Extending causal inferences from a randomized trial to a target population

Issa Dahabreh

Center for Evidence Synthesis in Health,
Brown University

issa_dahabreh@brown.edu

January 16, 2019

Disclaimer

- PCORI contract ME-1502-27794
- This presentation does not reflect the views of PCORI or its Methodology Committee
- Parts of what I will talk about summarize joint work with Sarah Robertson, Iman Saeed, Elisabeth Stuart, Miguel Hernan, Eric Tchetgen Tchetgen, Jamie Robins, ... All mistakes are my own.
- Work in progress

Overview

1. The problem of extending trial findings
2. Study designs for extending trial findings
   - Nested trial designs
   - Non-nested trial designs
3. Estimating the effect of treatment on non-participants
   - Identification
   - Estimation by modeling the outcome and the probability of trial participation
4. Simulation study
5. Application to the CASS study
6. Sensitivity analysis

The problem of extending trial findings

The problem

Trial participants often do not represent the population of (trial-eligible) patients seen in practice – the *target population*.

The problem

Trial participants often do not represent the population of (trial-eligible) patients seen in practice – the *target population*.

Even when evidence is available from a high-quality randomized trial, average treatment effects do not "transport" / "generalize" / "apply" to the target population.

The problem

Trial participants often do not represent the population of (trial-eligible) patients seen in practice – the *target population*.

Even when evidence is available from a high-quality randomized trial, average treatment effects do not "transport" / "generalize" / "apply" to the target population.

The problem

Trial participants often do not represent the population of (trial-eligible) patients seen in practice – the *target population*.

Even when evidence is available from a high-quality randomized trial, average treatment effects do not "transport" / "generalize" / "apply" to the target population.

We need methods to extend trial findings to the target population, under reasonable causal assumptions.

The problem

Trial participants often do not represent the population of (trial-eligible) patients seen in practice – the *target population*.

Even when evidence is available from a high-quality randomized trial, average treatment effects do not "transport" / "generalize" / "apply" to the target population.

We need methods to extend trial findings to the target population, under reasonable causal assumptions.

And to conduct sensitivity analysis when the assumptions fail.

Study designs for extending trial findings

Nested trial designs

Consider a *trial nested within a cohort of eligible patients* (including those who refuse randomization):

1. Identify patients meeting selection criteria
2. Collect baseline data on all patients
3. Ask for consent to randomization and randomize (marginally or conditionally)
4. Follow-up patients who consented to randomization

Data structure

Unit (i)	S	A	Y	X
1	1	0	y_1	x_1
$\vdots$	$\vdots$ 1	$\vdots$	$\vdots$	$\vdots$
n_0	1	0	y_{n_0}	x_{n_0}
$1+n_0$	1	1	y_{1+n_0}	x_{1+n_0}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n_1+n_0=n_{\text{RCT}}$	1	1	$y_{n_1+n_0}$	$x_{n_1+n_0}$
$1+n_{\text{RCT}}$	0	–	–	$x_{n_1+n_0+1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n_{\text{obs}}+n_{\text{RCT}}=n$	0	–	–	x_n

S is the indicator for consent to randomization; A is the random assignment indicator; X are baseline covariates; Y are observed outcomes. A and Y are missing for $S=0$

Perfect adherence; no dropout, no measurement error.

Non-nested trial designs

Append trial data to separately obtained sample from the target population.

Create an artificial composite dataset with the same data structure as in nested trial designs.

Causal quantities

Define Y_i^a as the "potential" ("counterfactual") outcome = the outcome that would be observed for the ith individual under treatment a.

For nested trials, 2 targets of inference:
$\mathbb{E}[Y^1 - Y^0]$ and $\mathbb{E}[Y^1 - Y^0|S = 0]$.

For artificial composite datasets, $\mathbb{E}[Y^1 - Y^0]$ is not identifiable;
but $\mathbb{E}[Y^1 - Y^0|S = 0]$ is.

In this talk, we focus on $\mathbb{E}[Y^1 - Y^0|S = 0]$.

