A Regularized Hidden Markov Model for Analyzing Multiple Tobacco Product Use in the US

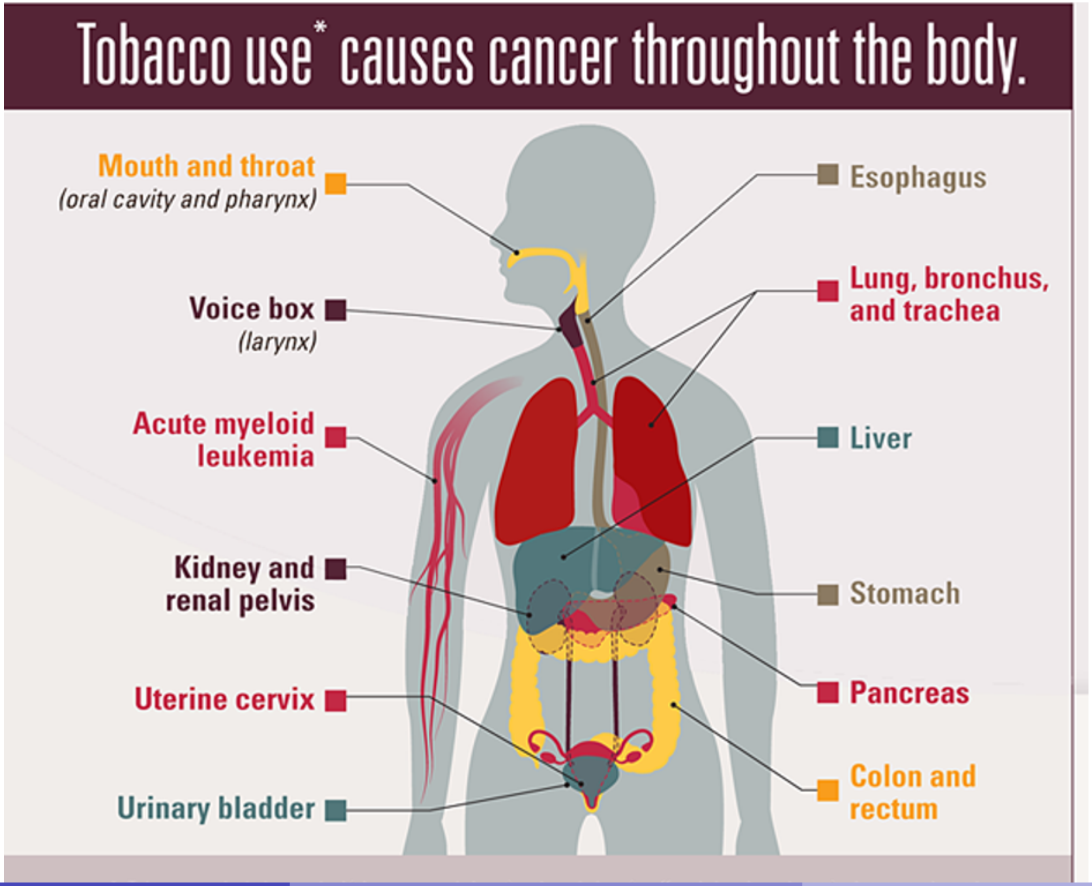
Xinyu Yan

University of Florida

August 7, 2024

Background

- Tobacco use remains the leading preventable cause of premature death.



Background

- Youth uptake of nicotine and vaporized products is increasing.

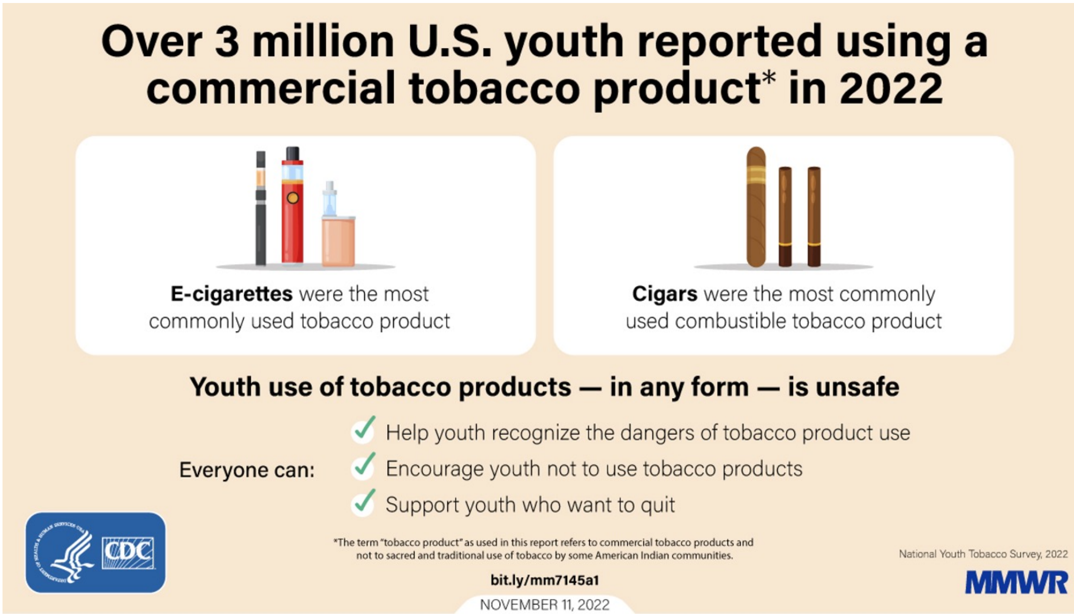


Figure 2: Youth use of tobacco products, 2022 (Source: CDC)

