Modeling time-varying behavioral variables in HIV prevention trials

James Y. Dai

Fred Hutchinson Cancer Research Center

2012 Atlantic Causal Inference Conference
Features of HIV prevention trials

Healthy, sexually active participants

- typically placebo controlled and double blind, but some open label
- time-to-HIV acquisition endpoint, quarterly or monthly visits
- extensive sexual risk taking data and product use data, collected longitudinally
- the prevention effect depends on sexual risk taking and adherence

Secondary analysis involves adjusting for sexual risk taking and adherence
Challenges when modeling behavioral data in prevention trials

- post-randomization selection bias:
  - risk compensation/behavioral disinhibition (open label)
  - differential adherence because of adverse effect (blind trials),

- measurement error, over-report of adherence, biomarker

- longitudinal covariates with time-to-event endpoint for causal inference
HPTN 035 trial

- HPTN 035 was a multi-site phase 2B trial that tested the safety and effectiveness of two vaginal microbicide gel:
  - BufferGel: maintain normal PH
  - 0.5% PRO 2000 Gel: inhibit viral attachment and entry
  - Placebo Gel
  - No-Gel

- The motivation is to let women have self-initiated prevention method,

- topical use of gel; before sex act

- Rationale of adding the No-Gel arm; but it is unblinded
Differential condom use

- All four arms receive standard treatment: condom counseling and provision
- Differential condom-use post treatment assignment

South Africa – Durban

- No−Gel arm
- Three gel arms

Quarterly Visit

% women report condom use in the last sex

1 2 3 4 5 6 7 8 9 10

James Y. Dai

Modeling time-varying behavioral data
Because male partners know whether women are getting a gel or not, they could change sexual risk taking - self-reported condom use in last sex act

**perceived reduction in risk → altered behavior**

- Consistent use of condoms offers over 90% protection

- Treatment (gel) effect may be offset by decreased condom-use!
### ITT analyses

#### Hazard Ratio vs Placebo Gel (95% CI), p value

<table>
<thead>
<tr>
<th></th>
<th>Hazard Ratio (95% CI), p value</th>
<th>Hazard Ratio (95% CI), p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO 2000 Gel</td>
<td>0.70 (0.46, 1.08), p=0.10</td>
<td>0.67 (0.44, 1.02), p=0.06</td>
</tr>
<tr>
<td>BufferGel</td>
<td>1.10 (0.75-1.62), p=0.63</td>
<td>1.05 (0.72-1.55), p=0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BufferGel</th>
<th>PRO 2000</th>
<th>Placebo</th>
<th>No Gel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>71.8%</td>
<td>71.9%</td>
<td>71.3%</td>
<td>80.7%</td>
</tr>
</tbody>
</table>

How to adjust for differential condom-use?
Related works in modeling longitudinal post-randomization variables

For the MIRA trial, Rosenblum et al (JRSSA 2009) estimated the direct effects of assignment to the diaphragm arm.

- “controlled” direct effect if participants in both arms are constrained to use condoms consistently, or not use at all.
- Assume sequential ignorability that conditional on covariates, condom use can be viewed as if “randomized”.
- Inverse probability weighted estimators
Our thinking of this problem

- The ability to manipulate behavioral intermediates, such as condom use, is limited.

- The behavior of condom-use is not well understood, nor it is well predicted.

- There are subgroups in the population who do not use condom. This is precisely the reason to develop microbicide for women,

James Y. Dai
Modeling time-varying behavioral data
Principal Stratification at a single time point

<table>
<thead>
<tr>
<th>principal stratum</th>
<th>$U$</th>
<th>$X(0)$</th>
<th>$X(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Condom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Risk Gambling</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>High Condom</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- assume a monotonicity assumption
- Risk gambling stratum drives the phenomena of risk disinhibition
- estimate the principal effect within low condom user, and the effect in high condom user.

How to model longitudinal condom use by principal stratification?
Our approach

- The number of possible strata is $3^k$, $k$ is the number of visits.

- Average condom use seems stable over time, but women could switch condom use status.

- We put constrait on time-varying strata by a partially Hidden Markov Model (pHMM).
Our approach

- The number of possible strata is $3^k$, $k$ is the number of visits.
- Average condom use seems stable over time, but women could switch condom use status.
- We put constraint on time-varying strata by a partially Hidden Markov Model (pHMM).

