Joint Models of Longitudinal and Time-to-Event Data for Informing Multi-Stage Decision Making in mHealth

Presented by
Walter Dempsey, Ph.D.
University of Michigan
Outline

• Mobile health: an introduction
• Joint models: an introduction
• Assessing association (Decision making, stage 1)
• Defining risk strata (Decision making, stage 2)
• Assessing time-varying treatment effects (Decision making, stage 3)
• Conclusions: towards incorporating joint models in mHealth decision making
Mobile Health

“The delivery of healthcare services via mobile communication devices” – Foundation for the National Institutes of Health (FNIH)
Active data

- Active data involves explicitly asking participants for information, preferences, and/or opinions

- **Ecological Momentary Assessments (EMAs)**
  - Administrative: fixed self-report times
    - E.g., Daily dairy
  - Random:
    - 1 EMA uniformly sent in 4-hour time blocks
    - Can depend on observed history
  - Event-contingent: Participant initiated
    - Engages in NSSI
    - Engages in smoking

- How many questions and how often you ask depends on
  - Attendant burden and intrusiveness
  - Behavioral constructs of interest
    - E.g., subset of a baseline battery assessment asked multiple times per day
Passive data

- Passive data, participant has little awareness of the data collection effort

- How many sensors and data depends on
  - Attendant burden and intrusiveness
  - Behavioral constructs of interest
Why are we collecting data?

• Example: Sense2Stop
• To...
  o Monitor stress continuously
  o Decide whether/how to intervene based on level of stress
  o Deliver the intervention as soon as stress occurs
    • In the person’s natural environment
• And...
  o Monitor Context
  o Ensure that the context enables the person to receive and employ the intervention
Why are we collecting so much data?

• Goal is delivery of the right type of support at the right time (JITAIs) while minimizing disruptions
  o Data-driven decision making
    • Both using collected data and in experimental design for intervention development
  o Recall mHealth most useful when we minimize attendant burden and maximize therapeutic effects of JITAIs

• Research Questions:
  o How strong is association between behavioral biomarker and event outcome of interest?
    • How strong is the association between stress and the risk of lapse?
  o Can we use passive and active data to discriminate between patients of low and high risk?
  o Are the digital phenotyping biomarkers good?
    • E.g., is stress a good biomarker?
    • If treatment improves stress, does it also lower risk of lapse?
Example: A-CHESS

- Smartphone application to support recovery from alcoholism (A-CHESS)
  - Mobile phone administered
- A-CHESS had both static content (e.g., audio-guided relaxation) and interactive features
- If an individual is near a high-risk location (e.g., a bar she used to frequent), GPS initiated alert will ask if patient wants to be there.
- **Event can be passively-detected via GPS.**

(Gustafson et al., 2011 & 2014)
Joint models for longitudinal and survival data

Methods overview for mHealth
Research questions: Multiple outcomes

• **Association**: How strong is the association between behavioral biomarker and instantaneous risk rate of event?
  - Does the association vary over time?
  - How are behavioral biomarkers related to each other?

• **Prediction**: Can we improve predictions of event risk by considering behavioral biomarkers?

• **Treatment effect**: Does treatment improve the behavioral biomarker?
  - What is the direct and indirect effect of treatment on hazard rate of event?
• Repeated measurements on the same individual are expected to be (positively) correlated

\[ y_i = X\beta + Z_i b_i + \epsilon_i, \]

\[ b_i \sim N(0, D) \text{ and } \epsilon_i \sim N(0, \sigma^2 I_{n_i}) \]

• Where
  o \( y_i \) is the outcome
  o \( X \) is the design matrix for fixed effects
  o \( Z \) is the design matrix for random effects
  o \( b_i \perp \epsilon_i \)
• Assume there exists subject-specific coefficients: $\beta_i \sim N(\beta, D)$
Time-to-event data

• Notation:
  - $T_i^*$: “true” time-to-event
  - $C_i$: censoring time

• Observable data
  - $T_i = \min(T_i^*, C_i)$: observed event time
  - $\Delta_i = 1[T_i = T_i^*]$: censoring indicator

• Goal:
  
  Valid inference for $T_i^*$ using observed data
Time-to-event analysis

- Distribution functions
  \[ F(t) = P(T^* < t); \quad f(t) = \frac{d}{dt} F(t) \]

- Survival function
  \[ S(t) = P(T^* \geq t) \]

- Hazard function
  \[ h(t) = \lim_{\delta \to 0} \frac{P(t \leq T^* < t + \delta \mid T \geq t)}{\delta} \]
  \[ \Rightarrow S(t) = \exp \left(- \int_0^t h(u) \, du \right) \]
Time-to-event analysis

