Causal Inference in the Presence of Interference

Michael Hudgens
UNC Dept of Bios
mhudgens@bios.unc.edu
Acknowledgments

NIH NIAID R01 AI085073

Betz Halloran, NIH NIAID R01 AI032042

Lan Liu

Carolina Perez-Heydrich, NIEHS T32ES007018
Outline

Causal Inference Introduction

Interference

Two-stage Randomized Experiments

Observational Studies
Causal Effect

• Let \( z = 1 \) for treatment (exposure); \( z = 0 \) for no treatment (control)
• Let \( y_i(z) \) denote the potential outcome for unit \( i \) given treatment \( z \)
  Each individual has a set of potential outcomes \( \{y_i(0), y_i(1)\} \)
• Individual causal effect is some comparison between individual potential outcomes, e.g.
  \[
  y_i(0) - y_i(1) \text{ or } 1 - \frac{y_i(1)}{y_i(0)}
  \]
  If \( y_i(0) \neq y_i(1) \), treatment has causal effect. O/w, no effect

• **Fundamental problem of causal inference:**
  Cannot observe both potential outcomes for same person
Summary or Population Causal Effects

- Typical unit-level causal effects

\[ ACE \equiv \frac{1}{n} \sum_{i=1}^{n} \{ y_i(0) - y_i(1) \} \]

\[ \text{median}\{y_i(0) - y_i(1) : i = 1, \ldots, n\} \]

- Marginal causal effects

\[ \text{median}\{y_i(0) : i = 1, \ldots, n\} - \text{median}\{y_i(1) : i = 1, \ldots, n\} \]

\[ ACE = \bar{y}(0) - \bar{y}(1) \]
Randomization based inference (Neyman)

- Potential outcomes \( (y_i(0), y_i(1)) \) for \( i = 1, \ldots, n \) fixed
- Treatment assignment \( Z_i \) for \( i = 1, \ldots, n \) random
- Define \( Y_i^{obs} \equiv y_i(Z_i) = (1 - Z_i)y_i(0) + Z_iy_i(1) \); random
- Let
  \[
  \hat{ACE} \equiv \frac{\sum_i Y_i^{obs}(1 - Z_i)}{\sum_i (1 - Z_i)} - \frac{\sum_i Y_i^{obs}Z_i}{\sum_i Z_i}
  \]
- For completely randomized experiment
  \[
  E(\hat{ACE}) = ACE
  \]
  Expectation taken w/r/t \( Z \) (hypothetical re-randomizations)
Randomization based inference

- Estimator of variance
  \[
  \hat{\text{Var}}(\hat{ACE}) \equiv \frac{\hat{\sigma}_0^2}{n_0} + \frac{\hat{\sigma}_1^2}{n_1}
  \]
  where \( \hat{\sigma}_0^2 \equiv \text{Var}\{Y_{\text{obs}}^i : Z_i = 0\}, \ldots \)

- Positively biased unless effect is additive, i.e.,
  \[
  E\{\hat{\text{Var}}(\hat{ACE})\} = \text{Var}(\hat{ACE}) + \frac{1}{n} \sigma_{(0-1)}^2
  \]
  where \( \sigma_{(0-1)}^2 \equiv \text{Var}\{y_i(0) - y_i(1)\} \)

- Appeal to finite-sample central limit theorem, approximate \((1 - \alpha)\) CI
  \[
  \hat{ACE} \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}(\hat{ACE})}
  \]
Assumptions

• Two assumptions:

  Random treatment assignment $Z$, SUTVA

• Randomization based inference:

  No assumptions about underlying distribution of the outcomes
  No assumption of random sampling from super population
Stable Unit Treatment Value Assumption

- That the notation \( \{y_i(0), y_i(1)\} \) sufficiently describes all possible potential outcomes requires SUTVA (Rubin)
- 1. No different versions of treatment (exposure)
- 2. No interference between units
  Outcome of one individual assumed to be unaffected by treatment assignment of others

Clearly not true in some settings
  - Infectious diseases (vaccines), educational interventions, social sciences (housing mobility)
  - Plant variety, cross-over studies

Phenomenon of interest versus a nuisance
Risk of cholera in recipients of killed oral cholera vaccines or placebo, by level of coverage of the bari during one year of follow-up (Ali et al. *Lancet* 2005)