Related work: Lin, Ten Have and Elliott (JASA 2008)

- Longitudinal adherence and longitudinal outcomes.
- The large number of principal strata were collapsed into superclasses in which principal effects were defined.
- Bayesian MCMC to fit the model.
for subject $i$ in $1 : n$

$Z_i$ treatment assignment 0, 1

$a_j$ discrete time for visits, $j = 1, \ldots J$

$T_i$ time to HIV infection

$C_i$ time to censor

$Y_i = \min(T_i, C_i)$ observed infection-free time

$\Delta_i = 1_{[T_i < C_i]}$ Indicator of censoring

$X_{ij}$ condom use in last sex act in the $j^{th}$ visit

$\bar{X}_{ij} = \{X_{i1}, \ldots, X_{ij}, a_j = Y_i\}$ the observed condom-use history

$W_i$ Baseline Covariates

Observed data:

$$(Z_i, Y_i, \Delta_i, \bar{X}_{ij}, W_i).$$
Principal stratification on longitudinal condom-use

latent pattern of potential condom use

\[
\bar{U}_{ij} = \left( \begin{array}{c}
X_{i1}(0) & X_{i2}(0) & \cdots & X_{iJ}(0) \\
X_{i1}(1) & X_{i2}(1) & \cdots & X_{iJ}(1)
\end{array} \right).
\]

Consider \( \{U_{ij}, j = 1, \ldots, J\} \) a sequence of sexual risk taking behavior under either treatment assignment.

Table: At \( j^{th} \) visit

<table>
<thead>
<tr>
<th>principal state</th>
<th>( U_{ij} )</th>
<th>Condom-only</th>
<th>Gel + Condom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low - Condom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Risk Gambling</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>High-Condom</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Cumulative relative risk (CRR)

\[
\text{CRR}(j)^1 = \frac{1 - \Pr(T_i(1) > a_j|\bar{U}_{ij} = 1)}{1 - \Pr(T_i(0) > a_j|\bar{U}_{ij} = 1)}
\]

\[
\text{CRR}(j)^3 = \frac{1 - \Pr(T_i(1) > a_j|\bar{U}_{ij} = 3)}{1 - \Pr(T_i(0) > a_j|\bar{U}_{ij} = 3)}
\]

For women whose partner uses condom all the time regardless treatment assignment, how much protection she will get if she was assigned to PRO2000 arm as compared to No-Gel arm?

For women whose male partner never use condom regardless treatment assignment, how much protection she will get if she was assigned to PRO2000 arm as compared to No-Gel arm?
$U_{ij}$ is a partially observed, latent process that determines $X_{ij}$ and at-risk probability.

The Markov chain stops when HIV +, or censoring.
HMM assumptions

- Markov property on the partially hidden states $U_{ij}$

$$
\Pr(U_{ij} = s | U_{ij-1} = r, ..., U_{i1}) \\
= \Pr(U_{ij} = s | U_{ij-1} = r) \\
= q_{sr},
$$

- Conditional independence

$$
\Pr(T_i(z) = a_j | T_i(z) \geq a_j, U_{ij}, U_{ij-1}, ..., U_{i1}) \\
= \Pr(T_i(z) = a_j | T_i(z) \geq a_j, U_{ij}),
$$

so that

$$
\Pr(T_i(z) > a_j | \bar{U}_{ij}) = \prod_j [1 - \Pr(T_i(z) = a_j | T_i(z) \geq a_j, U_{ij})]
$$
Checking Markov assumption

partially HMM, some states are observable!

Condom to Condom in the gel arms

No-condom to No-condom in the no-gel arm
additional standard assumptions

- randomization

\[ Z \perp (\bar{U}_i, T_i(0), T_i(1), C_i(0), C_i(1), Y_i(0), Y_i(1), \Delta_i(0), \Delta_i(1)) \mid W_i \]

- SUTVA (Stable Unit Treatment Value Assumption)

- Independent censoring
Identifiability of parameters

- denote $\Theta$ the set of parameters in HMM
  - $p_s$ the initial probability for state $s$
  - $q_{st}$ the transition matrix
  - $\pi_{sj}$ the probability of HIV acquisition for $s$ state at $a_j$
  - $\beta_s$ the treatment effect for $s$ state

- denote the $\Omega$ the set of parameters that describe the observed process $X_{ij}$ and $Y_i|X_{ij}$