- **Relative risk model**

\[
h_i(t) = h_0(t) \exp(\gamma_1 w_{i1} + \cdots + \gamma_p w_{ip})
\]

\[
\log h_i(t) = \log h_0(t) + \gamma_1 w_{i1} + \cdots + \gamma_p w_{ip}
\]

- **Where**
  - \( h_i(t) \) denotes the hazard for patient \( i \) at time \( t \)
  - \( h_0(t) \) denotes the baseline hazard
  - \( w_{i1}, \ldots, w_{ip} \) a set of covariates
Cox model: baseline covariates

- Cox model: no assumptions for baseline hazard

\[
pl(\gamma) = \sum_{i=1}^{n} \delta_i \left[ \gamma^T w - \log \left( \sum_{j:T_j \geq T_i} \exp(\gamma^T w_j) \right) \right]
\]

- Example: PBC dataset

\[
h_i(t) = h_0(t) \exp(\gamma_1 Treatment_i + \gamma_2 Female_i + \gamma_3 Age_i)
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>HR</th>
<th>Std. Err.</th>
<th>Z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma_1)</td>
<td>-0.138</td>
<td>0.871</td>
<td>0.156</td>
<td>-0.882</td>
<td>0.378</td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>-0.493</td>
<td>0.611</td>
<td>0.207</td>
<td>-2.379</td>
<td>0.017</td>
</tr>
<tr>
<td>(\gamma_3)</td>
<td>0.021</td>
<td>1.022</td>
<td>0.008</td>
<td>2.784</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Rizopolous (2016), Tutorial: Joint model for Longitudinal and Survival Data
Time dependent covariates

• We often wish to answer questions related to time-dependent biomarker measures

• There are two types of time-dependent covariates (Kalbfleisch and Prentice, 2002):
  
  o Exogenous/External: Future trajectory of the variable is not affected by occurrence of an event:
    \[ P(Y_i(t)|Y_i(s), T_i^* \geq s) = P(Y_i(t)|Y_i(s), T_i^* = s) \]
  
  o Endogenous: not external
    • Biomarkers are almost ALWAYS endogenous
Extended Cox Model

• Extend the Cox model to handle time-dependent covariates

\[ h_i(t \mid Y_i(t), w_i) = h_0(t)R_i(t) \exp(\gamma^T w_i + \alpha y_i(t)) \]

• Where
  o \( R_i(t) \) is the at-risk indicator
  o \( \exp(\alpha) \) is relative risk translating a one unit increase in \( y_i(t) \) at the same time point

• Parameters can be estimated based on partial likelihood
  o Assumes no measurement error
  o Step-function path
  o Existence of covariate not related to failure status

• Extended Cox model valid ONLY for exogenous variables
  o Treating endogenous variables as exogenous may produce spurious results
Joint modeling framework

- Special feature of endogenous covariates leads to a new class of models

**Joint models for longitudinal and time-to-event data**

- Intuitive idea behind the standard joint model
  - Use an appropriate model to describe the evolution of the marker in time in each patient
  - The estimated evolutions are then used in a Cox model
Method 1: Shared Random Effect Model

• Assume a latent trajectory (typically referred to as “true biomarker trajectory”)

• Build longitudinal model for biomarkers given true trajectory

• Build survival model depends on true trajectory

• Conditional independent processes given the true trajectory

• Requires knowledge of trajectory and association

• Typically flexible parametric modeling assumptions
Method 1: Shared Random Effect Model

• Assume a true and unobserved biomarker trajectory:

\[ M_i(t) = \{m_i(s), 0 \leq s < t\} \] longitudinal history

• Define relative risk model

\[ h_i(t \mid M_i(t)) = h_0(t) \exp(\gamma^T w_i + \alpha m_i(t)) \]

• Define mixed effects model

\[ y_i(t) = m_i(t) + \epsilon_i(t) = x_i(t)^T \beta + z_i(t)^T b_i + \epsilon_i(t) \]
Method 2: Random Sampling Method

- Use random EMA design instead of modeling assumptions
- Construct a design-unbiased estimator of the cumulative hazard
- Requires knowledge of the EMA design
- Parametric modeling assumptions
- Quantifiable sources of variance
Simulation comparison

- Assume incorrectly specified biomarker model (linear when truth is quadratic)

