Causal inference methods for combining randomized trials and observational studies

WCM Computing Club

Caroline Andy, MS

Recent research focus

- Division of Comparative Effectiveness and Outcomes Research - Jalali Lab
- Substance use research
- Methods development bridging econometrics and biostatistics
  - Data fusion causal inference to transport inferences from RCTs onto more representative observational cohorts
  - Considerations for data missingness

Problem setting

- RCTs isolate the effects of the treatment from that of unwanted co-occurring effects (confounding), but they may suffer from unrepresentativeness, and thus lack external validity.
- Observational studies, while representative, open the door to confounding bias, and therefore may lack internal validity.
- Colnet et al. 2024 summarizes current methods for combining insights gained from RCTs and observational research. This opens the door to new tools for:
  - Transporting inferences from RCTs onto an external representative sample of a target population of interest.
  - Making observational evidence more credible using RCT to ground observational analysis, such as detecting a confounding bias.
  - Improving statistical efficiency of ATE estimation.

Methods for when we have observational covariates but no treatment or outcomes

Estimation methods to generalize trial findings to a target population of interest

- Inverse probability of sampling weighting (IPSW) estimator
- Stratification
- Plug-in g-formula estimator
- Calibration weighting
- Doubly robust estimators
  - Augmented IPSW (AIPSW)
  - Augmented CW (ACW)

Inverse probability of sampling weighting (IPSW)

- IPSW can be seen as the counterpart of IPTW method of estimating the ATE in observational studies by controlling for confounding.
- The IPSW estimator for the ATE is defined as the weighted difference of average outcomes between the treated and control group in the trial.
- The observations are weighted by the inverse odds $1/\alpha(x) = \pi_O(x)/\pi_R(x)$ to account for the shift of the covariate distribution from the RCT sample.
- In the IPTW estimator, each observation is weighted by the probability to be treated, whereas in IPSW, each observation is weighted by the inverse of the odds of the probability to be in the trial sample.
- **Definition 1** (Inverse probability of sampling weighting - IPSW). The IPSW estimator is defined as follows:

$$\hat{\tau}_{\text{IPSW},n,m} = \frac{1}{n}\sum_{i=1}^{n} \frac{n}{m} \frac{Y_i}{\hat{\alpha}_{n,m}(X_i)} \left(\frac{A_i}{e_1(X_i)} - \frac{1 - A_i}{1 - e_1(X_i)} \right)$$

where $\hat{\alpha_{n,m}}$ is an estimate of the odds of the indicatrix being in the RCT.

Inverse probability of sampling weighting (IPSW)

- Similar to IPTW estimators, IPSW estimators are known to be highly unstable, especially when the weights are extreme.
  - This can occur if the observational study contains units with very small probability of being in the trial.
- Normalized weights can be used to overcome this issue.
- Still, the major challenge remains that IPSW estimators require a correct model specification of the weights.
- Avoiding this problem requires either very strong domain expertise or turning to doubly robust methods.

Stratification

- Stratification has been proposed as a solution to mitigate the risks of extreme weights in the IPSW formula.
- First, one estimates the conditional odds $\hat{\alpha}_{n,m}$ in the same manner as for the IPSW previously detailed. Then based on the values of the conditional odds obtained, L strata are defined.
- For each strata l one has to compute the average effect on this strata defined as $Y(1)_l - Y(0)_l$, where $Y(a)_l$ denotes the average value of the outcome for units with treatment a in stratum l in the RCT.
- The generalized ATE is defined by the aggregation of the treatment effect estimates on each strata weighted by the proportion of the strata in the target population m_l/m, where m_l is the number of individuals in strata l in the target sample.

$$\hat{\tau}_{\text{strat},n,m} = \sum_{l=1}^{L} \frac{m_l}{m} \underbrace{\left(\overline{Y(1)}_l - \overline{Y(0)}_l\right)}_{from\ RCT}.$$

Plug-in g-formula estimators

- G-formula estimators fit a model of the conditional outcome mean among trial participants, rather than modeling the probability of trial participation. Then the outcomes are averaged (marginalized) over the empiral covariate distribution of the target population.
- **Definition 2** (Plug-in g-formula). The plug-in g-formula (or outcome model-based) estimator is defined as:

$$\hat{\tau}_{G,n,m} = \frac{1}{m} \sum_{i=n+1}^{n+m} \left(\hat{\mu}_{1,1,n}(X_i) - \hat{\mu}_{0,1,n}(X_i) \right),$$

where $\hat{\mu}_{a,1,n}(X_i)$ is an estimator of $\mu_{a,1}(X_i)$ fitted using the RCT data.

