Constructing Inverse Probability Weights for Static Interventions

Kunjal Patel, DSc MPH
Senior Research Scientist
Harvard T.H. Chan School of Public Health
Acknowledgement

- Slides contributed by Miguel Hernán, Ellie Caniglia, or adapted from *Causal Inference* (Chapman & Hall/CRC, 2017) by Miguel Hernán and Jamie Robins
- Any mistakes are my own


- You can “like” Causal Inference at [https://www.facebook.com/causalinference](https://www.facebook.com/causalinference)
Summary of day 1

- Well-defined intervention
- Static vs. dynamic interventions
- Definition of an average causal effect
- Why is randomization important?
- Conditional exchangeability assumption to identify a causal effect
- When standard adjustment methods fail
- IP weights for treatment
Formulation of a well-defined study question

- Well-defined causal inference questions can be mapped into a target trial

- Specify the protocol of the target trial including:
  - Eligibility criteria
  - Treatment strategies
  - Randomized treatment assignment
  - Follow-up period
  - Outcome
  - Causal contrast of interest
  - Analysis Plan

Classification of sustained treatment strategies

- **Static**
  - a fixed strategy for everyone
  - *Example:* treat with 150mg of daily aspirin during 5 years
  - *Case example:* initiate HAART

- **Dynamic**
  - a strategy that assigns different values to different individuals as a function of their evolving characteristics
  - *Example:* start aspirin treatment if coronary heart disease, stop if stroke
  - *Case example:* initiate HAART if CD4 drops below 500 cells/mm³
Definition of an average causal effect

- Each person has two counterfactual outcomes:
  - Outcome Y if treated - $Y_{i,a=1}$
  - Outcome Y if untreated – $Y_{i,a=0}$

- Individual causal effect:
  - $Y_{i,a=1} \neq Y_{i,a=0}$
  - Cannot be determined except under extremely strong assumptions

- Average (population) causal effect:
  - $E[Y_{a=1} = 1] \neq E[Y_{a=0} = 1]$
  - Can be estimated under:
    - No assumptions (ideal randomized experiments)
    - Strong assumptions (observational studies)
Why is randomization important?

- When group membership is randomly assigned, risks are the same.
- Both groups are comparable or exchangeable.
- Exchangeability is the consequence of randomization.
Within levels of the covariates, L, exposed subjects would have had the same risk as unexposed subjects had they been unexposed, and vice versa.

Counterfactual risk is the same in the exposed and the unexposed with the same level of L.

$$\Pr[Y_a=1 \mid A=1, L=l] = \Pr[Y_a=1 \mid A=0, L=l] \iff A \perp Y_a \mid L=l \iff Y_a \perp A \mid L=l$$

Equivalent to randomization within levels of L.

Implies no unmeasured (residual) confounding within levels of the measured covariates L.
Methods to compute causal effects

- Stratification
- Regression
- Matching
- Standardization
- Inverse probability weighting

⇒ ALL assuming conditional exchangeability
Choice of method depends on type of strategies

- Comparison of strategies involving point interventions only
  - All methods work
  - if all baseline confounders are measured

- Comparison of sustained strategies
  - Generally only causal inference methods work
  - Time-varying treatments imply time-varying confounders
    - possible treatment-confounder feedback
  - Conventional methods may introduce bias even when sufficient data are available on time-varying treatments and time-varying confounders
Problem with stratified analytic approach

- Interested in the cumulative effect of treatment.
- $L_1$ is affected by $A_0$; if $A_0$ is adjusted for, then $L_1$ is confounded.
- Also could induce selection bias (collider).
Numerator: The probability that the subject received his/her observed treatment at week k, conditional on past treatment history and baseline covariates.

