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

Ravi Varadhan

Division of Biostatistics & Bioinformatics
Department of Oncology
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
Baltimore, MD, USA

December 02, 2015

Evidence Gap

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)

Evidence from RCTs

- 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
- Premise: significant heterogeneity of treatment effect (HTE)

Motivation

- Older adults, with multiple diseases, are poorly represented in RCTs
- Evidence for most interventions is lacking in older adults
- For example, effectiveness of ACE-inhibitors for treatment of CHF in women older than 75 years of age
- *Applicability* of trial evidence to a specific target group (cf. *generalizability*)
- Need to incorporate information from non-RCTs

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)*.

Applicability of Evidence

- 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, although further considerations might be needed apart from an absence of HTE.
- On the other hand,$\# E$ is relatively large and that we found significant HTE. We would really question the applicability of evidence from E to P.
- A Solution: Standardization approach of Cole and Stuart (AJE 2010)

Applicability of Evidence

- What if evidence of lesser validity is available in P? One reason might be that the assignment of intervention Z was confounded.
- 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

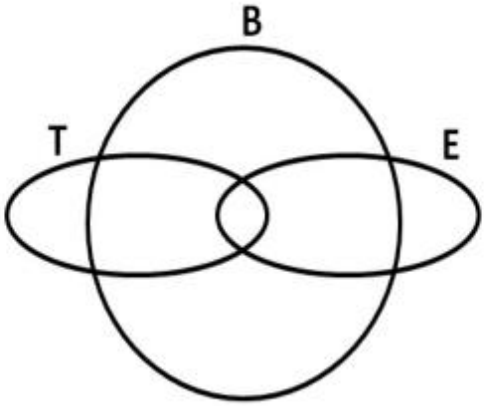
Cross-Design Synthesis

- 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

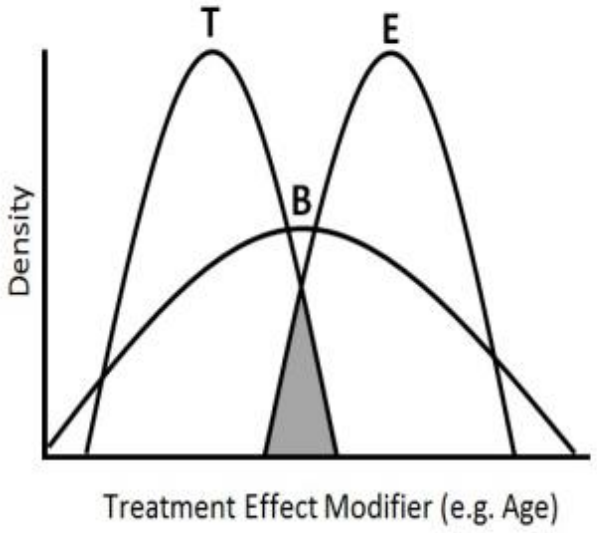
Essential Idea in CRAM: Calibration

- 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)



(B)

The 3 Studies

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

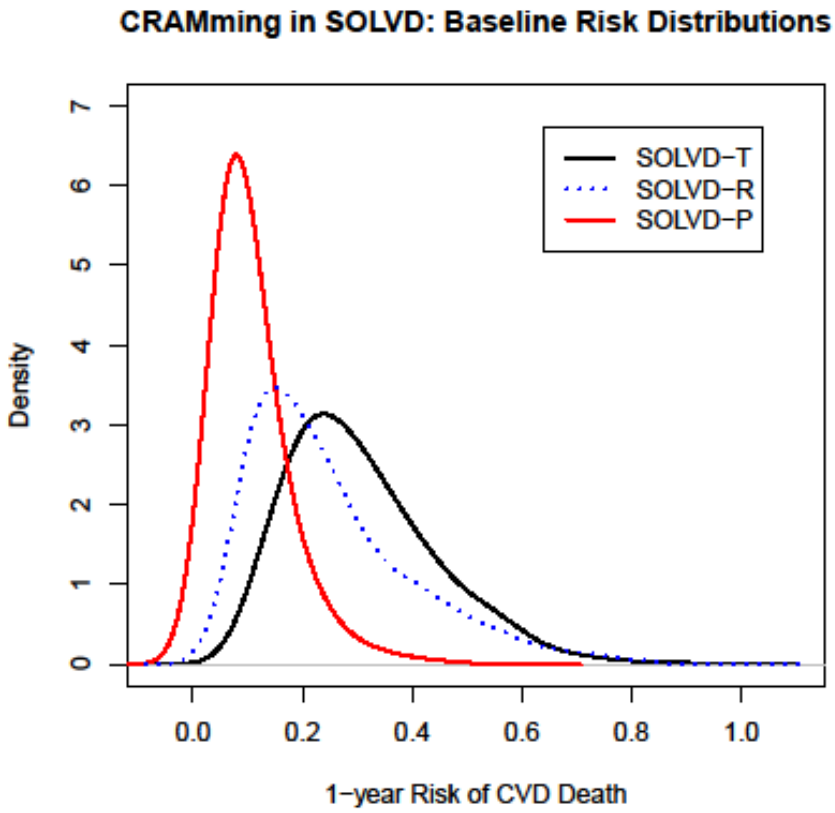
Major Steps in CRAM

- 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
- Estimate parameters of unmeasured confounding (solve an optimization problem) - calibration
- 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



