Statistical Science Tools

Todd MacKenzie, PhD
Biostatistician
Professor of Biomedical Data Science
Geisel School of Medicine

Primary Roles of Statistical Science

Causal Inference

Prediction

Causal Inference

- Does a treatment work
- Is this treatment superior to the standard of care
- Is this exposure detrimental

Prediction

- Will this inpatient
  survive hospital stay?
  discharge to a skilled nursing facility?
- Will this patient
  require CABG within 5 years ?
  have a stroke in 10 years?

Equivalence between Causal inference and Prediction

- “Treatment A is better than B” same as “We predict patient will do better on A than B”

Statistical Tools:

Causal Inference vs Prediction

Causal Inference Statistics: Basic

Comparison of two groups of subjects	• Those who received treatment A vs those who received treatment B
Compare proportions	• Pearson chi-square or Fisher exact test
Compare means	• 2-sample T-tests and ANOVA
Probability a subject who received A does better than subject on B	• Wilcoxon-Mann-Whitney

Causal Inference Statistics: Basic Paired Design

Compare outcomes within same subject

- Pre and post
- Cross-over (randomized sequence)

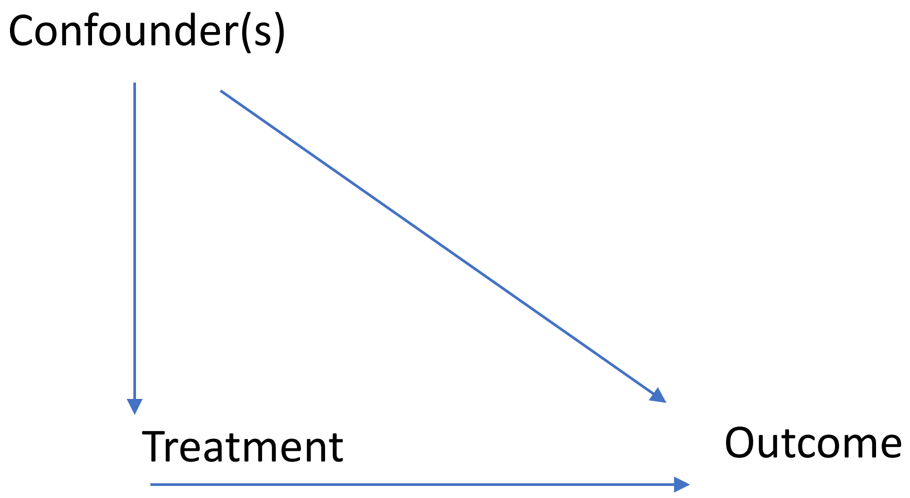
Compare outcomes between two matched subjects

Paired t-test to compare means

McNemar test for binary outcomes

Confounding

- Selection bias
- Subjects selecting or being selected to receive treatment A may differ from those receiving treatment B with respect to factors known to influence the outcome
    - Older
    - Sicker
    - Whiter



Overcoming confounding by Design

Randomization

- Guarantees those receiving A are from same population as those receiving B
- Analysis in presence of non-compliance
  - Intention-to-treat is the best (or closely related CACE is even better)
  - “per-protocol” is ok
  - As-treated is inferior

Types of randomization

- Parallel
- Cross-over
- Cluster

Natural Experiments

- Instrumental variables

Overcoming Confounding by Analysis

Sometime randomization is not feasible

- Too costly
- Not ethical

Observational studies require analytic tools to overcome confounding

- They utilized *measured* confounders
- Unmeasured confounders cannot be addressed in analytic strategy except possibly the approach of instrumental variables

Statistical Tools for Addressing Confounding

- Statistical modeling, aka Regression
- Propensity score adjustment
- Propensity score weighing

Statistical Modeling

- Multivariable Linear Regression
  - Continuous outcome
- Multivariable Logistic Regression
  - Binary outcome
- Multivariable Poisson Regression
  - Counts and binary outcomes
- Multivariable Cox Regression
  - Time-to-event subject to *censoring*
  - Proportional hazards Cox model

- Multivariable: treatment and confounders

Implementing and Interpretation of Multivariable Models

- In R
  - glm(Y ~ X + Z1 + Z2 + Z3 + …)
- Stata
  - regress Y X Z1 Z2 Z3 …
- It yields "coefficients" (estimates of) for the treatment effect that can be interpreted as
  - "<u>controlling</u> for Z1,Z2,… the effect of X was …"
  - "<u>adjusting</u> for Z1,Z2,… the effect of X was …"

Limitations of multivariable models

Parametric assumptions

- linear regression assumes confounders have linear effects
- Logistic regression assumes confounders have logit-linear effect
- Cox regression assumes confounders have log-linear effects on hazard

But … one can address that readily by being less "parsimonious"

- Throw in second order terms (interactions and quadratics)

Unmeasured confounders

Alternatives to multivariable models

- Propensity (Score) Matching
- Inverse Weighted Propensities

- These approaches take a step back and examine treatment selection

What is a propensity?

- The probability of receiving the treatment based on:
  - Characteristics of the patient (or their clinician or their healthcare center)
  - Demographics, Comorbidities, Other treatments, …
- Should it be limited to confounders?
  - What about characteristics that affect treatment choice but not the outcome?