<table>
<thead>
<tr>
<th>Level of vaccine coverage</th>
<th>Target population</th>
<th>Vaccine recipients</th>
<th>Placebo recipients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Risk per 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Cases</td>
</tr>
<tr>
<td>&gt; 50%</td>
<td>22394</td>
<td>12541</td>
<td>16</td>
</tr>
<tr>
<td>41-50%</td>
<td>24159</td>
<td>11513</td>
<td>26</td>
</tr>
<tr>
<td>36-40%</td>
<td>24583</td>
<td>10772</td>
<td>17</td>
</tr>
<tr>
<td>28-35%</td>
<td>25059</td>
<td>8883</td>
<td>22</td>
</tr>
<tr>
<td>&lt; 28%</td>
<td>24954</td>
<td>5627</td>
<td>15</td>
</tr>
</tbody>
</table>
General Approach

- Population of groups of individuals
- Assume **partial interference**: Possibly interference between individuals within a group, but not between groups
- Define direct, indirect, ... causal effects
- Two-stage randomization
  1. Groups to allocation strategies, \( S \in \{ \alpha_0, \alpha_1 \} \)
  2. Given 1, individuals to treatment/control, \( Z \in \{0, 1\} \)
- Randomization-based inference from observable data
Example: Vaccine Trial

- Goal: Study direct, indirect, ... effects of vaccination in school aged children
- Groups: schools sufficiently separated geographically
- Individuals: students
- Assignment mechanism
  1. Randomized some schools to 50% vaccine coverage ($\alpha_1$), others to 25% vaccine coverage ($\alpha_0$)
  2. Randomized students to vaccine or placebo conditional on school assignment from step 1
Notation

- $N$ groups; $n_i$ individuals in group $i = 1, \ldots, N$

- $Z_i = (Z_{i1}, \ldots, Z_{ini})$ treatments received for $n_i$ individuals in group $i$
  
  $Z_{ij} = 0$ or $1 \Rightarrow Z_i$ can take on $2^{n_i}$ possible values

  $Z_{i(j)}$ the $n_i - 1$ subvector of $Z_i$ with the $j^{th}$ entry deleted

  $z_i$ and $z_{ij}$ denote possible values of $Z_i$ and $Z_{ij}$

- $R^m$ set of possible treatment assignments vectors of length $m$
  
  Eg, $R^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}, z_i \in R^{n_i}$
Assignment Mechanism

• $S_i = 1$ if the group $i$ assigned to $\alpha_1$, 0 if assigned $\alpha_0$

$S \equiv (S_1, \ldots, S_N)$

$C \equiv \sum_i S_i$

$\nu$ parameterization for $S$

$\nu$ is a completely randomized group assignment strategy if $C$ fixed

$\Pr_\nu(S = s) = C!(N - C)!/N!$ if $\sum_i s_i = C$, 0 otherwise

• Let $K_i \equiv \sum_j Z_{ij}$

• Similarly define $\alpha_0$, $\alpha_1$ to be completely randomized individual assignment strategies
Potential Outcomes

- $y_{ij}(z_i)$ potential outcome of individual $j$ in group $i$ under $z_i$
- Allows for interference between individuals within group $i$
- Also write $y_{ij}(z_i)$ as $y_{ij}(z_{i(j)}, z_{ij} = z)$
- Have $2^{ni}$ potential outcomes per individual, instead of usual 2 potential outcomes per individual in the absence of interference
Average Potential Outcomes

- **Individual average potential outcome**
  \[ \bar{y}_{ij}(z; \alpha_1) \equiv \sum_{\omega \in R^{n_i-1}} y_{ij}(z_{i(j)} = \omega, z_{ij} = z) \Pr_{\alpha_1}(Z_{i(j)} = \omega | Z_{ij} = z) \]

- **Group average potential outcome**
  \[ \bar{y}_i(z; \alpha_1) \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{y}_{ij}(z; \alpha_1) \]

- **Population average potential outcome**
  \[ \bar{y}(z; \alpha_1) \equiv \frac{1}{N} \sum_{i=1}^{N} \bar{y}_i(z; \alpha_1) \]
Average Potential Outcomes

• **Marginal** individual average potential

\[ \bar{y}_{ij}(\alpha_1) \equiv \sum_{z \in R_{ni}} y_{ij}(z) Pr_{\alpha_1}(Z_i = z) \]

