How Much Do Effects Vary Across Sites?
Evidence From Existing Multisite Randomized Trials

Stanford Workshop
July 18, 2016

Michael J. Weiss¹, Howard S. Bloom¹, Natalya Verbitsky-Savitz², Himani Gupta¹, Alma Vigil², Dan Cullinan¹

¹MDRC
²Mathematica Policy Research
Outline

• Why do Effects Vary and Why Should we Care?
• Cross-site Distribution of Effects Defined
• Data
• Estimation
• Empirical Results
• Discussion
  – Implications for designing studies
  – When to expect a lot of x-site impact variation
Why do effects vary: The three C’s

1. **Treatment Contrast**
   1. **Program Group**: the services received by the program group
   2. **Control Group**: the counterfactual services received

2. **Client Characteristics**

3. **Program Context**
Why are about cross-site impact variation?

• Overall average impacts can mask heterogeneity in impacts across sites

• This information...
  – has substantive implications
  – is necessary for planning multi-site experiments
Site-level distribution of impacts

Let:

\( B_j \) = True average treatment effect at site \( j \)

Then:

\[ \beta \equiv \lim_{J^* \to \infty} \frac{\sum_{j=1}^{J^*} B_j}{J^*} \]

\[ \tau \equiv \lim_{J^* \to \infty} \sqrt{\frac{\sum_{j=1}^{J^*} (B_j - \beta)^2}{J^*}} \]

\( \beta = 0.20 \)
Data
# Data from large multi-site RCTs

<table>
<thead>
<tr>
<th>Early Childhood-Element. School</th>
<th>Middle School-High School</th>
<th>Post-secondary Education</th>
<th>Labor Market Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start Program (<em>Abt</em>)</td>
<td>Charter Middle Schools (<em>Mathematica</em>)</td>
<td>Learning Communities (<em>MDRC</em>)</td>
<td>Job Corps (<em>Mathematica</em>)</td>
</tr>
<tr>
<td>After School – Reading Program (<em>MDRC</em>)</td>
<td>Teach for America – Math (<em>Mathematica</em>)</td>
<td>Performance-based Scholarships (<em>MDRC</em>)</td>
<td>Welfare-to-Work Programs (<em>MDRC</em>)</td>
</tr>
<tr>
<td>After School – Math Program (<em>MDRC</em>)</td>
<td>Enhanced Reading Opportunity (<em>MDRC</em>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teach for America – Pooled (i3, National) (<em>Mathematica</em>)</td>
<td>Small Schools of Choice (<em>MDRC</em>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee STAR</td>
<td>Career Academies (<em>MDRC</em>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Early College H.S. (<em>Abt</em>)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation Model
Estimation Model

Level 1 (clients):

\[ Y_{ij} = \sum_{r=1}^{R} \alpha_r RA_{Block_{rij}} + B_j T_{ij} + \sum_{l=1}^{L} \gamma_l X_{lij} + e_{ij} \]

Level 2 (sites):

\[ B_j = \hat{\beta} + b_j \]

Where:

\[ e_{ij} \sim N \left(0, \sigma^2_{\alpha}(T_{ij})\right) \]
\[ b_j \sim N(0, \tau^2) \]
\[ Cov(e_{ij}, b_j) = 0 \]

\[ \hat{\beta} \], an estimate of the treatment effect for the average site

\[ \hat{\tau} \], an estimate of the cross-site standard deviation of site-average treatment effects
Results
## Selected Results

<table>
<thead>
<tr>
<th>Intervention</th>
<th>$\hat{\beta}$ - mean</th>
<th>$\hat{t}$ - s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start Impact Study (ES - Read)</td>
<td>0.20***</td>
<td>0.30***</td>
</tr>
<tr>
<td>After School Reading (ES - Read)</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>After School Math (ES - Math)</td>
<td>0.07***</td>
<td>0.00</td>
</tr>
<tr>
<td>Teach for America - Pooled (ES - Math)</td>
<td>0.10**</td>
<td>0.05*</td>
</tr>
<tr>
<td>Tennessee STAR (ES - Read)</td>
<td>0.15***</td>
<td>0.23***</td>
</tr>
<tr>
<td>Charter Middle Schools (ES - Read)</td>
<td>-0.07</td>
<td>0.16***</td>
</tr>
<tr>
<td>Enhanced Reading Opp’s (ES - Read)</td>
<td>0.07***</td>
<td>0.08**</td>
</tr>
<tr>
<td>Teach for America - Math (ES - Math)</td>
<td>0.08***</td>
<td>0.10***</td>
</tr>
<tr>
<td>Small High Schools of Choice (% on track)</td>
<td>10.3 ***</td>
<td>15.3 ***</td>
</tr>
<tr>
<td>Career Academies (avg yearly $, yrs 1-4)</td>
<td>1,883.00***</td>
<td>0.0</td>
</tr>
<tr>
<td>Early College High School (% on track)</td>
<td>3.4 *</td>
<td>8.2 ***</td>
</tr>
<tr>
<td>Learning Communities (credits, 1.5yrs)</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Perform-based Scholarship (credits, 3yrs)</td>
<td>1.8 **</td>
<td>1.3 *</td>
</tr>
<tr>
<td>Job Corps (avg yearly $, yr 4)</td>
<td>1,415.00***</td>
<td>1,687.00**</td>
</tr>
<tr>
<td>Welfare-to-Work (avg yearly $, yrs 1-2)</td>
<td>670.00***</td>
<td>601.00***</td>
</tr>
</tbody>
</table>

