Huanlin Dai
ADSP 32029, Autumn 2025
Jeong-Yoon Lee

Estimating the Causal Effect of Early Childhood Intervention on Cognitive Development

1. Question and Motivation

What is the effect of participation in an intensive early childhood intervention program on cognitive test scores among low birth weight, premature infants?

Early childhood is an important period of life where interventions can dramatically influence developmental disparities. Premature infants with lower birthweight are at higher risk of cognitive and developmental delays, further motivating targeted intervention. However, knowing the causal effect of such programs proves difficult due to potential systematic differences in families who enroll in intervention programs from those who do not.

Understanding early intervention's effects can help shape better policy decisions and possibly funding/scaling of similar programs. This analysis aims to contribute to that understanding by applying causal inference methods to quantify treatment effects while addressing confounding factors.

2. Dataset Selection and Description

The Infant Health and Development Program (IHDP) dataset houses a randomized controlled trial conducted across eight sites in the United States, evaluating an intensive early intervention program for low birth weight, premature infants. This analysis will use version of the IHDP data that introduces selection bias, making it appropriate for demonstrating observational causal inference methods. The data contain features such as:

- Treatment variable:
  - *treatment*: Binary indicator of participation in the early intervention program
- Outcome variable*:*
  - *iqsb_36m*: Cognitive test score (Stanford-Binet IQ) at 36 months of age
- Many other confounding variables, such as birth weight.

The IHDP dataset contains a lot of covariate information about child health and family socioeconomic factors, which are plausible confounders that affect both the likelihood of program participation and child cognitive outcomes. We can use multiple causal inference methods for stable estimation. The data also includes known treatment effects, allowing us to validate methodological choices against ground truth in preliminary analyses.

Huanlin Dai
ADSP 32029, Autumn 2025
Jeong-Yoon Lee

3. Proposed Methods

I will use the following methods to estimate the Average Treatment Effect (ATE) of the intervention program.

Method 1: Propensity Score Matching with Overlap Weighting

Propensity score methods address confounders by balancing treated and control groups on observed covariates. Overlap weighting is appropriate here because it looks at regions of common support where treated and control units are comparable.

Propensity score overlap weighting provides an intuitive method for handling the positivity assumption. It is less sensitive to model misspecification than outcome regression alone and automatically underweigh observations with extreme propensity scores that might unfairly influence estimates.

Method 2: Augmented Inverse Propensity Weighting (AIPW / Doubly Robust Estimation)

Doubly robust estimators combine outcome regression with propensity score weighting. The estimate remains consistent if either the outcome model or the propensity score model is correctly specified, although both are not necessary at the same time. This property makes AIPW appealing for this case.

AIPW offers robustness to model misspecification while potentially achieving lower variance than IPW alone when the outcome model captures meaningful variation.

4. Evaluation

I will evaluate the above methods and address key assumptions/potential bias sources.

Covariate Balance Assessment

For propensity score methods, I will evaluate covariate balance before and after weighting using standardized mean differences (SMD) for all covariates between treatment groups and visualizations showing SMD reduction after weighting. The goal is to see SMD below 0.1 for all covariates indicates adequate balance. If balance is inadequate, refine propensity score model (e.g., add interactions, polynomials)

Overlap and Common Support Diagnostics

The positivity assumption requires that all covariate patterns have non-zero probability of both treatment and control. I will assess this through visualization of propensity scores by treatment group to visualize distributional overlap, density plots showing regions of

common support, identification of observations with extreme propensity scores (e.g., <0.05 or >0.95), and sensitivity analysis: re-estimate effects after trimming observations outside common support region [0.1, 0.9].

5. Expected Challenges and Next Steps

Challenge 1: Limited Overlap in Propensity Score Distributions

If certain covariate combinations strongly predict treatment assignment, we may have regions of non-overlap where treated units have no comparable controls (or vice versa). We can try using overlap weighting because it automatically underweights units in regions of poor overlap, conduct sensitivity analysis by trimming observations with extreme propensity scores and comparing results, and be transparent about the target estimand—overlap weighting estimates the ATE for the overlap population, which may differ from the full sample ATE. We can also visualize the overlap population characteristics to help understand the issue better.

Challenge 2: Potential Unobserved Confounding

Despite rich covariates, other unmeasured factors (e.g., parental motivation, home environment quality) could influence both treatment participation and child outcomes. I will try conducting formal sensitivity analyses using E-values to quantify robustness

Compare the strength of observed confounders (birth weight, maternal education) to gauge plausibility of unmeasured confounding strong enough to overturn results

Acknowledge limitations transparently in reporting—causal claims are conditional on the unconfoundedness assumption

If possible, explore whether instrumental variable approaches could provide bounds or alternative estimates

Challenge 3: Treatment Effect Heterogeneity and Subgroup Analysis

If treatment effects vary substantially across subgroups, aggregate ATE estimates may mask heterogeneity. However, extensive subgroup analyses risk false discoveries due to multiple testing. To mitigate, this, we can pre-specify key subgroups of interest (e.g., by birth weight, maternal education) in this proposal to avoid data-driven fishing. We can also try meta-learners to flexibly estimate heterogeneous effects without requiring artificially creaed subgroups.

Huanlin Dai
ADSP 32029, Autumn 2025
Jeong-Yoon Lee

Challenge 4: Model Specification for Doubly Robust Methods

While doubly robust methods offer protection against misspecification, they still require careful modeling. Poor choices for both propensity score and outcome models could yield biased estimates. To strategize against this, we can explore multiple specifications for both components (e.g., linear, GAM, machine learning models), use cross-validation within each component to select well-performing models, and/or compare results across specifications to assess sensitivity

Next Steps

Following this proposal, I must prepare the data, run EDA, do data cleaning/encoding as needed. Then, I can begin implementing the core methods and begin preliminary comparison of estimates. I will then run appropriate evaluation and sensitivity analyses outlined in Section 4. Finally, I will synthesize the results into meaningful conclusions and prepare a presentation.