Background

- Diversification of tobacco products has created new challenges for regulatory actions.

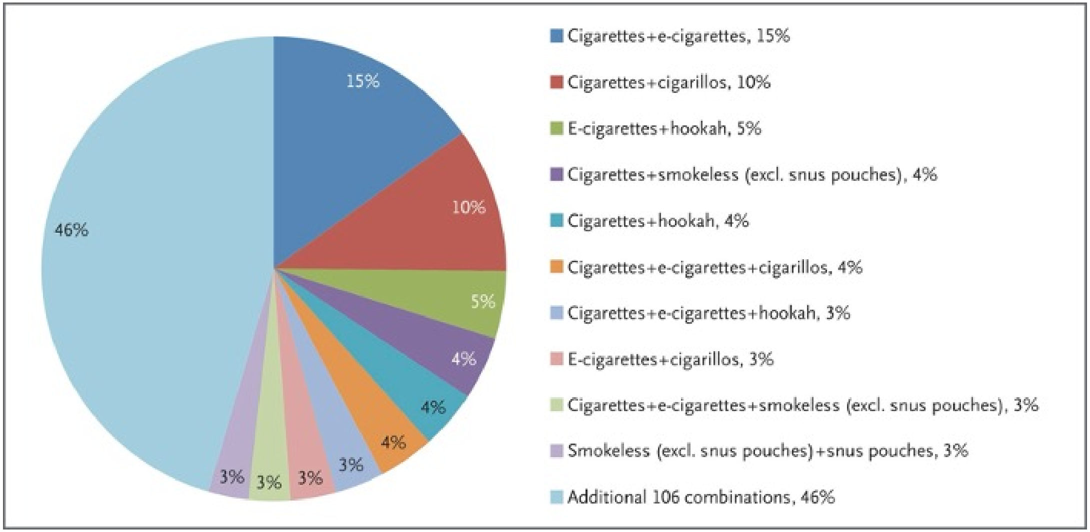


Figure 3: Most Common Combinations of Tobacco Products among Youth Multiple-Product Users (Kasza et al., 2017)

Background

- E-cigarette use was associated with initiation of tobacco cigarette smoking among teenagers.
- It is important to understand the patterns of transition between tobacco products and risk factors.

