Causal Inference in Experiments with Interference

David Choi
Carnegie Mellon University
September 12, 2021
Problem Setting

- In some experiments, each unit may be influenced by the treatment of others ("interference between units")
  - Vaccination studies (herd immunity)
  - Social networks (peer influence, sharing of information)
  - etc.

- In such settings, estimation generally requires bounds on the strength or structure of the interference. Examples:
  - Units can be divided into groups that are independent
  - Interference is "weak" – most units do not interfere with each other (even indirectly)

- Validity of such assumptions may sometimes be difficult to judge
What if no assumptions are made?

Unbounded interference poses major challenges even for randomized experiments:

- Observed outcomes might not be representative of any counterfactual
  - e.g., outcomes under 50% vaccination rate may not be representative of those under 90% vaccination

- If dependencies are unbounded, law of large numbers may not apply to an estimator using the observed outcomes

To address these challenges, our approach will be quite different from standard causal inference
Our Approach

- New class of estimands – “differences in attributable effects”
  - Ex: In a vaccination study, we might estimate the extent to which units who received the placebo were more protected in neighborhoods where the vaccination rate was high.

- Pro: Confidence intervals can be found with **no assumptions beyond randomization of treatment**
  - No bounds on interference are needed

- Con: Intervals are wider, and causal interpretation will be more limited
  - Describes the effect under the observed treatment assignment, but not others (such as if the vaccination rate was increased)

- Approach gives results that are weaker, but easier to believe
  - May complement other approaches by offering **reassurance**
  - May be of theoretical interest in its own right
Related Works

The weaker the assumptions, the less can be inferred:

- Exposure model: every unit depends only on treatment of its known neighbors
  - Can estimate any counterfactual intervention\(^1\)

- Weak dependence: dependency graph may be unknown but is sparse, so that most units do not interfere with each other (even indirectly)
  - Can estimate effects “at the intensive margin”\(^2\)

- No assumptions beyond randomization: unlimited, even pathological coordination possible
  - Can estimate attributable effects on ranks, hypothesis testing\(^3\)
  - We’ll extend this considerably, allowing for familiar tools (regression, matching, weighting) to be applied

---

**Setup**

**Notation:** Consider an randomized experiment where

- $X_i$: Treatment assignment of $i$th unit
- $Y_i$: Observed outcome of $i$th unit, affected by all $N$ treatments:
  \[ Y_i = f_i(X_1, \ldots, X_N), \quad f_i \text{ unknown} \]
- $\theta_i$: Counterfactual outcome of $i$ if nobody is treated ("uniformity trial")

**Definition:** The **individual-level attributable treatment effect** is given by

\[ Y_i - \theta_i, \]

e.g., $Y_i - \theta_i$ is the effect on unit $i$ of administering treatment according to $X = (X_1, \ldots, X_N)$.
Example

Consider a vaccination trial

- $X_i$: did unit $i$ receive vaccine or placebo?
- $Y_i$: did unit $i$ get the disease? (1=Yes)
- $\theta_i$: counterfactual for unit $i$ if all units received the placebo

$Y_i = 0, \theta_i = 1$: unit $i$ did not get the disease, but would have if all units had received the placebo

- The vaccines prevented the disease for unit $i$
- This can occur even if unit $i$ received the placebo in the trial (herd immunity)

We do not assume that $\theta$ is observed for any unit (e.g., not even if unit and all known friends receive the placebo)

- $Y$ may be uninformative as to $\theta$
Discussion

The vector $Y - \theta$ clearly has a causal interpretation

- Answers the question “What did the treatments do?”
- Effect of the treatment under the observed treatment assignment

Knowing $Y - \theta$ does not answer questions such as

- “What if all units received the treatment?”
- “What if a single additional unit was treated?”

