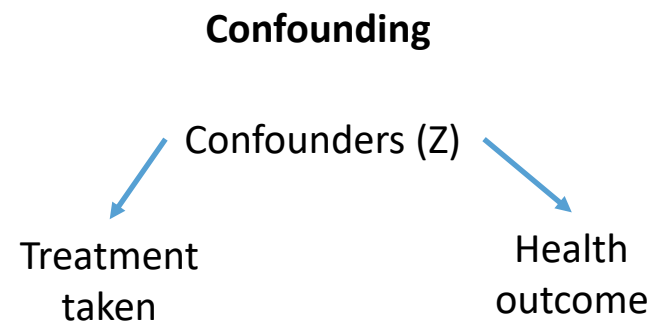
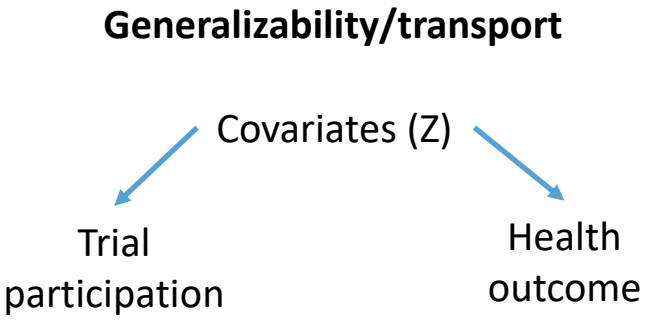
Analytic methods to estimate effects in target populations

Karolinska Institutet
Tuesday, November 16, 2021

Estimating effects in target populations

- Confounding and generalizability/transport problems often share a common causal structure
- There are multiple ways to estimate effects in target populations from a study sample just as there are multiple ways to reduce confounding bias
- Broadly speaking, you can model treatment (confounding) or trial participation (transport) or you can model the health outcome



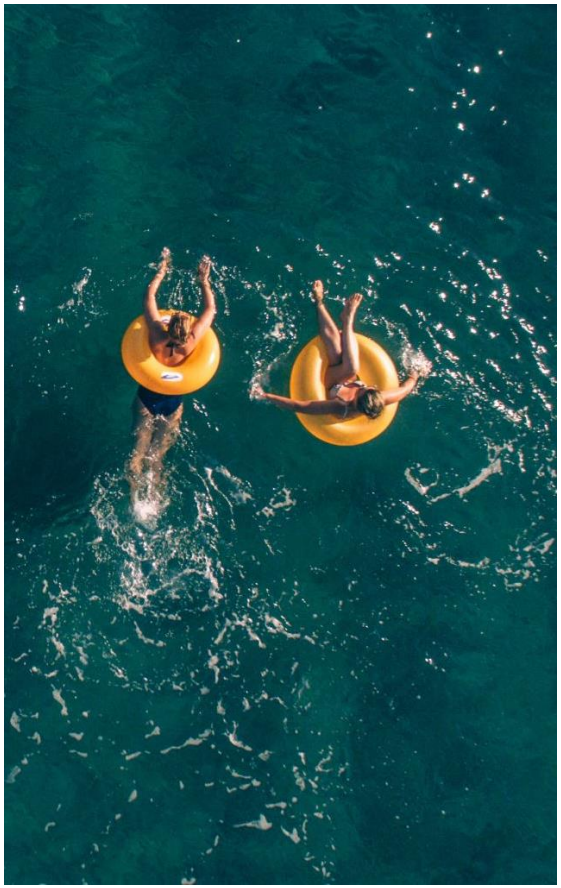


Available methods

- Doing nothing
- Restriction
- Matching
- Weight-based standardization
- Outcome modeling
- Doubly-robust approaches

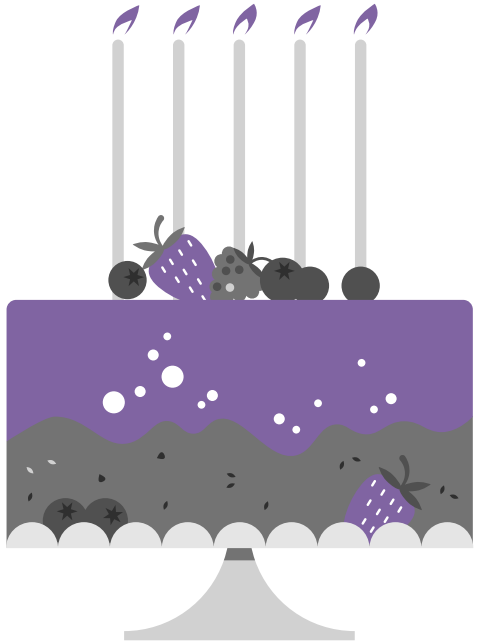
Doing nothing

- Easiest solution
- Involves strong assumptions about effect measure modifiers
  - Which cannot hold on both the RD and RR scale for risk factors
- Not ideal



