We will start by motivating the methods presentation with challenging problems found in the study of HIV.

Study Objective
To determine the impact of multiple risks on the overall health status of HIV+ homeless and unstably housed adults

Potential Challenges
- HIV is increasingly characterized as a chronic condition that can be managed through adherence to a healthy lifestyle, complex drug regimens and treatment monitoring;
- however, social and structural factors can be significant determinants of an individual’s ability to meet these requirements and achieve better overall health status.
- The effects of exposures that change over time and influence one another, such as drug use and housing status, now have the opportunity to influence a longer disease course.
Outcome of Interest

- Overall Physical Health Status (SF-36)
- Overall Mental Health Status (SF-36)

Exposures of Interest

- Age, race, education
- Employment, income
- Subsistence needs (housing, food, clothing, hygiene needs)
- Incarceration
- Drug use, alcohol use
- Victimization, social support
- Insurance status
- Adherence to antiretroviral therapy
- CD4 cell count, viral load

Physical Health Status (N=288)

<table>
<thead>
<tr>
<th>Main Effect</th>
<th>Adjusted Population Effect</th>
<th>Adjusted 95% Confidence Interval</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmet subsistence needs</td>
<td>-3.83</td>
<td>(-5.27, -1.56)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Caucasian race/ethnicity</td>
<td>-3.71</td>
<td>(-6.03, -1.39)</td>
<td>.002</td>
</tr>
<tr>
<td>No source of instrumental support</td>
<td>-1.56</td>
<td>(-2.88, -0.24)</td>
<td>.020</td>
</tr>
<tr>
<td>Viral load</td>
<td>-0.00018</td>
<td>(-0.00038, -0.000003)</td>
<td>.0410</td>
</tr>
</tbody>
</table>

Mental Health Status (N=288)

<table>
<thead>
<tr>
<th>Main Effect</th>
<th>Adjusted Population Effect</th>
<th>Adjusted 95% Confidence Interval</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmet subsistence needs</td>
<td>-3.51</td>
<td>(-5.08, -1.94)</td>
<td>.00036</td>
</tr>
<tr>
<td>Close friend/Confidant</td>
<td>3.19</td>
<td>(1.54, 4.83)</td>
<td>.00045</td>
</tr>
<tr>
<td>Any drug use</td>
<td>-3.67</td>
<td>(-5.53, -1.81)</td>
<td>.0052</td>
</tr>
<tr>
<td>No Sources of Instrumental Support</td>
<td>-2.2</td>
<td>(-3.62, -0.81)</td>
<td>.0312</td>
</tr>
<tr>
<td>&gt;90% ART Adherence</td>
<td>1.66</td>
<td>(0.07, 3.27)</td>
<td>.043</td>
</tr>
</tbody>
</table>

Cautionary Note

- Analyses do not necessarily indicate the highest health priorities for specific individuals; instead they indicate exposures with the largest population-level effects on the health of unstably housed HIV-positive adults and that the biggest population-wide impact on health would be made by focusing on these issues.

Causal Modeling

Let's Take Step Back...
When we ask scientific questions, we frequently collect data in an attempt to answer these questions. We are often, but not always, interested in causal effects. We prefer not to merely conclude that there is an association or correlation between two variables. Instead, we want to know that \( A \) causes \( Y \).

**What is the Goal of the Study?**

Key Take Home Message:

* Causal assumptions allow us to interpret the parameter of interest as a causal effect.
  1. These additional assumptions are untestable; we cannot use the data to verify their accuracy.
  2. The causal modeling assumptions are separate from the estimation procedure.
  3. If we choose not to make causal assumptions, perhaps because we know they do not hold, the parameter has a statistical interpretation, just not a causal one.

**Causal Modeling**

*The Neyman-Rubin Causal Model assumes:*

  * No unmeasured confounders
  * Consistency
  * No interference (the counterfactual outcome of one subject should not be affected by the treatment assignment of other subjects)

**Multiple Frameworks**

Causal Modeling

Structural Causal Models of Pearl:

  * Describes each endogenous variable \( X \) as a deterministic function of other endogenous variables and an exogenous error. (The errors are never observed.)
  * For each \( X \), the deterministic function depends on the other endogenous variables only through the parents of \( X \).
  * The exogenous variables have a particular joint distribution.

**Causal Modeling**

We could specify the following SCM:

\[
W = f_W(U_W),
A = f_A(W, U_A),
Y = f_Y(W, A, U_Y).
\]

Recall that we assume for the full data:

- for each \( X_i \), \( X_i = f(Pa(X_i), U_{X_i}) \) depends on the other endogenous variables only through the parents \( Pa(X_i) \).
- the exogenous variables have a particular joint distribution \( P_U \).

**Causal Modeling**

[Figure: Causal graphs with various assumptions about the distribution of \( P_U \)]
What is a Causal Effect?

- How would outcomes change in the population under different exposures/treatments?
- In the SCM framework, we want to know what happens when we intervene on the system to, for example, set $A=a$.

  Recall Elise’s Cautionary Note....

Cautionary Note

- Analyses do not necessarily indicate the highest health priorities for specific individuals; instead they indicate exposures with the largest population-level effects on the health of unstably housed HIV-positive adults and that the biggest population-wide impact on health would be made by focusing on these issues.

Why Not Randomized Studies?

- Ethical issues
- Time
- Cost
- Randomization may not occur perfectly

Data

- Our study is an experiment where we draw a random variable from our population $n$ times.
- The data we observe are realizations of these $n$ random variables, and the random variables have an underlying probability distribution.

