Combining estimators in the pursuit of robustness

Michele Jonsson Funk, PhD
UNC Chapel Hill

September 6-7, 2012
ACF/OPRE Methodological Advancement Meeting
Innovative Directions in Estimating Impact
Funding and Disclosure

• Current work on doubly robust estimators funded by the Agency for Healthcare Research & Quality (AHRQ) (K02HS017950)

• This materials does not represent the views of Agency for Healthcare Research & Quality.
Methods for estimating causal effects

A. Propensity scores used for inverse probability of treatment weighting (IPTW)
B. G-computation
C. Standardized mortality/morbidity ratio (SMR) weights
D. All of the above = Doubly robust estimator
Propensity Score (PS)

• Rosenbaum & Rubin, 1983
• Probability of treatment (or exposure), given a set of characteristics/conditions
• Balances risk of the outcome between treated and untreated groups
• Estimated from the observed data
IPTW

- Weight observations by inverse probability of actual treatment, given covariates
  - Treated (exposed): $1/\text{PS}$
  - Untreated (unexposed): $1/(1-\text{PS})$
- After weighting, ‘crude’ effect in the ‘pseudopopulation’ should be unconfounded
- Effects: Risk, difference, ratio
- Target Population: Total
- PS model must be specified correctly
SMR weights

• Standardized mortality/morbidity ratio (SMR)
• Weight observations by
  – 1 in the treated
  – Propensity odds in the untreated, PS/(1-PS)
• Target Population: Treated
• Effects: Risk (observed), difference, ratio
• Assumes PS model correctly specified
Hypothetical Distribution of Propensity Scores

Adapted from MA Brookhart, Counterfactuals, 2012
IPTW

SMR
G-computation

• Usual generalized linear outcome model
• Marginalizes the treatment effect by estimating each individual’s expected response (counter factual) under both treatment conditions
• Effects: Risk, difference, ratio
• Target Population: Total, treated, untreated
• Assumes outcome model is correctly specified
G-computation: Implementation

• Fit outcome regression model(s) to obtain parameter estimates
• Using the individual’s characteristics, calculate predicted outcomes for each patient with and without treatment
• Calculate average response across all patients under each treatment condition
DR Estimator: Conceptual description

• Doubly robust (DR) estimation uses two models:
  – Propensity score model for the confounder - exposure (or treatment) relationship
  – Outcome regression model for the confounder – outcome relationship, under each exposure condition

• These two stages can use:
  – different subsets of covariates, and
  – different parametric forms.

• If either model is correct, then the DR estimate of treatment effect is unbiased.
Doubly robust estimator

\[
\hat{\Delta}_{DR} = n^{-1} \sum_{i=1}^{n} \left[ \frac{X_i Y_i}{e(Z_i, \hat{\beta})} - \frac{\{X_i - e(Z_i, \hat{\beta})\}}{e(Z_i, \hat{\beta})} m_1(Z_i, \hat{\alpha}_1) \right] - n^{-1} \sum_{i=1}^{n} \left[ \frac{(1 - X_i) Y_i}{1 - e(Z_i, \hat{\beta})} + \frac{\{X_i - e(Z_i, \hat{\beta})\}}{1 - e(Z_i, \hat{\beta})} m_0(Z_i, \hat{\alpha}_0) \right]
\]

\[
\hat{\Delta}_{DR} = [E(Y_1) + \text{augmentation}] - [E(Y_0) + \text{augmentation}]
\]

\[
\hat{\Delta}_{DR} = [E(Y_1)] - [E(Y_0)]
\]
## DR estimator translated

<table>
<thead>
<tr>
<th>DR estimator</th>
<th>General Form</th>
<th>Among exposed (X=1)</th>
<th>Among unexposed (X=0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR_{1i}</td>
<td>$\frac{Y_{X=1} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS}$</td>
<td>$\frac{Y_{x=0}(1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS}$</td>
<td></td>
</tr>
<tr>
<td>DR_{0i}</td>
<td>$\frac{Y_{X=0} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS}$</td>
<td>$\frac{Y_{x=0}(1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X: exposure (0,1)</th>
<th>Z: covariates</th>
<th>$\hat{Y}_1$: predicted Y setting X to 1</th>
<th>$\hat{Y}_0$: predicted Y setting X to 0</th>
<th>$Y_{x=1}$: observed Y given X=1</th>
<th>$Y_{x=0}$: observed Y given X=0</th>
</tr>
</thead>
</table>
DR estimator translated

