Estimating the Health Consequences of Playing Football using Observational Data:
Challenges, Lessons Learned, and New Directions

Sameer K. Deshpande

04 June 2019
Bayesian Causal Inference Workshop 2019
The Ohio State University
Collaborators

Raiden Hasegawa
Dylan Small
Jordan Weiss
Concern over Football Safety

Increasing concern about the safety of playing football

- High-profile suicides among former NFL and college-level players
- Chronic traumatic encephalopathy
Concern over Football Safety

**Increasing concern about the safety of playing football**
- High-profile suicides among former NFL and college-level players
- Chronic traumatic encephalopathy

**Awareness/concern ⇒ changes to the game and policy**
- NFL concussion protocol
- Ivy League banned tackling in practice
- Possible state bans on youth tackle football
Concern over Football Safety

Increasing concern about the safety of playing football

- High-profile suicides among former NFL and college-level players
- Chronic traumatic encephalopathy

Awareness/concern ⇒ changes to the game and policy

- NFL concussion protocol
- Ivy League banned tackling in practice
- Possible state bans on youth tackle football

> 1M kids play HS football. What’s going to happen to them?

- HS players experience less head trauma than professionals
- Single season of play induces white matter changes
  
Determine the causal effect of FB participation

Randomized control trial:

😊 Arguably the gold standard

😊 Unethical: can’t force someone to continue to play...

😊 Impractical: long follow-up required

Use existing longitudinal data:

😊 Large, representative study populations

😊 Control for many potential confounders measured in adolescence

🍀 Non-random treatment assignment
Our Experience with Add Health

Data Source: National Longitudinal Study of Adolescent to Adult Health
- Followed \( \sim 10k \) subjects since 1994
- Last available follow-up in 2008 (subjects aged 25–35)
- Primary outcome: score on depression scale (CES-D)
- Secondary outcomes: diagnoses of depression, anxiety, PTSD, substance abuse, personality traits

General Strategy: Matched Observational Study
- Design (before looking at outcome data): Matching, pre-registration
- Analysis: covariance adjustment + randomization inference
Design Phase: Matching

**Variable ratio matching**: based on entire number

- \( \nu(x) = \frac{P(Z=0,X=x)}{P(Z=1,X=x)} \)

- Avg. number of controls available for matching with covariates \( x \)

- Implementation: optimal 1 : \( K \) matching for subjects with prop. score in \( ((K + 1)^{-1}, K^{-1}) \).

  [Pimentel et al. (2015), *Statistics in Medicine*]

- Several choices for propensity model and \( K_{max} \)

- Pick match with adequate balance and largest sample size
HS athletes may differ from non-athletes in important ways

Control group contains both athletes & non-athletes!

- Convincing demonstration of FB effect suggests 4 comparison:
  - FB vs All Controls (AC)
  - FB vs Sport Controls (SC)
  - FB vs Non-Sport Controls (NSC)
  - Sport Controls (SC) vs Non-Sport Controls (NSC)

- Ordered Testing: no loss of power from multiple testing correction.

[Rosenbaum (2008), *Biometrika*; Hasegawa et al. (2019+), *submitted*]
Analysis: Covariance Adjustment + Randomized Inference

- Residual imbalance w/in matched set might bias comparison
- Idea: Regress out covariates and compare residuals
- Validity of inference does not require correct regression model

[Rosenbaum (2002), *Statistical Science*]

**Our Procedure:** To test $H_0 : \tau = \tau_0$

1. For treated, compute potential outcome under control $Y_i - \tau_0$
2. Align covariates w/in matched set
3. Regress adjusted responses onto aligned covariates w/ BART
4. Run permutational t-test on residuals
Effect of FB on CES-D Scores

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Est. (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB vs AC</td>
<td>-0.26 (-0.52, 0.02)</td>
</tr>
<tr>
<td>FB vs SC</td>
<td>-0.24 (-0.57, 0.02)</td>
</tr>
<tr>
<td>FB vs NSC</td>
<td>-0.21 (-0.47, 0.11)</td>
</tr>
<tr>
<td>SC vs NSC</td>
<td>0.02 (-0.25, 0.27)</td>
</tr>
</tbody>
</table>

- CES-D scores: 0 (least depressed) – 15 (most depressed)
- Point estimate: τ with maximal p-value
- Pooled SD = 2.3; CI’s contains small effect sizes (cut-off = ±0.46)
- **Notable**: No evidence of large harmful effect (cut-off = ±1.84)
Effect of FB on CES-D Scores

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Est. (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB vs AC</td>
<td>-0.26 (-0.52, 0.02)</td>
</tr>
<tr>
<td>FB vs SC</td>
<td>-0.24 (-0.57, 0.02)</td>
</tr>
<tr>
<td>FB vs NSC</td>
<td>-0.21 (-0.47, 0.11)</td>
</tr>
<tr>
<td>SC vs NSC</td>
<td>0.02 (-0.25, 0.27)</td>
</tr>
</tbody>
</table>

• CES-D scores: 0 (least depressed) – 15 (most depressed)
• Point estimate: \( \tau \) with maximal \( p \)-value
• Pooled SD = 2.3; CI’s contains small effect sizes (cut-off = ±0.46)
• **Notable**: No evidence of large harmful effect (cut-off = ±1.84)
• **Necessary caveat**: we’re not saying FB is totally safe!
Limitations and Ways Forward

Noisy and incomplete FB participation data

- Subjects asked about participation or intention to participate that year
- Missing measures of prior exposure, position played, injury history
  Currently piloting survey to collect this

Heterogeneous Treatment Effects

- Assumed constant additive treatment effect model
- Missing modifiers: years played, position played, # concussions

Subgroups

- Subgroups most likely determined by unmeasured covariates
- Simultaneous subgroup detection of subgroups & effect estimation?
  Current idea: use sample splitting
ATE assuming heterogeneous effects

- Our procedure (RI + BART): more conservative than others
- Other intervals contain very small harmful effect sizes
- Qualitative results essentially the same!
Thanks, y’all!

sameerd@alum.mit.edu
people.csail.mit.edu/sameerd
@skdeshpande91 on Twitter