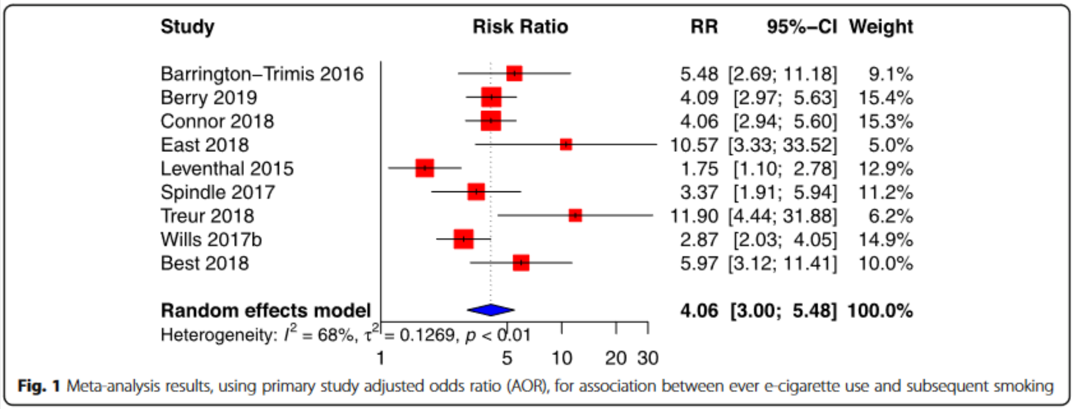
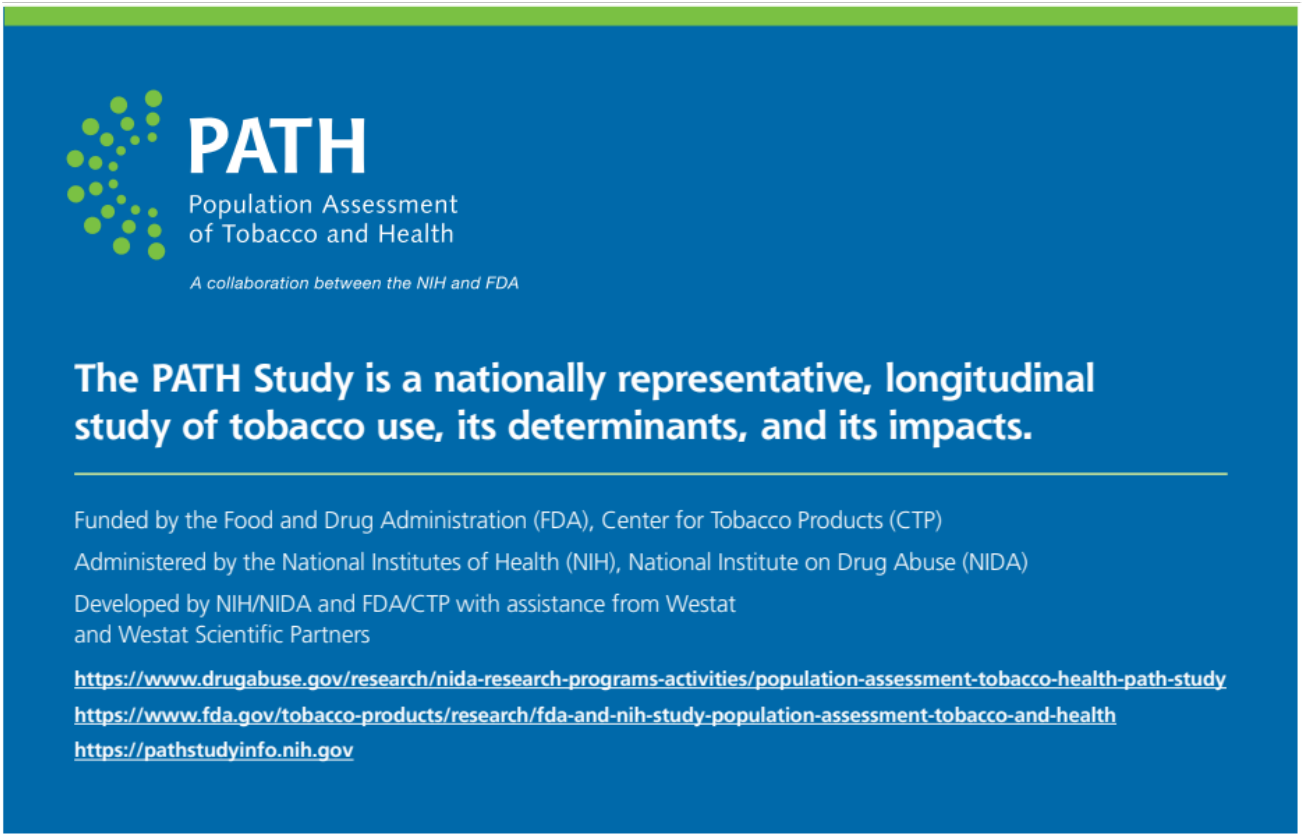
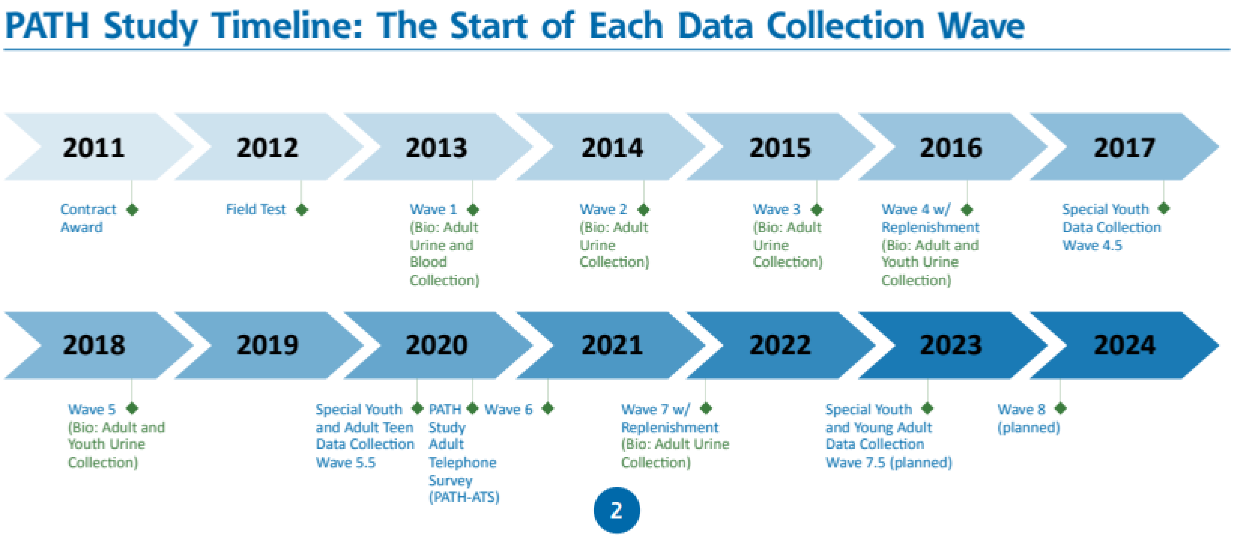


