an exponent on ϱ is an indicator that equals 0 or 1, not a power.)
- M– Number of observed outcome variables.
- r_m– Number of categories for each outcome.
- T– Number of time points observed.

The probability of an adolescent girl or young women belonging to latent status s at time T of re-assessment is given by:

$$\delta_{s_1/g}(x) = P\left(S_{1i} = s \mid X_i = x, G_i = g\right). \quad (2)$$

This term represents the probability that individual i starts in latent state s_1 at the first time point (this is the initial-state probability at $t=1$; the corresponding endpoint state at $t=T$ is s_T with an analogous conditional probability $\delta_{s_T|g}(x)$), given their observed covariates $X_i = x$ and latent class $G_i = g$. Thus, this is the initial state probability in the latent Markov model. Abarda et al. [29] noted that covariates, denoted by X, can be included in the model, with the parameters expressed in terms of β coefficients. These β parameters are then estimated from the fitted multinomial logistic regression. In a study by Kam et al, [10] on stability and predictability of transitions among latent classes, transitional probabilities were computed, and logistic regression coefficients were obtained by fitting a model with relevant covariates. The latent vulnerability classes are developed through a latent class model, which assumes that the observed behavioral and psychosocial indicators of each Adolescent Girl or Young Woman, denoted by $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iM})$ are manifestations of a latent vulnerability state S_i. To determine the number of vulnerability classes, statistical criteria such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) are used to choose the optimal number of latent classes n_s. Typically, the model is fitted iteratively with increasing numbers of latent classes, and the solution with the best model fit and ease of interpretability is selected. The model estimates the probability that an individual belongs to a particular latent class, conditional on observed responses:

$$\delta_{s/g}(x) = P\left(S_{1i} = s/X_i = x, G_i = g\right) \quad (3)$$
$$= \frac{\exp\left(\beta_{0s/g} + x\beta_{1s/g}\right)}{1 + \sum_{j=1}^{n_s-1} \exp\left(\beta_{0s/j} + x\beta_{1s/j}\right)},$$

where:

- $P\left(Y_i = y_i \mid S_i = s\right)$, conditional additionally on $G_i = g$, is the measurement model, linking the latent state s to the observed outcomes.
- $P\left(S_i = s\right)$ represents the latent class prevalence that is the prior probability of being in class s.
- n_s is the number of distinct latent vulnerability states identified.

To capture how an AGYW's characteristics shape her likelihood of belonging to a particular vulnerability class, both δ and τ are modeled as functions of covariates x using multinomial logistic regression. The regression designates the n_s^{th} latent status as the reference category, so every other status is interpreted relative to it. In practical terms, the model estimates the log-odds that an AGYW belongs to latent status s rather than the reference status n_s, based on her observed covariates such as source of income, location, and marital status.

Let $x = (x_1, \ldots, x_p)^T$ denote the vector of covariates, excluding an intercept term. For a multinomial response with categories $0, 1, \ldots, J$, where category 0 serves as the reference, the model expresses the log-odds of category s relative to the reference as a linear function of x:

$$\log\left[\frac{P(Y = s \mid x)}{P(Y = 0 \mid x)}\right] = \beta_{0s} + x^T \beta_s = \beta_{0s} + \sum_{j=1}^{p} \beta_{js} x_j. \quad (4)$$

Here $\boldsymbol{\beta}_s = (\beta_{1s}, \ldots, \beta_{ps})^T$ is the coefficient vector associated with the contrast between category s and category 0, while β_{0s} denotes the corresponding intercept. Equivalently, if the intercept is absorbed into the covariate vector by defining $\mathbf{x} = (x_1, \ldots, x_p)^\top$ and $\beta_s = (\beta_{0s}, \beta_{1s}, \ldots, \beta_{ps})^T$, the log-odds simplifies to the inner product $\tilde{x}^T \beta_s$.

More generally, coefficients may follow a two-index notation such as β_{1s} denoting the contrast between category s and a reference category g, possibly within a nested grouping structure. The latent transition model uses multinomial logistic regression to estimate the probability of an individual belonging to a specific vulnerability state. Each covariate has a coefficient that quantifies its effect on the log-odds of being in a state relative to a reference state. These coefficients capture the unique influence of each covariate while controlling for others, allowing multiple predictors to be considered simultaneously. Thus, this analysis adopted the logistic regression with covariates that include rural/urban classification for the 10–14-year-old AGYW as well as source of income for the 15–19-year-old AGYW. These two covariates were retained from the wider set of program variables because they were theorized, based on the DREAMS program design, to shape both access-to-services (residence) and economic vulnerability (income source), and because they had adequate cell counts across all four latent classes; other candidate covariates were dropped either because they duplicated information already encoded in the latent indicators or because sparse cells produced unstable estimates.

The model diagnostic was conducted by assessing invariance across latent class variables and across multiple observed groups as pointed out by Ryoo et al. [22]. The log-likelihood ratio test (LRT) was used for diagnostic tests where the Maximum Likelihood (ML) with robust standard errors was used. The general form of the log-likelihood ratio test (LRT) can be expressed as follows:

$$LRT = -2\left[\log L\left(\widehat{\theta}_r\right) - \log L\left(\widehat{\theta}_p\right)\right], \qquad (5)$$

where:

- $\widehat{\theta}_r =$ maximum likelihood estimates under the restricted model
- $\widehat{\theta}_p =$ maximum likelihood estimates under the full model
- $\log L(\cdot) =$ log-likelihood evaluated at the respective estimates

In Latent Markov modeling, missing data can occur due to non-response or loss to follow-up, which has implications for parameter estimation. Let δ_{it} denote the missing-data indicator for vulnerability to HIV among adolescent girls and young women (AGYW) enrolled in the DREAMS program. Let δ_{it} takes the value 1 if AGYW i provides information on vulnerability at time t, and 0 if this information is missing. Here i indexes the individual AGYW ($i = 1, \ldots, n$) and t indexes the observation occasion ($t = 1, 2, \ldots, T$). Proper handling of missing data is essential, as ignoring it may bias estimates of the initial-state probabilities, transition probabilities, and the effects of covariates in the model. Latent Markov models can accommodate missingness under the assumption that data are missing at random by integrating over the unobserved outcomes in the likelihood function. To cater for missing values in eq. (2) will be modified to admit an observation indicator so that only observed cells contribute to the likelihood and this yields the missing data adjusted analogue of eq. (2):

$$P\left(y_i \mid w, x_0, x_1, \ldots, x_T, z_i\right) = \prod_{t=0}^{T}\left[P\left(y_{it} \mid x_t, w, z_{it}\right)\right]^{\delta_{it}}, \qquad (6)$$

where:

- $y_{it} =$ observed outcome for individual i at time t,
- $x_t =$ covariates at time t,
- $w =$ covariates or parameters affecting the transition (as defined previously),
- $z_{it} =$ additional covariates or random effects,
- $\delta_{it} =$ indicator for missingness: 1 if y_{it} is observed, 0 if missing.