- In practice, any model can be used to fit $\mu_{a,1}(X_i)$, for e.g. standard ordinary least squares (OLS).
- If model is correctly specified, the estimator is consistent.

Calibration weighting: balancing covariates

- Beyond propensity scores, other schemes use sample reweighting. The RCT sample is calibrated such that after calibration, the covariate distribution of the sample empirically matches the target population.
- **Definition 3** (Calibration weighting - CW). Let g(X) be a vector of functions of X to be calibrated, e.g., the moments, interactions, and non-linear transformations of components of X. Then assign a weight w_i to each subject i in the RCT sample (obtained from solving optimization problem for w_i parameters). The CW estimator is given by:

$$\hat{\tau}_{CW,n,m} = \sum_{i=1}^{n} \hat{\omega}_{n,m}(X_i) Y_i \left(\frac{A_i}{e_1(X_i)} - \frac{1 - A_i}{1 - e_1(X_i)} \right),$$

where weights w_i are assigned to each subject i in the RCT sample, estimating by solving a calibration optimization problem for the balancing constraint

Calibration weighting: balancing covariates

- For an intuitive understanding of the calibration weighting framework, consider g(X) = X. In such a setting, the balancing constraint is forcing the means of the observational data and of the RCT to be equal after reweighting.
- More complex constraints can enforce balance on higher-order moments.
- The CW estimator $\hat{\tau}_{CW,n,m}$ is doubly robust in that it is a consistent estimator for τ if the selection score of RCT participation follows a log-linear model or if the CATE is linear in g(X), though not necessarily both.

Doubly robust estimators

- Doubly robust estimators will product a consistent estimator of the underlying ATE if at least one of the two models is specified correctly.
- The model for the expectation of the outcomes among randomized individuals (plug-in g-formula estimator) and the model for the probability of trial participation (IPSW estimator) can be combined to form an Augmented IPSW estimator (AIPSW).
- **Definition 4** (Augmented IPSW - AIPSW). The augmented IPSW estimator, denoted $\hat{\tau}_{\text{AIPSW},n,m}$ is defined as

$$\hat{\tau}_{AIPSW,n,m} = \frac{1}{n}\sum_{i=1}^{n} \frac{n}{m\,\hat{\alpha}_{n,m}(X_i)} \left(\frac{A_i\left(Y_i - \hat{\mu}_{1,1,n}(X_i)\right)}{e_1(X_i)} - \frac{(1-A_i)\left(Y_i - \hat{\mu}_{0,1,n}(X_i)\right)}{1-e_1(X_i)} \right) + \frac{1}{m}\sum_{i=n+1}^{m+n} \left(\hat{\mu}_{1,1,n}(X_i) - \hat{\mu}_{0,1,n}(X_i)\right),$$

where $\hat{\mu}_{a,1}$ are estimated on the RCT sample and $\hat{\alpha}_{n,m}$ on the concatenated RCT and observational samples.

Doubly robust estimators

- This estimator is also shown to be asymptotically normal when both the outcome mean and conditional odds model are consistently estimated at least at rate $n^{1/4}$.
- ML tools are tempting to avoid model misspecification when estimating nuisance parameters. However, this requires caution, such as using cross-fitting due to overfitting and regularization.
- **Definition 4** (Augmented CW - ACW). The augmented IPSW estimator, denoted $\hat{\tau}_{ACW,n,m}$ is defined as

$$\hat{\tau}_{ACW,n,m} = \sum_{i=1}^{n} \hat{\omega}_{n,m}(X_i) \left(\frac{A_i \left(Y_i - \hat{\mu}_{1,1,n}(X_i)\right)}{e_1(X_i)} - \frac{(1-A_i)\left(Y_i - \hat{\mu}_{0,1,n}(X_i)\right)}{1-e_1(X_i)} \right) + \frac{1}{m} \sum_{i=n+1}^{m+n} \left(\hat{\mu}_{1,1,n}(X_i) - \hat{\mu}_{0,1,n}(X_i)\right),$$

where $\hat{\mu}_{a,1,n}$ are estimated on the RCT sample.

