

Model selection

Outcome type

- Continuous → ANCOVA / ANHECOVA
- Binary → Logistic regression with g-computation
- Time-to-event → Cox model

Non-linearity and large sample size?

→ TMLE or DML with machine learning as a secondary analysis.

Covariate selection

- **Baseline outcome measure** – strongly recommended when available
- **Established prognostic factors** – strongly recommended (disease-specific risk factors)
- **Randomization factors** – recommended; if many levels, adjust for coarsened groups
- **Other demographics** – optional

Covariate handling

- **Avoid including too many covariates** (rule of thumb: ≥ 10 observations per covariate)
- **Pre-specify** the adjustment set in the protocol and statistical analysis plan
- **Impute** missing covariates using simple methods (median, mean, or mode)