Our goal: partially characterize $Y - \theta$, with no assumptions on $\theta$

- Next: a class of estimands to do so
Estimand 0: Attributable Effect

- The attributable effect is the average of $Y - \theta$,

\[
\text{attributable effect} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \theta_i),
\]

e.g., “outcomes caused minus outcomes prevented, per capita”

- Cannot be estimated without further assumptions
  - We don’t know the average value of $\theta$
Estimand 1: Difference in Attributable Effects

Here are some examples that we can estimate:

- Let $\tau$ denote the **difference in attributable effects** for treated ($X_i = 1$) and control ($X_i = 0$):

\[
\tau = \frac{1}{N_1} \sum_{i:X_i=1} (Y_i - \theta_i) - \frac{1}{N_0} \sum_{i:X_i=0} (Y_i - \theta_i)
\]

\[
\begin{align*}
\text{Attr. effect for treated} \quad &\quad \text{Attr. effect for control}
\end{align*}
\]

- Interpretation if $\tau < 0$: “Vaccinated units had more cases prevented (and/or fewer cases caused) per capita, compared to those who received the placebo”
Estimand 2: Simple Regression

What if treatment is non-binary? Here is another partial characterization of $Y - \theta$:

- $\beta_1$: The slope of the best fitting line to $Y - \theta$ as a function of the treatment $X$:

$$\text{slope } \beta_1 = \left( \sum_{i=1}^{N} (X_i - \bar{X}) \right)^{-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \theta_i)$$

- Interpretation: “A unit difference in treatment between two individuals is associated with a $\beta_1$ difference in their attributable treatment effects”

Note: the association here is between treatment and effect, not treatment and outcome

- The treatment may have affected everyone (including the control) – did it do so heterogeneously?
More generally, the estimand can be a linear function of $Y - \theta$,

$$\text{estimand} = \sum_{i=1}^{N} w_i(X) \cdot (Y_i - \theta_i),$$

where examples include $\tau$ and $\beta_1$

This includes various approaches

- Other regression specifications
- Matching-based and weighted comparisons of units

Many examples are discussed in paper
Inference

Inference is easy for $\tau$ when the treatments $X$ are assigned by simple random sampling:

$$\tau = \frac{1}{N_1} \sum_{i: X_i = 1} (Y_i - \theta_i) - \frac{1}{N_0} \sum_{i: X_i = 0} (Y_i - \theta_i)$$

$$= \left( \frac{1}{N_1} \sum_{i: X_i = 1} Y_i - \frac{1}{N_0} \sum_{i: X_i = 0} Y_i \right) - \left( \frac{1}{N_1} \sum_{i: X_i = 1} \theta_i - \frac{1}{N_0} \sum_{i: X_i = 0} \theta_i \right),$$

where $\sigma^2$ can be bounded by $1/4$ since $\theta$ is binary\(^4\).

\(^4\)Note: in this case, letting $\sigma^2 = 1/4$ will cover $\tau$ even if CLT doesn’t hold.
Inference (General Case)

For the general estimand $w(X)^T(Y - \theta)$, with complex randomization of $X$, finding a CI is similar if central limit theorem can be applied.

\[
\text{estimand} = \underbrace{w(X)^T Y}_{\text{point est.}} - \underbrace{w(X)^T \theta}_{\approx N(m, \nu^2)}
\]

where bias is $m$ and error has confidence interval bounded by

\[
\max_{\theta} \mathbb{E} \left[ w(X)^T \theta \right] + z_{1 - \frac{\alpha}{2}} \sqrt{\text{Var} \left[ w(X)^T \theta \right]},
\]

which can be bounded and **globally** solved* as integer quadratic program (using branch-and-bound).

* may require non-trivial amount of problem structuring.

* note that local optimum is insufficient for coverage.
The point estimate $w(X)^T Y$ is usually quite familiar:

- If $w(X)$ is a regression, then $w(X)^T Y$ is just the regression specification applied to the outcomes, instead of the effects.
- Similar if $w(X)$ is a weighted or matching-based comparison.

This point estimate may be biased.

- We will return a bound on the bias (and CIs will always account for bias as well).
- Correct usage of regression adjustment, matching, weights $\rightarrow$ smaller bias and tighter CIs.
  - These things result in $w(X)$ having mean zero.
Remark

Connection to Community Detection

- For binary $\theta$ and zero-mean $w$, bounding the confidence interval is equivalent to bounding the variance

$$\max_\theta \text{Var} \left[ w(X)^T \theta \right],$$

which divides the units into two groups whose values of $w(X)$ are highly correlated.

- This is equivalent/closely related to the Ising Blockmodel problem

$$\max_{\theta \in \{-1, +1\}} \theta^T \mathbb{E} \left[ w w^T \right] \theta \text{ such that } \sum_i \theta_i = 0$$

Example: $w$ is voting records of congress and you want to identify political factions.