Denominator: The probability that the subject received his/her own observed treatment at week k, given past treatment history and covariate history (baseline and time-dependent).
Directed Acyclic Graph in pseudopopulation with SW
Estimating IPW and fitting the MSM

- Estimate SW for both treatment and censoring:
  - Fit logistic regression models for treatment and censoring
  - Use predicted values from the models to calculate stabilized weights

- Estimate the IPW estimate of HAART on mortality:
  - Fit weighted pooled logistic model using the estimated stabilized weights.
  - Use “robust” variance estimators (GEE) to allow for correlated observations created by weighting – conservative 95% CI.
IPW for Selection Bias
Atazanavir exposure in utero and neurodevelopment in infants: a comparative safety study

Ellen C. Caniglia\textsuperscript{a}, Kunjal Patel\textsuperscript{a}, Yanling Huo\textsuperscript{a}, Paige L. Williams\textsuperscript{a}, Suad Kapetanovic\textsuperscript{b,c}, Kenneth C. Rich\textsuperscript{d}, Patricia A. Sirois\textsuperscript{e}, Denise L. Jacobson\textsuperscript{a}, Sonia Hernandez-Diaz\textsuperscript{a}, Miguel A. Hernán\textsuperscript{a,f}, George R. Seage III\textsuperscript{a}, for the Pediatric HIV/AIDS Cohort Study

\textit{AIDS} 2016, 30:1267–1277
Introduction/background

- The use of antiretroviral drugs (ARVs) during pregnancy has dramatically decreased the incidence of perinatal transmission of HIV.

- The effects of in utero exposure to ARVs on neurodevelopment in perinatally HIV-exposed but uninfected (PHEU) infants requires further study.

- Previous research evaluating developmental outcomes in PHEU infants identified atazanavir as a safety concern.

- A comparative safety study was needed to confirm these findings.
Objective

- To evaluate the effect of in utero exposure to ARV regimens containing atazanavir compared to non-atazanavir-containing regimens on neurodevelopment at 9-15 months of age

- using observational data from a cohort of PHEU infants

- with a comparative safety design
Study population

- SMARTT protocol of PHACS
- Pregnant women living with HIV enrolled in the dynamic cohort
  - Not on ARVs at their last antepartum menstrual period
  - Initiated ARVs during pregnancy
- Excluded sites in Puerto Rico
- Excluded if infant less than 15 months of age by July 1, 2014
Exposure ascertainment

- Start ARVs Trimester 1
  - Atazanavir regimen
  - Non-atazanavir regimen

- Start ARVs Trimester 2-3
  - Atazanavir regimen
  - Non-atazanavir regimen
Outcome ascertainment

- Bayley Scales of Infant and Toddler Development – Third Edition (Bayley-III)
  - Administered at 9-15 months of age
  - Only available in English
  - Provides 5 scores:
    - Cognitive
    - Language
    - Motor
    - Social-emotional
    - General adaptive
Secondary outcomes

- Neonatal outcomes
  - Low birth weight (≤2500 grams)
  - Gestational age
  - Prematurity (gestational age <37 weeks)
  - Neonatal hearing

- Head circumference z-scores at 9-18 months
Analysis

- Conducted separately for each of the five Bayley-III domains

- Multivariable adjusted linear regression models
  - To estimate the mean difference in each domain score comparing atazanavir-containing to non-atazanavir-containing regimens
  - Estimated separately by trimester of ARV initiation
  - Adjusted for baseline maternal characteristics
    - maternal education, CD4 cell count, HIV RNA, calendar year, race, ethnicity, language spoken at home, income, age, maternal Full Scale Intelligence Quotient, and maternal illicit substance, alcohol, and tobacco use
~40% had incomplete or invalid results for one or more Bayley-III domains

- 917 age-eligible for Bayley-III
- 678 attended 1 year study visit
- 575 at least one Bayley-III domain completed
Options for analysis

- Analyze observed non-missing outcome data
  - Any problems with this approach?

- 917 age-eligible for Bayley-III

- 678 attended 1 year study visit

- 575 at least one Bayley-III domain completed
Selection bias

- Bias that arises when the parameter of interest in a population differs from the parameter in the subset of individuals from the population that are available for analysis.

- Selection bias for descriptive measures (e.g., prevalence) because of non-random sampling.

- Selection bias for effect measures (e.g., causal risk ratio) because of differential loss to follow-up.
Selection bias for effect measures

- Differential loss to follow-up/censoring
- Missing outcome/Non-response
- Healthy worker bias
- Self-selection/volunteer bias
Structure of selection bias (under the null)

- Bias arises as the consequence of conditioning on a common effect of treatment and outcome
  - Or on a common effect of a cause of the treatment and a cause of the outcome
- That is, the design or the analysis is conditioned on “being selected for analysis” C=0
Is bias due to differential loss to follow-up possible in randomized experiments?