Estimating the effect of treatment on non-participants

Identification

Because, $\mathbb{E}[Y^1 - Y^0|S=0] = \mathbb{E}[Y^1|S=0] - \mathbb{E}[Y^0|S=0]$, we just need to worry about the "potential outcome means" $\mathbb{E}[Y^a|S=0]$, $a=0,1$.

We need identifiability conditions that will allow us to express the potential outcome means as functions of the observed data.

Identifiability conditions

In the trial, to identify $\mathbb{E}[Y^a|S=1]$:

1. Consistency: if $A_i = a$, then $Y_i = Y_i^a$, for $a = 0, 1$
2. Conditional exchangeability: $Y^a \perp\!\!\!\perp A|X, S=1$
3. Positivity of treatment assignment: $0 < \Pr[A=a|X=x, S=1] < 1$ for every x that occurs with positive density in the trial

Additional conditions about the relationship of trial participants and non-participants, to identify $\mathbb{E}[Y^a|S=0]$:

4. Conditional transportability: $Y^a \perp\!\!\!\perp S|X$
5. Positivity of trial participation: $\Pr[S=1|X=x] > 0$ for every x that occurs with positive density in the non-randomized individuals

What do these conditions mean?

Consistency: if $A_i = a$, then $Y_i = Y_i^a$, for $a = 0, 1$
Trial participation does not have a direct effect on the outcome (e.g., no Hawthorne effect).

Conditional transportability: $Y^a \perp\!\!\!\perp S|X$
We know enough factors that determine the outcome so that trial participation itself is unimportant.

Positivity of trial participation: $\Pr[S = 1|X = x] > 0$
No subgroups of patients excluded systematically on the basis of effect modifiers.

Identification

Under our identifiability conditions,

$$\mathbb{E}[Y^a|S=0] = \mathbb{E}\left[\mathbb{E}[Y|X,S=1,A=a]\Big|S=0\right] \equiv \mu(a).$$

The treatment effect among non-randomized individuals can also be expressed as a function of the observed data,

$$\mathbb{E}[Y^1 - Y^0|S=0] = \mu(1) - \mu(0).$$

Estimation of $\mu(a)$

Estimators that rely on modeling

- the expectation of the outcome among $S = 1$ and $A = a$, or
- the probability of $S = 1$ and $A = a$,

conditional on covariates.

Estimation by outcome modeling

Modeling the expectation of the outcome among $S = 1$ and $A = a$,

$$\widehat{\mu}_{\text{OM}}(a) = \left\{ \sum_{i=1}^{n} (1 - S_i) \right\}^{-1} \sum_{i=1}^{n} (1 - S_i) g_a(X_i; \widehat{\beta}),$$

where $g_a(X; \widehat{\beta})$ is an estimator of $\mathbb{E}[Y|X, S = 1, A = a]$, $a = 0, 1$.

Converges in probability to $\mu(a)$ when $g_a(X; \beta)$ is correctly specified.

"Regression-based extrapolation."

Estimation by trial participation modeling

Modeling the probability of $S = 1$, $A = a$,

$$\widehat{\mu}_{\text{IPW1}}(a) = \left\{ \sum_{i=1}^{n} (1 - S_i) \right\}^{-1} \sum_{i=1}^{n} \widehat{w}_a(S_i, X_i, A_i) Y_i,$$

where

$$\widehat{w}_a(S_i, X_i, A_i) = S_i I(A_i = a) \frac{1 - p(X_i; \widehat{\gamma})}{p(X_i; \widehat{\gamma}) e_a(X_i; \widehat{\theta})},$$

$p(X; \widehat{\gamma})$ is an estimator for $\Pr[S = 1|X]$, and
$e_a(X; \widehat{\theta})$ is and estimator for $\Pr[A = a|X, S = 1], a = 0, 1$.

Converges in probability to $\mu(a)$ when $p(X; \gamma)$ is correctly specified.
$e_a(X; \theta)$ is never misspecified; "true" value can be used.

Estimation by trial participation modeling

A variant where the weights are normalized to sum to 1 often works better:

$$\widehat{\mu}_{\text{IPW2}}(a) = \left\{ \sum_{i=1}^{n} \widehat{w}_a(S_i, X_i, A_i) \right\}^{-1} \sum_{i=1}^{n} \widehat{w}_a(S_i, X_i, A_i) Y_i.$$

Can be obtained by weighted least squares regression of Y on A among trial participants, with weights $\widehat{w}_a(S, X, A)$.