2.3. Data analysis

Modelling transitions between latent classes at different time periods for change in vulnerability to HIV among AGYW was done using LMest (version 3.0.3), an R (Version 4.5.2) package for fitting Latent Markov models by considering different time periods. LMest was preferred over alternative latent-variable modeling software because it accommodates multivariate categorical outcomes and missing responses within a single integrated framework, and allows for computing parameter estimates together with their standard errors through suitable parameterizations [23, 24]. Latent classes were fitted at time $t = 0$, where descriptions for vulnerability to HIV among AGYW were derived and used to inform the initial latent classes. In developing the structural model, multinomial logistic regression of a latent class variable on another was used for different time periods, where re-assessments were conducted to check the effectiveness of the DREAMS interventions on vulnerability to HIV among AGYW. Comparison of each class at different time intervals was done to explore similarities or differences in the emergent classes, which then led to measurement and testing of the variation, with covariates incorporated to model the transitions [25]. The best model was selected by comparing Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC), where the best fit was indicated by the lowest BIC and AIC values [26]. In the analysis, there were missing data for AGYW who were longitudinally tracked for 12 months; missing data accounted for less than 4% at the reassessment period. Following the approach of Tupin et al. [30], who applied a stochastic imputation method for missing variables in a comparable longitudinal design (with less than 5% item-level missingness and under 1% missingness in the association structure between variables), a similar stochastic imputation approach was adopted for this analysis. Spearman's rank correlation was used to assess associations among variables prior to imputation, with item-level missingness of less than 0.9% informing the use of intra-scale stochastic imputation.

3. Results and Discussion

Records for 4,341 adolescent girls and young women (AGYW) enrolled in the DREAMS program were assessed at enrolment with 62.6% (2,716/4,341) aged 10-14 years, whilst 37.4% were aged 15-19 years. After 12 months, 4,106 AGYW were assessed where 65.8% (2,703/4,106) were aged 10-14 years and 34.2% were aged 15-19 years old. Over 80% of the AGYW resided in rural areas at enrolment into the DREAMS intervention and at the 12 months' reassessment period (Table 1).

All the 10–14-year-old AGYW relied on parents or family as a source of income, whilst for the 15–19-year-olds husbands and boyfriends as source of income increased from 13.8% (224/1,625) to 21.1% (295/1,403). Similarly, informal work as source of income for the 15–19-year-olds increased from 16.8% to 29.1% whilst parental or family support decreased from 55.8% to 49.9%.

Of the 4,341 AGYW enrolled at baseline, 4,106 (94.6%) were reassessed at the 12-month follow-up. This attrition was uneven across age groups where among the 10–14-year-old AGYW, the sample declined from 2,716 to 2,703, reporting a loss of 13 par-

Table 1. Profiling AGYW vulnerable to HIV.

Characteristic	10–14-year-old AGYW T1 Screening $N = 2{,}716$	T2 12-Month $N = 2{,}703$	15–19-year-old AGYW T1 Screening $N = 1{,}625$	T2 12-Month $N = 1{,}403$
Residence				
Rural	2,377 (87.5%)	2,364 (87.5%)	1,355 (83.4%)	1,161 (82.8%)
Urban	339 (12.5%)	339 (12.5%)	270 (16.6%)	242 (17.2%)
Substance Use				
Alcohol use	325 (12.0%)	80 (3.0%)	—	—
Alcohol misuse	—	—	174 (10.7%)	118 (8.4%)
Sexual Behaviour				
Ever had sex	30 (1.1%)	32 (1.2%)	—	—
Multiple sexual partners	—	—	103 (6.3%)	102 (7.3%)
No/irregular condom use	—	—	1,046 (64.4%)	1,006 (71.7%)
Transactional sex	7 (0.3%)	1 (< 0.1%)	174 (10.7%)	30 (2.1%)
Violence & Safety				
Experience of violence	523 (19.3%)	529 (19.6%)	—	—
Sexual violence	—	—	22 (1.4%)	14 (1.0%)
Reproductive Health				
History of pregnancy	—	—	457 (28.1%)	474 (33.8%)
STI symptoms	9 (0.3%)	1 (< 0.1%)	35 (2.2%)	24 (1.7%)
Socioeconomic Factors				
Orphan status	889 (32.7%)	1,088 (40.3%)	369 (22.7%)	345 (24.6%)
In-school dropout risk	2,372 (87.8%)	2,098 (77.6%)	—	—
Income Source				
Husband/Boyfriend	0 (0%)	0 (0%)	224 (13.8%)	295 (21.1%)
Informal work	0 (0%)	0 (0%)	273 (16.8%)	408 (29.1%)
Parent/Family	2,716 (100%)	2,703 (100%)	906 (55.8%)	700 (49.9%)

Note. Values are n (%). T1 = Screening; T2 = 12-Month Reassessment. — = not assessed in this age group. AGYW = Adolescent Girls and Young Women; STI = Sexually Transmitted Infection.

Table 2. Class composition at T1.

State/ Class	Description of state	Alcohol Use	Experience of any violence	In-school Dropout Risk	Orphan
1	Very high risk	**0.983**	0.181	**0.663**	0.245
2	Low risk	0.017	0.000	**0.998**	0.185
3	Medium risk	0.005	0.025	**0.587**	**0.999**
4	High risk	0.059	**1.00**	**0.778**	0.166

ticipants, which is 0.48%, whereas among the 15–19-year-old, the sample declined from 1,625 to 1,403 reporting a loss of 222 participants that is 13.66%. Loss to follow-up in this cohort was primarily attributable to relocation out of the DREAMS program catchment areas and to discontinuation from the intervention before the 12-month reassessment.

Age groups that is 10–14 versus 15–19 were fixed at baseline and held constant through follow-up as evidenced by each cohort being assessed on the same distinct item set at both T1 and T2 in Table 1, rather than switching indicator sets as participants aged. This fixed-cohort design is appropriate for LTA, since consistent indicators across time are required to interpret transitions.

3.1. Analysis for 10–14-year-old AGYW

3.1.1. Initial class composition

At screening for enrolment into the DREAMS program, the variables used for 10-14-year-old AGYW were school status (out of school or in-school), risk of dropping out of school, orphanhood, ever had sex, engagement in transactional sex, had sexually transmitted infections (STI), experience of any violence (10-14), alcohol use, and history of pregnancy. At time 1 (T1), four vulnerabilities to HIV classes were identified and these were State 1, that is, very high risk with 11.0% of the total AGYW sample, state 2 was low risk with 56.4% of the sample, state 3 was medium risk with 15.6% and state 4 was high risk which constituted 17.0% of the AGYW assessed (Table 2). Results in Table 2 show that Class 1 had the vulnerabilities alcohol use and in-school dropout risk with probabilities 0.983 and 0.663 respectively, whilst class 2 had in-school dropout risk with probability 0.998. Class three had vulnerabilities of in-school dropout risk with probability 0.587 and orphanhood with probability 0.999 whilst class 4 had in-school dropout risk with probability 0.778.

The four-state classification is justified by the distinct probability profiles across the risk indicators. Class 1 (Very high risk) is characterised by a very high probability of alcohol use (0.983) and substantial in-school dropout risk (0.663), indicating multiple overlapping vulnerabilities. Class 2 (Low risk) has a very low

Table 3. Log-Odds ratios of belonging to a class at T1.

Classes	Covariate	Class	Odds ratios	LogOdds	Std Error	P-Value
2/1	(Intercept)	2	5.22	1.653	0.120	0.000
	Rural/Urban	2	0.879	-0.129	0.207	0.531
3/1	(Intercept)	3	1.500	0.405	0.144	0.004
	Rural/Urban	3	0.599	-0.513	0.289	0.075
4/1	(Intercept)	4	1.376	0.319	0.140	0.023
	Rural/Urban	4	2.060	0.723	0.226	0.001

Table 4. Model diagnostics.