- Fast exact methods under certain assumptions (that don’t hold for us, so we branch-and-bound)
Examples

1. Simulated Vaccinations
2. Real Vaccinations
3. Social encouragement to purchase crop insurance
4. Anti-conflict intervention in schools
Vaccination Study

- Cholera vaccine trial was conducted in 1985
- Participants were randomized to receive vaccine or placebo
- Participation rates varied widely by neighborhood, and hence vaccinations rates varied widely as well
- Better outcomes (both treated and control) were observed in neighborhoods where vaccination rate was high
- Previous analyses assumed (we’ll remove both):
  - No interference between neighborhoods (strongly questioned)
  - Decision to participate was random conditioned on observables
- Data: Simulated individual-level and aggregated real data are available

---

Example 1: Simulated Data

Simulated results for this study were released\(^6\) – models were partially fit to actual results of trial.

<table>
<thead>
<tr>
<th>Vaccination Rate</th>
<th>Cases, placebo</th>
<th>Cases, vaccine</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-33%</td>
<td>51% (19/37)</td>
<td>27% (24/90)</td>
</tr>
<tr>
<td>33-67%</td>
<td>32% (59/181)</td>
<td>20% (74/366)</td>
</tr>
<tr>
<td>68-100%</td>
<td>19% (75/378)</td>
<td>9% (66/742)</td>
</tr>
</tbody>
</table>

\(^6\)in interferference R package
Let $\mu_\ell$ denote the effects of vaccine on placebo units in the $\ell$th neighborhood $\eta_\ell$:

$$
\mu_\ell = \frac{1}{N_{0\ell}} \sum_{i \in \eta_\ell : X_i = 0} (Y_i - \theta_i),
$$

and regress $\mu_\ell$ on vaccination rates, controlling for participation:

$$
\mu_\ell \sim \text{ParticipationRate}_\ell + \text{VaccinationRate}_\ell
$$

**Result**: Given two neighborhoods with equal participation rates, a 1% difference in vaccination rates was associated with 0.64% difference in protective effects for the placebo units (bias $\pm 0.05\%$, 90% CI: [0.03%, 1.2%])

(i.e., 0.64 more cases prevented (fewer cases caused) per 100)
Nonsignificant result

Let $\tau_\ell$ denote the difference in effects for vaccinated and placebo units in $\eta_\ell$

$$
\tau_\ell = \frac{1}{N_{1\ell}} \sum_{i \in \eta_\ell: X_i = 1} (Y_i - \theta_i) - \frac{1}{N_{0\ell}} \sum_{i \in \eta_\ell: X_i = 0} (Y_i - \theta_i)
$$

and regress $\tau_\ell$ on vaccination rates

$$
\tau_\ell \sim \text{VaccinationRate}_\ell
$$

**Result:** Given two units in same neighborhood, receiving vaccine was associated with

$$
0.2 - 0.27 \cdot \text{VaccinationRate}
$$

more cases prevented (or fewer cases caused) per capita. However, slope is not significant at 90% confidence:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Point Est.</th>
<th>Bias</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.2</td>
<td>0</td>
<td>[0.04, 0.37]</td>
</tr>
<tr>
<td>VaccinationRate</td>
<td>-0.27</td>
<td>0</td>
<td>[-0.60, 0.07]</td>
</tr>
</tbody>
</table>
Example 2: Real results from Vaccination Trial

Aggregated results from trial are publicly available. Table shows results aggregated by vaccination rate into two groups:

<table>
<thead>
<tr>
<th>Group</th>
<th>Vaccination Rate</th>
<th>Cholera Rate, per 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treated</td>
<td>Control</td>
</tr>
<tr>
<td>1</td>
<td>less than 45%</td>
<td>2.1 (54/25K)</td>
</tr>
<tr>
<td>2</td>
<td>$\geq$ 45%</td>
<td>1.7 (42/25K)</td>
</tr>
</tbody>
</table>

Let $\tau_\ell$ denote the difference in effects between vaccinated and placebo units in group $\ell = 1, 2$:

$$
\tau_\ell = \frac{1}{N_{1\ell}} \sum_{\text{group } \ell: X_i = 1} (Y_i - \theta_i) - \frac{1}{N_{0\ell}} \sum_{\text{group } \ell: X_i = 0} (Y_i - \theta_i),
$$

and estimate $\tau_1, \tau_2$, and their difference $\delta = \tau_1 - \tau_2$.
## Results

