COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Accounting for effect modification by latent disease state for individual causal estimation among bipolar participants in mobile health studies

Charlotte Fowler

Mentor: Linda Valeri

Collaborators: Xiaoxuan Cai, Justin Baker, Dost Ongur, Jukka-Pekka Onnela

August 8th, 2023

Motivation

Clinically, bipolar disorder is characterized by moving through **disease states** such as mania, euthymia, and depression

The effect of an intervention is likely dependent on the individual's current disease state

- E.g. Increased social activity lessens loneliness when in a depressed state but not in a manic state

Mobile health

MHealth studies collect information using smartphone and wearable devices

Offer rich longitudinal data including

- **Passive data**: accelerometer reports, GPS location, call/text logs, temperature data
- **Active data**: self reported mood, symptoms, activity, voice recordings

Mobile health for studying Bipolar patients

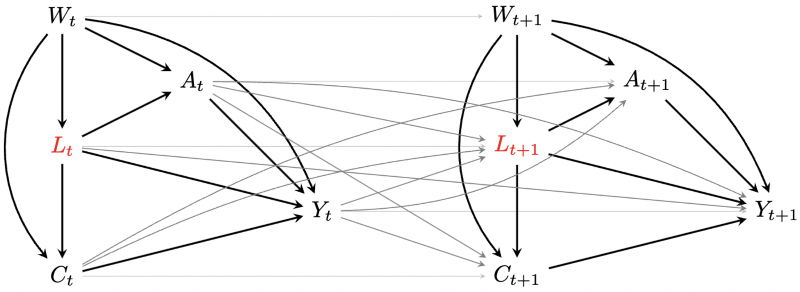
Strengths:

- ability to track at home behavior and symptoms
- observe raw information rather than potentially unreliable reports
- data collection at frequent time intervals

Challenges:

- unable to directly observe disease state
- missing data
- extreme heterogeneity between individuals
- complex causal relationships and pathways between variables

Hypothesized causal relationships



Y_t: Loneliness
A_t: Social activity
L_t : Latent disease state

C_t: Physical activity,
Self-reported mood
W_t: Environmental temperature

Objectives

Y_t: Loneliness
A_t: Social activity
L_t: Latent disease state

C_t: Physical activity,
Self-reported mood
W_t: Environmental temperature

Principal Objective: Estimate individual causal effect of $A_t = a_1$ versus $A_t = a_0$ on Y_t (and on future Y_{t+k}) among different levels of L_t

Secondary Objective: Predict latent class L_t given observed information

Existing Literature

Heterogeneous treatment effects

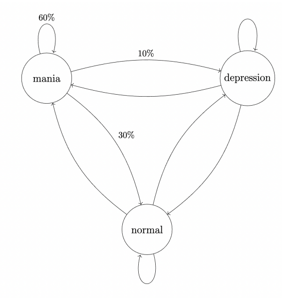
- Common to identify a latent modifier to explain the differences in observed effects between individuals in the sample (Pearl 2022, van den Ameele 2020)
- Focus on identification of latent subgroups among individuals in the sample, rather than time points for a given individual

Latent variable detection in time series data

- Goal of latent class prediction, not causal estimation (Chen et. al. 2020)

Hidden Markov Model (HMM)

Estimates and predicts latent disease state using transition probability framework:



Elements:

π: Initial latent state probabilities

- $\pi_i = P(L_1 = i)$

$\mathbf{A_t}$: Transition Probability Matrix

- $a_{ij} = P(L_t = j | L_{t-1} = i, X_{L,t} = x_{L,t})$

B_t: Response model

- $b_{1,j}(y_t) = P(Y_t = y_t | L_t = j, X_{Y,t})$
- $b_{2,j}(a_t) = P(A_t = a_t | L_t = j, X_{A,t})$
- $b_{3,j}(c_t) = P(C_t = c_t | L_t = j, X_{C,t})$