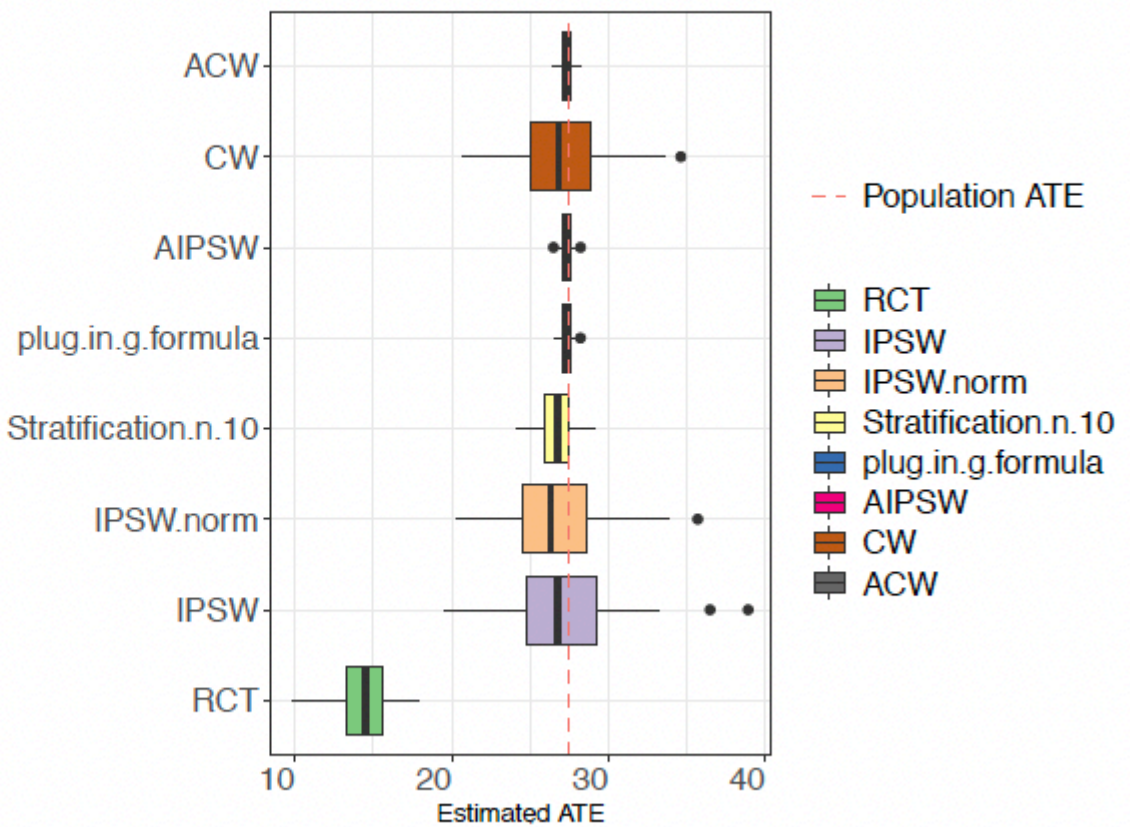
Doubly robust estimators

- Relatively new method proposed by Lee et al. (2021).
- The $\hat{\tau}_{ACW,n,m}$ achieves double robustness and local efficiency.
- Moreover, the convergence rate of the ACW estimator corresponds to the product of the convergence rates of the nuissance parameters, enabling the use of ML estimation of nuisance functions while preserving the √n-consistency of the ACW estimator, when both the outcome mean and calirbation weights model are consistently estimated at a rate $n^{1/4}$.
- **Therefore, when attempting to update RCT data given only the covariates (no treatment, outcomes) in an observational dataset, seems like the best approach would be using augmented CW (ACW) as our doubly robust causal effect estimator. (To be formally tested using simulation data).**

