Propensity score analysis [Reporting guideline]

ehsan.karim@ubc.ca

Oct 10, 2021

SPPH 504/007

Ref

Propensity Scores and Matching Methods

EA Stuart - The Reviewer's Guide to Quantitative Methods in the …, 2018 - taylorfrancis.com

Many studies aim to estimate causal effects of risk factors, interventions, or programs, on outcomes of interest. While randomization is generally seen as the preferred design for estimating causal effects it is not always possible to randomize the "treatments" of interest, especially in the social sciences. Propensity scores are a useful tool that can help yield better estimates of causal effects in non-experimental studies by ensuring that the treatment and comparison groups are similar with respect to the observed covariates. The propensity …

☆ 99 Import into BibTeX »

[HTML] Propensity score weighting compared to matching in a study of dabigatran and warfarin

JD Seeger, K Bykov, DB Bartels, K Huybrechts… - Drug safety, 2017 - Springer

Introduction Comparing medications in observational settings requires differences in patient characteristics to be accounted for. Propensity score (PS) methods can address these differences, but PS weighting approaches may introduce bias. Methods Within a cohort study of anticoagulant initiators from October 2010 through to December 2012, PS values for dabigatran relative to warfarin were estimated, and study outcomes (stroke or major bleeding) among the cohort were identified. The PS was used to match initiators and results …

☆ 99 Cited by 7 Related articles All 8 versions Import into BibTeX

Which is NOT the reason why Propensity score matching can be better than regression?

Diagnostics (balance checking) much easier compared to residual plot/influence

Exposure and outcome models are separate

Non-parametric (ML) approaches can be used to relax linearity assumption in estimating PS.

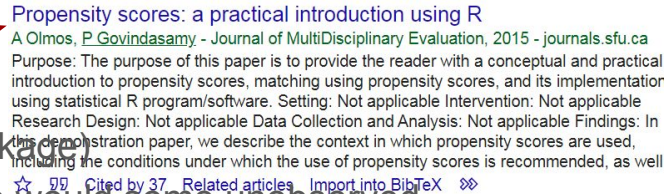
Propensity score matching can deal with unmeasured confounder problem

1. When PS match is preferable to weight?

- PS (probability of patient receiving treatment) model
  - predictive model to mitigate confounding.
  - Assumption 1: No unmeasured confounding
  - Assumption 2: Non-overlap
  - Assumption 3: Positivity: 0 < PS < 1
- (A1) If prediction is inadequate
  - comparisons (treated vs. untreated) remain confounded.
- (A2) If finding close match is challenging
  - OR may have some residual bias.
- (A3) If prediction is extreme
  - >95% or <5%

1. When PS match is preferable to weight?

- Matching: Unmeasured confounding
  - Some confounder is absent in the study
  - Proposed solution:
    - Add more and more covariates that are correlated
    - Add proxy variables to those who are unmeasured
    - High-dimensional propensity score (hdPS)
      - It uses variables associated with outcome
      - This does not have design/analysis stage separation
  - Is the proposed solution enough?
    - Sensitivity analysis:
      - Rosenbaum bounds (*psens()* in *rbounds* R package)
      - "how strongly related to treatment and outcome would some unobserved confounder have to be to change the study conclusions?"



1. When PS match is preferable to weight?

- Matching: Non-overlap or (near) non-positivity
  - majority of match will come from central part of PS distribution.
  - Non-overlap = no counterfactual
    - Matching far apart subjects (treated vs untreated) will induce bias in estimated OR
    - Proposed solution:
      - Set small caliper
    - Too small caliper may cause very small matched sample
  - Non-positivity: extreme patients may be present in one arm/both
    - Proposed solution:
      - Trim patients with PS < 0.05 or PS > 0.95
    - Trimming makes generalizability a problem
      - who are we excluding?

1. When PS match is preferable to weight?

- Weighting: Unmeasured confounding
  - Impacts PS model model building
  - As weights are direct estimates from this PS model, OR estimates are much more sensitive to model misspecification
    - Matching is affected by model mis-specification, but could be at a lesser extent
- Weighting: Non-overlap
  - Non-overlap not a major problem: nobody is excluded
    - Every patients is assigned a weight (IPTW) based on PS

1. When PS match is preferable to weight?

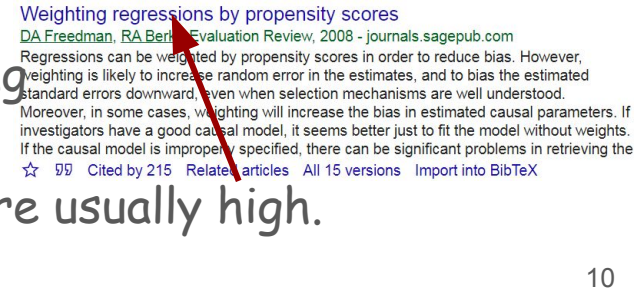
- Weighting: (near) non-positivity
  - includes extreme patients; may exacerbate confounding
    - Resulting OR can get heavily influenced by these extreme patients.
  - Easy to detect this problem:
    - assess max weight of IPTW
    - No specific cut-point (20 or 50); depends on sample size
      - IPTW = 50 when N = 100? vs. IPTW = 50 when N = 1,000,000 (one million)?
  - Proposed solution:
    - Truncate IPTW with IPTW < 5% or IPTW > 95%
  - degree of truncation
    - bias and variance tradeoff
    - Small amount of truncation suggested (5% or preferably 1%)