Hidden Markov Model Implementation

1. Frequentist Approach (HMM-F)
    - Baum Welch forward/backward EM algorithm
    - Viterbi algorithm for prediction
2. Bayesian Approach (HMM-B)
    - Forward algorithm for latent state identification
    - Optimized in STAN
    - Post convergence Viterbi algorithm for prediction

Columbia University | Mailman School of Public Health

Hidden Markov Model Adaptations

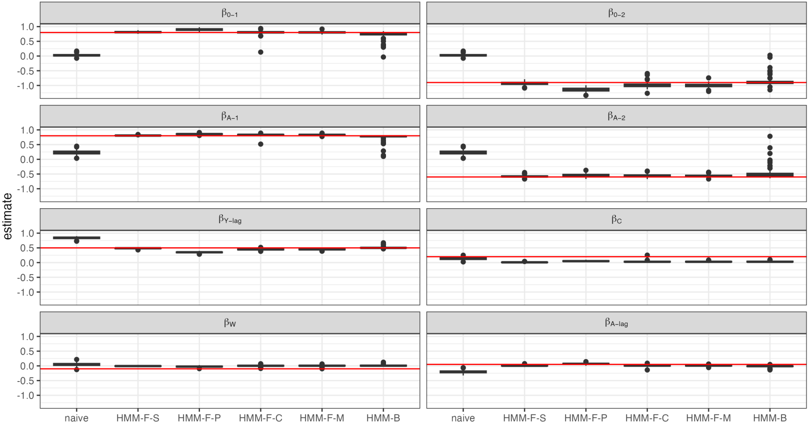
1. **Auto Regressive HMM**: allows observed time series (Y_t, A_t, C_t) to depend on lagged values even when conditioning on L_t

2. **Missing data in Y_t (and other observed variables)**:
   - **HMM-F-S**: Frequentist method that for missing Y_t singularly imputes $\hat{Y}_t$ when used as a regressor but marginalize over Y_t when treated as an outcome
   - **HMM-F**: Frequentist method that for missing Y_t multiply imputes $\hat{Y}_t$ when used as a regressor but marginalize over Y_t when treated as an outcome, pool results across multiple imputations within each E-M step
   - **HMM-B**: Bayesian method where missing Y_t is treated as additional parameter to be sampled

Hidden Markov Model Adaptations

3. **Latent L_t:**
   - **HMM-F-P/HMM-F-S**: $P(L_t = i)$ and $P(L_t = i, L_{t+1} = j)$ and use as probabilistic weights when updating parameter estimates
   - **HMM-F-M**: Multiply impute L_t from marginal $P(L_t = i)$ and pool estimates across MI within each E-M step
   - **HMM-F-C**: Multiply impute L_t from conditional $P(L_t = i|L_{t-1} = j)$ and pool estimates across MI within each E-M step
   - **HMM-B**: $P(L_t = i)$ treated as additional parameter to be sampled

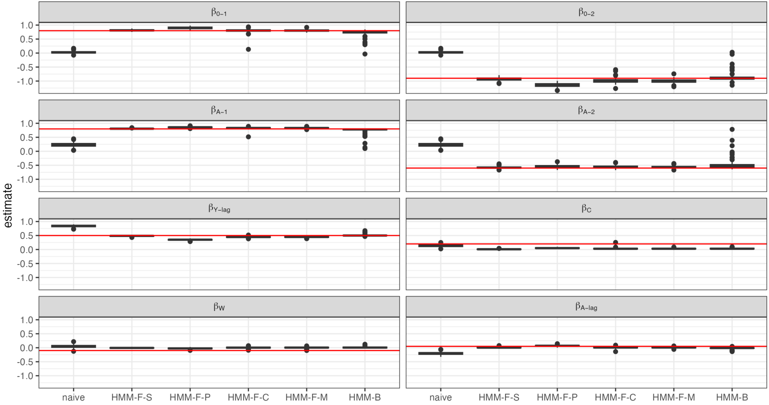
Simulation Results



$$L_t = 1 : Y_t = \beta_{0-1} + \beta_{A-1}A_t + \beta_{\text{Y-lag}}Y_{t-1} + \beta_C C_t + \beta_W W_t + \beta_{\text{A-lag}}A_{t-1} + \epsilon_t$$
$$L_t = 2 : Y_t = \beta_{0-2} + \beta_{A-2}A_t + \beta_{\text{Y-lag}}Y_{t-1} + \beta_C C_t + \beta_W W_t + \beta_{\text{A-lag}}A_{t-1} + \epsilon_t$$