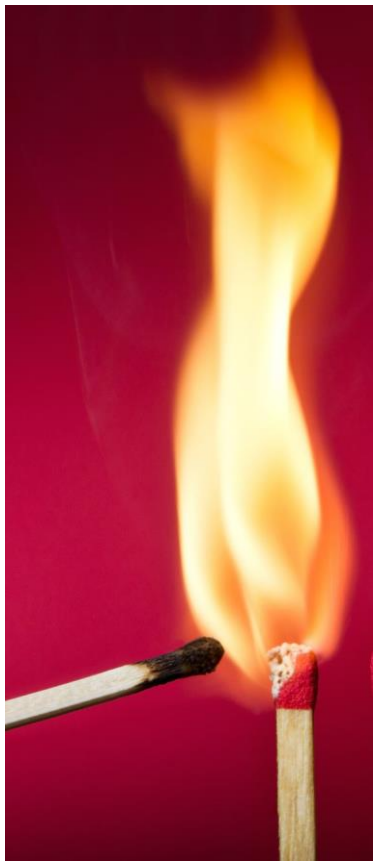
Restriction

- We can also **restrict** the trial or target population
- Includes:
  - Limiting trial analyses to key subgroups (e.g. >65)
  - Limiting the target population or indicated population based on trial exclusion criteria
- A good first step; still makes assumptions



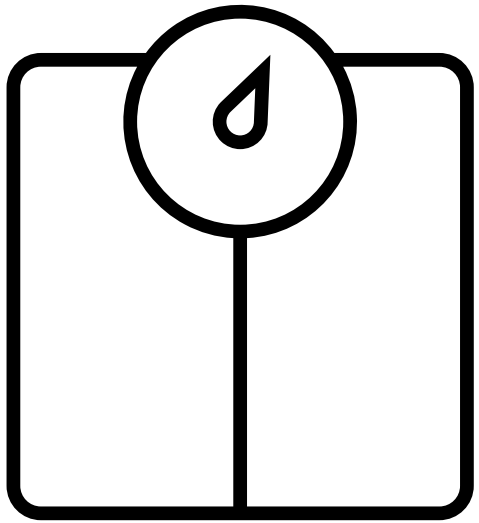
Matching

- Individually match trial participants to people from the target population
- Has the same advantages + disadvantages of exact matching for confounding
  - Difficult to do with many covariates
  - Must make decisions about replacement (especially since trials are usually smaller)



Weighting

- Creates a “pseudopopulation” with similar covariate distribution to the target from the trial participants
- Uses “probabilities of sampling” that are very similar to the propensity score
- Still makes assumptions
- More discussion soon!



Outcome modeling

- Estimate an outcome model in the trial based on covariates
- Apply the model to the target
- Obtain effect estimate of interest
- Most precise option, but the outcome may be harder to model than sampling

Doubly robust approaches

- Leverage both a “sampling” and “outcome” model
- If either model is correct, the estimate is unbiased
- Often integrates machine learning
- Most useful with large numbers of covariates/complex models

The goal of inference odds weights

- Use the **trial** population to create a new target population with the same distribution of covariates as the **target**
- Use the **outcome** data from this weighted **trial** population in analyses
- Perform the analysis you want to do in the weighted trial (Cox regression, intent-to-treat, etc.)

Inverse odds weights-estimating probabilities

- Combine study and target population individual-level data, with a new variable representing the population of origin
- Use some method to estimate each participant's probability of membership in the **study sample** based on covariates *Z*
  - Logistic regression
  - Machine learning (e.g. pseudo-HDPS)
- This probability can be represented as:
  - $Pr(sample_i = 1 | Z_i)$