<table>
<thead>
<tr>
<th>General Form</th>
<th>DR$_{1i}$</th>
<th>DR$_{0i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among exposed (X=1)</td>
<td>[ \frac{Y_{X=1} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS} ]</td>
<td>[ \frac{Y_{X=0}(1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS} ]</td>
</tr>
<tr>
<td>Among unexposed (X=0)</td>
<td>[ \frac{Y_{X=1}}{PS} - \frac{\hat{Y}_1(1 - PS)}{PS} ]</td>
<td>[ \hat{Y}_0 ]</td>
</tr>
</tbody>
</table>

X: exposure (0,1)  
Y: outcome  
Z: covariates  
PS: p(X=1|Z)  
\( \hat{Y}_1 \): predicted Y setting X to 1  
\( \hat{Y}_0 \): predicted Y setting X to 0  
\( Y_{X=1} \): observed Y given X=1  
\( Y_{X=0} \): observed Y given X=0
**IPTW estimator**

<table>
<thead>
<tr>
<th>General Form</th>
<th>DR_{1i}</th>
<th>DR_{0i}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among exposed (X=1)</td>
<td>( \frac{Y_{X=1} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS} )</td>
<td>( \frac{Y_{X=0} (1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS} )</td>
</tr>
<tr>
<td>Among unexposed (X=0)</td>
<td>( \hat{Y}_1 )</td>
<td>( \frac{Y_{X=0}}{1 - PS} - \frac{\hat{Y}_0 \times PS}{1 - PS} )</td>
</tr>
</tbody>
</table>

**Symbols:**
- \( X \): exposure (0,1)
- \( Y \): outcome
- \( Z \): covariates
- \( PS \): \( p(X=1|Z) \)
- \( Y_{X=1} \): observed \( Y \) given \( X=1 \)
- \( Y_{X=0} \): observed \( Y \) given \( X=0 \)
- \( \hat{Y}_1 \): predicted \( Y \) setting \( X \) to 1
- \( \hat{Y}_0 \): predicted \( Y \) setting \( X \) to 0
## G-computation

### General Form

<table>
<thead>
<tr>
<th>Exposure Status</th>
<th>DR&lt;sub&gt;1i&lt;/sub&gt;</th>
<th>DR&lt;sub&gt;0i&lt;/sub&gt;</th>
</tr>
</thead>
</table>
| **Among exposed** (X=1) | \[
\frac{Y_{X=1} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS}
\] | \[
\frac{Y_{X=0}(1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS}
\] |
| **Among unexposed** (X=0) | \[
\frac{Y_{X=1}}{PS} - \frac{\hat{Y}_1(1 - PS)}{PS}
\] | \[
\frac{Y_{X=0}}{1 - PS} \times \frac{\hat{Y}_0(1)}{1 - PS}
\] |

**X**: exposure (0,1)  
**Y**: outcome  
**Z**: covariates  
**PS**: \( p(X=1|Z) \)  
\( \hat{Y}_1 \): predicted \( Y \) setting \( X \) to 1  
\( \hat{Y}_0 \): predicted \( Y \) setting \( X \) to 0  
\( Y_{X=1} \): observed \( Y \) given \( X=1 \)  
\( Y_{X=0} \): observed \( Y \) given \( X=0 \)
## Counterfactual outcomes

### General Form

<table>
<thead>
<tr>
<th></th>
<th>$\text{DR}_{1i}$</th>
<th>$\text{DR}_{0i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among exposed</td>
<td>$\frac{Y_{X=1} \times X}{\text{PS}} - \frac{\hat{Y}_1(X-\text{PS})}{\text{PS}}$</td>
<td>$\frac{Y_{X=0}(1-X)}{1-\text{PS}} + \frac{\hat{Y}_0(X-\text{PS})}{1-\text{PS}}$</td>
</tr>
<tr>
<td>(X=1)</td>
<td>$\frac{Y_{X=1}}{\text{PS}} - \frac{\hat{Y}_1(1-\text{PS})}{\text{PS}}$</td>
<td>$\hat{Y}_0$</td>
</tr>
<tr>
<td>Among unexposed</td>
<td>$\hat{Y}_1$</td>
<td>$\frac{Y_{X=0}}{1-\text{PS}} - \frac{\hat{Y}_0 \times \text{PS}}{1-\text{PS}}$</td>
</tr>
</tbody>
</table>