- Yes?
- No?
Aside: Is bias due to self-selection possible in randomized experiments?

- Yes?
- No?
Aside: Internal vs. external validity in randomized experiments

- **Internal validity**
  - the estimated association has a causal interpretation in the studied population
  - i.e., no selection bias, no confounding

- **External validity**
  - the estimated association has a causal interpretation in another population
  - i.e., generalized or transportability

- In randomized experiments
  - There is internal validity
  - Perhaps not external validity
Simplified case example

- HIV-exposed uninfected infants

- Variables:
  - A=1: In utero exposure to ATV
  - L=1: Low maternal CD4 count at delivery
  - C=1: Missing 1-year Bayley exam
  - Y=1: Neurocognitive deficit

- Treatment status randomized
  - No confounding

- Under the null: No effect of *in utero* ATV exposure and neurocognitive function
Case example: Directed Acyclic Graph

- Where:
  - L: Maternal CD4 count at delivery
  - A: Maternal exposure to ATV
  - C: Censored
  - Y: Neurocognitive deficit in infant at 1 year
  - U: Unmeasured covariate – Maternal underlying immune function
Problem with stratified approach to adjust for censoring?

- Conditioning on descendent of a common effect (collider)
  - Only including those with C=0 in analyses (non-missing data)
- Observe biased association between A and Y through $A \rightarrow L \leftarrow U \rightarrow Y$
Alternative structure of selection bias due to differential loss to follow-up/non-response or missing data

Where:
- \( L \): Smoking intensity at baseline
- \( A \): Smoking cessation
- \( C \): Censored
- \( Y \): Weight gain
- \( U \): Lifetime history of smoking

Stratified approach will not cause bias if measure and adjust for \( L \)
Approaches for adjustment for selection bias

- Stratification
- Regression
- Inverse probability weighting

Approach depends on the structure of selection bias
Simplified case example original data

- Among A=0: No in utero exposure to ATV
  - L=1: Low maternal CD4 count at delivery
  - C=1: Missing 1-year Bayley exam
  - Y=1: Neurocognitive deficit
Case example pseudopopulation

- Among $A=0$: No in utero exposure to ATV
  - $L=1$: Low maternal CD4 count at delivery
  - $C=1$: Missing 1-year Bayley exam
  - $Y=1$: Neurocognitive deficit

$W = 1/\Pr[C=0/A,L]$:
- $1/0.5=2$
- $1/0.5=2$
- $1/0.1=10$
- $1/0.1=10$
Directed Acyclic Graph in pseudopopulation
What is an assumption are we making?

- Conditional exchangeability
  - Average outcome in the uncensored participants is the same as the average outcome in the censored participants with the same values of A and L

\[ Y^{a,c=0} \perp C|A,L \quad \text{for } c=0 \]

- Or selection is randomized within levels of A,L
Use of models for IPW

- Reality is we deal with high-dimensional data with multiple covariates (Ls), some with multiple levels
  - Cannot obtain meaningful non-parametric estimates of the weights
  - Model the probability of being uncensored with Ls (and A) as the covariates
- Some individuals may contribute a really high weight due to their a relatively small probability of being uncensored given their exposure and covariate history
  - Stabilize the weights by using the probability of being uncensored given treatment and baseline covariates in the numerator
  - Apply stabilized weights (SW) to estimate the parameters of a marginal structural model
    - reduce variance in model for the outcome
Stabilized inverse probability of censoring weights

\[ SW(V) = \prod_{k=0}^{K} \frac{\Pr \{ C(k) = 0 / \bar{A}(k), V \}}{\Pr \{ C(k) = 0 / \bar{A}(k), \bar{L}(k) \}} \]

- Numerator: The probability that the subject was uncensored at week \( k \), conditional on past treatment history and baseline covariates.