Model	*LogLik*	n_p	k	AIC	BIC	n	TT
2 classes	$-9,458.86$	14	2	18,945.72	19,028.35	2,703	2
3 classes	$-8,875.98$	28	3	17,807.96	17,973.22	2,703	2
★ 4 classes	$\mathbf{-8,655.81}$	**46**	**4**	**17,403.63**	**17,675.12**	**2,703**	**2**
5 classes	$-8,584.07$	68	5	17,304.14	17,705.48	2,703	2
6 classes	$-8,492.59$	94	6	17,173.19	17,727.99	2,703	2

Note. ★ Denotes the selected 4-class model based on the lowest BIC. Fit indices:
LogLik — Log-Likelihood — measures overall model fit; higher (less negative)
values indicate better fit. n_p — Number of free parameters estimated in the model.
k — Number of latent classes specified. AIC — Akaike Information Criterion — penalizes model complexity;
lower values preferred. BIC — Bayesian Information Criterion — applies a stronger complexity penalty;
lower values preferred. n — Total analytic sample size.
TT — Number of random starting-value sets used to verify convergence.

probability of alcohol use (0.017) and violence (0.000), although it has a very high probability of in-school dropout risk (0.998), suggesting a more specific educational vulnerability. Class 3 (Medium risk) is distinguished by an almost certain probability of orphanhood (0.999), combined with moderate dropout risk (0.587), reflecting psychosocial vulnerability. Class 4 (High risk) is defined primarily by a certain probability of experiencing violence (1.000), alongside a high probability of dropout risk (0.778), representing significant exposure to multiple risks.

Results in Table 3 show that, the baseline log-odds of being in Class 2 compared to Class 1 were high and statistically significant ($p < 0.001$). Rural/Urban residence is coded 1 = rural and 0 = urban (reference), thus positive coefficients therefore indicate higher log-odds of class membership, relative to Class 1, for rural AGYW compared to urban AGYW. Adolescent girls and young women (AGYW) residing in rural areas were less likely to belong to Class 2 than Class 1, relative to those in urban areas, although this effect was not statistically significant (OR = 0.879, p = 0.531). The baseline odds of being in Class 3 relative to Class 1 were 1.500 times higher and statistically significant (p = 0.004). However, AGYW in rural areas were less likely to be in Class 3 than Class 1, relative to those in urban areas (OR = 0.599), and this association was marginally significant (p = 0.075). The baseline odds of being in Class 4 compared to Class 1 were 1.376 times higher and statistically significant (p = 0.023). Moreover, AGYW in rural areas were 2.060 times more likely to belong to Class 4 than Class 1, relative to those in urban areas, and this effect was statistically significant (p = 0.001).

3.1.2. Model diagnostics

Model comparison was conducted using the log-likelihood (LogLik), number of parameters (n), number of classes (k), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample size (n), and number of iterations (TT). The log-likelihood values improved steadily from –9458.86 for the two-class model to –8492.59 for the six-class model, indicating progressively better fit (Table 4). The number of estimated parameters in the model n_p increased with the number of classes, reflecting greater model complexity. Both AIC and BIC decreased substantially from the 2 to 4 class model, after which the BIC began to increase slightly. The 4-class model achieved the lowest BIC value of 17675.12 and a relatively low AIC which is 17403.63, suggesting the best model fit and parsimony. The sample size, n, is 2703 and this remained constant across models, and two iterations TT = 2 were used for convergence checks. Based on the combination of these criteria, the four-class model was selected as the optimal representation of the data structure among adolescent girls and young women.

The unrestricted model (LogLik $= -8,643.81$, $n_p = 62$) and time-invariant model (LogLik $= -8,655.81$, $n_p = 46$) were compared with LRT: $\chi^2(16) = 24.00$, $p = .087$ which were not statistically significant (Table 5). The AIC (17,403.63 compared to 17,411.63) and BIC (17,675.12 compared to 17,778.65) both favoured the invariant model. Full longitudinal measurement invariance is supported with no partial-invariance model that was critical.

3.1.3. Transition probabilities for latent classes

At enrolment, most AGYW (56.4%) started in the low-risk state (state 2), linked to in-school dropout vulnerability, while 11.0% began in state 1, 15.6% in state 3, and 17.0% in state 4. State 3 that is the medium-risk class marked by dropout risk and orphanhood was the most stable, with the highest self-transition probability (0.908) (Figure 1).

Figure 1 depicts the full set of estimated transitions: transitions between all pairs of states are possible in both directions,

Table 5. Longitudinal measurement invariance: unrestricted vs. time-invariant four-class models.

Model	LogLik	n_p	LRT (χ^2)	Δdf	p	AIC	BIC
Unrestricted (item-response probabilities free across T1 and T2)	$-8,643.81$	62	—	—	—	17,411.63	17,778.65
★ Time-invariant (item-response probabilities constrained equal across T1 and T2)	$-8,655.81$	46	24.00	16	.087	17,403.63	17,675.12

Note. LogLik = log-likelihood; n_p = number of freely estimated parameters; LRT $= -2[\log L(\text{restricted}) - \log L(\text{unrestricted})]$, distributed χ^2 with Δdf degrees of freedom. ★ The time-invariant model was retained: the LRT was non-significant ($p = .087$) and both AIC and BIC favored the more parsimonious time-invariant model, supporting full longitudinal measurement invariance across T1 and T2; no partial-invariance model was required. The retained model's LogLik, n_p, AIC, and BIC reproduce the 4-class model.

Table 6. Log-odds for changing latent classes over time among AGYW.

	Source Class 1			Source Class 2			Source Class 3			Source Class 4		
Covariate	$\to 2$	$\to 3$	$\to 4$	$\to 2$	$\to 3$	$\to 4$	$\to 2$	$\to 3$	$\to 4$	$\to 2$	$\to 3$	$\to 4$
Intercept	2.255*** (0.414) $p = 5.19\text{E}{-08}$ OR=9.535	1.553** (0.419) $p = 0.001$ OR=4.726	1.404** (0.448) $p = 0.002$ OR=4.071	-4.017*** (0.289) $p = 0.000$ OR=0.018	-1.373*** (0.101) $p = 0.000$ OR=0.253	-1.981*** (0.108) $p = 0.000$ OR=0.138	-20.339*** (0.269) $p = 0.000$ OR=NR†	-15.852*** (0.042) $p = 0.000$ OR=NR†	-3.837*** (0.839) $p = 4.8\text{E}{-06}$ OR=0.022	-3.123*** (0.367) $p = 0.000$ OR=0.044	-1.078*** (0.153) $p = 1.98\text{E}{-12}$ OR=0.340	-1.495*** (0.195) $p = 1.75\text{E}{-14}$ OR=0.224
Rural/Urban	-0.711 (0.856) $p = 0.406$ OR=0.491	-2.150 (1.263) $p = 0.088$ OR=0.116	0.586 (0.861) $p = 0.496$ OR=1.797	1.039* (0.569) $p = 0.038$ OR=2.826	-0.109 (0.322) $p = 0.736$ OR=0.897	0.551* (0.267) $p = 0.039$ OR=1.735	18.816*** (0.270) $p = 0.000$ OR=NR†	-15.852*** (0.042) $p = 0.000$ OR=NR†	1.129 (3.075) $p = 0.713$ OR=3.093	-0.204 (0.903) $p = 0.821$ OR=0.815	0.365 (0.288) $p = 0.206$ OR=1.441	-0.258 (0.491) $p = 0.601$ OR=0.773