Figure 4: Association between ever e-cigaratte use and subsequent smoking (O'Brien et al., 2021)

Data source


PATH
Population Assessment
of Tobacco and Health
A collaboration between the NIH and FDA
The PATH Study is a nationally representative, longitudinal study of tobacco use, its determinants, and its impacts.
Funded by the Food and Drug Administration (FDA), Center for Tobacco Products (CTP)
Administered by the National Institutes of Health (NIH), National Institute on Drug Abuse (NIDA)
Developed by NIH/NIDA and FDA/CTP with assistance from Westat and Westat Scientific Partners
https://www.drugabuse.gov/research/nida-research-programs-activities/population-assessment-tobacco-health-path-study
https://www.fda.gov/tobacco-products/research/fda-and-nih-study-population-assessment-tobacco-and-health
https://pathstudyinfo.nih.gov

Data source


PATH Study Timeline: The Start of Each Data Collection Wave
2011
2012
2013
2014
2015
2016
2017
Contract Award
Field Test
Wave 1 (Bio: Adult Urine and Blood Collection)
Wave 2 (Bio: Adult Urine Collection)
Wave 3 (Bio: Adult Urine Collection)
Wave 4 w/ Replenishment (Bio: Adult and Youth Urine Collection)
Special Youth Data Collection Wave 4.5
2018
2019
2020
2021
2022
2023
2024
Wave 5 (Bio: Adult and Youth Urine Collection)
Special Youth and Adult Teen Data Collection Wave 5.5
PATH Study Adult Telephone Survey (PATH-ATS)
Wave 6
Wave 7 w/ Replenishment (Bio: Adult Urine Collection)
Special Youth and Young Adult Data Collection Wave 7.5 (planned)
Wave 8 (planned)
2

PATH Data

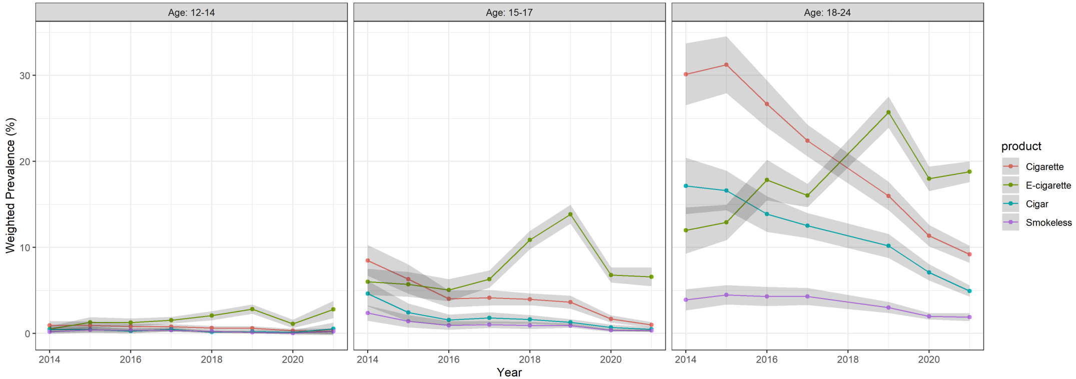


Figure 5: Weighted prevalence of past-30-day tobacco use in youth (aged 12-14 and 15-17) and young adults (aged 18-24) from Waves 1 to 6 (2014 - 2021) of the PATH Study.

Prior studies: Latent transition analysis

- Strength: A latent transition analysis with covariates
- Limitations:
  - Two-stage model fitting process leads to sub-optimal fit
  - Did not consider complex survey design
  - Did not account for unequal spacing in time between observations
  - Assumed constant probability of tobacco use over time