- Denominator: The probability that the subject was uncensored at week \( k \), given past treatment history and covariate history (baseline and time-dependent).
Estimating IPW and fitting the MSM

- Estimate SW for censoring:
  - Fit logistic regression models for being uncensored
  - Use predicted values from the models to calculate stabilized weights

- Estimate the IPW estimate of *in utero* ATV exposure on neurocognitive scores at 1-year:
  - Fit weighted linear regression models using the estimated stabilized weights.
  - Use “robust” variance estimators (GEE) to allow for correlated observations created by weighting – conservative 95% CI.
Summary: IP weights

- To adjust for confounding
  - Use IP weights for treatment – IPTW

- To adjust for selection bias
  - Use IP weights for censoring – IPCW

- To adjust for both biases
  - Multiply IPTW x IPCW
Case Example: Predictors of Censoring

- Baseline covariates: maternal education, CD4 cell count, HIV RNA, calendar year, race, ethnicity, language spoken at home, income, age, maternal Full Scale Intelligence Quotient, and maternal illicit substance, alcohol, and tobacco use

- Post-baseline covariates: mother’s last CD4 in pregnancy, positive test for STI in pregnancy, infant low birth weight, and gestational age at delivery
Primary effect estimates of interest

- Effect of in utero ATV exposure during the 1\textsuperscript{st} trimester on the following Bayley scores:
  - Cognitive
  - Language
  - Motor
  - Social-emotional
  - General adaptive

- Effect of in utero ATV exposure during the 2\textsuperscript{nd}/3\textsuperscript{rd} trimester on the following Bayley scores:
  - Cognitive
  - Language
  - Motor
  - Social-emotional
  - General adaptive
Results
### Characteristics of Study Population

<table>
<thead>
<tr>
<th>Atazanavir-containing regimen (n=167)</th>
<th>Non-atazanavir-containing regimen (n=750)</th>
</tr>
</thead>
</table>

Results
## Characteristics of Study Population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Atazanavir-containing regimen (n=167)</th>
<th>Non-atazanavir-containing regimen (n=750)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARV initiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First trimester</td>
<td>55 (33%)</td>
<td>227 (30%)</td>
</tr>
<tr>
<td>Second or third trimester</td>
<td>112 (67%)</td>
<td>523 (70%)</td>
</tr>
</tbody>
</table>
### Characteristics of Study Population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Atazanavir-containing regimen (n=167)</th>
<th>Non-atazanavir-containing regimen (n=750)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARV initiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First trimester</td>
<td>55 (33%)</td>
<td>227 (30%)</td>
</tr>
<tr>
<td>Second or third trimester</td>
<td>112 (67%)</td>
<td>523 (70%)</td>
</tr>
<tr>
<td>Age</td>
<td>older (mean 29 years)</td>
<td>younger (mean 27 years)</td>
</tr>
<tr>
<td>Cognitive scores</td>
<td>lower (mean 84.3)</td>
<td>higher (mean 86.5)</td>
</tr>
<tr>
<td>Initiate ARVs 2011-2014</td>
<td>more likely (40%)</td>
<td>less likely (26%)</td>
</tr>
</tbody>
</table>
## Common Regimens

<table>
<thead>
<tr>
<th>Type of regimen</th>
<th>Number of initiators</th>
<th>Type of regimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atazanavir-containing regimens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atazanavir, emtricitabine, tenofovir, ritonavir</td>
<td>126 (75%)</td>
<td>Boosted PI with 2 NRTIs</td>
</tr>
<tr>
<td>Non-atazanavir-containing regimens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopinavir, zidovudine, lamivudine, ritonavir</td>
<td>335 (45%)</td>
<td>Boosted PI with 2 NRTIs</td>
</tr>
<tr>
<td>Zidovudine, lamivudine, abacavir</td>
<td>134 (18%)</td>
<td>3 NRTIs</td>
</tr>
</tbody>
</table>
## Bayley-III - First Trimester

<table>
<thead>
<tr>
<th>Domain</th>
<th>No. of infants</th>
<th>Adjusted mean difference (95% CI) comparing atazanavir regimens with non-atazanavir regimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>182</td>
<td>-1.50 (-6.20, 3.20)</td>
</tr>
<tr>
<td>Language</td>
<td>182</td>
<td>-3.30 (-7.64, 1.04)</td>
</tr>
<tr>
<td>Motor</td>
<td>181</td>
<td>-2.92 (-7.75, 1.90)</td>
</tr>
<tr>
<td>Social-Emotional</td>
<td>173</td>
<td>0.14 (-6.16, 6.43)</td>
</tr>
<tr>
<td>Adaptive Behavior</td>
<td>173</td>
<td>-0.13 (-4.31, 4.05)</td>
</tr>
</tbody>
</table>
### Bayley-III - Second/Third Trimester