Note. Each set of three columns under a "Source Class" heading reports transitions FROM that source class TO the destination class named by the column arrow (e.g., "$\to 3$" under "Source Class 1" = the log-odds of transitioning from Class 1 to Class 3). Each cell reports β (SE), p-value, and OR $= \exp(\beta)$. OR=NR† where estimates were flagged as potentially unstable/separated (this occurs mainly for transitions out of Class 3); such coefficients should not be substantively interpreted. Rural/Urban: 1 = rural. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

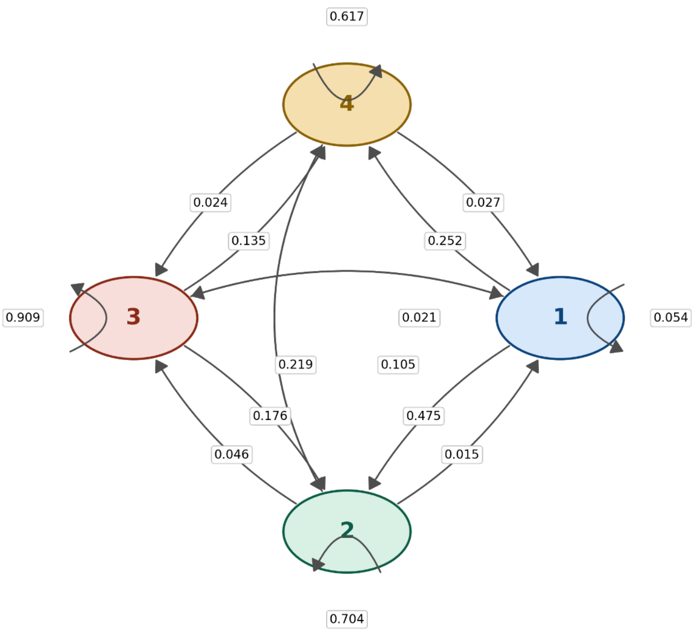


Figure 1. Transition probabilities.

and the probability of remaining in the same state (self-transition) is also shown for each state. Self-transition probabilities were high for the low risk group (state 2, 0.703) and the high-risk group (state 4, 0.618), showing relative stability in both. Movement into state 1 was rare across the board with 0.015 from low-risk, 0.021 from medium-risk (state 3), and 0.027 from high-risk with transitions concentrated among AGYW with histories of violence or orphanhood moving away from alcohol use. A moderate 0.221 probability marked movement from high-risk to low-risk, reflecting reduced use, while 25.2% moved from state 1 to high-risk and 10.5% moved from low-risk to high-risk.

3.1.4. Log-Odds of transitioning from a latent class to another

Rural–urban residence was significantly associated with selected latent-class transitions. Rural participants had higher odds of remaining in Class 2 (OR = 2.827, 95% CI: 0.908–8.803, $p =$ .038) and of transitioning from Class 2 to Class 4 (OR = 1.735, 95% CI: 1.028–2.930, $p = .039$) compared with urban participants (Table 6). These findings suggest that residential setting may influence the persistence and movement of individuals between latent classes.

3.1.5. Latent class transition between different times

Generally, there was a decline in the proportion of AGYW in the low-risk state, characterized by vulnerability to in-school dropout, from Time 1 (T1) to Time 2 (T2) (Figure 2). A similar decrease was also observed in State 1, representing the very high-risk group with vulnerabilities related to alcohol use and in-school dropout, indicating an overall reduction in these high-risk behaviors among AGYW over the study period.

State 3 and State 4 that is medium risk with vulnerabilities orphanhood and in-school dropout risk as well as high-risk with vulnerabilities experience of violence and in-school dropout risks slightly increased from T1 to T2.

3.2. Analysis for 15–19-year-old AGYW

3.2.1. Initial class composition

At screening for enrolment into the DREAMS program for the 15–19-year-old AGYW, the variables used were engagement in transactional sex, had sexually transmitted infections (STI), have multiple sexual partners, no or irregular condom use, ex-

Table 7. Description of classes at T1.

State/ Class	Description of state	History of Pregnancy	Multiple Partners	Irregular Condom Use	Transactional Sex	Orphan
1	Low risk	0.067	0.032	0.275	0.003	0.760
2	Medium risk	0.070	0.090	0.718	0.036	0.018
3	High risk	0.794	0.071	0.916	0.004	0.065
4	Very high risk	0.305	0.528	0.892	0.693	0.331

Table 8. Log-odds ratios of belonging to a class at T1.

	Class 2 vs. Class 1				Class 3 vs. Class 1				Class 4 vs. Class 1			
Covariate	*OR*	β	*SE*	p	*OR*	β	*SE*	p	*OR*	β	*SE*	p
Intercept	6.20	1.824**	0.650	0.004	22.36	3.106***	0.569	4.84E-08	9.71	2.273***	0.597	0.000
Rural/Urban	3.25	1.179***	0.277	2.1E-05	0.96	-0.039	0.376	0.916	1.22	0.199	0.437	0.649
Income: Informal work	0.28	-1.279	0.706	0.070	0.12	-2.166***	0.599	0.000	0.11	-2.234***	0.638	0.000
Income: Parent/Family	0.18	-1.727*	0.699	0.013	0.02	-4.065***	0.584	3.41E-12	0.02	-3.708***	0.609	1.19E-09

Note. Each cell reports the odds ratio (OR), log-odds coefficient β, standard error (SE), and p-value for class membership relative to Class 1 (reference). Negative β in red; positive in blue. Income reference category = formal employment; Rural/Urban: 1 = rural. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

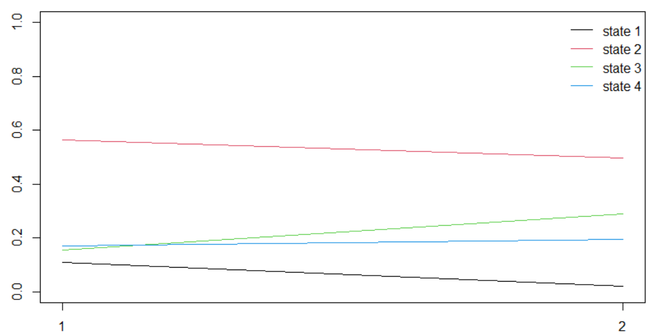


Figure 2. Overview of transition between states from T1 to T2.

Key: State 1 is Very high risk, State 2 is Low risk, State 3 is Medium risk, State 4 is High risk.

perience of sexual violence, alcohol misuse, and history of pregnancy. At the initial screening and enrolment stage (T1), the vulnerabilities to HIV among AGYW aged 15–19 years were categorized into four distinct classes. State 1 represented the low-risk group, characterized mainly by vulnerability related to orphanhood. State 2 denoted the medium-risk group, associated with irregular condom use with non-marital partners (Table 7). State 3 represented the high-risk group, defined by a history of pregnancy and irregular condom use with non-marital partners. Lastly, State 4 represented the very high-risk group, characterized by multiple sexual partners, irregular condom use with non-marital partners, and engagement in transactional sex. At Time 1 (T1), 24.8% of AGYW aged 15–19 years were classified in State 1, representing the low-risk group, while 38.6% were in the medium-risk group, 24.4% in the high-risk group, and 12.2% in the very high-risk group. The low-risk state was primarily characterized by vulnerability related to orphanhood, with an associated probability of 0.760. In contrast, State 2, representing the medium-risk group, was defined by no or irregular condom use with non-marital partners, with a probability of 0.718 (Table 8).