- $\Omega$ is reparameterization of $\Theta$, and $\Omega$ is identifiable
\[
\prod_{i=1}^{n} \left[ f \Pr(Y_i, \Delta_i, \bar{X}_i1_{[a_j < Y_i]} | Z_i, W_i) \right]^{Z_i} \left[ (1 - f) \Pr(Y_i, \Delta_i, \bar{X}_i1_{[a_j < Y_i]} | Z_i, W_i) \right]^{1 - Z_i}
\]

\[
= \prod_{i=1}^{n} \left[ f \sum_{\bar{U}_i \in \bar{X}_i} \Pr(Y_i, \Delta_i | \bar{U}_i, Z_i, W_i) \Pr(\bar{U}_i | Z_i, W_i) \right]^{Z_i}
\]

\[
[ (1 - f) \sum_{\bar{U}_i \in \bar{X}_i} \Pr(Y_i, \Delta_i | \bar{U}_i, Z_i, W_i) \Pr(\bar{U}_i | Z_i, W_i) ]^{1 - Z_i}
\]

\[
\propto \prod_{i=1}^{n} \left\{ f \sum_{\bar{U}_i \in \bar{X}_i} \left[ \prod_{j=1}^{J_i - 1} (1 - \lambda_{ij}(1)) \lambda_{ij}(1) \right]^{\Delta_i} \left[ \prod_{j=1}^{J_i} (1 - \lambda_{ij}(1)) \right]^{1 - \Delta_i} p_{U_{i1}} \prod_{j=2}^{J_i} q_{U_{ij}U_{ij-1}} \right\}^{Z_i}
\]

\[
\times \left\{ (1 - f) \sum_{\bar{U}_i \in \bar{X}_i} \left[ \prod_{j=1}^{J_i - 1} (1 - \lambda_{ij}(0)) \lambda_{ij}(0) \right]^{\Delta_i} \left[ \prod_{j=1}^{J_i} (1 - \lambda_{ij}(0)) \right]^{1 - \Delta_i} p_{U_{i1}} \prod_{j=2}^{J_i} q_{U_{ij}U_{ij-1}} \right\}^{1 - Z_i}
\]
- modified Baum-Welch (backward-forward) algorithm

- baseline predictors are incorporated in both Markov-state and disease risk models

- bootstrap to get the variance estimate
Model diagnostics

The proportion of condom use in the last sex over visits.

- No-treatment arm: observed
- Gel arms: observed
- No-treatment arm: estimated
- PRO2000 arm: estimated

The cumulative probability of HIV infection over visits.

- No-treatment arm: observed
- PRO2000 arm: observed
- No-treatment arm: estimated
- PRO2000 arm: estimated

Modeling time-varying behavioral data
The ITT effect of PRO/2000 gel assignment compared to the No-gel assignment

<table>
<thead>
<tr>
<th>principal stratum</th>
<th>estimate</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low - Condom</td>
<td>0.56</td>
<td>[0.16, 2.04]</td>
</tr>
<tr>
<td>Risk Gambling</td>
<td>1.02</td>
<td>[0.22, 4.70]</td>
</tr>
<tr>
<td>High - Condom</td>
<td>0.70</td>
<td>[0.45, 1.09]</td>
</tr>
</tbody>
</table>
Baseline distribution of condom-use states

<table>
<thead>
<tr>
<th></th>
<th>Low-Condom</th>
<th>Risk Gambling</th>
<th>High-Condom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Condom</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Gambling</td>
<td></td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>High-Condom</td>
<td></td>
<td></td>
<td>0.70</td>
</tr>
</tbody>
</table>

Transition matrix of Markov states

<table>
<thead>
<tr>
<th></th>
<th>Low-Condom</th>
<th>Risk Gambling</th>
<th>High-Condom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Condom</td>
<td>0.532</td>
<td>0.270</td>
<td>0.198</td>
</tr>
<tr>
<td>Risk Gambling</td>
<td>0.282</td>
<td>0.003</td>
<td>0.715</td>
</tr>
<tr>
<td>High-Condom</td>
<td>0.084</td>
<td>0.090</td>
<td>0.826</td>
</tr>
</tbody>
</table>
We proposed a principal stratification approach to adjust for time-varying post-treatment behavior.

Using HMM models, we exploits the pattern of behavioral risk taking, and the transient nature of prevention effect.

limitation: condom-use is self reported.

could be extended to deal with longitudinal compliance
Aknowledgement

Peter Gilbert
Ben Mâsse
Jim Hughes