From SATE to PATT
Combining Experimental with Observational Studies to Estimate Population Treatment Effects

Erin Hartman
(with Richard Grieve, Roland Ramsahai, Jasjeet Sekhon)

Johns Hopkins
Biostatistics and Bloomberg School of Public Health
Ross-Royall Symposium: From Individuals to Populations

02-26-2016
The Problem

How to combine information from randomized controlled trials (RCTs) and non-random studies (NRSs) in order to provide evidence for treatment effects in a full population of interest.

- RCTs raise issues of randomization bias which leads to poor external validity (Heckman and Smith 1995)
- NRSs raise issues of selection bias, or non random assignment to treatment, which leads to poor internal validity
  - How do we define the target population? Is it changing?
The Opportunity

- Explosion of data sources: administrative, electronic medical records (EMR), online behavior
- Population data is becoming more common, more precise, and more widely available, which is particularly helpful for determining cost effectiveness in practice
- Policy makers: “let’s just use the big data to make causal inferences”
- Tension between identification and machine learning/predictive methods
- How can we leverage the identification of RCTs with this explosion of data sources?
Roadmap

Goal: To determine the effect of one medical treatment in the target population

- Develop a theoretical decomposition of the bias of going from the Sample Average Treatment Effect (SATE) to the Population Average Treatment Effect on the Treated (PATT)
- Derive the assumptions needed to identify PATT from RCT and NRS data (agnostic to estimation strategy)
- Introduce a new estimation strategy to combine RCTs and NRSs
- Most importantly, provide a set of placebo tests to validate the identifying assumptions
- Results for applied example: Pulmonary Artery Catheterization (PAC)
Pulmonary Artery Catheterization (PAC)

- PAC is an invasive cardiac monitoring device for critically ill patients (ICU patients)—e.g. myocardial infarction (ischaemic heart disease)
- It is a diagnostic device: allows for simultaneous measurement of pressures in the right atrium, right ventricle, pulmonary artery, and filling pressure of the left ventricle.
- Widely used in the past 30 years: spend $2 billion / year in the U.S.
Pulmonary Artery Catheterization (PAC)

- A series of NRS found PAC was associated with increased mortality and increased costs (e.g. Chittock et al, 2004, Connors et al, 1996)

- This prompted a series of randomized controlled trials and meta-analyses, all of which found no statistically significant differences in mortality rate between the PAC and no-PAC groups (e.g. Harvey et al, 2005)
PAC-Man study

- Randomized controlled trial, publicly funded, pragmatic design conducted in 65 UK ICUs in 2000-2004
  - 1,014 subjects, 506 randomly assigned to receive PAC
- No difference in hospital mortality \( (p = 0.39) \) (e.g. Harvey et al, 2005)
- Some heterogeneity in effect by subgroup (e.g. Harvey et al, 2008)
- Non-representative nature of patient mix could mean unadjusted estimates don’t apply to the target population
ICNARC Case Mix Program database

- Non-random study: prospective in nature, conducted between May 2003 and December 2004
- ICNARC CMP database contains information on: case-mix, patient outcomes, resources use for 1.5 million admissions and 250 critical care units in the UK (e.g. Harrison et al, 2004)
- Same inclusion and exclusion criteria for individual patients as the corresponding PAC-Man study
- 1,052 cases with PAC and 32,499 controls in 57 critical care units
- Target Population: The 1,052 NRS cases that received PAC in practice
Definitions:

- $T_i \in (0, 1)$ - Treatment indicator for unit $i$
- $S_i \in (0, 1)$ - Indicator for whether or not unit $i$ was in the RCT (vs the target population)
- $Y_{ist}$ - Potential outcomes for subject $i$
- $W$ - Set of observable covariates
Extrapolating experimental findings to target populations

Schematic showing adjustment of sample effect to identify population effect.

Double arrows indicate exchangeability of potential outcomes.