Class 3 was characterized by vulnerabilities that included a history of pregnancy, with a probability of 0.794, and irregular condom use with non-marital partners, having a probability of 0.916. The very high-risk class (Class 4) was defined by multiple sexual partners, irregular condom use, and engagement in transactional sex, with corresponding probabilities of 0.527, 0.892, and 0.693, respectively.

3.2.2. Log-Odds ratios of belonging to a class at T1

The intercept log-odds of 1.824 (OR = 6.20) indicates that, without considering covariates, AGYW were about 6.20 times more likely to be in Class 2 (medium risk) than in Class 1 (low risk) (Table 8). Also, AGYW residing in rural areas were 3.25 times more likely to be in the medium-risk class (Class 2) relative to the low-risk class (Class 1), compared to those in urban areas. In contrast, having an income source from informal work (log-odds = -1.279; OR = 0.28) or from parents or family (log-odds = -1.727; OR = 0.18) reduced the likelihood of being in Class 2 relative to Class 1, suggesting that reliance on these income sources decreased the probability of being in the low-risk group.

For Class 3 (high risk), the intercept log-odds of 3.106 (OR = 22.36) reflects a high baseline likelihood of membership relative to Class 1. Area of residence was not significant (log-odds = -0.039; OR = 0.96), indicating similar probabilities for rural and urban AGYW. Income from informal work (log-odds = -2.166; OR = 0.12) and from parents or family (log-odds = -4.065; OR = 0.02) strongly reduced the odds of being in the high risk class relative to low risk.

The intercept log-odds of 2.273 (OR = 9.71) for Class 4 (very high risk), indicate a high baseline likelihood of membership relative to Class 1. Rural/urban residence was not significant (log-odds = 0.199; OR = 1.22). However, income from informal work (log-odds = -2.234; OR = 0.11) and from parents or family (log-odds = -3.708; OR = 0.02) substantially reduced the odds of belonging to this class compared to Class 1, suggesting that AGYW relying on these income sources were much less likely to be in the high-risk class relative to very low-risk.

Table 9. Model diagnostics.

Model	LogLik	n_p	k	AIC	BIC	n	TT
2 classes	-5796.802	22	2	11637.6	11753.02	1403	2
3 classes	-5603.817	47	3	11301.63	11548.21	1403	2
4 classes	-5479.19	80	4	11118.38	**11538.09**	1403	2
5 classes	-5401.614	121	5	11045.23	11680.04	1403	2
6 classes	-5374.087	170	6	11088.17	11980.06	1403	2

Note: LogLik is the Log-Likelihood; n_p is the total number of estimated parameters in the model; k is the number of latent classes specified in the model; AIC is the Akaike Information Criterion; BIC is the Bayesian Information Criterion; n is the total sample size used in the analysis (here, 1,403 participants); TT is the number of random sets of starting values or iterations used to ensure model convergence.

Table 10. Longitudinal measurement invariance: unrestricted vs. time-invariant four-class models.

Model	LogLik	n_p	LRT (χ^2)	Δdf	p	AIC	BIC
Unrestricted (item-response probabilities free across T1 and T2)	$-5,464.19$	100	—	—	—	11,128.38	11,674.88
★ Time-invariant (item-response probabilities constrained equal across T1 and T2)	$-5,479.19$	80	30.00	20	.071	11,118.38	11,538.09

Note. LogLik = log-likelihood; n_p = number of freely estimated parameters; LRT = $-2[\log L(\text{restricted}) - \log L(\text{unrestricted})]$, distributed χ^2 with Δdf degrees of freedom. ★ The time-invariant model was retained: the LRT was non-significant ($p = .071$) and both AIC and BIC favored the more parsimonious time-invariant model, supporting full longitudinal measurement invariance across T1 and T2; no partial-invariance model was required.

3.2.3. Model diagnostics

For the 4-class model, the log-likelihood of -5479.19 indicates a good fit (Table 9). The model estimates 22 parameters (n_p), capturing more detailed class specific probabilities and relationships among observed variables. With $k = 4$, the population for AGYW can be meaningfully divided into four distinct latent subgroups. The AIC (11118.38) and BIC (11538.09) values are notably lower than those for the 2 and 3 class models, suggesting a good fit. Based on data from $n = 1403$ individuals measured at two time points ($T = 1$ and $T = 2$), the 4-class solution appears to provide a richer yet still efficient representation of the underlying heterogeneity in the sample. The model estimates 80 parameters (n_p) for 4 classes, which include class probabilities and item-response probabilities for each observed variable. The 4-class solution is also substantively coherent with low risk (orphanhood), medium risk (condom use), high risk (pregnancy and condom use), very high risk (multiple partners, condom use and transactional sex), forming a clear severity gradient.

The unrestricted model (LogLik $= -5,464.19$, $n_p = 100$) and time-invariant model (LogLik $= -5,479.19$, $n_p = 80$) were compared with LRT: $\chi^2(20) = 30.00$, $p = .071$ and these were non-statistically significant (Table 10). The AIC (11,118.38 compared to 11,128.38) and BIC (11,538.09 compared to 11,674.88) both favoured the invariant model showing that full longitudinal measurement invariance is supported.

3.2.4. Transition probabilities

At screening and enrolment into the DREAMS program, most adolescent girls and young women (AGYW) aged 15–19 were in the medium-risk state (State 2), accounting for 38.6%, characterized by vulnerabilities or irregular condom use with non-marital partners (Figure 3). About 24.8% were in the low-risk state (State 1), mainly associated with orphanhood, while 24.4% were in the high-risk state (State 3), and 12.2% were in the very high-risk state (State 4). The highest transition probabilities were for remaining in the same risk state, particularly State 1 (low risk) with a probability of 0.714, and State 3 (high risk) with 0.819, indicating stability in these states over time. Moderate transitions were observed for maintaining State 2 (medium risk) with a probability of 0.488, and for movement from State 4 (very high risk) to State 3 (high risk) with a probability of 0.483, suggesting a tendency toward gradual risk reduction among the most vulnerable AGYW.

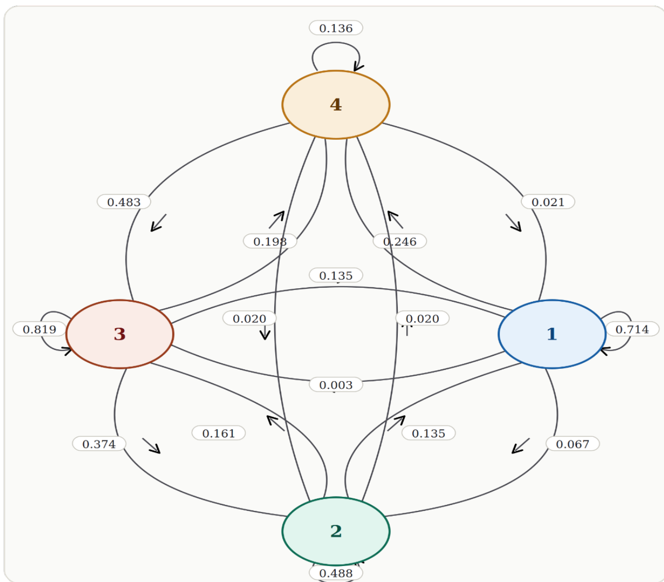


Figure 3. Transition probabilities matrix.

None of the AGYW transitioned from State 3 (high risk) to

Table 11. Log-odds of latent class transitions.