Simulation study of generalization estimators

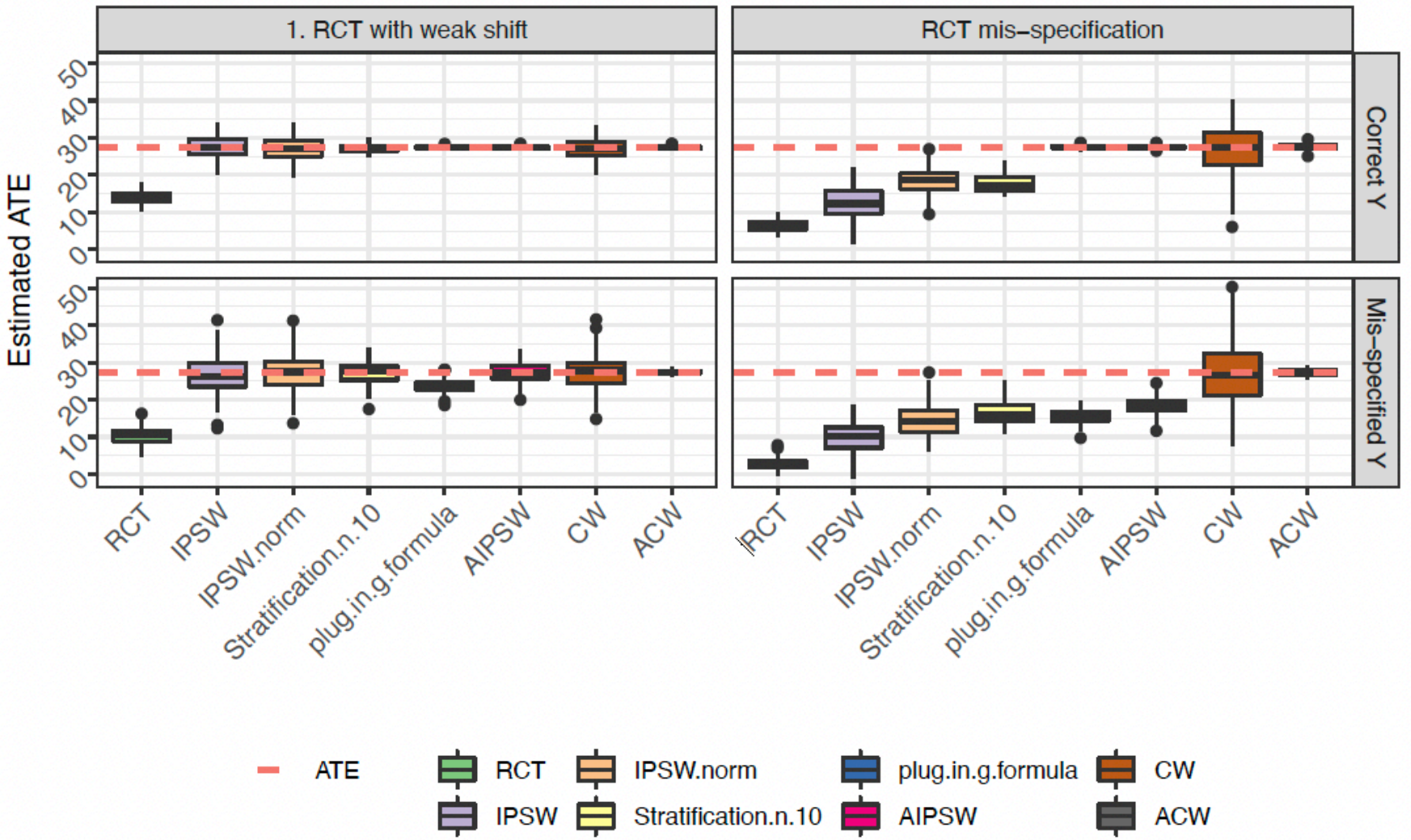
- Here we evaluate estimator performance under several misspecification patterns using simulated data.
- R packages for all reported estimators provided. Simulation protocols outlined.
- **Scenario 1. Well-specified models.**

Figure 5: **Well-specified model** Estimated ATE with the inverse propensity of sampling weighting with and without weights normalization (IPSW and IPSW.norm; Definition 2), stratification (with 10 strata; Definition 3), plug-in g-formula (Definition 4), calibration weighting (CW; Definition 5), augmented IPSW (AIPSW; Definition 6) and ACW (Definition 7)) over 100 simulations.