**X**: exposure (0,1)  
**Y**: outcome  
**Z**: covariates  
**PS**: $p(X=1|Z)$  
$\hat{Y}_1$: predicted $Y$ setting $X$ to 1  
$\hat{Y}_0$: predicted $Y$ setting $X$ to 0  
$Y_{X=1}$: observed $Y$ given $X=1$  
$Y_{X=0}$: observed $Y$ given $X=0$
Weighting relevant observed events

<table>
<thead>
<tr>
<th>General Form</th>
<th>$\text{DR}_{1i}$</th>
<th>$\text{DR}_{0i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among exposed $(X=1)$</td>
<td>$\frac{Y_{X=1} \times X}{PS} \cdot \hat{Y}_1 (1 - PS)$</td>
<td>$\frac{Y_{X=0} (1 - X)}{1 - PS} + \hat{Y}_0 (1 - PS)$</td>
</tr>
<tr>
<td>Among unexposed $(X=0)$</td>
<td>$\frac{Y_{X=1}}{PS}$</td>
<td>$\hat{Y}_0$</td>
</tr>
</tbody>
</table>

$X$: exposure (0,1)  
$Y$: outcome  
$Z$: covariates  
PS: $p(X=1|Z)$  
$\hat{Y}_1$: predicted $Y$ setting $X$ to 1  
$\hat{Y}_0$: predicted $Y$ setting $X$ to 0  
$Y_{X=1}$: observed $Y$ given $X=1$  
$Y_{X=0}$: observed $Y$ given $X=0$
### Subtracting?

<table>
<thead>
<tr>
<th>General Form</th>
<th>DR$_{1i}$</th>
<th>DR$_{0i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among exposed (X=1)</td>
<td>$\frac{Y_{X=1} \times X}{PS} - \frac{\hat{Y}_1(X - PS)}{PS}$</td>
<td>$\frac{Y_{X=0}(1 - X)}{1 - PS} + \frac{\hat{Y}_0(X - PS)}{1 - PS}$</td>
</tr>
<tr>
<td>Among unexposed (X=0)</td>
<td>$\frac{Y_{X=1}}{PS} - \frac{\hat{Y}_1(1 - PS)}{PS}$</td>
<td>$\frac{Y_{X=0}}{1 - PS} - \frac{\hat{Y}_0 \times PS}{1 - PS}$</td>
</tr>
</tbody>
</table>

X: exposure (0,1)  
Y: outcome  
Z: covariates  
PS: p(X=1|Z)  
$\hat{Y}_1$: predicted Y setting X to 1  
$\hat{Y}_0$: predicted Y setting X to 0  
$Y_{X=1}$: observed Y given X=1  
$Y_{X=0}$: observed Y given X=0
Net effect of combining weights

#1: IPTW * observed outcomes = Response under exposure standardized to the total population

#2: 1/SMR * predicted outcomes = Response under exposure standardized to the unexposed population

Subtract #2 from #1
For the net result:
Response under exposure standardized to the exposed population
Combining weights for relevant observed outcomes

<table>
<thead>
<tr>
<th></th>
<th>$\text{DR}_{1i}$</th>
<th>$\text{DR}_{0i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Form</td>
<td>$\frac{Y_{X=1} \times X}{\text{PS}} - \frac{\hat{Y}_1(X - \text{PS})}{\text{PS}}$</td>
<td>$\frac{Y_{X=0}(1 - X)}{1 - \text{PS}} + \frac{\hat{Y}_0(X - \text{PS})}{1 - \text{PS}}$</td>
</tr>
<tr>
<td>Among exposed (X=1)</td>
<td>$\frac{Y_{X=1}}{\text{PS}} - \frac{\hat{Y}_1(1 - \text{PS})}{\text{PS}}$</td>
<td>$\hat{Y}_0$</td>
</tr>
<tr>
<td>Among unexposed (X=0)</td>
<td>$\hat{Y}_1$</td>
<td>$\frac{Y_{X=0}}{1 - \text{PS}} - \frac{\hat{Y}_0 \times \text{PS}}{1 - \text{PS}}$</td>
</tr>
</tbody>
</table>