<table>
<thead>
<tr>
<th>n</th>
<th>π</th>
<th>Bias</th>
<th>SD</th>
<th>CR</th>
<th>Bias</th>
<th>SD</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>16</td>
<td>0.009</td>
<td>0.062</td>
<td>95.5</td>
<td>0.092</td>
<td>0.065</td>
<td>71.4</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.011</td>
<td>0.068</td>
<td>95.3</td>
<td>0.099</td>
<td>0.067</td>
<td>68.9</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.011</td>
<td>0.072</td>
<td>94.4</td>
<td>0.102</td>
<td>0.068</td>
<td>71.0</td>
</tr>
<tr>
<td>100</td>
<td>16</td>
<td>0.024</td>
<td>0.110</td>
<td>93.7</td>
<td>0.104</td>
<td>0.117</td>
<td>89.0</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.033</td>
<td>0.119</td>
<td>94.3</td>
<td>0.112</td>
<td>0.119</td>
<td>88.2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.033</td>
<td>0.134</td>
<td>91.1</td>
<td>0.109</td>
<td>0.122</td>
<td>89.7</td>
</tr>
</tbody>
</table>
Assessing association

Decision making, stage 1
Research Question: Assessing association

• Claim: Biomarker is strongly associated with a particular negative behavioral event outcome
  o Craving $\mapsto$ Drug Lapse
  o Stress $\mapsto$ Smoking Lapse
  o Suicidal ideation $\mapsto$ Distress, Attempts

• In support of: interventions aimed at biomarker
  o Coping with Craving and Urges module $\mapsto$ reduce craving $\mapsto$ reduce risk of substance use
  o Reminder to practice mindfulness $\mapsto$ reduce stress $\mapsto$ reduce risk of lapse
  o Automated call to increase social support $\mapsto$ reduce loneliness $\mapsto$ reduce risk of STBs
## Case Study: Smoking Cessation Study

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Z</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (0,1)</td>
<td>0.2129</td>
<td>0.1706</td>
<td>1.25</td>
<td>0.2120</td>
</tr>
<tr>
<td>Craving (1-11)</td>
<td>0.4234</td>
<td>0.0273</td>
<td>15.51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Negative Affect (-2.31, 5.85)</td>
<td>0.4941</td>
<td>0.0714</td>
<td>6.92</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Restlessness (1-11)</td>
<td>0.0631</td>
<td>0.0285</td>
<td>2.21</td>
<td>0.0268</td>
</tr>
<tr>
<td>Attention Disturbance (1-11)</td>
<td>-0.0573</td>
<td>0.0410</td>
<td>1.40</td>
<td>0.1602</td>
</tr>
<tr>
<td>Positive Affect (-3.43, 2.84)</td>
<td>0.0856</td>
<td>0.0872</td>
<td>0.98</td>
<td>0.3266</td>
</tr>
</tbody>
</table>

Rathbun et al. (2014), J. Royal Stat Soc., Series C
Building predictions/classifications

Decision making, stage 2
Method 3: Leveraging behavioral latent constructs

1. Latent behavioral constructs
   - Measurement-error models to connect observables to the latent states
   - GLM: \( E[0_t^{(J)}] = g(S_t \beta) = g(S_t^{(1)} \beta_1 + S_t^{(2)} \beta_2) \)
   - Use anchor measurements to make latent constructs identifiable

2. Latent risk model
   - \( R_t = S_t^{(1)} \tilde{\beta}_1 + S_t^{(2)} \tilde{\beta}_2 \)

3. Latent risk connects to observed event outcome of interest
   - Probability of reporting drug use = expit \((R_t)\)
Case Study 2: Recovery Support Services

30 minute lapse probability

- High risk, low engagement
- High risk, high engagement
- Low risk, low engagement
- Low risk, high engagement
Comparison of prediction accuracy

Comparison of prediction accuracy

- sHMM
- Current O
- Event at prior time
- Event in prior day
- Event in prior week
- Complete

Brier Score

Window length (in days)
Sense²Stop MRT for Stress Management in Newly Abstinent Smokers

Every minute of every day starting with quit date

For hour* after intervention is delivered

Measured via EMA and puffMarker over 10 days

Observations
- stress (via AutoSense sensor suite)
- motion (via accelerometer)
- smoking (via self report)