• **Marginal** group and population average potential outcomes

\[ \bar{y}_i(\alpha_1) \equiv \sum_{j=1}^{n_i} \bar{y}_{ij}(\alpha_1) / n_i \]

\[ \bar{y}(\alpha_1) \equiv \sum_{i=1}^{N} \bar{y}_i(\alpha_1) / N \]
Causal Estimands

- Population average direct causal effect
  \[ \overline{DE}(\alpha_1) \equiv \bar{y}(0; \alpha_1) - \bar{y}(1; \alpha_1) \]

- Population average indirect (spillover) causal effect
  \[ \overline{IE}(\alpha_0, \alpha_1) \equiv \bar{y}(0; \alpha_0) - \bar{y}(0; \alpha_1) \]
  similarly for \( z = 1 \)

- Population average total causal effect
  \[ \overline{TE}(\alpha_0, \alpha_1) \equiv \bar{y}(0; \alpha_0) - \bar{y}(1; \alpha_1) \]

- Population average overall causal effect
  \[ \overline{OE}(\alpha_0, \alpha_1) \equiv \bar{y}(\alpha_0) - \bar{y}(\alpha_1) \]
Causal Estimands: Remarks

- Total = direct + indirect
- Under no interference

\[ y_{ij}(z_i) = y_{ij}(z'_i) \text{ for all } z_i, z'_i \text{ such that } z_{ij} = z'_{ij} \]

- Indirect causal effects are zero
- Total causal effect equals direct causal effect
Inference: Estimators

- **Assumption 1.** \( v, \alpha_1, \alpha_0 \) completely randomized assignment strategies

- For \( z = 0, 1 \), let

\[
\hat{Y}_i(z; \alpha_1) \equiv \frac{\sum_j Y_{ij}^{obs} I[Z_{ij} = z]}{\sum_j I[Z_{ij} = z]} = \frac{1}{n_i} \sum_j \frac{Y_{ij}^{obs} I[Z_{ij} = z]}{\Pr[Z_{ij} = z]}
\]

\[
\hat{Y}(z; \alpha_1) \equiv \frac{\sum_i \hat{Y}_i(z; \alpha_1) I[S_i = 1]}{\sum_i I[S_i = 1]} = \frac{1}{N} \sum_i \frac{\hat{Y}_i(z; \alpha_1) I[S_i = 1]}{\Pr[S_i = 1]}
\]
Inference: Estimators

• **Proposition.** Under Assumption 1,

\[ E\{\hat{Y}(z; \alpha_1)\} = \bar{y}(z; \alpha_1) \]

• **Corollary.** Unbiased estimators for the population average direct, indirect, and total causal effects are

\[ \hat{DE}(\alpha_1) \equiv \hat{Y}(0; \alpha_1) - \hat{Y}(1; \alpha_1) \]
\[ \hat{IE}(\alpha_0, \alpha_1) \equiv \hat{Y}(0; \alpha_0) - \hat{Y}(0; \alpha_1) \]
\[ \hat{TE}(\alpha_0, \alpha_1) \equiv \hat{Y}(0; \alpha_0) - \hat{Y}(1; \alpha_1) \]
Inference: Estimators

• Let

\[ \hat{Y}_i(\alpha_1) \equiv \frac{\sum_j Y_{ij}^{obs}}{n_i} \]

and

\[ \hat{Y}(\alpha_1) \equiv \frac{\sum_i \hat{Y}_i(\alpha_1) I[S_i = 1]}{\sum_i I[S_i = 1]} \]

• Proposition. Under Assumption 1

\[ E\{\hat{Y}(\alpha_1)\} = \bar{y}(\alpha_1) \text{ for } z = 0, 1. \]

• Corollary. Unbiased estimator for the population average overall causal effect is

\[ \widehat{OE}(\alpha_0, \alpha_1) \equiv \hat{Y}(\alpha_0) - \hat{Y}(\alpha_1) \]
Inference: Variance

• Unbiased estimators of the variance of the estimators do not exist without further assumptions

• Assumption 2: Stratified Interference (SI)

\[ y_{ij}(z_i) = y_{ij}(z_{ij}, \sum_{k \neq j} z_{ik}) \]

• Eg, vaccine effects on children in a school. SI implies outcome for a child receiving vaccine will be the same when \( k - 1 \) schoolmates also receive vaccine, regardless of particular \( k - 1 \) schoolmates
Stratified Interference