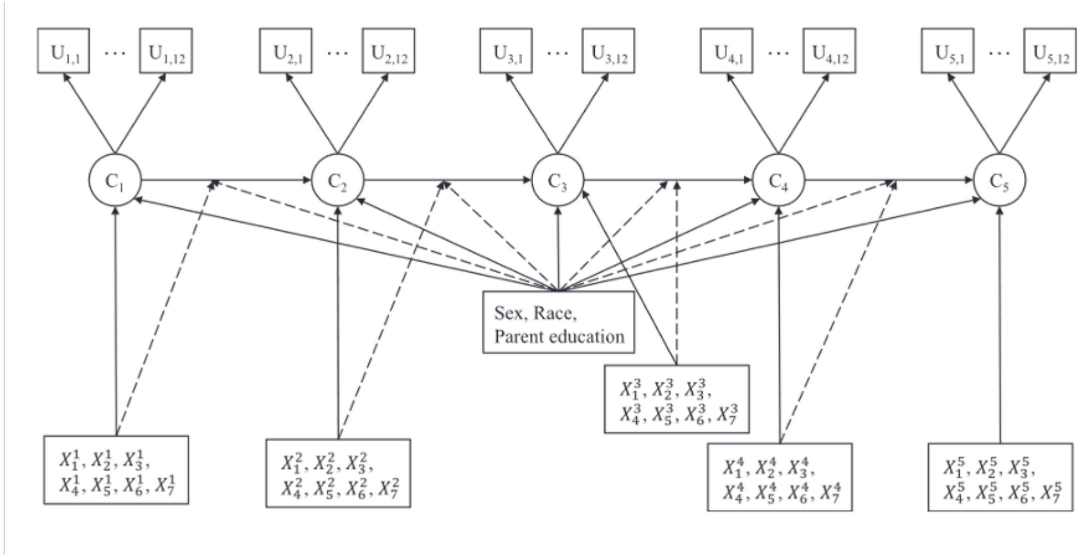


Figure 6: A latent transition model to study covariate effects on latent tobacco use behavior classes, PATH study, 2013–2019 (Huang et al., 2023)

Latent variable approach for analyzing tobacco use

- A latent variable modeling approach systematically categorizes individuals into meaningful, parsimonious groupings;
- Facilitates interpretable predictions of group membership and transitions (Lanza et al., 2010).

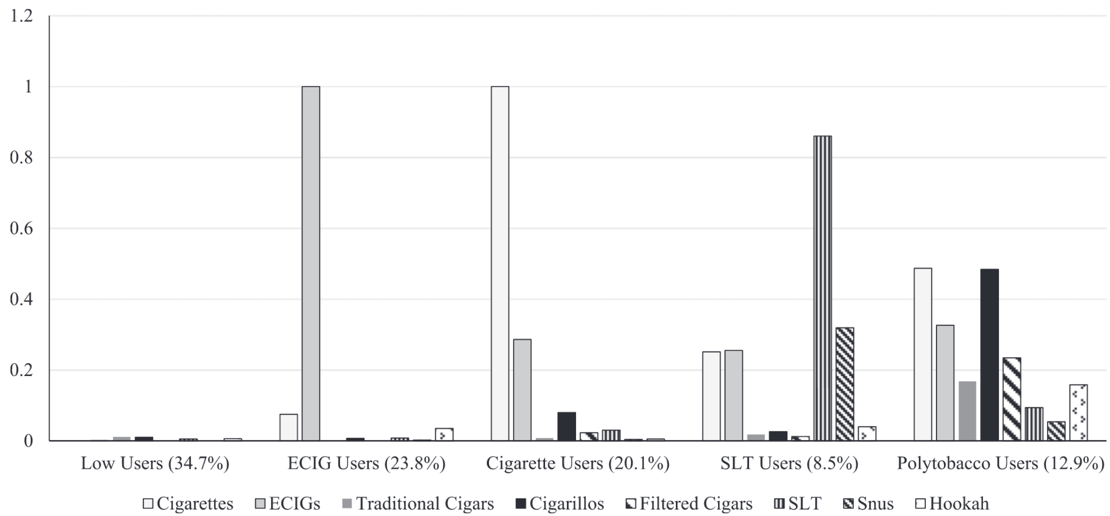


Figure 7: Conditional probabilities of endorsing tobacco use for the 5 latent classes of tobacco use (Romm et al., 2022)

Prior studies: continuous-time Markov transition model

- Strength: Continuous-time Markov transition model accounts for complex survey design
- Limitation: The manifest approach makes it hard to include more tobacco products; Markov assumption might be inappropriate

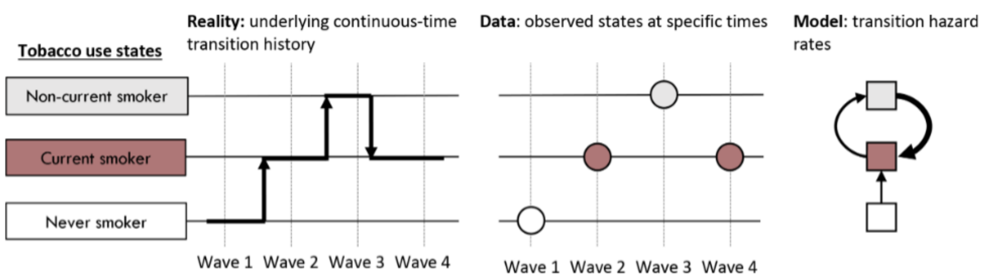


Figure 2 People transition between tobacco use states. We observe these states at fixed time points, but we do not know when the transitions occurred or if there were multiple transitions between observations. The multistate transition model estimates the underlying instantaneous transition hazard rates that best explain the observed data when they are combined to estimate probabilities of being in each state at future times.

Figure 8: Transitions occur between time of observation (Brouwer et al., 2020)

Motivation for the current study

- To address limitations of prior studies
- We want to use estimates from this study to forecast future tobacco use trends
    - Adding regularization to covariate effects can select variables, stabilize estimates, and promote generalizability of the model.