- $Y$: outcome
- $X$: exposure (0,1)
- $Z$: covariates
- $\text{PS}$: $p(X=1|Z)$
- $\hat{Y}_1$: predicted $Y$ setting $X$ to 1
- $\hat{Y}_0$: predicted $Y$ setting $X$ to 0
- $Y_{X=1}$: observed $Y$ given $X=1$
- $Y_{X=0}$: observed $Y$ given $X=0$
Effect measures

• Scale
  – Risk, mean response
  – Risk difference, difference in means
  – Relative risk
  – Odds ratio

• Target Populations
  – Total
  – Treated
  – Untreated
Assumptions

• Positivity
• Consistency
• No interference (aka independence)
• Exchangeability (aka ignorability)
  – Correct model specification for
    PS model or outcome regression models
  – No unmeasured confounding

See Cole & Hernan, AJE, 2008
Misspecified covariates

• Categorize continuous covariates (realistic scenario)
  – Simulated to mirror the distribution of common confounders
    • Age, BMI, LDL cholesterol, physical activity
  – Categories reflect ‘meaningful’ cutpoints

• True relationships known (simulated)
  – Linear or slightly quadratic
Residual confounding
Monte Carlo simulation

- Draw a random sample (n=5000)
- Fit a model (OLS or DR)
- Save the parameter estimate & standard error
- Repeat 1000 times

- For each of 11 scenarios x 4 treatment effects
# Root Mean Squared Error

<table>
<thead>
<tr>
<th>Scenario</th>
<th>True TX effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>2.963</td>
</tr>
<tr>
<td>True</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>DR</td>
</tr>
<tr>
<td>Misspecified Outcome Model</td>
<td></td>
</tr>
<tr>
<td>Categorize linear covariates</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.113</td>
</tr>
<tr>
<td>DR</td>
<td>0.047</td>
</tr>
<tr>
<td>Categorize nonlinear covariates</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.144</td>
</tr>
<tr>
<td>DR</td>
<td>0.054</td>
</tr>
<tr>
<td>Categorize linear &amp; nonlinear covariates</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.250</td>
</tr>
<tr>
<td>DR</td>
<td>0.064</td>
</tr>
</tbody>
</table>
### 95% CI coverage

<table>
<thead>
<tr>
<th>Scenario</th>
<th>True TX effect</th>
<th>0</th>
<th>-0.41</th>
<th>-1.10</th>
<th>-1.61</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unadjusted</strong></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>True</strong></td>
<td></td>
<td>94.4</td>
<td>95.7</td>
<td>95.7</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td><strong>OLS</strong></td>
<td>94.5</td>
<td>95.7</td>
<td>95.2</td>
<td>95.1</td>
</tr>
<tr>
<td></td>
<td><strong>DR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Missspecified Outcome Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorize linear covariates</td>
<td></td>
<td>19.9</td>
<td>18.5</td>
<td>19.2</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>DR</strong></td>
<td>95.9</td>
<td>95.3</td>
<td>94.8</td>
<td>96.2</td>
</tr>
<tr>
<td>Categorize nonlinear covariates</td>
<td></td>
<td>17.5</td>
<td>17.8</td>
<td>16.2</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>DR</strong></td>
<td>96.2</td>
<td>96.0</td>
<td>94.2</td>
<td>96.4</td>
</tr>
<tr>
<td>Categorize linear &amp; nonlinear covariates</td>
<td></td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>DR</strong></td>
<td>96.4</td>
<td>94.3</td>
<td>95.5</td>
<td>95.2</td>
</tr>
</tbody>
</table>
Limitations

• Two poorly specified models can be worse than a single wrong maximum likelihood regression

• Standard errors tend to be slightly larger compared to a single correctly specified regression model

• Residual confounding is modest in magnitude relative to bias of crude estimate

• DR estimation is not a panacea for unmeasured confounding

• Standard errors/confidence intervals require bootstrapping

• Best practices & diagnostics still under development
Conclusions

• Observational (non-experimental) studies depend on statistical models to disentangle causal effects from confounding
• We can never be certain that the statistical model we have chosen is correct
• DR estimator is unbiased if at least one of the two component models is right and therefore provides some protection against residual confounding
• Attractive properties of marginalized effect estimates with improved efficiency relative to IPTW
Resources


• **Recommended further reading**
  
  
  
  
SAS macro: www.unc.edu/~mfunk/dr
Michele Jonsson Funk
www.unc.edu/~mfunk
mfunk@unc.edu