Augmenting the weighted estimators

This is a missing data problem.

Think of $\widehat{\mu}_{\text{OM}}$ and $\widehat{\mu}_{\text{IPW}}$ as imputation and probability-of-missingness estimators, respectively.

We can "augment" the weighted estimator using the outcome model to gain efficiency and robustness.

In-sample one-step doubly robust estimator (AUG1)

Using the efficient influence function for $\mu(a)$, we obtain the augmented estimator

$$\widehat{\mu}_{\text{AUG1}}(a) = \left\{\sum_{i=1}^{n}(1-S_i)\right\}^{-1}\sum_{i=1}^{n}\left\{\widehat{w}_a(S_i,X_i,A_i)\left[Y_i - g_a(X_i;\widehat{\beta})\right] + (1-S_i)g_a(X_i;\widehat{\beta})\right\}$$

Remarkably, this converges in probability to $\mu(a)$, when either $p(X;\gamma)$ or $g_a(X;\beta)$ is correctly specified.

Another in-sample one-step doubly robust estimator (AUG2)

An alternative augmented estimator normalizes the weights to sum to 1:

$$\widehat{\mu}_{\text{AUG2}}(a) = \left\{ \sum_{i=1}^{n} w_a(S_i, X_i, A_i) \right\}^{-1} \sum_{i=1}^{n} w_a(S_i, X_i, A_i) \left[Y_i - g_a(X_i; \widehat{\beta}) \right] + \left\{ \sum_{i=1}^{n} (1 - S_i) \right\}^{-1} \sum_{i=1}^{n} (1 - S_i) g_a(X_i; \widehat{\beta}).$$

This also converges in probability to $\mu(a)$ when either $p(X; \gamma)$ or $g_a(X; \beta)$ is correctly specified.

Weighted multi-variable regression-based DR estimator (AUG3)

Use a regression model, $g_a(X;\beta)$, for $\mathbb{E}[Y|X, A=a, S=1]$ where Y belongs to the linear exponential family.

Estimating the parameters of this model using standard methods (e.g., OLS, MLE, quasilikelihood), weighting by $\widehat{w}_a(S, X, A)$, and standardizing over the covariate distribution of the non-participants, has the double robustness property.

Simulation study

Simulation set up

Generated data for a composite dataset of

- Observational study sample of 5000 or 10000
- Randomized trials of different sample sizes (500, 1000, 5000)

Settings chosen to generate

- strong confounding in the observational study and
- strong effect modification by a predictor of trial participation

Analyzed data using different estimators.

Data generation

n_{RCT} of 250, 500, or 1000

n_{obs} of 2,500, 5,000 or 10,000

Baseline covariates for $S = 1$: $X_{1i} \sim \mathcal{N}(\mu, 1)$ and $X_{ji} \sim \mathcal{N}(1, 1)$; $j = 2, 3$; $i = 1, ..., n_{\text{RCT}}$.

Baseline covariates for $S = 0$: $X_{ji} \sim \mathcal{N}(0, 1)$; $j = 1, 2, 3$; $i = n_{\text{RCT}} + 1, ..., n_{\text{RCT}} + n_{\text{obs}}$; μ controls selection on X_1 and we used values 0 and 1, representing no and strong selection.

Logistic regression of S on $(X_1, \ldots, X_3)$ is correctly specified.

Outcomes: $Y_i = \tau A_i + \phi X_{1i} A_i + \psi X_i + \epsilon_i, \psi = (1, 1, 1)$, and $\epsilon_i \sim \mathcal{N}(0, 1)$.

Examined scenarios with different levels of effect modification by setting ϕ to 0 or 1; we set the "main" treatment effect to $\tau = 1$ in all scenarios.