Methods: Data

- Data: Waves 1 to 6 of the PATH Data.
- Study population: Youth aged 12-17 and young adults 18-24 years old ($n = 12169$).
- Multivariate outcome of interest: Past 30 day use of tobacco products at interview (Yes/No).
- Tobacco products: Cigarettes, E-cigarettes, Cigars (traditional, filtered, cigarillo), smokeless tobacco (chewing and snus pouches), and hookah.

Methodological Approach

Elastic-net Regularized Log-Likelihood:

- Regularization formula:
  $\ell_{\text{pen}}(\theta) = \log(L(\theta)) - \lambda \left(\alpha \cdot \sum |\beta_k| + (1-\alpha) \cdot \sum \beta_k^2\right)$
- A data-driven way to select informative risk factors that drive change in tobacco use behavior

Incorporating Survey Weights:

- Weighted likelihood: $L = \prod L_i^{w_i}$, $\ell = \sum w_i \ell_i$
- To obtain nationally representative estimates from survey data

Identifying Optimal Components:

- Challenge of optimal component identification in HMMs
- Solution through cross-validation (CV) to select hidden states
- Evaluation based on the lowest normalized log-likelihood

These methods collectively address challenges in model fitting, account for complex survey design, and ensure optimal selection of model components.

Simulation Study I - Identifying Number of Hidden States

Objective

- To assess the capability of CV in accurately identifying the number of hidden states.
- To evaluate the effect of covariates in this process

Methodology Overview

- Base Model: 4-state HMM.
- Models fitted with state numbers varying from 2 to 7.
- Utilized 10-fold CV to evaluate model performance.

Simulation Scenarios

1. Data without covariate effects; model fitted similarly.
2. Data with covariate effects; model fitted including covariates.
3. Data with covariate effects; model fitted without considering covariates.

Simulation results I

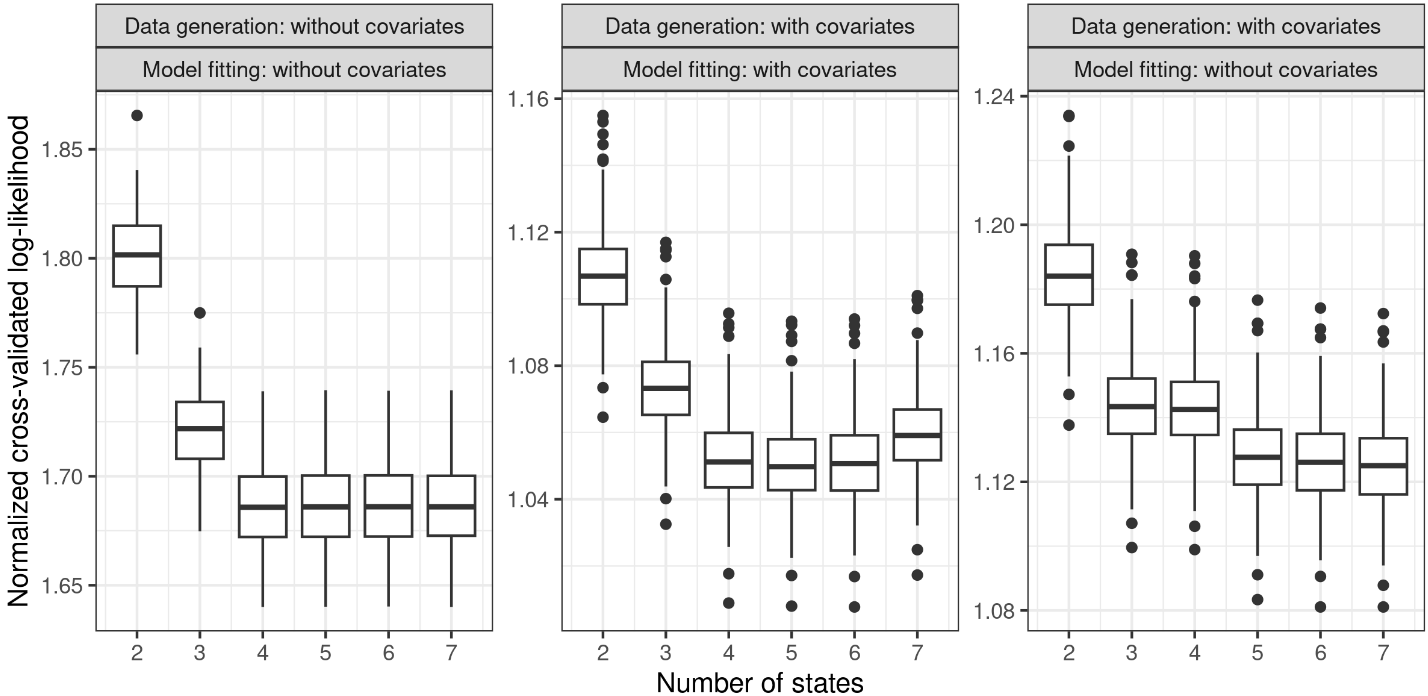


Figure 9: Cross-validation log-likelihood for HMMs of two to seven hidden states in three simulation settings