2. Does the choice of PS impact interpretation?

1. ATT (matching, and weighting)
   - Conditional estimate (patients may be more interested in this)
   - Treated group versus a group that that is similar to the treated group, but never got the treatment
     i. Age group 30-35, male, Mexican, University graduate, treated
     ii. Age group 30-35, male, Mexican, University graduate, untreated
2. ATE (weighting)
   - Marginal estimate (policy-makers may be more interested in this)
   - Compare groups in a purely counterfactual fashion
     i. All subjects (treated and untreated) got treatment vs.
     ii. All subjects (treated and untreated) did not get treatment

Summary of 1 & 2

- Matched sample size reducing a lot?
  - Try PS weighting
- Non-overlap in PS distributions visible?
  - Try different PS model-specification (interactions, polynomials, add more covariates/proxy)
  - Try weighting
- IPTW maximum value is very large?
  - Try truncation (small amount) in PS weighting
  - Still large weights? Try matching.
  - Note that CI/SE from weighting methods are usually high.



Summary of 1 & 2

- Worried about model misspecification (unmeasured confounding + model form, e.g., interactions/not)?
  - Use machine learning (tree based methods) to build PS model
  - PS matching may be more robust compared to PS weighting
  - If still unsatisfactory, try hdPS as a sensitivity analysis
- Want to make generalization about treated/whole?
  - If overlap is good, try ATT from matching (treated vs. counterfactual)
  - If weights are reasonable, try ATE from weighting (marginal estimate)

What should I do if the SMD is not satisfactory for some covariates?

Add more covariates

Remodel propensity score by changing model specification (interaction, polynomial)

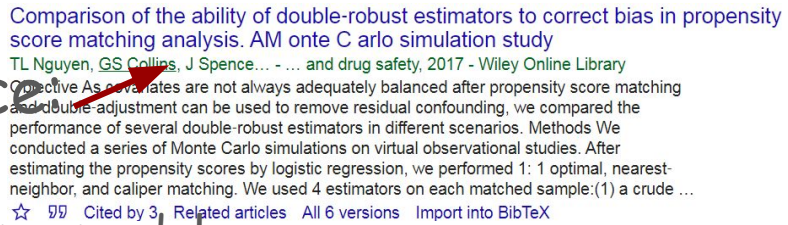
Estimate propensity score by different approaches (ML, ensemble)

Select the variables that are unbalanced, and then adjust them in the outcome model

Adjust all covariates (irrespective of balance by SMD) in the outcome model

3. Matched SMDs not satisfactory. What now?

1. Adjust imbalanced ones:
   a. Identify covariates that are associated with high SMD (SMD > 0.2, say)
   b. Adjust for those imbalanced covariates in the outcome model to get OR
2. Adjust all/some covariates twice:
   a. Put all the covariates in the PS model
   b. Adjust all the covariates in the outcome model
   c. "Doubly adjustment" models (2 chances to adjust confounding)
3. Important covariate handling by exact match
   a. Combination of exact & PS match, e.g., age/sex by exact, others by PS



4. Reporting guidelines for a PS analysis.

- What do you want to estimate? [Methods]
  - ATE / ATT
  - This will impact interpretation / implication
- PS model [Methods]
  - Which covariates?
  - Which covariates were proxies (of what unmeasured covariate)?
  - Post-treatment covariates? (if temporality can be established)
  - Model specification (interaction/ polynomials included?)
  - What model was used (logistic, ensemble methods)

4. Reporting guidelines for a PS analysis.

- How PS was used? [Methods]
  - Matching?
    - Full matching, variable ratio matching, fixed matching?
    - If fixed matching, at what ratio? 1:1? 1:3?
    - With replacement / without?
    - Caliper?
    - PS trimming? At what cut point? <|.05|?
  - Weighting?
    - IPTW truncation? At what %? 1%? Scaling?
- Outcome model [Methods]
  - What was adjusted in the outcome model
    - Matched sample IDs/clusters accounted in the analysis or not.
    - If matching was with replacement, how were SEs calculated? Bootstrap?

4. Reporting guidelines for a PS analysis.

- Sample size [Results]
  - Original vs. matched; a brief summary of who were excluded
- Plot of overlap [Results] (if matching)
  - Treated PS vs. untreated PS distributions/histograms/boxplot
- Covariate balance (both matching & weighting) [Results]
  - SMD before and after (at least a summary, e.g., max SMDs)
  - SMDs (before vs. after) in figure: "Love plot"
  - SMDs survey weighted?
  - What was chosen SMD cut-point?
    i. 0.1 or
    ii. 0.2?

4. Reporting guidelines for a PS analysis.

- Summary of PS / IPTW [Results]
  - Summary of PS (if matching)
  - summary of IPTW (if weighting)
  - Also report summary of IPTW * survey weights
    i. scaled,
    ii. truncated?
- Was PS used in multiple ways?
  - Often PS is used as matching and weighting
  - Goal is to find approach that provides best covariate balance

4. Reporting guidelines for a PS analysis.

- Implication [Discussion] (if appropriate)
  - Policymakers will be more in therested in this study?
  - A patient /caregiver will be more in therested in this study?
- Plausibility of assumption [Discussion]
  - No unmeasured confounding?
  - How was overlap?
  - Positivity was an issue?
    i. Estimated PS too close to 0 or 1?
    ii. Estimated IPTW very high?

Thanks!

ehsan.karim@ubc.ca
www.ehsankarim.com