Covariate	Source Class 1			Source Class 2			Source Class 3			Source Class 4		
	→ 2	→ 3	→ 4	→ 2	→ 3	→ 4	→ 2	→ 3	→ 4	→ 2	→ 3	→ 4
Intercept	-14.701*** (0.433) $p < .001$ OR=NR†	2.279** (1.000) $p = .025$ OR=9.767	0.455 (1.199) $p = .705$ OR=1.576	-18.884*** (0.185) $p < .001$ OR=NR†	0.947 (0.700) $p = .176$ OR=2.578	-15.394*** (0.774) $p < .001$ OR=NR†	-20.606*** (0.609) $p < .001$ OR=NR†	-32.300*** (0.382) $p < .001$ OR=NR†	-11.254 (268.058) $p = .967$ OR=NR†	-1.979** (0.940) $p = .035$ OR=0.138	-1.996 (2.797) $p = .476$ OR=0.136	0.173 (0.465) $p = .709$ OR=1.189
Rural/Urban	-0.537 (2.175) $p = .805$ OR=0.584	-2.371 (1.839) $p = .197$ OR=0.093	-16.882*** (< 0.001) $p < .001$ OR=NR†	-0.537 (0.554) $p = .332$ OR=0.584	-0.597 (0.471) $p = .205$ OR=0.550	-13.821*** (< 0.001) $p < .001$ OR=NR†	-14.049*** ($< .001$) $p < .001$ OR=NR†	16.133*** (0.000) $p < .001$ OR=NR†	-13.077*** (0.004) $p < .001$ OR=NR†	13.149*** (0.892) $p < .001$ OR=NR†	15.209*** (0.985) $p < .001$ OR=NR†	14.954*** (0.939) $p < .001$ OR=NR†
Income: Informal	12.448*** (0.833) $p < .001$ OR=NR†	-4.044*** (1.199) $p = .001$ OR=0.018	-19.091*** (0.000) $p < .001$ OR=NR†	4.173*** ($< .001$) $p < .001$ OR=NR†	-0.393 (0.579) $p = .498$ OR=0.675	-16.001*** ($< .001$) $p < .001$ OR=NR†	-0.679*** ($< .001$) $p < .001$ OR=NR†	-3.384*** ($< .001$) $p < .001$ OR=NR†	-14.996*** (0.004) $p < .001$ OR=NR†	-12.979*** (0.001) $p < .001$ OR=NR†	-13.051*** (0.000) $p < .001$ OR=NR†	2.886* (1.139) $p = .011$ OR=17.921
Income: Parent/Family	12.482*** (0.589) $p < .001$ OR=NR†	-5.761*** (1.399) $p < .001$ OR=0.003	-20.419*** (0.000) $p < .001$ OR=NR†	18.136*** (0.185) $p < .001$ OR=NR†	-2.677*** (0.891) $p = .004$ OR=0.069	10.836*** (0.774) $p < .001$ OR=NR†	17.849*** (0.609) $p < .001$ OR=NR†	30.927*** (0.382) $p < .001$ OR=NR†	-14.426*** (0.005) $p < .001$ OR=NR†	3.342*** (1.067) $p = .002$ OR=28.276	2.327 (2.717) $p = .392$ OR=10.247	-0.097 (0.837) $p = .908$ OR=0.908

Note. Each cell reports β (SE), p-value, and OR $= \exp(\beta)$. OR=NR† where estimates were flagged as potentially unstable/separated; Income reference category = formal employment; Rural/Urban: 1 = rural. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

State 4 (very high risk), indicating no worsening of risk among those initially high-risk individuals. Low transition probabilities were observed from State 2 to 4 (0.003), State 3 to 1 (0.020), and State 1 to 2 (0.067), suggesting limited movement across risk categories and overall stability in AGYW risk profiles over time.

3.2.5. Log-Odds of transitioning from a latent class to another

Across the four source-class models, income source was the most consistent predictor of category membership relative to the reference category. In Source Class 1, informal income significantly reduced odds of category 3 (OR = 0.018, p = .001), and parent/family income significantly reduced odds of category 3 (OR = 0.003, $p < .001$) (Table 11). In Source Class 2, parent/family income significantly reduced odds of category 3 (OR = 0.069, p = .004). In Source Class 4, the category 2 intercept was significant ($\beta = -1.979$, OR = 0.138, p = .035), informal income significantly increased odds of category 4 (OR = 17.921, p = .011), and parent/family income significantly increased odds of category 2 (OR = 28.276, p = .002). Rural/urban residence showed no significant effects with interpretable odds ratios in any source class.

3.2.6. Latent class transition between T1 and T2

Figure 4 tracks how the size of each latent class changes between the two time points in the model. State 3 shows the most notable shift, nearly doubling from about 0.24 to 0.45 and becoming the largest class by time 2, while state 4 shrinks steadily from about 0.12 to near zero. State 2 declines moderately (0.39 to 0.25), and state 1 stays roughly flat around 0.25–0.27. Overall, the population is redistributing away from states 2 and 4 and toward state 3, while state 1 remains stable reflecting the net effect of the individual-level transition probabilities in the underlying model.

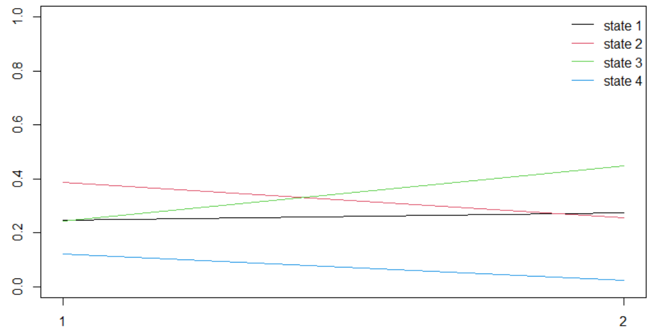


Figure 4. Latent class transition between T1 and T2.

Key: State 1 is Low risk, State 2 is Medium risk, State 3 is High risk, and State 4 is Very high risk

3.3. Discussion

The results provide evidence for the use of Latent Markov Modeling (LMM) in longitudinally tracking hidden categorical variables representing vulnerability levels to HIV among adolescent girls and young women (AGYW) over time. Notable changes in vulnerability were observed among 10–14-year-old and 15–19-year-old AGYW from screening and enrolment into the DREAMS program to reassessments conducted after 12 months. These findings align with studies on latent transitions in drug and substance use observed among men who have sex with men (MSM) tracked over two years in British Columbia by Noor et al. and Card et al. [5, 13]. The Latent Markov analysis documented transitions toward lower-risk vulnerability classes among AGYW during the 12-month period of DREAMS participation, evidenced by shifts in vulnerability states between the two time points that is enrolment and reassessment period. This pattern is broadly consistent with findings from Wambiya et al. [6], who also reported positive changes associated with the DREAMS program. The intervention's comprehensive layering of social, clinical, and financial empowerment services corresponds with the observed transitions to lower-risk classes, consistent with a study in Kenya and South Africa by Gourlay et al. [31]. Financial empowerment is associated with reduced engagement in transactional sex, which corresponds with the observed transitions to lower-risk classes.

Major transitions were observed among the 10–14-year-old cohort, who largely maintained membership in Class 3 which is medium risk. This stability could be attributed to the higher proportion of orphans in this group, as the intervention cannot alter orphanhood status, a finding consistent with the Civil Addict Program study in California by Xinyuan at al. [32]. Moreover, most 10–14-year-old AGYW were still in school, which is known to delay sexual debut and thus reduce HIV risk which is in line with a study in Eastern and Southern Africa by Gourlay et al. [31]. Overall, transitions in HIV vulnerability observed over the 12-month reassessment period are similar to those reported among MSM engaged in drug and substance use, who transitioned from high-risk to low-risk categories over 24 months in a study by Card et al. [13]. Incorporating covariates in the Latent Transition Analysis (LTA) model enhanced precision and reduced model error and improving accuracy as highlighted in a study by Kim et al. [2]. In this analysis, covariates included area of residence (rural or urban) and main source of income, consistent with covariates used in LTA studies on mental health and drug use by Turpin et al [30].