<table>
<thead>
<tr>
<th>Domain</th>
<th>No. of infants</th>
<th>Adjusted mean difference (95% CI) comparing atazanavir regimens with non-atazanavir regimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>383</td>
<td>0.39 (-3.19, 3.96)</td>
</tr>
<tr>
<td>Language</td>
<td>379</td>
<td>-3.37 (-6.23, -0.51)</td>
</tr>
<tr>
<td>Motor</td>
<td>376</td>
<td>0.27 (-2.88, 3.41)</td>
</tr>
<tr>
<td>Social-Emotional</td>
<td>374</td>
<td>-5.86 (-9.44, -2.28)</td>
</tr>
<tr>
<td>Adaptive Behavior</td>
<td>380</td>
<td>-2.53 (-5.86, 0.80)</td>
</tr>
</tbody>
</table>
## Secondary Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. of infants</th>
<th>No. of outcomes</th>
<th>Adjusted mean difference (95% CI) comparing atazanavir regimens with non-atazanavir regimens</th>
<th>Adjusted risk ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head circumference z-score</td>
<td>652</td>
<td>--</td>
<td>-0.45 (-0.66, -0.24)</td>
<td></td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>906</td>
<td>--</td>
<td>0.00 (-0.35, 0.36)</td>
<td></td>
</tr>
<tr>
<td>Hearing screen referral</td>
<td>898</td>
<td>31</td>
<td>1.21 (0.53, 2.80)</td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>911</td>
<td>163</td>
<td>1.06 (0.73, 1.53)</td>
<td></td>
</tr>
<tr>
<td>Prematurity (&lt;37 weeks)</td>
<td>911</td>
<td>161</td>
<td>1.00 (0.68, 1.48)</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions
Conclusions (1)

- Atazanavir-containing regimens may lower infants’ performance on the Language domain of the Bayley-III by about 3.4 points, regardless of trimester of initiation.

- Atazanavir-containing regimens may lower infants’ performance on the Social-Emotional domain by 5.9 points, when initiated in the second/third trimester.
Conclusions (2)

- The lack of an estimated effect of initiation of atazanavir in the first trimester on social-emotional development may be explained by a high proportion of women who switched to another ARV regimen later in pregnancy.
Atazanavir could affect neurodevelopment via hyperbilirubinemia

Clinical implications may be small, but future work should evaluate whether the differences observed in this study persist over time.
Acknowledgements

PHACS is funded by:

[NIH National Institutes of Health]

under cooperative agreements HD052104 (PHACS Coordinating Center, Tulane University School of Medicine) and HD052102 (PHACS Data and Operations Center, Harvard T. H. Chan School of Public Health).

Ellen Caniglia was supported by T32 AI007433 from NIAID

We thank the study participants, clinical sites, PHACS Community Advisory Board, Frontier Science & Technology Research Foundation, and Westat.
PHACS  US Clinical Sites

- Ann & Robert Lurie Children’s Hospital of Chicago
- Baylor College of Medicine
- Bronx Lebanon Hospital Center
- Children's Diagnostic & Treatment Center
- Children's Hospital, Boston
- Children’s Hospital of Philadelphia
- Jacobi Medical Center
- New York University School of Medicine
- St. Christopher’s Hospital for Children
- St. Jude Children's Research Hospital
- San Juan Hospital/Department of Pediatrics
- SUNY Downstate Medical Center
- SUNY Stony Brook
- Tulane University Health Sciences Center
- University of Alabama, Birmingham
- University of California, San Diego
- University of Colorado Health Sciences Center
- University of Florida/Jacksonville
- University of Illinois, Chicago
- University of Maryland, Baltimore
- Rutgers- New Jersey Medical School
- University of Miami
- University of Southern California
- University of Puerto Rico Medical Center