Available? — YES — Is stressed? — NO — R — No Intervention

Available? — NO — No intervention

Prompt use of stress-management exercises

Proximal Outcome
Fraction of time stressed*

Distal Outcome
Relapse or smoking abstinence

*Adjusted for current talk
Assessing treatment effects

Decision making, stage 3
## Case Study 1: Smoking Cessation Study

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Z</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (0,1)</td>
<td>0.2129</td>
<td>0.1706</td>
<td>1.25</td>
<td>0.2120</td>
</tr>
<tr>
<td>Craving (1-11)</td>
<td>0.4234</td>
<td>0.0273</td>
<td>15.51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Negative Affect (-2.31, 5.85)</td>
<td>0.4941</td>
<td>0.0714</td>
<td>6.92</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Restlessness (1-11)</td>
<td>0.0631</td>
<td>0.0285</td>
<td>2.21</td>
<td>0.0268</td>
</tr>
<tr>
<td>Attention Disturbance (1-11)</td>
<td>-0.0573</td>
<td>0.0410</td>
<td>1.40</td>
<td>0.1602</td>
</tr>
<tr>
<td>Positive Affect (-3.43, 2.84)</td>
<td>0.0856</td>
<td>0.0872</td>
<td>0.98</td>
<td>0.3266</td>
</tr>
</tbody>
</table>

Rathbun et al. (2014), J. Royal Stat Soc., Series C
Current issues in JMs in mHealth

Translating JMs into mHealth research
Topic 1: Multiple noisy measurements

- **Smoking Episodes**
- **Smoking Puffs**

**PuffMarker**

**Random EMA**

**Event-based EMA**

**End-of-day EMA**

**Various Measures Of Smoking Episodes And Puffs**

Multiple imputation to produce smoking episodes with error

Event-based EMA:
- Random EMA
- Event-based EMA
- End-of-day EMA
Topic 2: Sensors + Event outcomes

• Participant wears Empatic E4 bracelets

• Measure various physiological responses (including EDA)

• Self-report via button press moments of distress

• Research question: how can we map these high-frequency sensor curves onto instantaneous risk?
  \[ h_i(t | M_i(\cdot)) = h_0(t) \exp \left( \gamma^T w_i + \int_{t-\Delta}^{t} m_i(u)\alpha(u)du \right) \]
Conclusion

• Focused on mHealth (mobile health) studies in which both longitudinal and time-to-event data are recorded per participant.

• Discussed how joint models enter into various stages of the intervention development process:
  o Assessing levels of biomarker association with event risk,
  o Defining risk strata for a stratified micro-randomized trial,
  o Post-study analysis of the treatment effect on event risk

• Discussed how mHealth studies present novel methodological challenges for joint modeling and solutions in several case studies.

• Data collection and analysis geared toward informing multi-stage decision making in mHealth.
The End

Email me with questions:

Walter: wdem@umich.edu

Data science for dynamic decision making (d3-) lab
University of Michigan

Susan Murphy (Harvard University)
Inbal Nahum-Shani
Danny Almirall
Ambuj Tewari
Method 2: Random Sampling Method

- Assuming EMAs are sent at random times; that is,
  \[
  \pi_{\theta}(t) = \lim_{\delta \to 0} \delta^{-1} P(N_{i}^{c}[0, t + \delta) - N_{i}^{c}[0, t) | H_t)
  \]
  is a known intensity function for sending EMAs.

- Random EMAs then \( \pi_{\theta}(t) \propto 1 \)

- Self-correcting point process (Isham and Wescott, 1970):
  \[
  \pi_{\theta}(t) = \exp(\alpha + \beta(t - \rho N_{i}^{c}[0, t)))
  \]
  corrects when far from target = \( t/\rho \).
Method 2: Random Sampling Method

- Recall the likelihood

\[ l_n(\theta) = \sum_{i=1}^{n} \Delta_i \log h_i(T_i; \theta) - H_i(T_i; \theta) \]

- Where

\[ H_i(T_i; \theta) = \int_0^{T_i} h_i(u; \theta) du \]

- **Key idea**: approximate cumulative hazard using random-design:

\[ \hat{H}_i(T_i; \theta) = \sum_{j=1}^{N_{i}(T_i)} h_i(u_{ij}; \theta) \pi_i(u_{ij})^{-1} \]
Case Study 3: Sense2Stop

- Participant wears Autosense chest band + sensors on each wrist
- Measure various physiological responses and body movements to robustly assess physiological stress.
- Pattern-mining algorithm uses the sensor data to construct a binary time-varying stress classification.
- Participant is then classified at each minute as either “Stressed” or “Does not qualify as Stressed.”
• Recovery Support Services (RSS) Studies of individuals with Substance Use Disorders (SUDs) (PI: Scott)
  o 2 pilot RSS studies on individuals who have been discharged from outpatient, intensive outpatient, or residential treatment
  o EMA + EMI usage data
• Smoking cessation study (PI: Shiffman)
  o N=412 volunteer smokers
  o Randomized at quit date N=188 to nicotine patch, N=136 to placebo
  o Random EMAs, no passive data
• Sense2Stop (PI: Spring)
  o Stratified micro-randomized trial
  o N=75 (expected)
  o Various EMA + sensor data;
  o Sequential randomization to reminder to practice mindfulness
• Suicidal Inpatient study (PI: Nock)
  o N=91
  o Wrist sensor (measures EDA and accelerometer)
  o User-initiated button press to denote times of distress (i.e., recurrent event outcome)
What can digital phenotyping tell us?

