Calibrated Risk Adjusted Modeling (CRAM) With a Bridge Design for Extending the Applicability of Randomized Controlled Trials

Ravi Varadhan

Division of Biostatistics & Bioinformatics
Department of Oncology
Johns Hopkins University
Baltimore, MD, USA

Inaugural Ross-Royall Symposium
February 26, 2016
Dedication

- Professor Richard Royall from whom I learnt so much
- Late Professor Alan Ross
Evidence Gap

- Older adults, with multiple diseases, are poorly represented in RCTs (Zulman, JGIM 2011)
- Evidence for most interventions is lacking in older adults
- Effectiveness of ACE-inhibitors for treatment of CHF in women older than 75 years of age (Heiat, ArchIntMed 2002)
- CMS in 2009 refused to cover CT Colonography due to lack of relevant evidence (MedCAC)
- Solid organ transplant trials of immunosuppression agents - older participants are excluded (Blosser, Transpl. 2011)
The paradox of the clinical trial is that it is the best way to assess whether an intervention works, but is arguably the worst way to assess who benefits from it (Mant 1999)
Limitation of RCT

- Report overall or average treatment effects (OTE)
- Participants in RCTs are a select group, not representative of at-risk population
- Concern that OTE is not generalizable
- Why? Potential for significant heterogeneity of treatment effect (HTE)
Applicability of Evidence

Let $\beta_Z(E)$ be the estimate of efficacy of intervention $Z$ from an RCT with sample $E$.

Denote the larger at-risk population as $P$ and the target population as $Q$ (e.g., women older than 75 years).

**Generalizability**: Is the evidence from $E$ generalizable to $P$? Yes, if $E$ is a random sample of $P$.

**Applicability**: Is the evidence from $E$ applicable to $Q$? Yes, if $Q$ is well-represented in $E$ and if there is no relevant *heterogeneity of treatment effect (HTE)*.
Suppose that \( \#E \) is relatively large and that we did not find any significant HTE.

We might suspect that the evidence is applicable to \( P \).

On the other hand, suppose we found significant HTE - Does evidence from \( E \) apply to \( P \) or to \( Q \)?

A Solution: Standardization approach of Cole and Stuart (AJE 2010)
What if evidence of lesser validity is available in a representative sample of $P$? (observational database with confounded treatment selection)

Let us denote this as $b_Z(P)$, which differs from $\beta_Z(P)$ that would result if we enrolled a random sample from $P$ in the trial.

Can we make use of lesser quality evidence from $P$ in conjunction with that from $E$?

This is the problem that we address using CRAM, which is a method for cross-design synthesis
Goal

To extend the applicability of evidence on treatment effectiveness to target groups poorly represented in RCTs

Bring information from observational studies
Integrate trial and observational data to project treatment effect from a trial to a target group

RCT provides internally valid treatment effects but lacks broader applicability

Observational database (e.g. registry) has broader representation but lacks internal validity

Confounding in observational data (measured + unmeasured)

Methodology to exploit strengths and mitigate limitations of two study designs
Essential Idea in CRAM: Bridge Design

- Calibration adjustments for unmeasured confounding in the observational study: tweak unmeasured confounding parameters to match treatment effects
- Calibration adjustment performed where trial and observational data overlap
- Calibration makes it possible to estimate a treatment effect in observational data with adjustment for unmeasured confounding
- Extend applicability to target groups using models for heterogeneity
Bridge Study

(A) Venn diagram showing the overlap between treatment (T), baseline (B), and exposure (E).

(B) Graph showing the density of treatment effect modifier (e.g., age) with Treatment (T) and Exposure (E) categories, highlighting an overlap at B.
## The 3 Studies