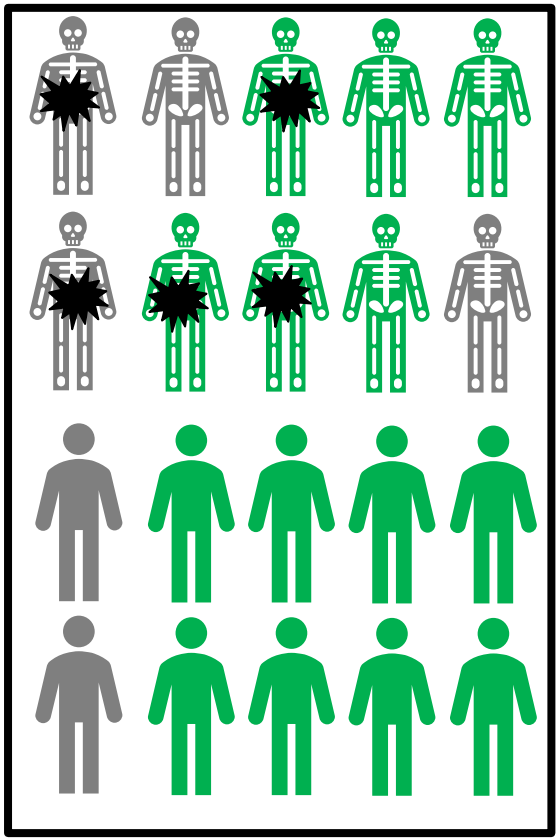
Inverse odds weights-creating weights

- These predicted probabilities can be transformed into **odds** and used to re-weight the study population individuals **only**
- The weights are the inverse odds:
  - $W_i = 1/Odds_i = \frac{Pr(sample_i=0|Z_i)}{Pr(sample_i=1|Z_i)}$
- Additional options include:
  - Stabilizing the weights
  - Resampling techniques based on the weights

A simple demonstration

1. Combine the study and target population

Combined population



2. Estimate the probability of study membership based on EMM

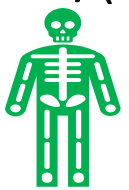
P(Study | Gray) = 4/6 = 0.67

P(Study | Green) = 6/14 = 0.43

3. Assign each study participant a weight of: $W_i = 1/Odds_i = \frac{Pr(sample_i=0|EMM_i)}{Pr(sample_i=1|EMM_i)}$

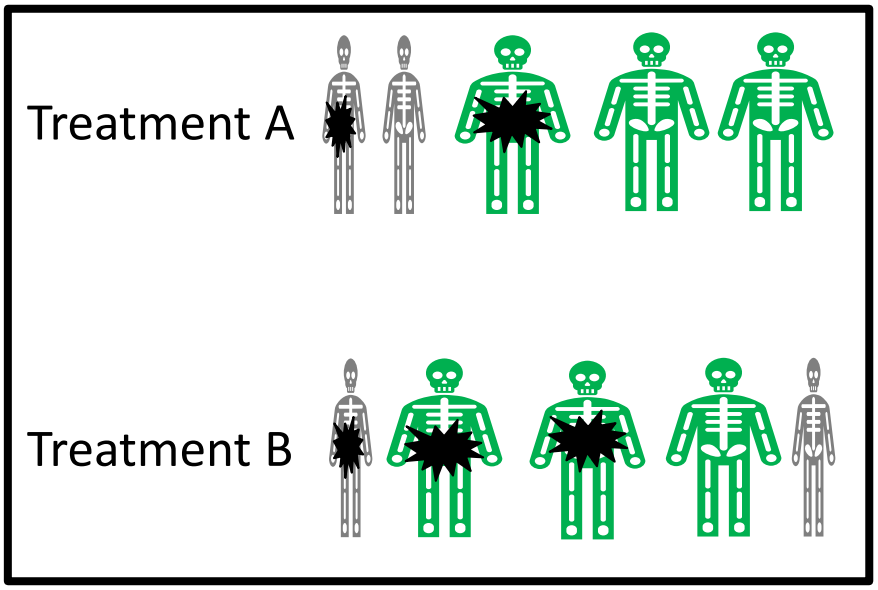
W_{gray} = 1/(4:2) = 0.5

W_{green} = 1/(6:8) = 1.33



Analyzing the weighted study population

Weighted study population



Original trial:
Risk in treatment A = 2/5= 40%
Risk in treatment B = 3/5 =60%
Risk difference A vs B = -20%
Risk ratio A vs B = 0.67

With weights:
Risk in treatment A = (0.5 + 1.33)/5 = 37%
Risk in treatment B = (0.5 + 2*1.33)/5 = 63%
Risk difference A vs B = -26%
Risk ratio A vs B = 0.59

An example using MOSAIC…

MOSAIC trial population

- Included patients after first colorectal cancer + surgery
- Randomized 2,246 patients to FOLFOX4 vs 5FU/LV
- Enrolled worldwide from 1998-2001

US Oncology real-world target

- Included patients in US Oncology's IKnowMed EHR
- Patients receiving mFOLFOX6
- Treated from 2008-2019 and eligible for the trial
- Assessed multiple targets:
  - All US Oncology patients
  - Stage III patients
  - Stage III patients w/ FOLFOX

Reweighted MOSAIC results

Trial / *Target Population*	Hazard ratio (95% CI)
MOSAIC – Crude, overall	**0.85 (0.71, 1.01)**
All US Oncology patients	1.05 (0.83, 1.33)
MOSAIC – Crude, stage III	**0.80 (0.65, 0.97)**
Stage III US Oncology patients	0.97 (0.74, 1.24)
Stage III US Oncology FOLFOX patients	0.93 (0.71, 1.22)

Questions?

mawcpharmdphd@gmail.com