Data Integration in Causal Inference

Xu Shi, Ziyang Pan, Wang Miao

September 12, 2023

Questions for Discussion

- What are some examples of causal inference that required data integration?
- Summarize the general ideas of the paper.
- What is confusing about this paper?
- What are the strengths of the paper?
- What are the paper's limitations?
- What are potential future directions for research?

Why did I choose this paper?

- Review of introductory material from last week
- Contains lots of good references
- I think that the topic of data integration is interesting

Who are these people?

- Xu Shi - Michigan Biostats
- Ziyang Pan - Michigan Biostats (Graduate Student)
- Wang Miao - Peking University
- Xu Shi and Wang Miao overlapped at Harvard around 2018.

Summary

This is a literature review.

1. Overview of Necessary Assumptions
2. Overview of Basic Estimators (OR, IPW, AIPW)
3. Problems to Study:
    - Combine RCT with External Data
    - Correct Observational Study Results

Notation

- A is a (binary) treatment (Takes values $\{0, 1\}$)
- Y is the observed outcome
- $Y(a)$ is the potential outcome of an observation under treatment $A = a$.
- X is a vector of covariates
- S denotes the study sample
- $S = 1$ indicates that an observation is in the sample; $S = 0$ means that an observation is NOT in the sample.
- For the $S = 0$ sample, we will sometimes use $D = 1$ as a missingness indicator to say if an observation has observed covariates X.

Notation

- The definition of S is unclear to me, but I think that we should think about S as a finite population in the survey sampling framework from which we take a random sample $S = 1$.
- In this setup, we want to make inference on the super-population model.

Recall: CI Assumptions

- Stable unit treatment value assumption (SUTVA):

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i).$$

- No Unmeasured Confounders (NUC):

$$Y(1), Y(0) \perp A \mid X.$$

- Positivity: For X with $p(X) > 0$,

$$0 < \Pr(A = 1 \mid X) < 1.$$

CI Assumptions

- Consistency: $Y = Y(a)$ if $A = a$ for $a = 0, 1$
- Treatment Exchangeability:

$$Y(a) \perp A \mid X, S = 1 \text{ for } a = 0, 1.$$

- Treatment Positivity:

$$\Pr(A = a \mid X, S = 1) > 0.$$

- Selection Exchangeability:

$$Y(a) \perp S \mid X \text{ for } a = 0, 1.$$

- Selection Positivity:

$$\Pr(S = s \mid X) > 0.$$

Why do we need all of these assumptions?

- Identification
- Estimation

The Identification Problem

- We want our analysis to make inference on what we want!
- Asks if we can get an answer to our question given an unlimited sample size.
- We want inference of $E[Y(a)]$ but we only get to observe $(Y, A = a)$.

The Identification Problem

- Using our assumptions, we can see that an estimator (such as the IPW estimator) is

$$
\begin{aligned}
E\left[\frac{AY}{e(X)}\right] &= E\left[\frac{AY(1)}{e(X)}\right] && \text{Consistency} \\
&= E\left[E\left[\frac{AY(1)}{e(X)} \mid Y(1), X\right]\right] && \text{Law of Iterated Expectations} \\
&= E\left[Y(1)E\left[\frac{A}{e(X)} \mid X\right]\right] && \text{Treatment Exchangeability} \\
&= E[Y(1)]. && \text{Treatment Positivity}
\end{aligned}
$$

Confounding Bias (Internal Validity)

- $E[Y(a) \mid A = a, S = 1] \neq E[Y(a) \mid A = a', S = 1]$.
- Example: In a study of learning outcomes, wealthier families are more likely to get into the treatment arm.
- This violates the assumption: Treatment Exchangeability

Selection Bias (External Validity)

- $E[Y(a) \mid S = 1] \neq E[Y(a) \mid S = 0]$.
- Example: There is a study on the effect of a drug on sleep habits. The sample consists of college students who have different sleep patterns than the general population.
- This violate the assumption: Selection Exchangeability