<table>
<thead>
<tr>
<th>Sample</th>
<th>Source</th>
<th>Bridge</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Name</td>
<td>SOLVD Treatment Trial (n=2,569)</td>
<td>SOLVD Registry (n=5,100)</td>
<td>SOLVD Prevention Trial (n=4,228)</td>
</tr>
<tr>
<td>Study Type</td>
<td>RCT</td>
<td>Observational</td>
<td>RCT(^a)</td>
</tr>
<tr>
<td>Proportion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>19.6</td>
<td>28.8</td>
<td>11.3</td>
</tr>
<tr>
<td>Age (\geq 75) years</td>
<td>5.8</td>
<td>16.1</td>
<td>4.2</td>
</tr>
<tr>
<td>Female and age (\geq 75) years</td>
<td>1.4</td>
<td>8.0</td>
<td>0.6</td>
</tr>
<tr>
<td>History of diabetes mellitus</td>
<td>25.8</td>
<td>24.6</td>
<td>15.3</td>
</tr>
<tr>
<td>History of myocardial infarction</td>
<td>65.8</td>
<td>76.0</td>
<td>80.1</td>
</tr>
<tr>
<td>History of atrial fibrillation</td>
<td>10.8</td>
<td>15.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Dependent edema</td>
<td>16.8</td>
<td>29.0</td>
<td>4.4</td>
</tr>
<tr>
<td>Pulmonary edema</td>
<td>25.7</td>
<td>40.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Lung crackles</td>
<td>12.1</td>
<td>36.3</td>
<td>2.6</td>
</tr>
<tr>
<td>History of COPD</td>
<td>10.0</td>
<td>17.7</td>
<td>5.4</td>
</tr>
<tr>
<td>History of stroke</td>
<td>7.7</td>
<td>8.9</td>
<td>5.9</td>
</tr>
<tr>
<td>Mean / (std. dev.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>60.4(9.9)</td>
<td>62.8(12.2)</td>
<td>58.7(10.3)</td>
</tr>
<tr>
<td>LVEF, %</td>
<td>24.9</td>
<td>31.9</td>
<td>28.3</td>
</tr>
<tr>
<td>Unadjusted treatment effect, log hazard ratio (SE)</td>
<td>-0.51 (0.080)</td>
<td>0.47 (0.05)</td>
<td>-0.28 (0.10)</td>
</tr>
</tbody>
</table>
Major Steps in CRAM

- 3 samples: trial, observational ("bridge"), target
- Model the baseline risk of outcome (the basis of CRAM)
- **Assumption**: same baseline risk $\Rightarrow$ same treatment effect (w/o confounding)
- Test for presence of HTE using an interaction test
- Standardize Tx effect from the RCT to the observational sample
- Find parameters of unmeasured confounding (solve an optimization problem)
- Using the calibrated model, estimate Tx effect in the target sample
CRAM: Application

- To estimate the effect of ACE-Inhibitors for women older than 75 years of age
- There are few women > 75 years of age in RCTs
- Studies of Left Ventricular Dysfunction (SOLVD): prevention (P), treatment (T), and registry (R)
- P and T are RCTs and R is observational
- Uniform protocols and measurement across studies
- CRAM strategy: calibrate R with T, and then project onto P
- Validation by comparing the CRAM estimate to truth in SOLVD-P
- Another validation with a low-risk subset in SOLVD-P
Comparison of Baseline Risk Distributions

1-year Risk of CVD Death

SOLVD-T
SOLVD-R
SOLVD-P
CRAM Results - ≥ 75 yr Women

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardization</th>
<th>Standardization</th>
<th>CRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Covariate-based</td>
<td>Risk-based</td>
<td></td>
</tr>
<tr>
<td>Estimate $\psi_{t=1}$</td>
<td>-0.094 (0.44)</td>
<td>-0.64 (0.13)</td>
<td></td>
</tr>
<tr>
<td>$\mu_1=-0.5$</td>
<td>--</td>
<td>--</td>
<td>-0.43 (0.08)$^d$</td>
</tr>
<tr>
<td>$\mu_1=-1.0$</td>
<td>--</td>
<td>--</td>
<td>-0.44 (0.09)$^f$</td>
</tr>
</tbody>
</table>
## CRAM Results - Distant Target Sample

<table>
<thead>
<tr>
<th>Model</th>
<th>True Effect</th>
<th>Standardization, Covariate-based</th>
<th>Standardization, Risk-based</th>
<th>CRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate $\psi_{t=1}$</td>
<td>-0.35 (0.19)</td>
<td>-0.55 (0.43)</td>
<td>-0.11 (0.19)</td>
<td>--</td>
</tr>
<tr>
<td>$\mu_1=-0.5$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-0.31 (0.18)</td>
</tr>
<tr>
<td>$\mu_1=-1.0$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-0.28 (0.18)</td>
</tr>
</tbody>
</table>
Weights for Distant Target Sample
CRAM Limitations

- Results are encouraging, but ...
- Requires an appropriate bridging (observational) sample
- Modeling assumptions pertaining to risk-based HTE
- Computationally demanding, especially, bootstrapping for standard errors
Acknowledgements

- Agency for Healthcare Research Quality (Dr. Parivash Nourjah)
- Brookdale Leadership in Aging Fellowship
- Carlos Weiss, Michigan State University
- Organizers: Liz Stuart, Michael Rosenblum, Tom Louis