<table>
<thead>
<tr>
<th>Estimand</th>
<th>Point Est.</th>
<th>Bias</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_1 ) (Vaccination &lt; 45%)</td>
<td>-3.5</td>
<td>0</td>
<td>[-5.6, -1.4]</td>
</tr>
<tr>
<td>( \tau_2 ) (Vaccination ≥ 45%)</td>
<td>-1.3</td>
<td>0</td>
<td>[-3.5, 0.9]</td>
</tr>
<tr>
<td>( \delta = \tau_1 - \tau_2 )</td>
<td>-2.2</td>
<td>0</td>
<td>[-4.4, 0.0]</td>
</tr>
</tbody>
</table>

(negative \( \tau_\ell \): more cases prevented for vaccinated units, per thousand)

- Vaccinated units were more protected than placebo units, but difference significant only for < 45% vaccinated group
- Difference in differences \( \tau_1 - \tau_2 \) was significant at 90% conf.
- These CIs used an additional assumption on \( \theta \)
  - This assumption reduced the CI widths by factor of roughly 5
  - Let’s discuss this assumption now
Additional Assumption

- Trial covered Matlab, Bangladesh, 1985
- Using historical rates, let’s assume that in absence of trial, 1985 cholera rate would not have exceeded 7 cases per 1000:

\[
\frac{1}{N} \sum_{i=1}^{N} \theta_i \leq 0.007
\]
Justifying our Assumption

- Making no assumptions would allow for average $\theta$ to be as high as 0.5
  - i.e., in the absence of the trial, 50% of population might have gotten cholera
  - This is 100s of times larger than the observed rate and seems unrealistic

- Our bound is significantly higher than the highest historical rates (pre-vaccine) of cholera
  - This assumption may be easier to consider than a bound on interference between neighborhoods

- Sensitivity analysis may also be of interest
  - E.g., $\delta$ significant at 95% conf. level if bound is tightened to 0.005 (the highest observed rate)
Comparison with Existing Approaches

- Other approaches tell similar story of herd immunity, with stronger conclusions:
  - Predict the effects of intervening on the vaccination rate
  - Smaller confidence intervals

- However, strong assumptions were required:
  - No interference (i.e., infections) between neighborhoods
  - Nonparticipants are statistically identical to participants, conditioned on observables

- Our assumption based on historical rates might be easier to consider than theirs
  - If so, then our weaker (but more robust) results may offer reassurance
例3：社会影响的实验分析

实验设计（非常简化）

- 全部单位：n = 3331
- 第一回合，低信息：n = 1425
- 第一回合，高信息：n = 673
- 第二回合（3天后）：n = 1425

- 农民随机分配到信息会话，内容为高或低信息
- 第一和第二回合相隔3天，允许沟通和可能的社会影响
- 在第二回合，保险参与率更高
- Q：这是第一回合会话导致的吗？

---

Let’s apply our approach:

- $i = 1, \ldots, n$ denotes a 2nd round unit
- $Z_i$: how many of $i$’s 1st round friends\(^8\) received high info?
- $Y_i$: did $i$ purchase insurance?
- $\theta_i$: outcome if all 1st round sessions were not held
- $Y_i - \theta_i$: the effect on $i$ of holding the 1st round sessions

Assume that first round units assigned to high/low sessions by random sampling within each village

- Actual randomization was stratified by variables that are missing for many units – pedagogical example only
Regression Results

- Regress treatment effect $Y - \theta$ on exposures $Z$, which are not identically distributed

$$Y_i - \theta_i \sim \beta_1 Z_i + \text{propensity class}$$

- Given two farmers in the same propensity class, a unit difference in the number of high information friends is associated with an 8% difference in treatment effect (95% CI: [0, 0.17])

<table>
<thead>
<tr>
<th>Point Est.</th>
<th>Bias</th>
<th>$\geq$ 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.08</td>
<td>± 0.002</td>
</tr>
</tbody>
</table>

The 1st round sessions were more effective on farmers who had more 1st round friends that received high information (weakly significant)

$^9$The distribution of $Z_i$ depends on $i$’s village and # of first round friends.
Nonparametric Fit