Calculation of the Propensity

- Apply a prediction model
- Usually, a multivariable logistic regression
  - Treatment (yes or no) regressed on patient characteristics
  - The “predicted values” (“fitted values”) are the propensities
- Could be a *random forest* or other ML approach

Two Steps in Propensity Scores

Propensity is probability of receiving treatment based on patient characteristics (e.g. confounders)

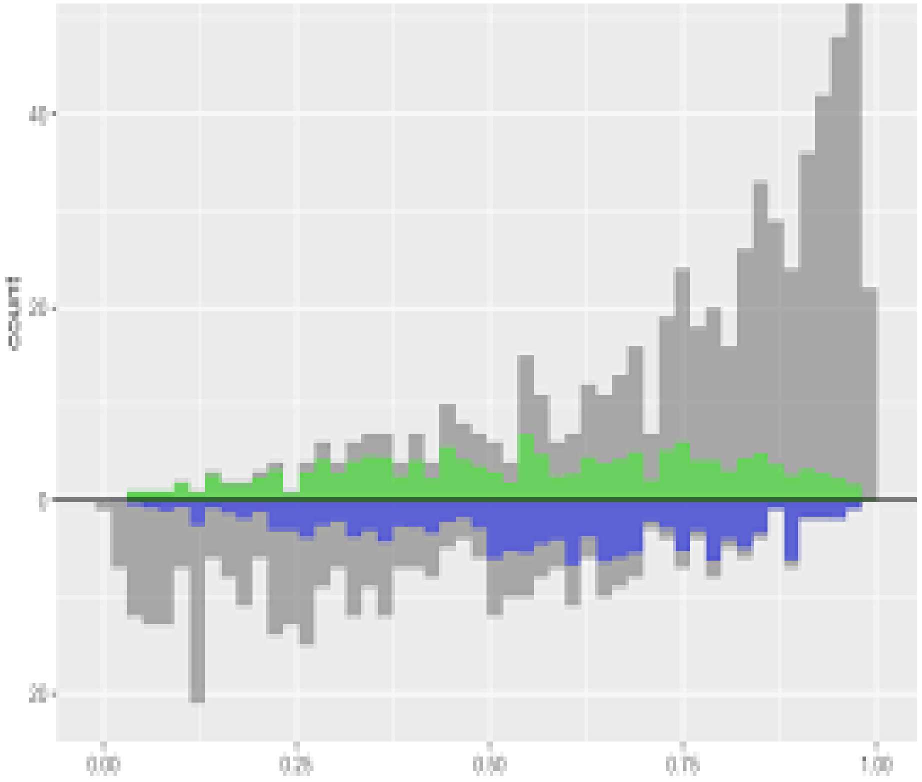
Step One

- Build a prediction model for this using logistic regression or other "ML" technique

Step Two

- Use the propensities to adjust for confounding using
  - Matching
  - Binning
  - Weighting

Propensity Score <u>Matching</u>



- Create matched pairs of subjects matching on the propensity
    - A treated subject and untreated subject are matched if they have a similar probability of treatment
    - “Caliper” – the maximum difference allowed for matches
- It can be 1-to-1 or 1:2 or even variable ratios
- Requires an algorithm

Propensity Score **Matching:** Comparing the matches

- Evaluate the treatment effect by comparing treated and untreated matches
- Using statistical methods for matched studies
  - Paired t-test if 1:1 and outcome continuous
  - McNemar's test if 1:1 and outcome binary
  - GEE for clusters (matches)
  - Mixed effects models (matches as random effects)
  - Conditional logistic regression

Propensity Score Binning

Create bins of the propensities (groups of subjects)

- Deciles (10 groups)
  - 10% of subjects with lowest propensities, …, 10% with highest
- Ventiles (20 groups)
  - 5% of subjects with lowest propensities, …, 5% with highest
- Or even finer resolution if you have enough data

Then use regression model to adjust for this categorized propensity

Equivalent to comparing the treated and untreated in each bin and aggregating the results

- Bins with exclusively treated or exclusively untreated are tossed

Propensity Score **Weighting**

You calculate *weights*

e.g. Reciprocal of the propensity of the treatment received

Then use weighted two-sample statistics or regression

Inverse Weighting

- Idea: Make analysis representative of the population of interest
- Suppose 5%, 10%, 5% and 80% of your patient sample is Black, Asian, Other and White but you want to make your analysis representative of the entire population in which those frequencies are 15%, 10%, 5% and 70% then weight subjects by 15/5=3, 10/10=1, 5/5=1 and 70/80=0.875
- This idea is used in the design of surveys
  - Minorities often over-sampled but then must be under-weighted
- Also used to adjust for non-response
  - e.g if older people are twice as likely to respond then young then their weights should be half

Propensity Score Weighting

- Suppose blacks are half as likely as whites to receive the treatment then they should be weighted twice as much
- Likewise: the blacks in the untreated group should be weighted half as much as whites in the untreated group
- Same thinking holds for all patient characteristics
- All the patient characteristics are rolled up into a patient's propensity

Propensity Score **Weighting**

Weight subjects in inverse proportion to the probability of what they received

Treated subjects are weighted by the reciprocal of the probability of treatment: 1/Propensity

Untreated subjects are weighted by the reciprocal of the probability of non-treatment: 1/(1 – Propensity)

Propensity Score **Weights**

This approach weights to the treatment effect in the population

Average Treatment Effect (ATE)

Different weighting schemes are possible

Weight to the treatment effect in the treated (ATET)

The effects may be different especially if patients have been selected to the best treatment for them

Doubly Robust Approaches

- Use a multivariable regression model in the second step in which you add covariates to the model, as well as the propensity
- Double dipping but in a good way (can’t hurt)
- Covers two possible failures
  1. The model for the treatment selection has poor fit
  2. The model for the outcome has poor fit

Summary

- Estimating treatment effects
- Addressing confounding in observational studies
- Multivariable regression models
- Propensity Score Approaches