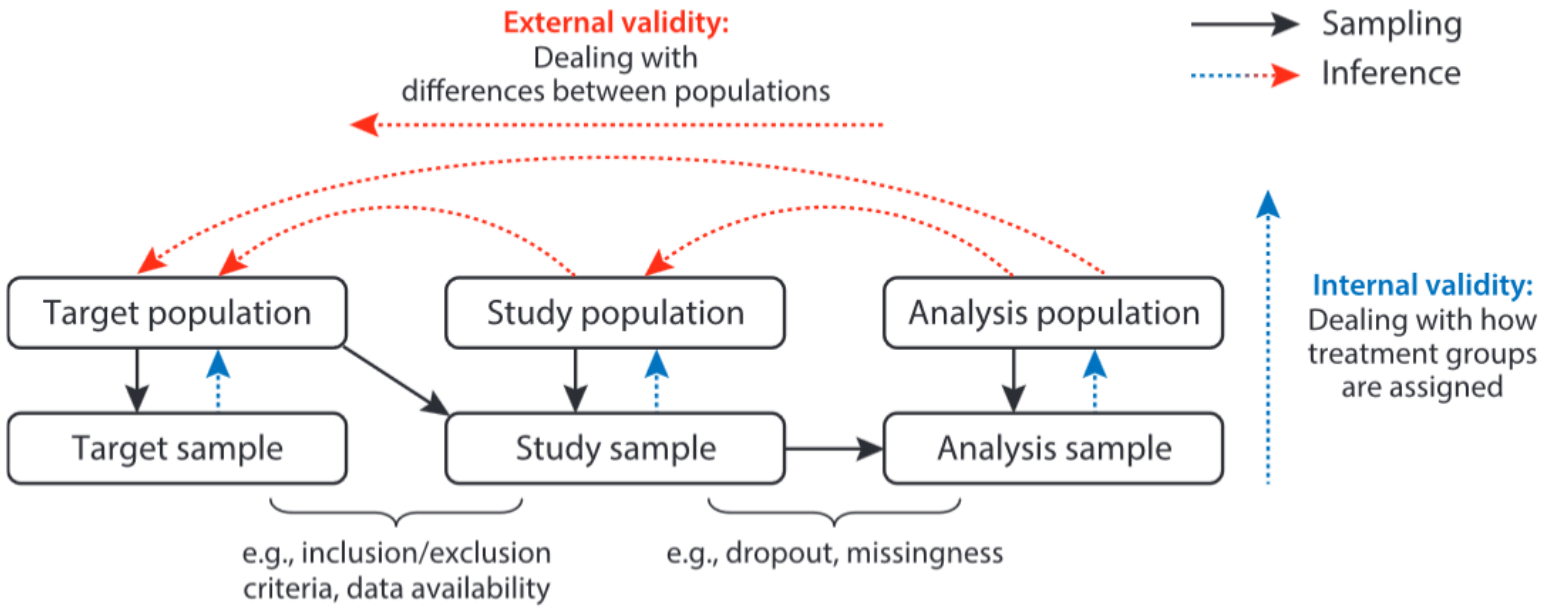

Figure 1

Internal versus external validity biases as they relate to target, study, and analysis populations.

Goals of Data Integration

- Generalizability
- Transportability

Generalizability

- Extending causal knowledge from a study to a target population when the study population is a subset of the target population.
- Example: Generalizing results from a particular state to the nation.

Transportability

- Extending causal knowledge from a study to a target population when the study population is (at least partly) external to the target population.
- Example: Generalizing from one country to another.

Comment

- Both of these techniques are a form of extrapolation!

Goals for Causal Inference

- Average Treatment Effect (ATE)

$$\tau = E[Y(1) - Y(0)].$$

- Average Treatment Effect in the Sample

$$\tau_1 = E[Y(1) - Y(0) \mid S = 1].$$

- Average Treatment Effect Not in the Sample

$$\tau_0 = E[Y(1) - Y(0) \mid S = 0].$$

Common Estimators

- Outcome Regression
- Inverse Probability Weighting
- Augmented Inverse Probability Weighting

Outcome Regression

- Use a regression function $m_a(x) = E[Y \mid A = a, X = x, S = 1]$ for the analysis where this is estimated on the data for $S = 1$.
- The estimator is then

$$\hat{\theta}_{OR} = n^{-1} \sum_{i=1} \hat{m}_a(x_i).$$

- In the case where unsampled observations still have known covariates, this estimator is $E[m_a(X)]$ for τ and $E[m_a(X) \mid S = 0]$ for τ_0.