Simulation study of generalization estimators

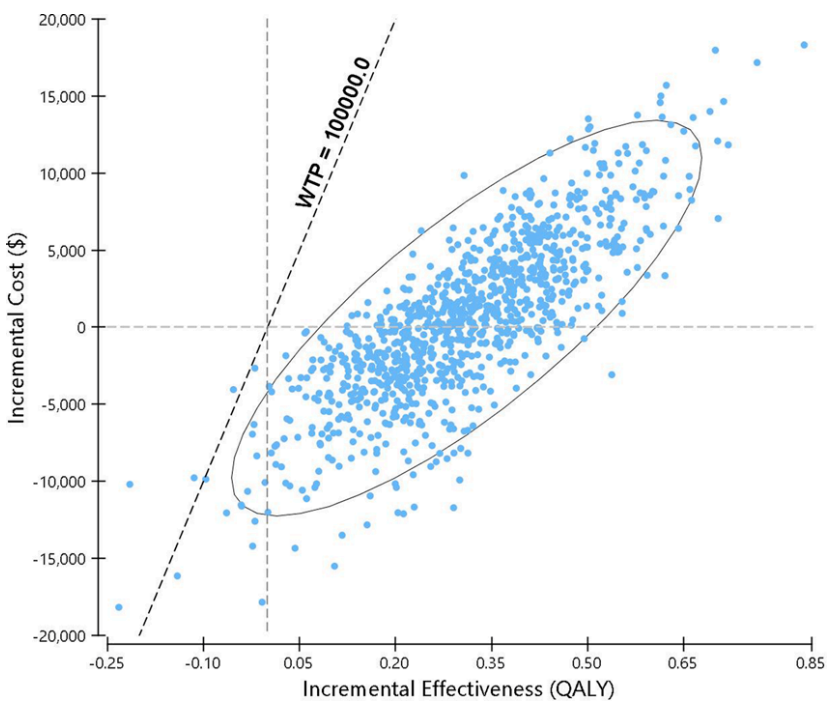
- **Scenario 2. Misspecification of the sampling propensity score or outcome model.**



Simulation study of generalization estimators

- When sampling propensity score model is misspecified, IPSW estimators are biased; when outcome model is misspecified, the plug-in g-estimator is biased. In both settings, the doubly robust AIPSW is unbiased and robust to model misspecification.
- When both models are misspecified, all estimators are biased except the CW and ACW estimators.
- **Based on the simulated results, the best estimator to use will be the ACW estimator.**