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Simulation Results



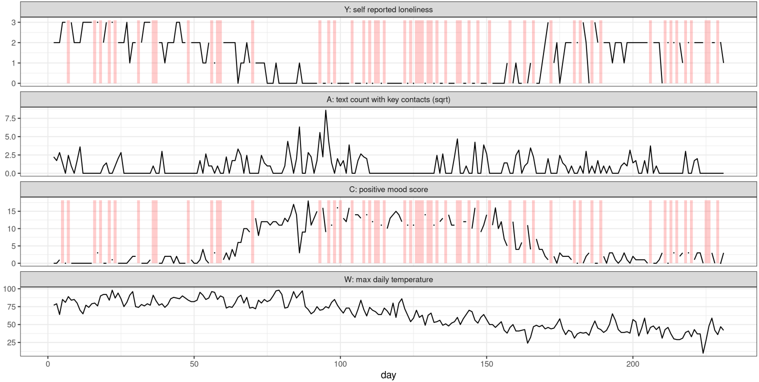
Latent state prediction accuracy:

naive	HMM-F-S	HMM-F-P	HMM-F-C	HMM-F-M	HMM-B
-	92.7%	87.9%	82.7%	83.4%	79.2%

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

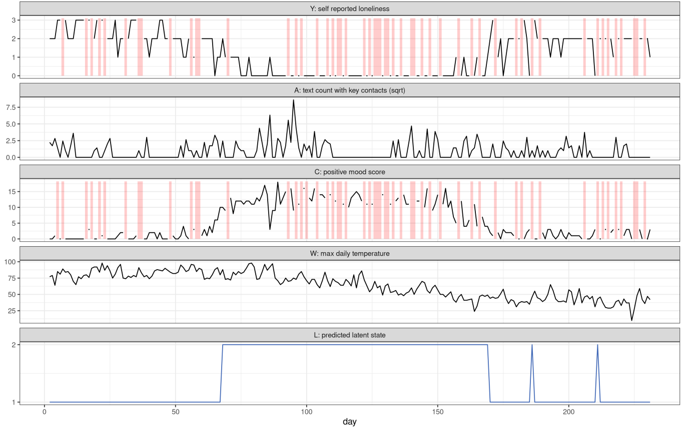
Bipolar Longitudinal Study (BLS)

Ongoing mHealth study from McLean Hospital with participants with bipolar disorder or schizophrenia followed for up to five years



Observed time series of participant with bipolar spectrum disorder

BLS Initial Results



$$\text{OR} = \frac{\text{odds}(Y_t^{a1} \leq j | L_t = i, X_{Y,t})}{\text{odds}(Y_t^{a0} \leq j | L_t = i, X_{Y,t})}$$

The observed effect of digital socialization on loneliness is more beneficial when in a depressed state (OR = 0.63) compared to a non-depressed state (OR = 0.94).

Further Directions

- Additional application results
  - Statistical inference and missing data in covariates
  - Interest in measuring impact of physical activity on sleep (both observed from passive data)
- Sensitivity analysis to address potential unmeasured confounding
  - Informative prior for Bayesian model
- Expand modeling framework to leverage information from similar individuals

Acknowledgements

Thank you to my advisor Linda Valeri, co-authors Xiaoxuan Cai, Justin Baker, Dost Ongur and Jukka-Pekka Onnela, and lab members Melanie Mayer and Zilan Chai for their help, suggestions, and feedback.

Contact: crf2147@cumc.columbia.edu