Evaluating the Use of Propensity Scores with Multilevel Data

Noah Greifer[1] and Felix Thoemmes[2]

[1]Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill [2]College of Human Ecology, Cornell University

Introduction

Propensity scores are a popular method and are also used with multi-level data.

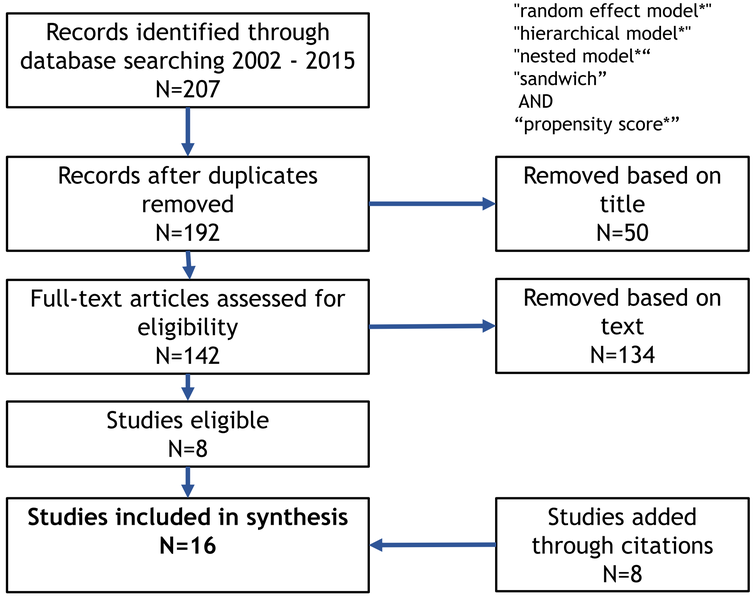
- **Six steps for propensity score analysis:**
  1. Identify outcome, treatment, relevant covariates, sample, and causal estimand (ATE, ATT, etc.)
  2. Specify propensity score model and estimate propensity scores
  3. Condition on propensity scores (matching, weighting, etc.)
  4. Assess balance on covariates; if poor, repeat 2-4
  5. Estimate treatment effect on conditioned sample
  6. Perform sensitivity test

- **In multilevel contexts (units clustered in groups), additional considerations apply:**
  - Considering cluster effects in propensity score model
  - Conditioning on propensity scores within or across clusters
  - Attaining and assessing balance on cluster covariates, including aggregate unit-level covariates
  - Addressing SUTVA when clusters may introduce spillover or interference
  - Accounting for residual dependence in outcome analysis

- **Systematic review of studies that use propensity score analysis in clustered data, with recommendations for reporting and methodological practice**

Methods

Study Selection:
"multilevel model*"
"mixed model*"
"random effect model*"
"hierarchical model*"
"nested model*"
"sandwich"
AND
"propensity score*"



Results

Treatment Level
- Level 1: Units within clusters can have different treatment statuses
- Level 2: Units within clusters have the same treatment status

Modeling the Propensity Score
- SL1 - Level 1 covariates only, no random effects (RE)
- SL2 - Level 2 covariates only (Level 2 treatment)
- SL - Level 1 and Level 2 covariates, no RE
- FE - Level 1 covariates only, dummies for clusters
- MLM-1 - Level 1 covariates only, RE for clusters
- MLM-2 - Level 1 and Level 2 covariates, RE for clusters

PS Model \ T Level	1	2	Total
SL1	2*	1	3
SL2	0	2	2
SL	2	2	4
FE	3	0	3
MLM-1	2	0	2
MLM-2	1	1	2
Total	10	6	16

Conditioning on the Propensity Score
- Matching (N=8)
- Stratification (N=5)
- Weighting (N=3)

Conditioning Within/Across Clusters
- Conditioning within clusters (CWC) (N=2)
- Conditioning across clusters (CAC) (N=14)

Balance Assessment

Achieving Balance	N
Standardized differences*	6
P-value	6
Unknown	4

*3 used .1, 1 used .2, 2 used .25, 1 did not specify

Balanced Assessed at Level 2	N
Yes	8
No	6
N/A (CWC)	2

Strong Ignorability/Sensitivity

SI/Sensitivity	N
Y	2
N	14

SUTVA

SUTVA	N
Mentioned	2
Unmentioned	14

Outcome Analysis
- Random Effects (N=11)
  - 4 included covariates used in PS model
  - 7 included additional covariates not used in PS model
- OLS (N=3)
  - 2 used robust SE

Conclusions

- Generally good reporting
- Some suboptimal/poorly reasoned methods
- Balance on group level covariates is lacking
- Balance assessment is often not reported or p-values are used
- SI and SUTVA are largely unaddressed; "strong ignorability strongly ignored"

Recommendations
- Specify several propensity score models to find the one that yields the best balance either within or across groups
- Ensure balance on group level covariates, including on aggregate unit level covariates; may be best achieved by CWC when possible
- Use standardized mean differences for balance checking, using low threshold ~.1
- Ensure and address plausibility of SUTVA; perform Rosenbaum Sensitivity Analysis

Further Research
- Include non-continuous (e.g., binary, count) outcomes in review
- Examine which methodological articles are most cited
- Expand search terms to be more inclusive

Acknowledgements

Kirsten Kainz, Daniel Bauer, Ashley Givens, The L. L. Thurstone Psychometric Laboratory

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Cornell University