Closing the Evidence Gap: Calibrated Risk Adjusted Modeling (CRAM) for Cross-Design Synthesis

Ravi Varadhan\textsuperscript{1,2}

\textsuperscript{1}Division of Geriatric Medicine & Gerontology
\textsuperscript{2}Department of Biostatistics
Johns Hopkins University
Baltimore, MD, USA

Brookdale Foundation Retreat
Teaneck, NJ
June 07, 2011
Evidence from RCTs

- RCTs provide the most reliable evidence for approval of new treatments, informing clinical practice, and coverage decisions.
- Report average treatment effects (ATE) (the "evidence").
- Participants in RCTs are a select group, not representative of at-risk population.
- Concern that ATE is not generalizable.
- Significant heterogeneity of treatment effects (HTE).
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)

- 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 congestive heart failure in women older than 75 years of age
- Need to incorporate information from non-RCTs
Cross-Design Synthesis

- A method to project the treatment effect from a trial to a target group
- This is achieved by integrating trial and observational data
- RCT provides internally valid treatment effects
- Lack of applicability in RCT
- Target group is well-represented in an observational database (e.g. registry)
- Confounding in observational data (measured + unmeasured)
- Cross-design synthesis to exploit respective strengths
- Cross-design synthesis to mitigate respective limitations
Let $\beta_Z(E)$ be the estimate of efficacy of intervention $Z$ from an RCT conducted in sample $E$. Denote the larger at-risk population as $P$.

Is the evidence from $E$ applicable to $P$? Yes, if $E$ is “exchangeable” with $P$, i.e., it is reasonable to conceive of $E$ as a random sample of $P$.

Complete exchangeability of $E$ with $P$, which ensures applicability of evidence, is highly unlikely.

How can we, then, apply evidence from $E$ to $P$?
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 did find significant HTE. We would really question the applicability of evidence from \( E \) to \( P \).
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.
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
- Method for cross-design synthesis
- An extension of the sensitivity analysis approach for unmeasured confounding
Simulation results are encouraging

- Optimization algorithm (**bobyqa**) works well, but computationally intensive (ca. 5-10 mins for $N = 5000$)
- Able to obtain/recover proper UMC settings
- Can handle both binary and continuous UMCs
- Can have different UMC effects in exposed and unexposed groups
CRAM - Application

- Applying it to a real problem: 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
Acknowledgements

- Mentors: Drs. Jeremy Walston & Constantine Frangakis
- Dr. Karen Bandeen-Roche
- Drs. Chris Durso & Neal Fedarko
- Collaborators: Drs. Carlos Weiss, Cynthia Boyd, Jodi Segal, Albert Wu, Dan Scharfstein
- Agency for Healthcare Research Quality: Dr. Parivash Nourjah
Major Steps in CRAM

- Estimate the baseline risk of outcome (the basis of CRAM)
- Assumption: same baseline risk $\Rightarrow$ same treatment effect (w/o confounding)
- Determine interval of overlap, i.e. the calibration interval (we use a density quantile algorithm)
- Estimate treatment effect in calibration interval of RCT
- Find parameters of unmeasured confounding (solve an optimization problem)
- Using the CRAM parameters, estimate Tx effect in the projection interval of Observational study
Let the data generating model be:

\[ g(E[Y_i]) = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \gamma U_i \]

where \( Z_i \) = baseline risk, \( X_i \) = treatment indicator, and \( U_i \sim \text{Bernoulli}(Z_i p_1(X_i) + (1 - Z_i) p_0(X_i)) \), is an unmeasured confounder.

Given: \( \{ Y_i, X_i, Z_i \} \), and the RCT Tx effect, \( \hat{\beta} \)

- \( \hat{\alpha} = f_{X,Y,Z}(p_0, p_1, \gamma) \)
- Find \( \alpha_* = \arg\min_{p_0, p_1, \gamma} D(\hat{\alpha}, \hat{\beta}) = (\hat{\alpha} - \hat{\beta})^T S^{-1} (\hat{\alpha} - \hat{\beta}) \)

This is a non-trivial optimization problem since the objective function is a bit noisy.