State Transition Models

Roch Nianogo & Ashley Buchanan

Outline

- Definitions
- Key Properties
- Main uses
- Strengths and limitations
- Case examples

Definitions

Definition and history

- Simulation models that allow the representation and analysis of disease progression over time in cycles.
- Models disease progression pathways
  - Health states and events/transitions
  - Cycle by cycle until the end of the total time horizon

History

- Developed in the 1970s–80s to model transitions between health states (e.g., healthy → ill → dead)
- Initially used in demography and chronic disease research with Markov assumptions
- Broader adoption in the 1990s–2000s for HIV, disability, and smoking studies
- Recent extensions include non-Markov models, time-varying covariates, and causal inference
- Now widely used in evaluating disease progression and public health interventions

Cox & Miller, The theory of stochastic processes, 1965. Putter et al., *Stat Methods Med Res*, 2007.

Assumptions

Assumptions

- **Markov Assumption:** Future transitions depend only on the current state (memoryless), not on the path taken to reach that state.
- **Homogeneity of Transition Rates:** Transition intensities are constant over time (time-homogeneous) or vary only with known covariates (semi-Markov or time-inhomogeneous variants).
- **Independent Censoring:** The probability of censoring is independent of the unobserved transitions and states.
- **Accurate State Classification:** Individuals are correctly classified into discrete states without measurement error.
- **Non-informative Observation Process:** The timing and frequency of observations do not depend on the underlying transition process.
- **Absorbing and Recurrent States:** Models may include absorbing (e.g., death) or recurrent (e.g., hospitalization) states depending on the research question.

Key Properties

Key Properties - States

- Disease conditions or states.
  - States are mutually exclusive (i.e., individuals can only be in one state at a time)
  - Collectively exhaustive (e.g. , probability of being in each health state must sum to 1).
- Assume to be memoryless (Markov): probability of a health state in t+1 depends only on the state at time t.
- Individual-based STMs do not need the population to be homogeneous or memoryless.

Key Properties (1)

- Transition probabilities and cycle length: The population in each state moves from one state to the other in one cycle.
- Decision rules: Movement from one state to another is dictated by transition probabilities. Achieved using matrix multiplication.

Key Properties (2)

- Aggregate vs individual. Can be aggregate (population-level, Markov or Cohort STM). Can be *individual-based* simulation models and as such are referred to as microsimulation STM.
- Dynamic/Time. STM are generally modeled dynamically and time is generally modeled discretely.
  - When time is continuous—STM are mathematically indistinguishable from an SDM.

Main Uses

- One of the more common and basic simulation models
- Often used to model chronic diseases dynamic, especially those that have some recurrences.
- Example studies using an STM framework are as follows:
  - *Lee. Beyond binary retention in HIV care: predictors of the dynamic processes of patient engagement, disengagement, and re-entry into care in a US clinical cohort. AIDS 2018.*
  - *Weinstein. Forecasting CHD Incidence Mortality and Cost. The Coronary Heart Disease Policy Model. AJPH 1987.*
  - *Kim. Cost Effectiveness of Nutrition Policies on Processed Meat-Implications for Cancer Burden in the U.S. AJPM 2019.*
  - *Kowada. Cost-effectiveness and health impact of lung cancer screening with low-dose computed tomography for never smokers in Japan and the United States. BMC Pul 2022.*
  - *Murillo. Novel Application of a Multistate Model to Evaluate the Opioid Use Disorder Care Cascade: A Retrospective Cohort Study. Preprint 2025.*