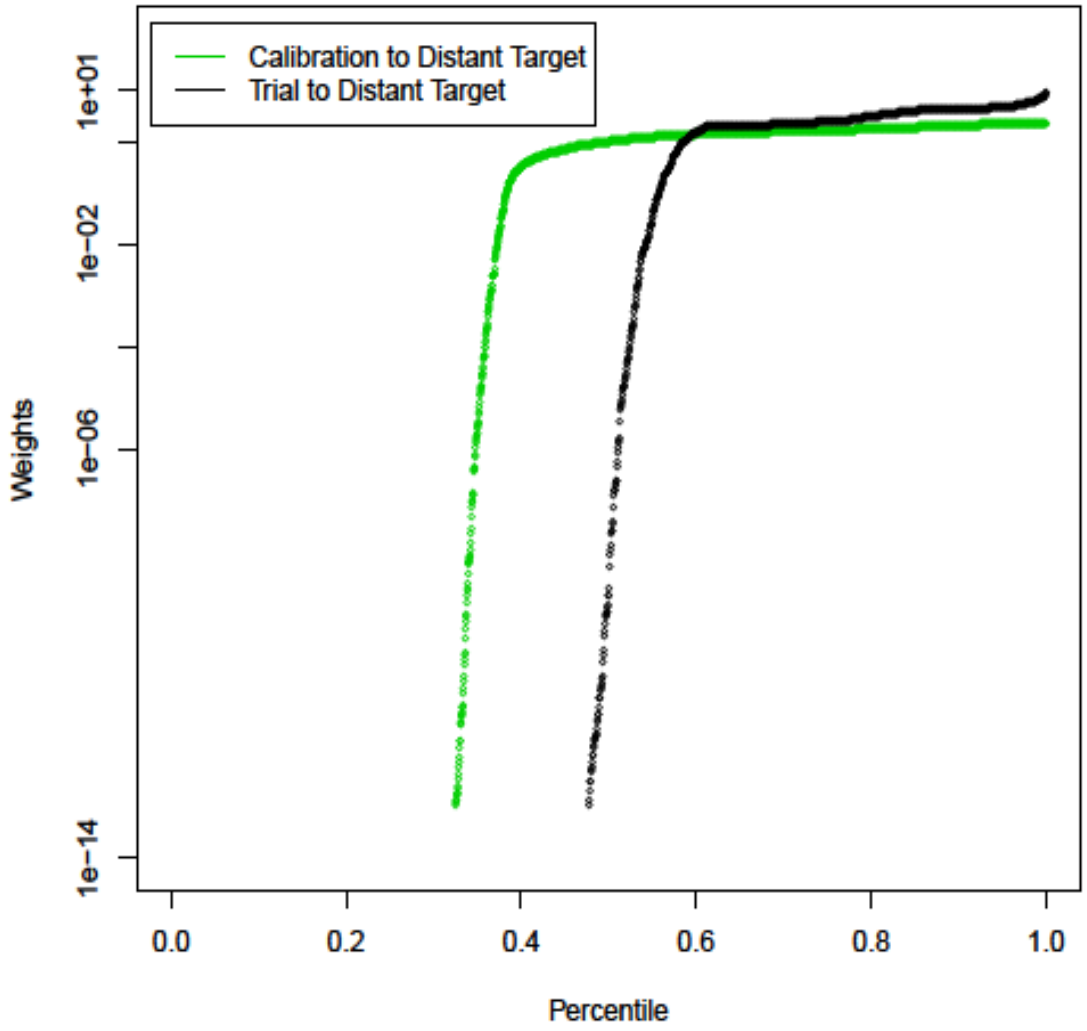
CRAM Results - ≥ 75 yr Women

Model	Standardization Covariate-based	Standardization Risk-based	CRAM
Estimate $\psi_{t=1}$[a]	-0.094 (0.44)	-0.64 (0.13)	
μ_1=-0.5	--	--	-0.43 (0.08)[d]
μ_1=-1.0	--	--	-0.44 (0.09)[f]

CRAM Results - Distant Target Sample

Model	True Effect	Standardization Covariate-based	Standardization, Risk-based	CRAM
Estimate $\psi_{t=1}$	-0.35 (0.19)	-0.55 (0.43)	-0.11 (0.19)	--
μ_1=-0.5		--	--	-0.31 (0.18)
μ_1=-1.0		--	--	-0.28 (0.18)

Weights for Distant Target Sample



CRAM Limitations

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

Acknowledgements

- Dr. Carlos Weiss, Michigan State University
- Agency for Healthcare Research Quality (Dr. Parivash Nourjah)
- Brookdale Leadership in Aging Fellowship