- SI be viewed as an intermediate assumption between
  1. assuming no interference within a group and
  2. making no assumptions about interference within a group
- For a given $z_{ij} = z$, individual $j$ in group $i$ has
  - $1$ potential outcome assuming no interference
  - $n_i$ potential outcomes assuming SI
  - $2^{n_i-1}$ potential outcomes under no assumptions
Inference: Variance

• Under SI and two-stage completely-randomized experiment, begins to resemble classical causal inference setting

• Standard finite sampling results (simple random sampling and two-stage cluster sampling) yield unbiased estimators of variance of $\tilde{Y}_i(z; \alpha_1)$ and $\tilde{Y}(z; \alpha_1)$

• Variance estimators of causal effect estimators are unbiased when effect is additive, positively biased otherwise
Confidence Intervals

- Under what assumptions are Wald CIs justified?

\[
\hat{DE}(\alpha_1) \pm z_{1-\gamma/2} \sqrt{\text{Var}\{\hat{DE}(\alpha_1)\}}
\]

CLT-type justification, either as \( n_i \to \infty \) or \( N \to \infty \), relies on certain Hajek or Lindeberg conditions and mean/var homogeneity assumptions (next slide)

- Chebyshev inequality based \( 1 - \gamma \) CI

\[
\hat{DE}(\alpha_1) \pm \sqrt{\text{Var}\{\hat{DE}(\alpha_1)\}} / \gamma
\]

- For binary \( y \), exact CIs proposed by Tchetgen Tchetgen and VanderWeele (2012) using Hoeffding inequalities
For $n_i$ large, $N$ small, mixture Gaussian
Simulation Study: Empirical width and coverage [in brackets] of Wald (W), Chebyshev (C) and exact (E) 95% CIs of the direct effect $\overline{DE}(\alpha_1)$, indirect effect $\overline{IE}(\alpha_1, \alpha_0)$, total effect $\overline{TE}(\alpha_1, \alpha_0)$ and overall effect $\overline{OE}(\alpha_1, \alpha_0)$ with $N$ heterogeneous groups, $n_i = 1000$ individuals per group, and binary outcome.