Area of residence that is urban or rural and source of household income (parents or family members, informal sector, or husband/boyfriend) were identified as key factors associated with latent Markov transitions across classes over time, reflecting the vulnerability of AGYW to HIV. This finding aligns with studies by Zhou et al. [33] and Pennoni et al [34], which demonstrated that covariates significantly influence transition probabilities. The novelty of this study lies in the application of Latent Markov Modeling to examine transitions in HIV vulnerability among AGYW in low-resource settings. Employing this modeling approach is crucial for designing and prioritizing targeted interventions to enhance effectiveness among vulnerable populations. Within this context, the DREAMS interventions through Social Asset Building Clubs (SABCs) play a pivotal role by providing life skills training, economic empowerment, and linkages to clinical and social services, thereby strengthening resilience, alongside the observed transitions from high-risk to lower-risk latent classes over time.

4. Conclusion

Latent Markov modeling is a valuable approach for assessing longitudinal transitions of latent classes related to vulnerability to HIV among adolescent girls and young women (AGYW). This modeling technique enables the characterization of changes in key variables over time and the estimation of state transition probabilities, documenting the pattern and pace of change over the course of an intervention period. Findings indicated that latent classes for HIV vulnerability among AGYW showed transitions consistent with movement from higher-risk to lower-risk states during the period of DREAMS participation. Among AGYW aged 10–14 years, major transitions were from Class 3, identified as medium risk and characterized by orphanhood and school dropout vulnerability. The dominant transition for AGYW aged 15–19 years was from Class 3, a high-risk class associated with a history of pregnancy and irregular condom use with non-marital partners. Class numbers were estimated independently for each age group's model, so a given class number (e.g., Class 3) does not represent the same risk profile across the two age groups; the accompanying risk descriptor, rather than the numeral alone, should be used when comparing classes between age groups. Low transition probabilities were observed for movement from Class 2 (low risk) to Class 1 (very high risk) among the 10–14-year-olds, and from Class 3 (high risk) to Class 1 (low risk) among the 15–19-year-olds. Incorporating covariates such as area of residence (rural or urban) and household income source enhanced model precision. We recommend the use of Latent Markov modeling in interventions such as the DREAMS program to monitor transitions in vulnerability over time and to inform the design of targeted, group-specific interventions for greater effectiveness.

Author Contributions. **Fungai Hamilton Mudzengerere:** Conceptualization, methodology, formal analysis, investigation, writing original draft preparation. **Oliver Bodhlyera:** Methodology, supervision, writing review and editing. **Henry Mwambi:** Supervision, resources, writing review and editing. All authors have reviewed and approved the final version of this manuscript.

Acknowledgement. The authors are grateful for the support of Zimbabwe Health Interventions, DREAMS project staff, and beneficiaries who contributed to this study.

Funding. This research was not funded by any institution.

Conflict of interest. The authors declare no conflict of interest.

Data availability. The dataset includes sensitive, participant-level data on adolescent girls' and young women's HIV-related vulnerabilities and sexual and reproductive health behaviors, including data from minors in the 10–14-year-old cohort. Due to re-identification risk, the full dataset cannot be made publicly available. De-identified data may be shared with qualified researchers on reasonable request to the corresponding author, subject to a data-sharing agreement consistent with the original informed consent, parental/guardian consent, and participant assent under which the data were collected.

Ethical Considerations. The study was covered by the Zimbabwe Health Interventions (ZHI) non-research determination protocol approved by the Medical Research Council of Zimbabwe (MRCZ/E/216); the protocol provides for analysis, use, and dissemination of routinely collected program data; provision of consent was therefore waived. Routine program data were analyzed, and beneficiaries' personal identification information was excluded from the analysis.

Patient and Public Involvement Statement. The research was based on routinely collected monitoring and evaluation data captured in district health information systems (DHIS2). Thus, there was no direct contact with study participants and the public. During the collection of the routine monitoring and evaluation data, AGYW were concerned about being screened for eligibility into the DREAMS program. The program beneficiaries were not involved in the design and recruitment process as routine program data used. The findings from the study were disseminated to the program's provincial and district offices staff who in-turn cascaded the finding to local and community structures including the program beneficiaries for the purposes of program improvement and prioritising interventions for AGYW.

References

[1] G. J. Melendez-Torres, E. Allen, R. Viner, and C. Bonell, "Effects of a whole-school health intervention on clustered adolescent health risks: Latent transition analysis of data from the inclusive trial," *Prev. Sci.*, vol. 23, no. 1, pp. 1–9, 2022, doi: 10.1007/s11121-021-01237-4.

[2] E. A. Kim, H. Chung, and S. Jeon, "Joint latent class analysis for longitudinal data: An application on adolescent emotional well-being," *Commun. Stat. Appl. Methods (CSAM)*, vol. 27, pp. 241–254, 2020, doi: 10.29220/CSAM.2020.27.2.241.

[3] R. Turner, *hmm.discnp: Hidden Markov Models with Discrete Non-Parametric Observation Distributions*, 2016, r package version 0.2-4; url: https://CRAN.R-project.org/package=hmm.discnp.

[4] C. Bonell, E. Beaumont, M. Dodd, D. Elbourne, L. Bevilacqua, A. Mathiot,

and E. Allen, "Effects of school environments on student risk-behaviours: Evidence from a longitudinal study of secondary schools in england," *J. Epidemiol. Community Health*, 2019.

[5] S. W. Noor, T. A. Hart, C. N. Okafor *et al.*, "Staying or moving: Results of a latent transition analysis examining intra-individual stability of recreational substance use among msm in the multicenter aids cohort study from 2004 to 2016," *Drug Alcohol Depend.*, vol. 220, p. 108516, 2021, doi: 10.1016/j.drugalcdep.2021.108516.

[6] E. O. A. Wambiya, A. J. Gourlay, S. Mulwa, F. Magut, N. Mthiyane, B. Orindi, N. Chimbindi, D. Kwaro, M. Shahmanesh, S. Floyd, I. Birdthistle, and A. Ziraba, "Impact of dreams interventions on experiences of violence among adolescent girls and young women: Findings from population-based cohort studies in kenya and south africa," *PLOS Glob. Public Health*, vol. 3, no. 5, p. e0001818, 2023, doi: 10.1371/journal.pgph.0001818.

[7] USAID, "Dreams: Partnership to reduce hiv/aids in adolescent girls and young women," url: https://www.usaid.gov/global-health/health-areas/hiv-and-aids/technical-areas/dreams; accessed: May 21, 2024.

[8] F. Bartolucci, S. Pandolfi, and F. Pennoni, "Lmest: An r package for latent markov models for longitudinal categorical data," *J. Stat. Softw.*, vol. 81, no. 4, pp. 1–38, 2017, doi: 10.18637/jss.v081.i04.

[9] S. Bonhomme, K. Jochmans, and J.-M. Robin, "Estimating multivariate latent structure models," *Ann. Statist.*, vol. 44, pp. 540–563, 2016.

