**Simulation Pipeline Overview**

This project consists of a simulation pipeline to evaluate various methods under different data linkage scenarios. The pipeline is organized into four main components:

**1. Function Definitions – linkage\_functions.R**

This script defines all core functions used in the simulation:

* Data generation under LCAR, CLAR, LNAR, MAR mechanisms
* Estimators: Full cohort, Complete case, Augmented complete case, IPW, NLAC, AIPW
* AIPW-specific utility functions for gradient and variance estimation
* Simulation wrapper: all\_simulation()

**2. Simulation Execution – linkage\_example.R**

This script performs simulation experiments and saves intermediate results:

* Defines simulation parameters (n\_rep, method\_param, mar)
* Calls all\_simulation() in a loop for each replication
* Handles errors gracefully with tryCatch()
* Saves output vectors (e.g., beta\_1\_arr, beta\_hat\_se) to disk for downstream aggregation

**3. Result Aggregation & Summarization – simulation\_result\_collection.R**

This script loads simulation outputs and summarizes them:

* Reads saved .txt output arrays (betas, SEs, coverage)
* Aggregates across parallel runs (thread-aware)
* Computes performance metrics: bias, standard error, empirical SD, and coverage
* Prints a summary report suitable for tables or manuscripts

**4. Master Execution Script – run\_all.R**

This script coordinates the **entire pipeline** from start to finish:

* Loads all required packages
* Sources the above scripts
* Runs all simulations via all\_simulation()
* Aggregates results using simulation\_result\_collection.R
* Intended as a single-entry point for reproducible, end-to-end analysis