Applied study of generalization estimators

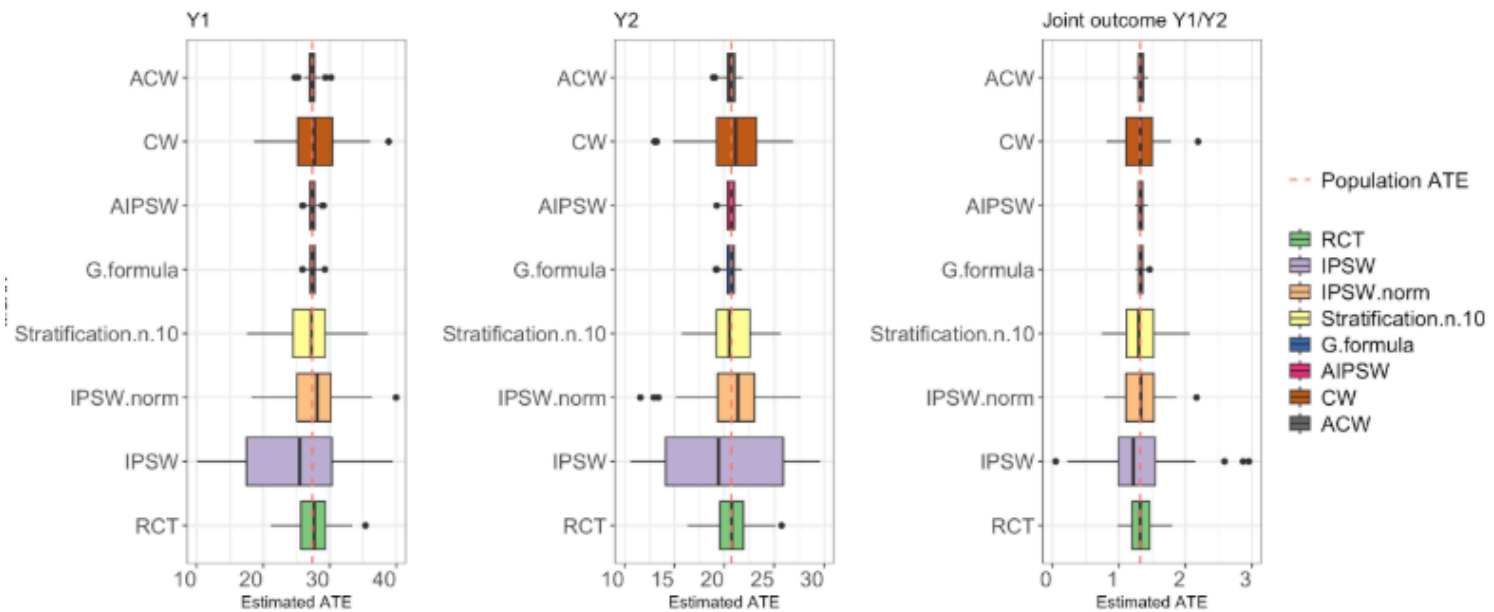
- Dealing with missing values when generalizing a treatment effect remains an open research question.
- The extent to which these data fusion causal effect estimators can be used to infer the joint distribution of incremental cost and effect outcomes is unclear.
- $$\text{ICER} = \frac{\Delta[\text{Cost}]}{\Delta[\text{Effectiveness}]}$$
- 