Data

- Formally, the data consists of $n$ i.i.d. copies of random variable $O \sim P$, where $P$ is the true underlying probability distribution for $O$.
- We’ll start with the same simple case used for the SCM, where $W$ is a vector of baseline variables, $A$ is an intervention, and $Y$ is an outcome.
- $O=(W,A,Y) \sim P$.

Statistical Model

- A statistical model represents the set of possible probability distributions of the data.
- You are likely familiar with parametric statistical models, where one assumes that the probability distribution underlying the data is known (up to a certain number of parameters).
- You can also assume nonparametric and semiparametric models.
The target parameter of interest will depend on your scientific question. One simple parameter is the risk difference.

\[ \psi_{RO} = \psi(P) = E[E(Y \mid A = 1, W) - E(Y \mid A = 0, W)] = E(Y_1) - E(Y_0) = P(Y_1 = 1) - P(Y_0 = 1) \]

In HIV research, we frequently address complicated research questions that require more complex parameters.

Marginal Structural Models

Marginal structural models (MSMs) are a useful tool to describe additional parameters.

MSMs are not estimators.

MSMs are simply a way to define parameters.

Example (Effect Modification):
One may be interested in the treatment-specific mean of an outcome conditional on a particular baseline covariate. Now we have a treatment effect that is a function of a baseline covariate.

We could use an MSM to define such a parameter:

\[ E(Y \mid V) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV, \]

with effect modifier \( V \) and a continuous \( Y \).


Example (High-Dimensional Treatment):
What if we are interested in the effect of a continuous treatment?

Recall our risk difference parameter...

\[ \psi_{RD} = \psi(P) = E(Y | A = 1, W) - E(Y | A = 0, W) = E(Y_1) - E(Y_0) = P(Y_1 = 1) - P(Y_0 = 1) \]

**Marginal Structural Models**

**Example (High-Dimensional Treatment):**
What if we are interested in the effect of a continuous treatment?
We could use an MSM to define such a parameter:
\[ E(Y_j) = \beta_0 + \beta_1 a_j \]
with continuous outcome \( Y \).


**Marginal Structural Models**

**Example (Dynamic Treatment Regimes):**
What if we are interested in the effect of a particular “rule” for assigning the intervention in response to baseline or intermediate variables?
We could use an MSM to define such a parameter.


**Marginal Structural Models**

**Examples (Dynamic Treatment Regimes):**
- Understanding adherence to combined antiretroviral therapy in HIV treatment


**Examples (Dynamic Treatment Regimes):**
- When to initiate combined antiretroviral treatment in therapy-naïve HIV-infected person

MSMs: Dynamic Regimes

- Consider data structure \( O=(L_0, A_0, L_1, A_1, L_2=Y) \)
- Let \( D=(d_1, \ldots, d_K) \) be the set of dynamic regimes we consider. These dynamic regimes define a set of rules for guiding intervention \( A(t) \) at each time point based on previous covariates and prior interventions.
- Thus, each rule \( d \) takes as input previous covariates and interventions to assign \( a(t) \).
One could estimate the regime-specific mean – the population mean of $Y$ had everyone followed regime $d$, and then do this for each of the $k$ regimes.

However, suppose we have a large number of regimes, or that we have a small number of subjects. In these cases, we may wish to smooth across regimes to obtain a summary measure.

We could define an MSM $m_{\beta}$, which is a known function of our parameter $\beta=\{\beta_0, \beta_1, \ldots, \beta_{k-1}\}$.

For example:

$$m_{\beta}=\logit^{-1}(\beta_01_{a=d_1} + \beta_11_{a=d_2} + \ldots + \beta_{k-1}1_{a=d_{k-1}})$$

We can also consider nonsaturated MSMs, working MSMs, as well as MSMs that include baseline covariates.

Assume a model $m_{\psi}$ for the parameter $\psi(a)$, for example:

$$\psi(a)=\logit^{-1}(\beta(a))$$

Here, we focus on estimating $\beta$, but have forced ourselves to make restrictive modeling assumptions that may not be true.

This is called a “working” marginal structural model.

If helps us define a parameter that allows for smoothing, but does not represent an additional statistical or causal assumption.

Estimation Approaches

a. Maximum-likelihood-based estimators
   - G-computation
     - requires estimate of outcome regression at each t
   - Targeted maximum likelihood estimation (TMLE)
     - requires estimate of outcome regression and treatment mechanism at each t

b. Estimating equation Estimators
   - Inverse probability weighted estimators (IPW)
     - requires estimate of treatment mechanism at each t
   - Augmented inverse probability weighted estimators (A-IPW)
     - requires estimate of outcome regression and treatment mechanism at each t


Study Objective

To determine the impact of multiple risks on the overall health status of HIV+ homeless and unstably housed adults.

Study Objective

Used marginal structural models to define
\[ \mathbb{E}(Y(a)) \mid m_1 = \beta_0 + \beta_1 a(t-1) \]
for their continuous outcome variables Y.

The authors used targeted maximum likelihood estimation to estimate the effect of each exposure, which led to a ranked list of “variable importance measures.”

References

- Riley et al. (2011) Population-level effects of uninterrupted health insurance on services use among HIV-positive unstably housed adults. AIDS Care, 23(7): 822-830.
- Rose & van der Laan, Chapter 2. (For introduction to causal assumptions)
- Rose & van der Laan, Chapter 6. (For comparison of different estimators)
- Rosenblum, Chapter 9. (For chapter on MSMs we reference in this talk)

Please feel free to email (srose@jhsph.edu and eriley@epi-center.ucsf.edu) for PDFs of papers or book chapters.