Average Treatment Effect

randomized controlled trials estimate the average treatment effect, which has a causal interpretation due to randomization

$$ATE = E(Y|A = 1) - E(Y|A = 0)$$

$$ATE = E(Y^{a=1}) - E(Y^{a=0})$$

$E(Y^{a=1})$ = counterfactual outcome mean if every patient had been assigned $A = 1$
$E(Y^{a=0})$ = counterfactual outcome mean if every patient had been assigned $A = 0$
ATE can be easily calculated unless have to adjust for baseline covariates

Baseline Covariates

quantitative variables expected to influence the outcome
must be measured before the start of the intervention
e.g. demographics, disease characteristics, prognostic factors, centers, baseline values of primary outcome
Table 1 in all randomized controlled trials describe the population enrolled in the study
examine if groups are comparable and control for confounding through adjusted analysis

Post-Randomization Variables

variables collected after randomization
never adjust for post-randomization covariates
crucial for per-protocol effect estimation and trials with adaptive design

Baseline Imbalance

randomization usually produces balance between groups with respect to all measured and unmeasured factors that may influence the outcome
randomization doesn't guarantee balance in any specific trial for any specific variable
e.g. a prognostic characteristics might by more common in one treatment group
Rule of Thumb: likelihood of baseline imbalance is small if $n > 200$
statistical testing to assess imbalance not recommended because multiple testing needed and hard to reject null hypothesis in small trials

<u>Randomization Stratification Factors</u>

analysis should always be adjusted for randomization stratification factors to improve the variance
e.g. sex, race, age group, study center
adjusted analysis should include covariates found between imbalanced between groups and stratified variables during randomization

Conditional Methods for Baseline Adjustment

<u>ANCOVA for Continuous Outcomes</u>

H_0: $\mu_{tretament} = \mu_{placebo}$
H_A: $\mu_{tretament} \neq \mu_{placebo}$

baseline covariates are known to be correlated with the primary outcome
e.g. baseline DNA level, initial depression scores
correlation coefficient > 0.3 indicated moderate to high correlation
analysis of covariance tests adjusts for baseline covariates to give more precise treatment effect estimates
adjusting for variables not correlated with the outcome will decrease the precision

Binary Outcomes
H_0: $\pi_{tretament} = \pi_{placebo}$
H_A: $\pi_{tretament} \neq \pi_{placebo}$
e.g. whether there was a decrease in depression scores at end of trial
baseline adjustment tries to deal with imbalance and stratification factors
adjusting for covariates that are moderately or strongly correlated with the outcome will give a less precise treatment effect estimate

Disadvantages of Conditional Approaches
regression models rely on assumptions made about the relationship between baseline covariates and the outcome
treatment effect is defined in strata of individuals with the same baseline characteristics
no direct causal interpretation

Conditional Methods for Baseline Adjustment

Inverse Probability Weighting
adjust for baseline covariates through weighting
Step 1: assign each subject a weight that's the inverse of being assigned to the treatment they received conditional on their baseline covariates L
Step 2: fit an unadjusted model in the weighted population that's not conditional on L

$$W^A = \frac{1}{f(A|L)}$$

denominator is the probability of having the assigned treatment giving L value
denominator not same for everyone with the same L value because it also depends on A value

each subject represents themselves and someone else who received the other treatment assignment but with the same baseline covariates
creates a population twice the size of the actual population where everyone appears once as treated and once as untreated
A and L are statistically independent in the pseudo-population