Simulation study II - Performance of Regularization on Transition Covariates

- We consider a scenario with 3 Bernoulli-distributed response variables, and underlying 2-state Markov chains. We have 40 covariates on transition intensities, 34 of which being noise covariates.
- We conduct 100 simulation runs, in each run generating $N = 800$ individuals with length of Markov chain $T = 7$.
- For the choice of tuning parameters α and λ, we consider a two-dimensional grid.
- We compare the performance of the elastic net estimators in the following fitting schemes:
    - HMM without regularization
    - elastic net HMM with α and λ selected by AIC
    - elastic net HMM with α and λ selected by BIC
    - LASSO HMM with α and λ selected by AIC
    - LASSO HMM with α and λ selected by BIC
    - elastic net HMM with α and λ selected by CV
    - LASSO HMM with α and λ selected by CV
- We calculate the mean squared error (MSE) of the coefficients.

Simulation results II

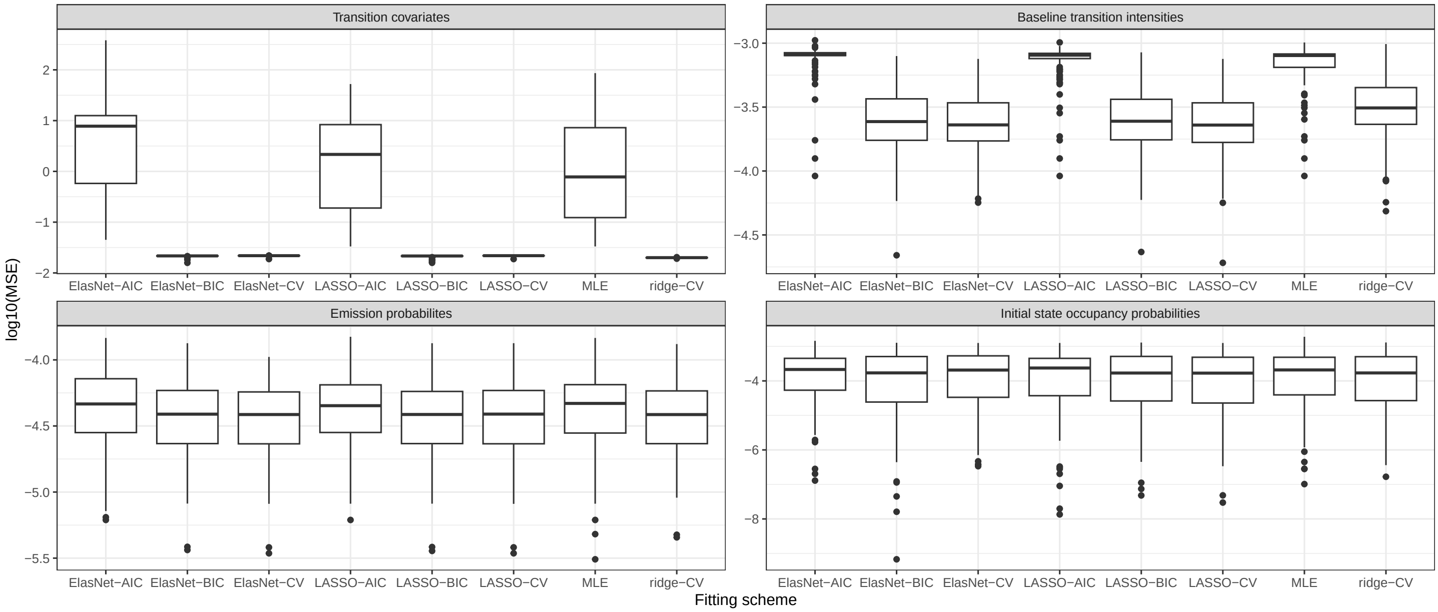


Figure 10: Log-MSE obtained in 100 simulation runs with independently distributed covariates