<table>
<thead>
<tr>
<th>$N$</th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>30</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{DE}(\alpha_1)$</td>
<td>W</td>
<td>1.73[0.67]</td>
<td>1.33[0.90]</td>
<td>0.99[0.79]</td>
<td>0.55[0.98]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3.26[0.67]</td>
<td>2.89[0.90]</td>
<td>2.25[0.99]</td>
<td>1.26[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>6.07[1.00]</td>
<td>4.96[1.00]</td>
<td>3.84[1.00]</td>
<td>2.22[1.00]</td>
</tr>
<tr>
<td>$\overline{IE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>1.22[0.67]</td>
<td>0.93[0.90]</td>
<td>0.70[0.99]</td>
<td>0.39[0.98]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.31[0.67]</td>
<td>2.02[0.90]</td>
<td>1.59[0.99]</td>
<td>0.89[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
</tr>
<tr>
<td>$\overline{TE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>1.22[0.67]</td>
<td>0.96[0.90]</td>
<td>0.70[0.99]</td>
<td>0.39[0.98]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.31[0.67]</td>
<td>2.08[0.90]</td>
<td>1.59[0.99]</td>
<td>0.90[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
</tr>
<tr>
<td>$\overline{OE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>0.03[0.85]</td>
<td>0.04[0.87]</td>
<td>0.02[0.92]</td>
<td>0.01[0.95]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.06[0.98]</td>
<td>0.09[1.00]</td>
<td>0.05[1.00]</td>
<td>0.03[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
</tr>
</tbody>
</table>
Simulation Study: Empirical width and coverage [in brackets] of Wald (W), Chebyshev (C) and exact (E) 95% CIs of the direct effect $\overline{DE}(\alpha_1)$, indirect effect $\overline{IE}(\alpha_1, \alpha_0)$, total effect $\overline{TE}(\alpha_1, \alpha_0)$ and overall effect $\overline{OE}(\alpha_1, \alpha_0)$ with $N$ homogeneous groups, $n_i = 1000$ individuals per group, and binary outcome.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>30</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{DE}(\alpha_1)$</td>
<td>W</td>
<td>0.06[0.94]</td>
<td>0.04[0.96]</td>
<td>0.03[0.97]</td>
<td>0.02[0.97]</td>
<td>0.01[0.97]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.12[1.00]</td>
<td>0.09[1.00]</td>
<td>0.08[1.00]</td>
<td>0.05[1.00]</td>
<td>0.03[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>6.07[1.00]</td>
<td>4.96[1.00]</td>
<td>3.84[1.00]</td>
<td>2.22[1.00]</td>
<td>1.21[1.00]</td>
</tr>
<tr>
<td>$\overline{IE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>0.04[0.80]</td>
<td>0.05[0.87]</td>
<td>0.03[0.90]</td>
<td>0.02[0.94]</td>
<td>0.01[0.94]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.08[0.95]</td>
<td>0.10[0.99]</td>
<td>0.07[1.00]</td>
<td>0.04[1.00]</td>
<td>0.02[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
<td>0.77[1.00]</td>
</tr>
<tr>
<td>$\overline{TE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>0.04[0.87]</td>
<td>0.04[0.88]</td>
<td>0.03[0.93]</td>
<td>0.02[0.95]</td>
<td>0.01[0.98]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.09[0.97]</td>
<td>0.10[0.99]</td>
<td>0.06[1.00]</td>
<td>0.04[1.00]</td>
<td>0.02[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
<td>0.77[1.00]</td>
</tr>
<tr>
<td>$\overline{OE}(\alpha_1, \alpha_0)$</td>
<td>W</td>
<td>0.03[0.85]</td>
<td>0.04[0.87]</td>
<td>0.02[0.92]</td>
<td>0.01[0.95]</td>
<td>0.01[0.96]</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.06[0.98]</td>
<td>0.09[1.00]</td>
<td>0.05[1.00]</td>
<td>0.03[1.00]</td>
<td>0.02[1.00]</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.84[1.00]</td>
<td>3.14[1.00]</td>
<td>2.43[1.00]</td>
<td>1.40[1.00]</td>
<td>0.77[1.00]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group assignment</th>
<th>Vaccine recipients ($Z_{ij} = 1$)</th>
<th>Placebo recipients ($Z_{ij} = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_i$</td>
<td>$\sum_j Z_{ij}$</td>
</tr>
<tr>
<td>$i$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>12541</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>11513</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>10772</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>8883</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>5627</td>
</tr>
</tbody>
</table>

* $S_i = 1$ (0) corresponds to 50% (30%) vaccine coverage
Estimates of population average effects per 1000 individuals per year

<table>
<thead>
<tr>
<th>Estimate</th>
<th>$\overline{DE}(\alpha_1)$</th>
<th>$\overline{DE}(\alpha_0)$</th>
<th>$\overline{IE}(\alpha_1, \alpha_0)$</th>
<th>$\overline{TE}(\alpha_1, \alpha_0)$</th>
<th>$\overline{OE}(\alpha_1, \alpha_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.30</td>
<td>3.64</td>
<td>2.81</td>
<td>4.11</td>
<td>2.37</td>
</tr>
<tr>
<td>$W$</td>
<td>1.30 [-0.52, 3.11]</td>
<td>3.64 [2.81, 4.46]</td>
<td>2.81 [-0.63, 6.25]</td>
<td>4.11 [2.50, 5.71]</td>
<td>2.37 [0.03, 4.71]</td>
</tr>
<tr>
<td>$C$</td>
<td>[-2.84, 5.43]</td>
<td>[1.75, 5.52]</td>
<td>[-5.03, 10.7]</td>
<td>[0.44, 7.77]</td>
<td>[-2.98, 7.72]</td>
</tr>
<tr>
<td>$E$</td>
<td>[-3540, 3543]</td>
<td>[2177, 2184]</td>
<td>[-2145, 2150]</td>
<td>[-2143, 2151]</td>
<td>[-2145, 2150]</td>
</tr>
</tbody>
</table>

- **Direct**($\alpha_0$): 3.6 fewer cases per 1000 pers-yrs among vaccinated compared to unvaccinated when coverage is 30%
- **Total**: 4.1 fewer cases per 1000 pers-yrs among vaccinated when coverage is 50% compared to unvaccinated when coverage is 30%
- **Simulation results suggest W may be too narrow; E non-informative**
Observational Studies w/ Interference

- What to do in partially randomized or observational studies?