Strengths and Limitations

Strengths and limitations

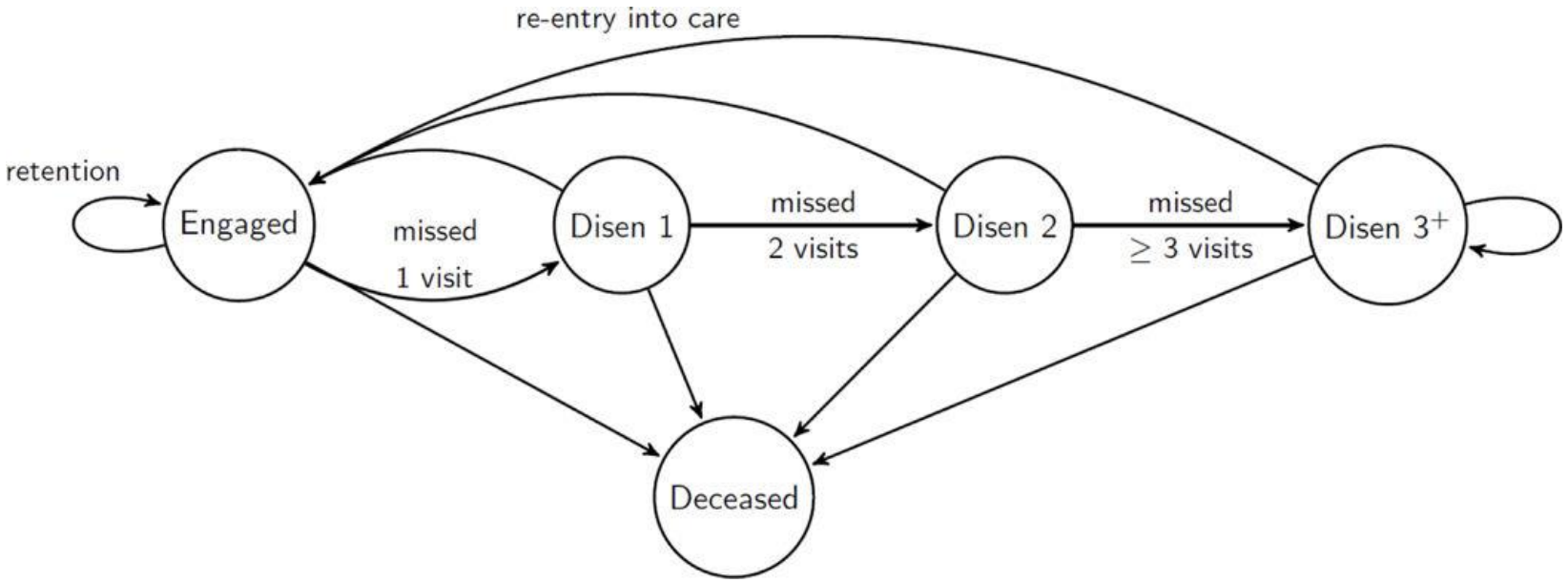
- Cohort STM assumes the population is homogeneous
- Markov STM is relatively simple to understand and implement.
- Time is generally modelled discretely.
- Suitable for modeling the progression of chronic diseases and can handle recurrences.
- In some cases, can be programmed using canned software in SAS or R
- Limitations
  - Individual-based STM can be computationally intensive and require technical knowledge
  - Cycle length is generally assumed to be fixed

Case Example: HIV Care

Case Example: HIV Care (1)

- Multistate framework: Represents dynamic patient transitions through states of HIV care:
  - Engaged (short, medium, long-term disengagement)
  - Transferred out or deceased
- Cohort Studies
  - AMPATH (Kenya): 92,215 patients (2001–2016)
  - CNICS (USA): 31,009 patients (1996–2014)
- State Assignment:
  - AMPATH: 200-day intervals
  - CNICS: 6-month intervals (based on viral load monitoring)
- Multinomial models estimate transition risks using RRRs
- Findings
  - The first 2 years are critical for preventing disengagement
  - Re-engagement likelihood declines with longer disengagement
  - Predictors of disengagement: low CD4 count, no ART, high viral load, absence of AIDS-defining illness

Case Example: HIV Care (2)



Lee et al. *AIDS*, 2018.

Disen X: Disengaged for X visit(s)

Lab Exercises

1. Modify the code so it is possible to have a 5% background probability of weight reduction (moving from susceptible to reduced weight without the intervention). Let this reduce the probability of transition from susceptible to susceptible to ensure the rows of the transition matrix sum to 1. How does this change the model results?
2. Using the original transition probabilities (rerun example code), modify the code so that there are 10% exposed and 5% adopted initially in the model. How does this change the model results?
3. Use the original transition probabilities by rerunning the example code. Vary the probability of transition from susceptible to exposed to half of the original value, then also double. In each case, let this reduce the probability of transition from susceptible to susceptible to ensure the rows of the transition matrix sum to 1. What happens to the output figure?

References

- Cox, D. R., & Miller, H. D. (1965). *The Theory of Stochastic Processes*. Chapman and Hall.
- Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine,* 26(11), 2389-2430.
- Weinstein, M. C., et al. (1987). Forecasting coronary heart disease incidence, mortality and cost: The CHD Policy Model. *American Journal of Public Health*, **77**(11), 1417–1426.
- Kim, D. D., et al. (2019). Cost-effectiveness of nutrition policies on processed meat: Implications for cancer burden in the United States. *American Journal of Preventive Medicine*, **57**(6), e143-e152.
- Kowada, A. (2022). Cost-effectiveness and health impact of lung cancer screening with low-dose computed tomography for never smokers in Japan and the United States. *BMC Pulmonary Medicine*, **22**, 1-13.
- Lee, H., et al. (2018). Beyond binary retention in HIV care: Predictors of the dynamic processes of patient engagement, disengagement, and re-entry. *AIDS*, **32**(15), 2217–2225.
- Murillo, A., et al. (2025). Novel application of a multistate model to evaluate the opioid use disorder care cascade: A retrospective cohort study. *medRxiv* [preprint].