Bias of estimators across different magnitudes of effect modification (ϕ) and sample sizes, under strong selection on the effect modifier ($\mu = 1$).

n_{RCT}	n_{obs}	ϕ	μ	Trial	OM	IPW1	IPW2	AUG1	AUG2	AUG3
250	2500	0	1	-0.002	-0.003	0.010	-0.004	-0.001	-0.001	-0.000
250	2500	1	1	0.998	0.002	0.008	0.081	-0.001	0.001	0.002
250	5000	0	1	0.000	0.003	-0.023	0.001	0.012	0.008	0.005
250	5000	1	1	1.003	-0.002	-0.007	0.068	-0.006	-0.007	-0.008
250	10000	0	1	-0.002	-0.001	-0.019	-0.003	-0.004	-0.005	-0.003
250	10000	1	1	0.999	0.005	0.000	0.078	-0.001	0.002	0.003
500	2500	0	1	0.003	0.002	-0.008	-0.001	0.001	0.003	0.003
500	2500	1	1	1.000	0.000	0.022	0.067	0.001	-0.000	0.000
500	5000	0	1	0.001	-0.001	-0.012	-0.005	-0.007	-0.005	-0.003
500	5000	1	1	1.003	-0.000	-0.004	0.054	-0.003	-0.002	-0.001
500	10000	0	1	-0.003	-0.001	-0.001	0.002	0.001	0.000	0.000
500	10000	1	1	1.000	-0.001	-0.013	0.045	-0.004	-0.003	-0.003
1000	2500	0	1	-0.001	-0.001	-0.014	-0.003	-0.000	-0.001	-0.002
1000	2500	1	1	1.000	-0.001	0.010	0.031	-0.001	0.001	0.002
1000	5000	0	1	-0.001	-0.003	0.005	0.006	-0.000	-0.001	-0.002
1000	5000	1	1	1.000	-0.001	-0.005	0.021	-0.004	-0.004	-0.005
1000	10000	0	1	-0.000	-0.001	0.013	0.010	0.004	0.003	0.002
1000	10000	1	1	1.000	-0.000	-0.011	0.014	-0.005	-0.003	-0.003

Variance of estimators across different magnitudes of effect modification (ϕ) and sample sizes, under strong selection on the effect modifier ($\mu = 1$).

n_{RCT}	n_{obs}	ϕ	μ	Trial	OM	IPW1	IPW2	AUG1	AUG2	AUG3
250	2500	0	1	0.063	0.067	4.032	0.999	0.328	0.164	0.177
250	2500	1	1	0.087	0.067	4.698	1.341	0.300	0.164	0.172
250	5000	0	1	0.063	0.067	2.760	0.961	0.276	0.164	0.174
250	5000	1	1	0.089	0.068	5.324	1.387	0.342	0.166	0.176
250	10000	0	1	0.063	0.067	2.897	0.988	0.298	0.162	0.174
250	10000	1	1	0.090	0.066	3.295	1.399	0.291	0.163	0.173
500	2500	0	1	0.032	0.034	1.572	0.712	0.169	0.098	0.092
500	2500	1	1	0.044	0.034	2.515	0.962	0.169	0.100	0.095
500	5000	0	1	0.032	0.033	1.443	0.699	0.157	0.098	0.094
500	5000	1	1	0.043	0.033	2.016	0.962	0.143	0.096	0.092
500	10000	0	1	0.032	0.033	1.431	0.700	0.150	0.099	0.093
500	10000	1	1	0.044	0.032	2.321	0.945	0.161	0.096	0.093
1000	2500	0	1	0.016	0.016	1.134	0.457	0.077	0.055	0.049
1000	2500	1	1	0.022	0.016	1.094	0.617	0.087	0.055	0.049
1000	5000	0	1	0.016	0.016	0.658	0.451	0.075	0.056	0.049
1000	5000	1	1	0.023	0.016	1.070	0.644	0.088	0.057	0.050
1000	10000	0	1	0.016	0.016	0.798	0.462	0.077	0.057	0.050
1000	10000	1	1	0.022	0.016	1.088	0.667	0.081	0.056	0.049

Application to the CASS study

The Coronary Artery Surgery Study (CASS)

Comprehensive cohort study of medical therapy vs. CABG for chronic coronary artery disease; 2099 patients met inclusion criteria