- In the absence of interference, one approach is to inverse weight observed outcomes by estimated probability of treatment received, i.e., propensity score → IPW estimator
Inverse Probability Weighted (IPW) Estimators

- Tchetgen and VanderWeele proposed IPW estimators

\[ \hat{Y}_{ipw}(z; \alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{Pr(\alpha(Z_{i(j)})I(Z_{ij} = z) Y_{ij}^{obs}}{Pr(Z_i|X_i; \hat{\beta})} \]

where \( X_i \) covariates for group \( i \)

- Goal: Apply to Matlab cholera vaccine trial, where individuals self-selected to participate, then randomized to vaccine conditional on participation

- Large sample distribution using M-estimation theory \( \rightarrow \) CAN, empirical sandwich variance estimator

- Simulation study to assess empirical bias of IPW estimators vs naive estimators when baseline covariate \( X \) affects treatment \( Z \) and outcome \( Y \) (confounding)
<table>
<thead>
<tr>
<th>Estimand</th>
<th>Truth</th>
<th>IPW</th>
<th>Bias</th>
<th>ESE</th>
<th>ASE</th>
<th>Cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{DE}(0.3)$</td>
<td>0.131</td>
<td>0.131</td>
<td>0.000</td>
<td>0.015</td>
<td>0.016</td>
<td>97%</td>
</tr>
<tr>
<td>$\overline{TE}(0.3,0.45)$</td>
<td>0.186</td>
<td>0.186</td>
<td>0.000</td>
<td>0.015</td>
<td>0.017</td>
<td>97%</td>
</tr>
<tr>
<td>$\overline{OE}(0.3,0.45)$</td>
<td>0.131</td>
<td>0.128</td>
<td>0.003</td>
<td>0.019</td>
<td>0.021</td>
<td>95%</td>
</tr>
<tr>
<td>$\overline{DE}(0.45)$</td>
<td>0.112</td>
<td>0.114</td>
<td>-0.002</td>
<td>0.008</td>
<td>0.008</td>
<td>95%</td>
</tr>
<tr>
<td>$\overline{TE}(0.45,0.6)$</td>
<td>0.153</td>
<td>0.152</td>
<td>0.002</td>
<td>0.010</td>
<td>0.010</td>
<td>92%</td>
</tr>
<tr>
<td>$\overline{OE}(0.45,0.6)$</td>
<td>0.099</td>
<td>0.096</td>
<td>0.003</td>
<td>0.013</td>
<td>0.015</td>
<td>98%</td>
</tr>
<tr>
<td>$\overline{DE}(0.6)$</td>
<td>0.096</td>
<td>0.094</td>
<td>0.002</td>
<td>0.012</td>
<td>0.011</td>
<td>94%</td>
</tr>
<tr>
<td>$\overline{TE}(0.3,0.6)$</td>
<td>0.132</td>
<td>0.131</td>
<td>0.001</td>
<td>0.018</td>
<td>0.020</td>
<td>97%</td>
</tr>
<tr>
<td>$\overline{OE}(0.3,0.6)$</td>
<td>0.227</td>
<td>0.224</td>
<td>0.003</td>
<td>0.017</td>
<td>0.018</td>
<td>96%</td>
</tr>
<tr>
<td>$\overline{OE}(0.3,0.6)$</td>
<td>0.230</td>
<td>0.224</td>
<td>0.006</td>
<td>0.026</td>
<td>0.027</td>
<td>96%</td>
</tr>
</tbody>
</table>
Generalized, Stabilized IPW Estimators (Liu 2013)

• Let $\tilde{Z}_i$ denote treatment received by the set of individuals that potentially interfere with individual $i$

• Generalized IPW estimators (HT)

$$\hat{Y}_{ipw}^i(z; \alpha) = \frac{\sum_{i=1}^{n} \Pr(\alpha(\tilde{Z}_i)I(Z_i = z)y_i(Z_i, \tilde{Z}_i)/\Pr(Z_i, \tilde{Z}_i|X_i; \hat{\beta})}{n}$$

• Stabilized IPW estimators (Hajek, Aronow and Samii). Replace $n$ with $\hat{n}$, eg,

$$\sum_{i=1}^{n} \frac{I(Z_i = z)}{\Pr(Z_i|X_i; \hat{\beta})} \text{ or } \sum_{i=1}^{n} \frac{\Pr(\alpha(\tilde{Z}_i)I(Z_i = z)}{\Pr(Z_i, \tilde{Z}_i|X_i; \hat{\beta})}$$