Andy/Jalali contribution

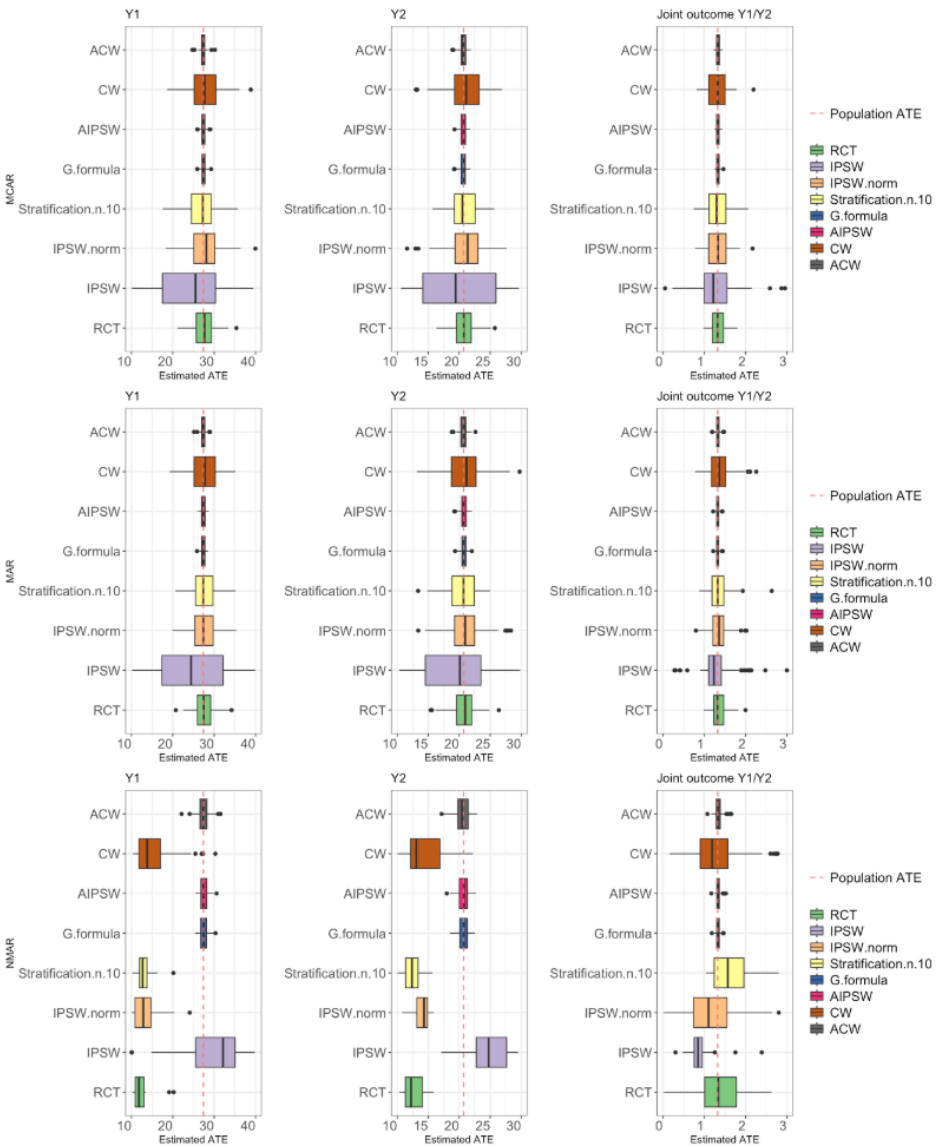
- Simulated estimator performance for a joint outcome (ie., ICER) = Y1/Y2 where Y1 and Y2 are random independent variables.
- Introduced missingness patterns (MCAR, MAR, NMAR).
- Introduced correlation between Y1 and Y2 components.