Dashed arrows indicate adjustment of the covariate distribution.
Extrapolating experimental findings to target populations

**Assumption 1:** Consistency Under Parallel Studies

\[ Y_{i01} = Y_{i11} \]
\[ Y_{i00} = Y_{i10} \]

**Assumption 2:** Strong Ignorability of Sample Assignment for Treated

\[ (Y_{i01}, Y_{i11}) \perp S_i \mid (W_i^T, T_i = 1) \]
\[ 0 < \Pr(S_i = 1 \mid W_i^T, T_i = 1) < 1 \]

**Assumption 3:** Strong Ignorability of Sample Assignment for Controls

\[ (Y_{i00}, Y_{i10}) \perp S_i \mid (W_i^{CT}, T_i = 1) \]
\[ 0 < \Pr(S_i = 1 \mid W_i^{CT}, T_i = 1) < 1 \]

**Assumption 4:** Stable Unit Treatment Value Assumption (SUTVA)

\[ Y_{ist}^{L_i} = Y_{ist}^{L_j} \quad \forall i \neq j \]
Placebo Tests

Assumptions imply that:

$$\mathbb{E}(Y_i|S_i = 0, T_i = 1) - \mathbb{E}_{01}\{\mathbb{E}(Y_i|W_i, S_i = 1, T_i = 1)\} = 0$$

- The difference between the mean outcome of the NRS treated and the mean outcome of the reweighed RCT treated should be zero.
- If this is not zero, then at least one assumption has failed.
- Similar placebo test for controls, but it is not as informative for identifying PATT (i.e., it could fail due to lack of overlap).
- Tested using equivalence tests (Hartman and Hidalgo, 2010)
Estimating PATT for PAC

- Using Genetic Matching to maximize the internal validity
  - SATE $\rightarrow$ SATT
  - Create match pairs within the randomized trial
  - New pairs created within subgroups for subgroup estimates

- Using Maximum Entropy Weighting to maximize the external validity
  - SATT $\rightarrow$ PATT
  - Weight using the distribution RCT treated $W$ to the distribution of NRS $W$
  - Weights applied to matched pairs

- Conduct validity check using equivalence placebo tests
## Baseline Characteristics and End-points

**Table:** Baseline characteristics and endpoints for the PAC-Man Study, and for patients in the NRS who received PAC. Numbers are N (%) unless stated otherwise

<table>
<thead>
<tr>
<th></th>
<th>RCT No PAC n=507</th>
<th>PAC n=506</th>
<th>NRS PAC n=1052</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline Covariates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admitted for elective surgery</td>
<td>32 (6.3)</td>
<td>32 (6.3)</td>
<td>98 (9.3)</td>
</tr>
<tr>
<td>Admitted for emergency surgery</td>
<td>136 (26.8)</td>
<td>142 (28.1)</td>
<td>243 (23.1)</td>
</tr>
<tr>
<td>Admitted to teaching hospital</td>
<td>108 (21.3)</td>
<td>110 (21.7)</td>
<td>447 (42.5)</td>
</tr>
<tr>
<td>Mean (SD) Baseline probability of death</td>
<td>0.55 (0.23)</td>
<td>0.53 (0.24)</td>
<td>0.52 (0.26)</td>
</tr>
<tr>
<td>Mean (SD) Age</td>
<td>64.8 (13.0)</td>
<td>64.2 (14.3)</td>
<td>61.9 (15.8)</td>
</tr>
<tr>
<td>Female</td>
<td>204 (40.2)</td>
<td>219 (43.3)</td>
<td>410 (39.0)</td>
</tr>
<tr>
<td>Mechanical Ventilation</td>
<td>464 (91.5)</td>
<td>450 (88.9)</td>
<td>906 (86.2)</td>
</tr>
<tr>
<td>ICU size (beds)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 or less</td>
<td>57 (11.2)</td>
<td>59 (11.7)</td>
<td>79 (7.5)</td>
</tr>
<tr>
<td>6 to 10</td>
<td>276 (54.4)</td>
<td>272 (53.8)</td>
<td>433 (41.2)</td>
</tr>
<tr>
<td>11 to 15</td>
<td>171 (33.7)</td>
<td>171 (33.8)</td>
<td>303 (28.8)</td>
</tr>
<tr>
<td><strong>Endpoints</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaths in Hospital</td>
<td>333 (65.9)</td>
<td>346 (68.4)</td>
<td>623 (59.3)</td>
</tr>
<tr>
<td>Mean Hospital Cost (£)</td>
<td>19,078</td>
<td>18,612</td>
<td>19,577</td>
</tr>
<tr>
<td>SD Hospital Cost (£)</td>
<td>28,949</td>
<td>23,751</td>
<td>24,378</td>
</tr>
</tbody>
</table>
### PAC Maxent Placebo Tests