Let $\gamma_0, \ldots, \gamma_{4+}$ regress $Y - \theta$ nonparametrically on exposure $Z$

$$Y_i - \theta_i = \gamma_{Z_i} + \text{propensity class}$$

Shown: point estimates of $\gamma_k - \gamma_0$, bias, 95% CI

- Effect of having 1 friend in 1st round w/high information was same as having zero such friends
- Effect visible at 2 friends. Large CI’s for 3+ due to small samples
Matching-based Differences in Treatment Effect

- Define thresholded exposures

\[ W_i = 1 \text{ if } i \text{ has } \geq 2 \text{ friends in 1st round with high info} \]

- Let \( \mathcal{M} \) denote a matched subset (units are paired within each propensity class), and estimate the difference in treatment effects for these units only:

\[
\tau_{\text{matched}} = \sum_{i \in \mathcal{M}} \left( \frac{W_i(Y_i - \theta_i)}{m} - \frac{(1 - W_i)Y_i - \theta_i}{m} \right)
\]

- The difference in effects was 20% (CI: [0.04, 0.37]).

<table>
<thead>
<tr>
<th>( \tau_{\text{matched}} )</th>
<th>Point Est.</th>
<th>Bias</th>
<th>( \geq 95% ) CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21</td>
<td>( \pm 0.01 )</td>
<td>[0.03, 0.37]</td>
<td></td>
</tr>
</tbody>
</table>

“Within \( \mathcal{M} \), the 1st round sessions were more effective on units who had \( \geq 2 \) high info 1st round friends, versus those with 0 or 1 such friends”
Weighting-based Differences in Treatment Effect

Estimate the difference in treatment effects within each propensity class $\Pi_1, \ldots, \Pi_K$, averaged together (weighted by class size):

$$\tau_{\text{weighted}} = \sum_k \frac{n_k}{N} \sum_{i \in \Pi_k} \left( \frac{W_i(Y_i - \theta_i)}{n_{k1}} - \frac{(1 - W_i)(Y_i - \theta_i)}{n_{k0}} \right)$$

Diff. in Effects within $\Pi_k$

Similar to matching: under this weighting, the difference in effects was 17% (CI: [0.03, 0.34])

<table>
<thead>
<tr>
<th>$\tau_{\text{weighted}}$</th>
<th>Point Est.</th>
<th>Bias</th>
<th>$\geq$ 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.19</td>
<td>± 0.02</td>
<td>[0.02, 0.36]</td>
</tr>
</tbody>
</table>
Remark

Why did we use propensity class adjustment, matching, and weighting in our estimands?

- Without this, point estimates would have larger bias (and hence CIs would be wider, as they include bias)
  - \( \theta \) could be correlated with probability of treatment exposure

Interesting to contrast our approach with others:

- In most approaches, failing to correctly adjust will cause erroneous inferences
- Our method detects the problem, and widens the CIs accordingly
  - Can safely use approximate matchings, etc.
  - Method can be very picky about regression specification
  - (This is possible because the randomization is known)
Example 4: Experiment in Schools

- Small subset (40-64) of students at each selected school (28) were randomized between treatment and control.
- Treated (‘seed’) students were invited to participate in bi-monthly meetings where they identify social conflicts and design/enact strategies to reduce schoolwide conflict.
- Schoolwide outcomes included self-reported wearing of wristband representing support for conflict reduction.
- Network information (“spent time with”) recorded for all students.

Setup

- $X_i$: was unit $i$ invited to participate in meetings?
- $Y_i$: did unit $i$ self-report wearing wristband?
- $\theta_i$: counterfactual if all eligible units were assigned to treatment (“full treatment”)
- $Y_i - \theta_i$: the effect of not inviting control units to participate

Inviting a unit may affect outcome of others, for example by:
- Altering the strategies chosen at the meeting
- Altering who is influenced to wear a wristband
Matched Differences in Treatment Effect

Shown: point estimates of $E[\tau^{\text{matched}}]$, bias, 95% CI

- Restricted to units with positive probability of all 4 groups
- Propensity classes: school, # of eligible friends
- Not inviting the control units to participate had the most negative effect on those who were control and whose eligible friends were also control.
Discussion

- We estimate differences in attributable effects – “relative attributable effects”
  - Answers the question “Did the treatment work?”

- Require no assumptions beyond randomization of treatment, allowing unbounded interference between units
  - To reduce CI’s further, assumptions based on historical outcomes (or other info?) may be considered instead of bounding interference