Estimator performance for joint outcome

- **Estimator performance for Y1, Y2, and joint outcome Y1/Y2**



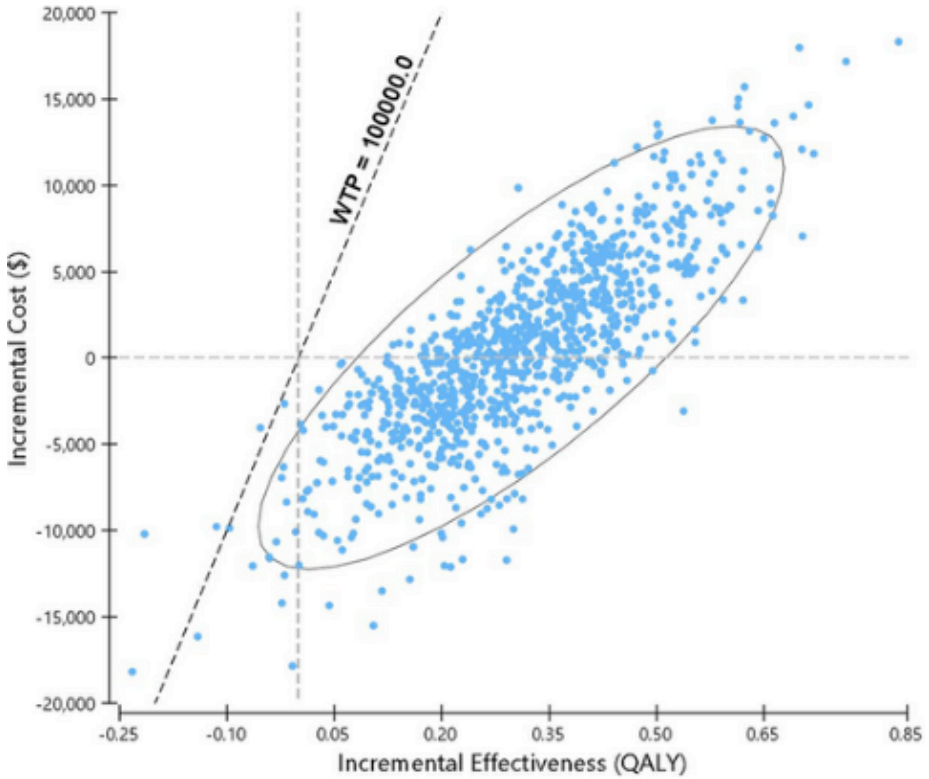
Joint outcome under missingness

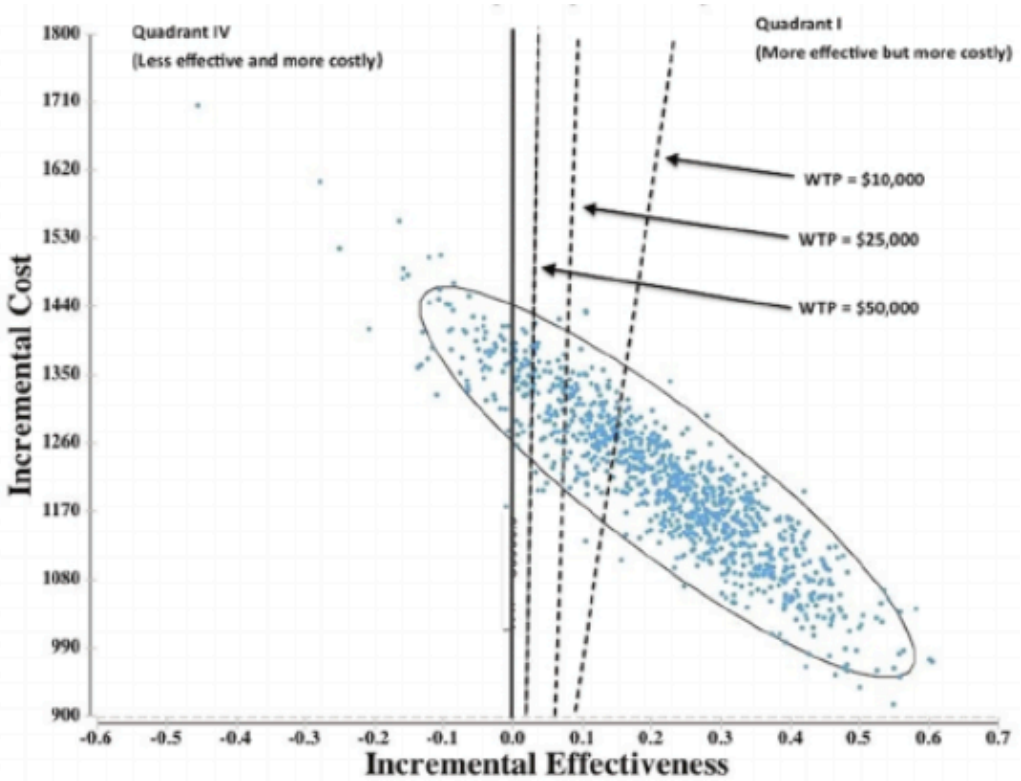


Correlation between Y1 and Y2

- Correlation between ICER components is a common consideration for comparative effectiveness work.
  - Positive correlation of Y1 and Y2: A new drug improves patient survival but is also more expensive than the standard treatment.
  - Negative correlation of Y1 and Y2: A minimally invasive surgery improves recovery time but is more expensive. As cost increases, the incremental gains in effectiveness slow down.
- Are the data fusion causal effect estimators considered sensitive to correlation?

Correlation between Y1 and Y2





Summary of findings

- ACW estimator is consistent and precise in estimating joint outcomes, even in settings of missingness and correlated components.
- Estimator variance is highly sensitive to correlation in outcome components.
- Next steps:
  - Compare estimator variance to true variance as calculated under Fieller's theorem (for ratio of two means given elliptical distribution) or others.
  - Explore settings of model misspecification.

Collaboration

- Contact me if you're interested in this work!
  - Theory (simulation) stuff, or let's find applied datasets (RCT with questionable representativeness and observational dataset) to update/transport inferences.