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Outcome</th>
<th>Difference Obs – Adj</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>survival</td>
<td>−0.03</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>733</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>Elective Surgery</td>
<td>survival</td>
<td>0.08</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>−3069</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>−11917</td>
<td></td>
</tr>
<tr>
<td>Emergency Surgery</td>
<td>survival</td>
<td>−0.07</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>2226</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>−1821</td>
<td></td>
</tr>
<tr>
<td>Non–Surgical</td>
<td>survival</td>
<td>−0.04</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>2566</td>
<td></td>
</tr>
<tr>
<td>Teaching Hospital</td>
<td>survival</td>
<td>−0.04</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>3934</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>5765</td>
<td></td>
</tr>
<tr>
<td>Non–Teaching Hospital</td>
<td>survival</td>
<td>−0.03</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>cost</td>
<td>−1635</td>
<td></td>
</tr>
<tr>
<td></td>
<td>net benefit</td>
<td>−3917</td>
<td></td>
</tr>
</tbody>
</table>

Outcome Difference

<table>
<thead>
<tr>
<th>Difference Obs – Adj</th>
<th>Naive p–value</th>
<th>FDR p–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>□</td>
<td>△</td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Placebo Test P–Values

<table>
<thead>
<tr>
<th>Placebo Test P–Values</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>△</td>
</tr>
</tbody>
</table>
Mortality Estimates

Effect on Survival Rate (% points)

Overall
Elective Surgery
Emergency Surgery
Non-Surgical
Teaching Hospital
Non-Teaching Hospital

-0.4  -0.2  0   0.2  0.4  0.6

Treatment Effect Estimates

Effect on Survival Rate (% points)
Cost Estimates

-25000 −15000 −5000 0 5000 10000 15000

Strata

Overall

Elective Surgery

Emergency Surgery

Non-Surgical

Teaching Hospital

Non-Teaching Hospital

Treatment Effect Estimates

Effect on Costs in £
Cost-Effectiveness Estimates

Treatment Effect Estimates

Non-Teaching Hospital

Teaching Hospital

Non-Surgical

Emergency Surgery

Elective Surgery

Overall

Strata

SATT

PATT

Effect on Incremental Net Benefit (Valuing £ 20K per Quality Adjusted Life Year (QALY))
Conclusions and Implications

- We pass placebo tests for both the costs and hospital mortality, as well as cost-effectiveness, thus validating our assumptions for identifying PATT
- Recover experimental benchmark of null results overall
- Evidence for future research for positive effects for elective surgery patients and negative effects in teaching hospitals
- Implications for cost-effectiveness analyses, since these are often based on observational studies
The value of placebo tests

- We used two alternative estimation strategies:
  - Inverse Propensity Score Weighting (IPSW) to construct the weights
  - Bayesian Additive Regression Trees (BART) to model the response surface and predict the outcome

- Placebo tests show that these methods were not appropriate for this example
Often policymakers are also interested in comparing results from disjoint experiments:

- Experiment 1: A vs. B
- Experiment 2: A vs. C
- We care about the effect of: B vs. C

- Extensions to Difference-in-Difference
- Extensions to Instrumental Variables
- Alternative estimands
Thank you!