- 780 randomized (390 medical – 390 CABG)
- 1319 declined randomization (745 medical – 570 CABG; 4 excluded)

The Coronary Artery Surgery Study (CASS)

Comprehensive cohort study of medical therapy vs. CABG for chronic coronary artery disease; 2099 patients met inclusion criteria

- 780 randomized (390 medical – 390 CABG)
- 1319 declined randomization (745 medical – 570 CABG; 4 excluded)

A near-perfect setting for transportability

- Same research centers (including treating surgeons)
- Common protocol (e.g., followup procedures; outcome ascertainment)
- Near complete followup
- A large number of baseline covariates with complete data
- Not too small

How do the estimates compare across methods?

Estimated 10-year survival probabilities for surgery vs. medical therapy among non-participants in CASS.

Estimator	**Survival probability (Surgery)**	**Survival probability (Medical)**
Trial-only	17.4% (13.6%, 21.4%)	20.4% (16.3%, 24.6%)
OM	18.9% (13.9%, 22.7%)	20.1% (15.9%, 24.5%)
IPW1	18.2% (13.9%, 22.7%)	20.1% (15.9%, 24.4%)
IPW2	18.2% (14.6%, 23.5%)	20.1% (16.0%, 24.4%)
AUG1	18.7% (14.5%, 23.3%)	20.1% (16.0%, 24.4%)
AUG2	18.7% (14.5%, 23.3%)	20.1% (16.0%, 24.4%)
AUG3	18.7% (14.4%, 23.2%)	20.0% (15.9%, 24.3%)

These results are similar to an analysis using data only from the non-randomized patients.

Sensitivity analysis

What if the assumptions fail?

Suppose that the conditional generalizability assumption does not hold, so that $Y^a \not\perp\!\!\!\perp S|X$, or, equivalently, for binary S,

$$f_{Y^a|X,S}(y|x, s = 0) \neq f_{Y^a|X,S}(y|x, s = 1).$$

We can parameterize violations of the conditional generalizability assumption using the exponential tilt model

$$f_{Y^a|X,S}(y|x, s = 0) \propto e^{\eta_a q(y)} f_{Y^a|X,S}(y|x, s = 1),$$

$\eta_a \in \mathbb{R}, a \in \mathcal{A}$, with q a fixed increasing function.

This condition replaces the conditional transportability condition.

Sensitivity analysis model

Under consistency of potential outcomes, exchangeability and positivity of treatment assignment in the trial, we obtain

$$f_{Y^a|X,S}(y|x,s=0) = \frac{e^{\eta_a q(y)} f_{Y|X,S,A}(y|x,s=1,a)}{\mathbb{E}[e^{\eta_a q(Y)}|X,S=1,A=a]}, \eta_a \in \mathbb{R}, a \in \mathcal{A}.$$

The data do not contain outcome information for non-randomized individuals $S=0$.

We cannot nonparametrically identify $f_{Y^a|X,S}(y|x,s=0)$ and η_a is not identifiable.

Sensitivity analysis model

We can now re-express $\mathbb{E}[Y^a|S=0]$ as

$$\mathbb{E}\left[\frac{\mathbb{E}[Ye^{\eta_a q(Y)}|X,S=1,A=a]}{\mathbb{E}[e^{\eta_a q(Y)}|X,S=1,A=a]}\Big|S=0\right] \equiv \mu(a,\eta_a).$$

We can use this re-expression to conduct sensitivity analyses for different, sufficiently dispersed, values of η_a.

Sensitivity analysis estimator

For a binary outcome, using the influence function for $\mu(a, \eta_a)$, and assuming q is the identity function, we obtain the following estimator