Mohr et al. (2017), Annual Reviews of Clinical Psychology
Latent trait shared-parameter mixed ecological momentary assessment

A shared parameter location scale mixed model for EMA data subject to informative missingness

John F. Cursio¹  |  Robin J. Mermelstein²  |  Donald Hedeke³

¹Department of Public Health Sciences, Biological Sciences Division, The University of Chicago, Chicago, Illinois
²Institute for Health Research and Policy, The University of Illinois at Chicago, Chicago, Illinois
³Correspondence
John F. Cursio, Department of Public Health Sciences, Biological Sciences Division, The University of Chicago, 5841 S. Maryland Ave, Room E317 MC320, Chicago, IL 60637.
Email: jfcursio@health.bsd.uchicago.edu

Funding Information
National Cancer Institute, Grant/Award Number: R01CA909262

Abstract
In this paper, we address the problem of account for the context of ecological momentary assessment studies (EMAs) longitudinal studies), where each study unit gets measured intermittently missing is usually present. We present a shared parameter model that links the primary longitudinal outcome with potentially common set of random effects that summarize subjects' mean (location) and variability (scale). The primary outcome effects are allowed to exhibit heterogeneity in terms of both variance. Unlike previous methods which largely rely on nonparametric estimation, we estimate the model by a full Bayesian approach. Carlo. An adolescent mood study example is illustrated together. Studies. Results in comparison to more conventional approaches for the common but unobserved random subject mean and variance also provide the benefit of understanding how missingness primary outcome.

Keywords
Ecological momentary assessments · Informative missing · Mixed effects 
Shared parameter model

Binary variable multiple-model multiple imputation to address missing data mechanism uncertainty: application to a smoking cessation trial

Juned Siddique,²†  |  Ofer Harel, ¹  |  Catherine M. Crespi²  |  Donald Hedeke³

¹Department of Public Health Sciences, Biological Sciences Division, The University of Chicago, Chicago, Illinois
²Institute for Health Research and Policy, The University of Illinois at Chicago, Chicago, Illinois
³Correspondence
Donald Hedeke, Department of Public Health Sciences, Biological Sciences Division, The University of Chicago, 5841 S. Maryland Ave, Room E317 MC320, Chicago, IL 60637.
Email: donald.hedeke@uic.edu

Funding Information
National Cancer Institute, Grant/Award Number: R01CA909262

Abstract
In this paper, we address the problem of account for the context of ecological momentary assessment studies (EMAs) longitudinal studies), where each study unit gets measured intermittently missing is usually present. We present a shared parameter model that links the primary longitudinal outcome with potentially common set of random effects that summarize subjects' mean (location) and variability (scale). The primary outcome effects are allowed to exhibit heterogeneity in terms of both variance. Unlike previous methods which largely rely on nonparametric estimation, we estimate the model by a full Bayesian approach. Carlo. An adolescent mood study example is illustrated together. Studies. Results in comparison to more conventional approaches for the common but unobserved random subject mean and variance also provide the benefit of understanding how missingness primary outcome.

Keywords
binary data · NMAR · nonignorable · not missing at random

References
Received: 16 March 2018 / Revised: 20 May 2018 / Accepted: 29 May 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018
Topic 4: Multiple treatment effects

$$\log h_i(u; \theta) = \beta_0 + \int_{u-c}^{u} a_i(t)w(u - t)dt + \cdots$$
Topic 5: EMA + Event outcomes

- Participant wears Empatic E4 bracelets
- EMA question on suicidal ideation also asked 5-times per day
- Research question: can we disentangle “actual risk” from the risk due to reminder based on EMA and study fatigue?
Randomization formula with soft-budget constraint

- Assume the randomization probability formula, $P(A_t = 1 \mid H_t) = p_t(H_t)$, is known.
Markov dynamics for latent constructs

1. Partially observable Markov decision process

2. Time-varying exogenous process affects risk (e.g., weather)

3. Actions can impact outcome directly and indirectly

4. Latent state can account for impact of continuing impact of actions on risk