Results: PATH Data Analysis

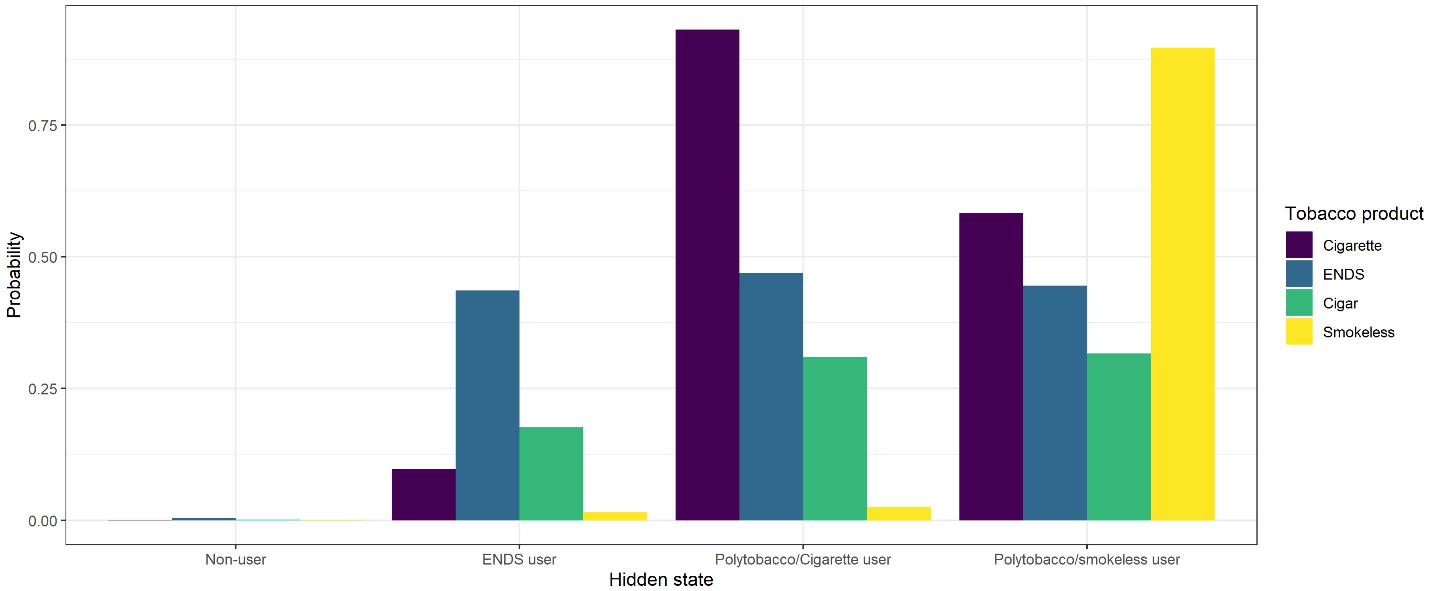


Figure 11: Conditional probabilities of endorsing tobacco use for the four hidden states.

Results: PATH Data Analysis

Table 1: One-year overall transition probabilities between hidden states

	Transition to			
Transition from	Non-user	ENDS user	Polytobacco/Cigarette user	Polytobacco/smokeless user
Non-user	0.969	0.030	0.001	0.001
ENDS user	0.254	0.716	0.030	0.000
Polytobacco/Cigarette user	0.129	0.168	0.703	0.001
Polytobacco/smokeless user	0.023	0.107	0.093	0.777

Effects of Covariates on Transition Rates

- Key covariates influencing transitions include age, education, gender, race/ethnicity, and substance use behaviors.
- Age and substance use (alcohol, marijuana, illicit drugs) significantly impact transitions to more intensive tobacco use states.
- Gender and race/ethnicity also play critical roles in transition dynamics, with notable differences in susceptibility and progression paths.

Note: Adjusted hazard ratios (HRs) for transitions between states reveal complex interactions between user characteristics and tobacco use patterns.

Reference I

Brouwer, A. F., Jeon, J., Hirschtick, J. L., Jimenez-Mendoza, E., Mistry, R., Bondarenko, I. V., Land, S. R., Holford, T. R., Levy, D. T., Taylor, J. M., et al. (2020). Transitions between cigarette, ends and dual use in adults in the path study (waves 1–4): Multistate transition modelling accounting for complex survey design. *Tobacco control*, *31*(3), 424–431.

Huang, S., Chen, Q., Griffin, P., Liu, G., & Azagba, S. (2023). Longitudinal transitions in tobacco use in youth and young adults: A latent transition analysis of the population assessment of tobacco and health study from wave 1 to 5. *Addictive Behaviors*, *138*, 107548.

Kasza, K. A., Ambrose, B. K., Conway, K. P., Borek, N., Taylor, K., Goniewicz, M. L., Cummings, K. M., Sharma, E., Pearson, J. L., Green, V. R., et al. (2017). Tobacco-product use by adults and youths in the united states in 2013 and 2014. *New England Journal of Medicine*, *376*(4), 342–353.

Reference II

- Lanza, S. T., Patrick, M. E., & Maggs, J. L. (2010). Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. *Journal of drug issues*, *40*(1), 93–120.
- O'Brien, D., Long, J., Quigley, J., Lee, C., McCarthy, A., & Kavanagh, P. (2021). Association between electronic cigarette use and tobacco cigarette smoking initiation in adolescents: A systematic review and meta-analysis. *BMC Public Health*, *21*(1), 1–10.
- Romm, K. F., Childers, M. G., Douglas, A. E., Bray, B. C., Dino, G., & Blank, M. D. (2022). Transitions in tobacco use profiles among adolescents: Results from the population assessment of tobacco and health (path) study waves 3 and 4. *Drug and alcohol dependence*, *232*, 109272.

Thank you!