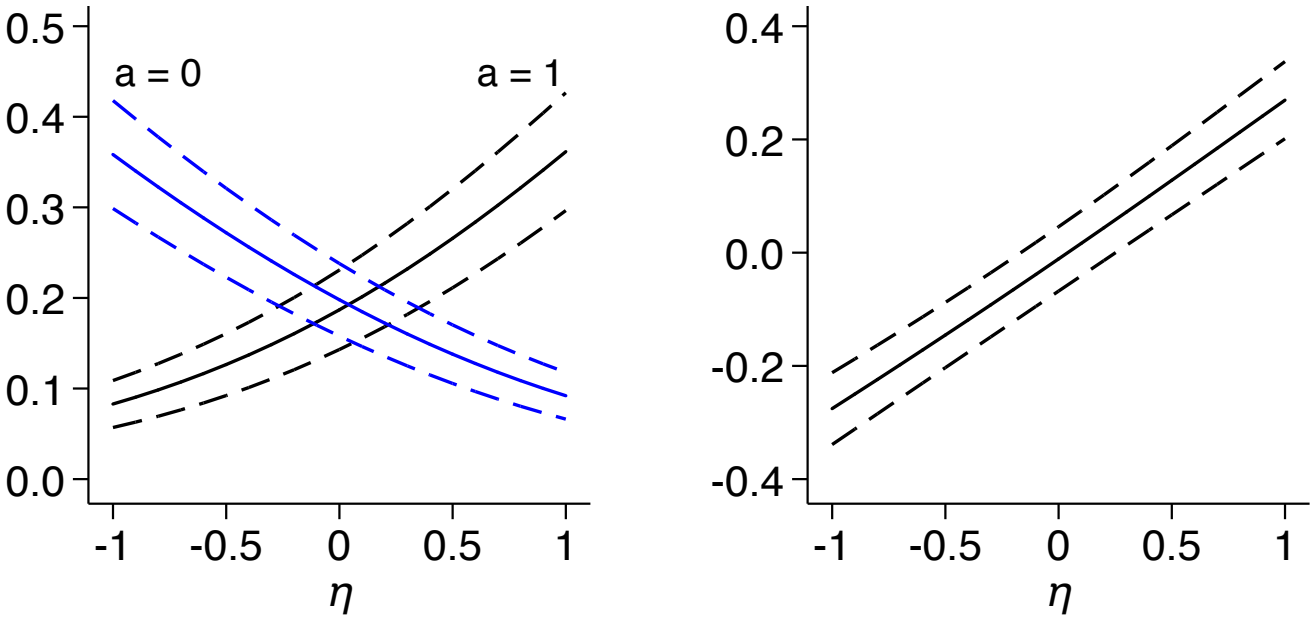
$$\widehat{\mu}_{\text{sens}}(a, \eta_a) = \left\{ \sum_{i=1}^{n} (1 - S_i) \right\}^{-1} \sum_{i=1}^{n} \left\{ \frac{(1 - S_i) e^{\eta_a} \widehat{g}_a(1|X_i)}{e^{\eta_a} \widehat{g}_a(1|X_i) + \widehat{g}_a(0|X_i)} \right.$$
$$\left. + \frac{\widehat{w}_a(S_i, X_i, A_i) e^{\eta_a Y_i}}{e^{\eta_a} \widehat{g}_a(1|X_i) + \widehat{g}_a(0|X_i)} \left\{ Y_i - \frac{e^{\eta_a} \widehat{g}_a(1|X_i)}{e^{\eta_a} \widehat{g}_a(1|X_i) + \widehat{g}_a(0|X_i)} \right\} \right\},$$

where $\widehat{g}_a(1|X)$ is an estimator for $\Pr[Y = 1|X, S = 1, A = a]$, $\widehat{g}_a(0|X) = 1 - \widehat{g}_a(1|X)$, and all other notation is as defined earlier.

Sensitivity analysis in the CASS

Sensitivity analysis curves for the potential outcome means (left panel) and the average treatment effect (right panel). Solid lines connect point estimates; dashed lines connect point-wise 95% bootstrap intervals.

We set $\eta_1 = -\eta_0 = \eta$: when randomized individuals have a higher outcome probability under $a = 1$ compared to non-randomized individuals, they also have lower outcome probability under $a = 0$.



Conclusions

Outcome model-based, probability of participation-based, and doubly robust estimators can be used to extend trial findings to a new target population.

Sensitivity analysis helps to control overconfidence.

Useful for estimating parameters for structural models that combine information from multiples sources.

More work needed to examine performance under model misspecification and methods for model selection.