Proof

- For the case of estimating $E[m_a(X)]$, note that by the Law of Iterated Expectations,

$$E[m_a(X)] = E[E[Y \mid A, X, S]] = E[E[Y(a) \mid A, X, S]] = E[Y(a)].$$

Inverse Probability Weighting

- Using a design-consistent approach, we can use an estimator

$$\hat{\theta}_{IPW} = n^{-1} \sum_{i=1}^{n} \frac{I(S_i = 1, A_i = a) Y_i}{\hat{\Pr}(A = a \mid S = 1, X_i)}.$$

Examples

- Correcting Bias in a Randomized Control Trial (RCT)
- Combining Clinical Trials with External Data

Correcting Bias in RCT

- The problem with a RCT is that the sample population might not be representative of the target population.
- Furthermore, we might be able to improve the efficiency of the estimator by including additional data.

Approach of Yang and Ding (2020)

- Observe a large data set with $(Y, A, X, S = 0)$.
- Observe a small (validation) data set with $(Y, A, X, U, S = 1)$.
- U is the set of confounders.

Approach of Yang and Ding (2020)

- Their approach is to estimate τ with $\hat{\tau}_1$ and then by taking the difference between $\hat{\tau}_{0,ep} - \hat{\tau}_{1,ep}$ where $\hat{\tau}_{ep}$ is some estimator of τ that does not include the unconfounding covariates.
- Then they combine these estimators using a normality assumption to find the optimal.
- This procedure is equivalent to using a calibration estimator in survey sampling where the calibration is from the error prone estimator $\hat{\tau}$.

Further Research Topics

- What do we do if the validation sample does not have information on confounding variables U?
- How can we extend this to three datasets: the RCT that has valid internal validity, an external data set that is a valid sample of the target population, and another large data set that can assist in reducing the variance of the combined estimator?

Combining Clinical Trial with External Data

- The problem with data from a (small) clinical trial is that it may not be representative of the intended target population.
- With additional information the efficiency of the estimator may be improved.

Approach of Li, Miao, Lu, Zhou (2020)

- Modified Assumption: Mean Exchangeability.
  $E[Y \mid X, S = 0] = E[Y \mid X, S = 1]$.
- This is a weaker version of Selection Exchangeability.

Approach of Li, Miao, Lu, Zhou (2020)

- Estimator is

$$\hat{\tau} = \frac{1}{n_1} \sum_{i=1}^{n} \left\{ S_i(m_1(X_i, \hat{\beta}_1) - m_0(X_i, \hat{\beta}_0)) + \frac{S_i A_i}{e(X_i, \hat{\phi})} \hat{R}_1 - \hat{W}(X_i, S_i, A_i) \hat{R}_0 \right\}$$

where

$$\hat{W}(X, S, A) = \frac{(S(1-A) + (1-S) r(X; \hat{\eta})) \pi(X_i, \hat{\alpha})}{\pi(X_i, \hat{\alpha})(1 - e(X_i, \hat{\phi})) + (1 - \pi(X_i, \hat{\alpha})) r(X; \hat{\eta})}$$

$$\hat{R}_t = Y - m_t(X, \hat{\beta}_t)$$

- $r(X; \hat{\eta})$ is an estimate of the variance ratio between the two data sets.
- $\pi(X, \alpha)$ is the selection propensity score
- $e(X, \phi)$ is the treatment propensity score

References

- Degtiar, Irina and Sherri Rose (2023). “A review of generalizability and transportability”. In: *Annual Review of Statistics and Its Application* 10, pp. 501–524.
- Li, Xinyu et al. (2023). “Improving efficiency of inference in clinical trials with external control data”. In: *Biometrics* 79.1, pp. 394–403.
- Shi, Xu, Ziyang Pan, and Wang Miao (2023). “Data integration in causal inference”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 15.1, e1581.
- Yang, Shu and Peng Ding (2019). “Combining multiple observational data sources to estimate causal effects”. In: *Journal of the American Statistical Association*.