*p<.10  **p<.05  ***p<.01
Consistent zero average impact across sites

Afterschool Reading Program – Reading, yr 1

\[ \hat{\beta} = -0.02 \]
\[ \hat{t} = 0.04 \]
Near zero average impact with a lot of cross-site variation
Charter Middle School – Reading

\[ \hat{\beta} = -0.07 \]

\[ \hat{t} = 0.16^{***} \]
Consistent positive impacts across sites

Career Academies – Average yearly earnings, yrs 1-4

\[
\hat{\beta} = $1,883*** \\
\hat{\tau} = $0
\]

Constrained EB impact Estimates
Large average impacts with a lot of cross-site variation

Welfare to Work – Average yearly earnings, yrs 1-2

\hat{\beta} = $670***
\hat{t} = $601***
Discussion
Minimum Detectable Effect Size (MDES)

\[ MDES_Z = M_{J-1} \sqrt{\left( \frac{1}{J} \right) \left( \frac{\tau^2_Z}{\bar{T}} + \frac{(1 - \rho_C)(1 - R^2_{within})}{n\bar{T}(1 - \bar{T})} \right)} \]

Where:

- \( M_{J-1} \) = a multiplier that rapidly approaches 2.8 as \( J \) increases (for a 2-tail test at the 0.05 significance level with 80 percent power)
- \( J \) = number of sites
- \( n \) = number of individuals per site (assumed constant across sites)
- \( \bar{T} \) = proportion of individuals randomized to treatment
- \( \tau_Z \) = cross-site standard deviation of site-average program effects on the z-score metric
- \( \rho_C \) = intra-class correlation for control group outcomes (i.e., the proportion of total outcome variance explained by site indicators)
- \( R^2_{within} \) = proportion of within-site outcome variance explained by our baseline covariates
(MDES) by (# of Sites) by (Tau)

Assuming: $R^2 = 0.50$, $\sigma_Z^2 = 1$, $n = 75$ and $\bar{T} = 0.5$
Check this...
Mike Weiss, 7/11/2016
When do effects vary across sites a little vs. a lot?

Hypothesis: When the site-average TCs varies a lot across sites, so will treatment effects.
When do effects vary across sites a little vs. a lot?

\[ \overline{TC}_j \equiv \bar{S}_{j|T=1} - \bar{S}_{j|T=0} \]

**Hypothesis:**

As \( \text{Var}(\overline{TC}) \) increases, so does \( \tau \)

\[
\text{Var}(\overline{TC}) = \text{Var}(\bar{S}_{T=1}) + \text{Var}(\bar{S}_{T=0}) - 2\text{Cov}(\bar{S}_{T=1}, \bar{S}_{T=0})
\]
When to expect a large $\text{Var}(\overline{TC})$ (and $\tau$)

\[
\text{Var}(\overline{TC}) = \text{Var}(\overline{S}_{T=1}) + \text{Var}(\overline{S}_{T=0}) - 2\text{Cov}(\overline{S}_{T=1}, \overline{S}_{T=0})
\]

- Low specificity of the program model
- A high proportion of formal education is altered by the intervention
- When treatment and control group members from the same “site” are served in a different setting for a high proportion of their formal education experience
## Selected Results

<table>
<thead>
<tr>
<th>Intervention</th>
<th>$\hat{\beta}$ - mean</th>
<th>$\hat{\sigma}$ - s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start Impact Study (ES - Read)</td>
<td>0.20***</td>
<td>0.30***</td>
</tr>
<tr>
<td>After School Reading (ES - Read)</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>After School Math (ES - Math)</td>
<td>0.07***</td>
<td>0.00</td>
</tr>
<tr>
<td>Teach for America - Pooled (ES - Math)</td>
<td>0.10**</td>
<td>0.05*</td>
</tr>
<tr>
<td>Tennessee STAR (ES - Read)</td>
<td>0.15***</td>
<td>0.23***</td>
</tr>
<tr>
<td>Charter Middle Schools (ES - Read)</td>
<td>-0.07</td>
<td>0.16***</td>
</tr>
<tr>
<td>Enhanced Reading Opp’s (ES - Read)</td>
<td>0.07***</td>
<td>0.08**</td>
</tr>
<tr>
<td>Teach for America - Math (ES - Math)</td>
<td>0.08***</td>
<td>0.10***</td>
</tr>
<tr>
<td>Small High Schools of Choice (% on track)</td>
<td>10.3***</td>
<td>15.3***</td>
</tr>
<tr>
<td>Career Academies (avg yearly $, yrs 1-4)</td>
<td>1,883***</td>
<td>0.0</td>
</tr>
<tr>
<td>Early College High School (% on track)</td>
<td>3.4*</td>
<td>8.2***</td>
</tr>
<tr>
<td>Learning Communities (credits, 1.5yrs)</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Perform-based Scholarship (credits, 3yrs)</td>
<td>1.8**</td>
<td>1.3*</td>
</tr>
<tr>
<td>Job Corps (avg yearly $, yr 4)</td>
<td>1,415***</td>
<td>1,687**</td>
</tr>
<tr>
<td>Welfare-to-Work (avg yearly $, yrs 1-2)</td>
<td>670***</td>
<td>601***</td>
</tr>
</tbody>
</table>

*p<.10  **p<.05  ***p<.01
What about Client Characteristics?

• For many characteristics (e.g., prior achievement) most variation is within sites

• We suspect cross-site impact variation driven by cross-site variation in client characteristics may be hard to predict
What about Context?

• We suspect contextual moderation often occurs through the treatment contrast
Funding

• Spencer Foundation
• IES

“The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D140012 to MDRC. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education”
How Much Do Effects Vary across Sites?
Evidence from Existing Multisite Randomized Trials

QUESTIONS?