[10] C. Kam, A. J. Morin, J. P. Meyer, and L. Topolnytsky, "Are commitment profiles stable and predictable? a latent transition analysis," *J. Manage.*, vol. 42, no. 6, pp. 1462–1490, 2016.

[11] K. Saengtrakul, S. Kanjanawasee, and N. Wiratchai, "Student factors affecting latent transition of mathematics achievement measuring from latent transition analysis with a mixture item response theory measurement model," *Procedia Soc. Behav. Sci.*, vol. 217, pp. 729–737, 2016.

[12] J. Zyberaj, C. Bakac, and S. Seibel, "Latent transition analysis in organizational psychology: A simplified 'how to' guide by using an applied example," *Front. Psychol.*, vol. 13, p. 977378, 2022, doi: 10.3389/fpsyg.2022.977378.

[13] K. G. Card, H. L. Armstrong, A. Carter, Z. Cui, L. Wang, J. Zhu, N. J. Lachowsky, D. M. Moore, R. S. Hogg, and E. A. Roth, "Assessing the longitudinal stability of latent classes of substance use among gay, bisexual, and other men who have sex with men," *Drug Alcohol Depend.*, vol. 188, pp. 348–355, 2018, doi: 10.1016/j.drugalcdep.2018.04.019.

[14] I. Zych, J. Rodríguez-Ruiz, I. Marín-López, and V. J. Llorent, "Longitudinal stability and change in adolescent substance use: A latent transition analysis," *Children Youth Serv. Rev.*, vol. 112, 2020.

[15] A. H. Weinberger, J. Platt, and R. D. Goodwin, "Is cannabis use associated with an increased risk of onset and persistence of alcohol use disorders? a three-year prospective study among adults in the united states," *Drug Alcohol Depend.*, vol. 161, pp. 363–367, 2016, doi: 10.1016/j.drugalcdep.2016.01.014.

[16] P. M. M. Okuda, W. Swardfager, P. S. Lucio, G. B. Ploubidis, T. Liu, M. Pangelinan, and H. Cogo-Moreira, "The trajectory of balance skill development from childhood to adolescence was influenced by birthweight: A latent transition analysis in a british birth cohort," *J. Clin. Epidemiol.*, vol. 109, pp. 12–19, 2019.

[17] Y. Ni, J. Y. Tein, M. Zhang, Y. Yang, and G. Wu, "Changes in depression among older adults in china: A latent transition analysis," *J. Affect. Disord.*, vol. 209, pp. 3–9, 2017.

[18] B. A. Hultgren, R. Turrisi, M. J. Cleveland, K. A. Mallett, R. Reavy, M. E. Larimer, I. M. Geisner, and M. M. Hospital, "Transitions in drinking behaviors across the college years: A latent transition analysis," *Addict. Behav.*, vol. 92, pp. 108–114, 2019, doi: 10.1016/j.addbeh.2018.12.021.

[19] F. Liao, E. Molin, H. Timmermans, and B. van Wee, "The impact of business models on electric vehicle adoption: A latent transition analysis approach," *Transp. Res. Part A Policy Pract.*, vol. 116, pp. 531–546, 2018.

[20] R. Di Mari, F. Dotto, A. Farcomeni, and A. Punzo, "Assessing measurement invariance for longitudinal data through latent markov models," *Struct. Equ. Modeling*, vol. 29, no. 3, pp. 381–393, 2021, doi: 10.1080/10705511.2021.1993857.

[21] H. L. Nguena Nguefack, M. G. Pagé, J. Katz, M. Choinière, A. Vanasse, M. Dorais, O. M. Samb, and A. Lacasse, "Trajectory modelling techniques useful to epidemiological research: A comparative narrative review of approaches," *Clin. Epidemiol.*, vol. 12, pp. 1205–1222, 2020, doi: 10.2147/CLEP.S265287.

[22] J. H. Ryoo, C. Wang, S. M. Swearer, M. Hull, and D. Shi, "Longitudinal model building using latent transition analysis: An example using school bullying data," *Front. Psychol.*, vol. 9, p. 675, 2018, doi: 10.3389/fpsyg.2018.00675.

[23] B. Muthén and T. Asparouhov, "Latent transition analysis with random intercepts (ri-lta)," *Psychol. Methods*, vol. 27, no. 1, pp. 1–16, 2022, doi: 10.1037/met0000370.

[24] L. V. D. E. Vogelsmeier, J. K. Vermunt, L. Keijsers, and K. De Roover, "Latent markov latent trait analysis for exploring measurement model changes in intensive longitudinal data," *Eval. Health Prof.*, vol. 44, no. 1, pp. 61–76, 2021, doi: 10.1177/0163278720976762.

[25] K. Nylund-Gibson, A. C. Garber, D. B. Carter *et al.*, "Ten frequently asked questions about latent transition analysis," *Psychol. Methods*, vol. 28, no. 2, pp. 284–300, 2023, doi: 10.1037/met0000486.

[26] O. Perra, "Latent transition analysis," in *SAGE Research Methods: Foundations*. SAGE Publications, 2020, doi: 10.4135/9781526421036878157.

[27] F. Li, A. Cohen, B. Bottge, and J. Templin, "A latent transition analysis model for assessing change in cognitive skills," *Educ. Psychol. Meas.*, vol. 76, no. 2, pp. 181–204, 2016, doi: 10.1177/0013164415588946.

[28] N. S. Ahanchi, F. Hadaegh, A. Alipour, A. Ghanbarian, F. Azizi, and D. Khalili, "Application of latent class analysis to identify metabolic syndrome components patterns in adults: Tehran lipid and glucose study," *Sci. Rep.*, vol. 9, no. 1, 2019.

[29] A. Abarda, M. Dakkon, M. Azhari, A. Zaaloul, and M. Khabouze, "Latent transition analysis (lta): A method for identifying differences in longitudinal change among unobserved groups," *Procedia Comput. Sci.*, vol. 170, pp. 1116–1121, 2020, doi: 10.1016/j.procs.2020.03.059.

[30] R. E. Turpin, T. V. Dyer, D. T. I. Dangerfield, H. Liu, and K. H. Mayer, "Syndemic latent transition analysis in the hptn 061 cohort: Prospective interactions between trauma, mental health, social support, and substance use," *Drug Alcohol Depend.*, vol. 214, p. 108106, 2020, doi: 10.1016/j.drugalcdep.2020.108106.

[31] A. Gourlay, S. Floyd, F. Magut, S. Mulwa, N. Mthiyane, E. Wambiya *et al.*, "Impact of the dreams partnership on social support and general self-efficacy among adolescent girls and young women: Causal analysis of population-based cohorts in kenya and south africa," *BMJ Glob. Health*, vol. 7, p. e006965, 2022, doi: 10.1136/bmjgh-2021-006965.

[32] X. Song, Y. Xia, and H. Zhu, "Hidden markov latent variable models with multivariate longitudinal data," *Biometrics*, vol. 73, no. 1, pp. 313–323, 2017, doi: 10.1111/biom.12536.

[33] X. Zhou, K. Kang, and X. Song, "Two-part hidden markov models for semi-continuous longitudinal data with nonignorable missing covariates," *Stat. Med.*, vol. 39, pp. 1801–1816, 2020.

[34] F. Pennoni, M. Barbato, and S. Del Zoppo, "A latent markov model with covariates to study unobserved heterogeneity among fertility patterns of couples employing natural family planning methods," *Front. Public Health*, vol. 5, p. 186, 2017, doi: 10.3389/fpubh.2017.00186.