• Typically results in decrease in variance, sometimes dramatically
Empirical Bias, empirical standard error (ESE) and the average estimated standard error (ASE) of IPW and Hajek estimators of $\bar{y}(1, \alpha)$ when the propensity scores are known (Known $f$), correctly modeled (Correct $f$), and incorrectly modeled (Mis $f$)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.5</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td><strong>Known $f$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{ipw}(1, \alpha)$</td>
<td>0.035</td>
<td>1.48</td>
<td>1.46</td>
<td>0.004</td>
<td>0.68</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^1(1, \alpha)$</td>
<td>0.050</td>
<td>1.61</td>
<td>1.60</td>
<td>0.008</td>
<td>0.58</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^2(1, \alpha)$</td>
<td>0.009</td>
<td>0.03</td>
<td>0.03</td>
<td>0.001</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Correct $f$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{ipw}(1, \alpha)$</td>
<td>0.036</td>
<td>1.14</td>
<td>1.06</td>
<td>0.151</td>
<td>0.34</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^1(1, \alpha)$</td>
<td>0.039</td>
<td>1.14</td>
<td>1.12</td>
<td>0.154</td>
<td>0.33</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^2(1, \alpha)$</td>
<td>0.008</td>
<td>0.03</td>
<td>0.03</td>
<td>0.023</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Mis $f$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{ipw}(1, \alpha)$</td>
<td>0.566</td>
<td>1.79</td>
<td>1.25</td>
<td>0.154</td>
<td>2.53</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^1(1, \alpha)$</td>
<td>0.721</td>
<td>1.43</td>
<td>1.23</td>
<td>0.387</td>
<td>1.21</td>
</tr>
<tr>
<td>$\hat{Y}_{haj}^2(1, \alpha)$</td>
<td>0.007</td>
<td>0.03</td>
<td>0.02</td>
<td>0.112</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Doubly Robust Estimators (Liu 2013)

- In the absence of interference
  \[ \hat{Y}^{DR}(z) = \hat{\mu}(z) + n^{-1} \sum_i I(Z_i = z) \hat{\epsilon}_i(z) / \Pr[Z_i|X_i; \hat{\beta}] \]

- With partial interference
  \[ \hat{Y}_i^{DR}(z, \alpha) = n_i^{-1} \sum_{j=1}^{n_i} \left\{ \sum_{z_{i(j)}} \hat{\mu}_{ij}(z, z_{i(j)}; X_i) \Pr_{\alpha}(z_{i(j)}) \right\} \]
  \[ + \frac{I(Z_{ij} = z)\{y_{ij}(Z_i) - \hat{\mu}_{ij}(Z_i; X_i)\}}{\Pr[Z_i|X_i; \hat{\beta}]} \Pr_{\alpha}(Z_{i(j)}) \right\} \]

- DR (as total number of groups \( N \to \infty \)) w/ empirical results as expected
<table>
<thead>
<tr>
<th>$\pi_{tru}$ $\mu_{mis}$</th>
<th>$\alpha$</th>
<th>$\alpha$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>ESE</td>
<td>ASE</td>
</tr>
<tr>
<td>$\hat{Y}_{ipw}(1, \alpha_0)$</td>
<td>0.032</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>$\hat{Y}_{reg}(1, \alpha_0)$</td>
<td>0.284</td>
<td>0.71</td>
<td>0.17</td>
</tr>
<tr>
<td>$\hat{Y}_{DR}(1, \alpha_0)$</td>
<td>0.013</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>$\pi_{mis}$ $\mu_{tru}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{ipw}(1, \alpha_0)$</td>
<td>0.706</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td>$\hat{Y}_{reg}(1, \alpha_0)$</td>
<td>0.004</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>$\hat{Y}_{DR}(1, \alpha_0)$</td>
<td>0.005</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\pi_{tru}$ $\mu_{tru}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{DR}(1, \alpha_0)$</td>
<td>0.006</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>$\pi_{mis}$ $\mu_{mis}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{DR}(1, \alpha_0)$</td>
<td>0.501</td>
<td>0.76</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Conclusion

Interference present in many settings, ID/vaccines, education, econometrics/social sciences, networks, spatial analysis, ...

Failure to consider interference may result in misleading inference, eg failure to detect/assess non-direct effects

Many open problems ...
References


