SA15

Propensity Score Matching: Role of Clinical Expertise in a Data-Driven Approach to Model Selection

Tyler Knight, MS; Laura Kauffman, MPH; and Shoshana Daniel, PhD
Market Access Consulting, Labcorp Drug Development Inc., Princeton, New Jersey, USA

Background

- Observational studies are not controlled and are subject to confounding factors that could influence outcomes
- In an open-label, non-randomized, prospective observational cohort study to assess post-procedural outcomes in two cohorts:
  - Some confounding factors were more critical and had more influence on the outcomes of interest, so more weight and importance needed to be placed on those variables than others
  - Clinical input was needed to properly categorize confounding factors
  - Wanted a thorough, data-driven and clinical expertise approach to control for potential bias

Objectives

- To provide further evidence of the value of clinical expertise alongside data-driven propensity score (PS) model selection in observational studies
- Share learnings on process for model selection from a real-world study

Methods

- Analysis utilized data from an open-label, non-randomized, prospective, observational study that evaluated the safety outcomes of a new surgical procedure (N=280) to standard of care (SoC) procedure (N=707)
- Clinical experts identified baseline confounders for the safety outcomes and classified them into tiers: Tier 1=definite risk factors, Tier 2=probable confounders and Tier 3=potential instruments
- Five logistic PS matching models (1:1 ratio) were developed from different combinations of the confounders
  - **Model 1:** Forced all Tier 1 covariates into model; Tier 2 and 3 covariates not included in model
  - **Model 2:** Forced all Tier 1 and 2 covariates into model; Tier 3 covariates not included in model
  - **Model 3:** Forced all Tier 1, 2 and 3 covariates into model
  - **Model 4:** Forced all Tier 1 covariates into model, and allow Tier 2 and 3 covariates to be selected by model selection procedure (entry/exit P-value: ≤0.25/>0.25)
  - **Model 5:** Forced all Tier 1 and 2 covariates into the model and allow Tier 3 covariates to be selected by model selection procedure (entry/exit P-value: ≤0.25/>0.25)
- PS model used a nearest neighbor approach with a caliper=0.2 standard deviations of the logit PS
- Final model selection was based on goodness-of-fit, maximizing proportions matched between cohorts, and balance across all confounders (i.e., absolute mean and maximum standardized difference [std diff] across all covariates <0.2)
- Following consultation with clinical experts, the following baseline covariates were chosen for the PS model

Tier 1 (definite risk factors)	– Age, body mass index, chronic pain condition, pre-existing abnormal uterine bleeding, disorders associated with hysterectomy, hormonal contraception use
Tier 2 (probable confounders)	– Pre-existing depression, history of allergic/hypersensitivity reactions, history of prior abdominal surgery
Tier 3 (potential instruments)	– Race, ethnicity, US geographical region, pre-existing autoimmune disease, pre-existing diabetes

Results

- Prior to match:
  - Patient demographics were not all well balanced (std diff≥0.20) between the two cohorts for (Table 1):
    - Age
    - U.S. geographical region
  - Patient clinical characteristics were fairly balanced (std diff<0.20) between the two cohorts with the exception of hormonal contraception use (Table 2)

Table 1. Pre-match Demographic Characteristics

Characteristic	New Procedure (N=280)	SoC Procedure (N=707)	Std Diff
Age (years)			0.4502
Mean (SD)	35.3 (5.47)	32.7 (5.77)	
Median	36.0	33.0	
Race, n (%)			0.1410
White	229 (81.8%)	562 (79.5%)	
Black or African American	31 (11.1%)	90 (12.7%)	
Asian	3 (1.1%)	5 (0.7%)	
Native Hawaiian or Other Pacific Islander	0 (0.0%)	2 (0.3%)	
American Indian or Alaska Native	0 (0.0%)	2 (0.3%)	
Multiple	1 (0.4%)	6 (0.8%)	
Not reported	16 (5.7%)	40 (5.7%)	
Ethnicity, n (%)			0.1711
Not Hispanic or Latino	182 (65.0%)	483 (68.3%)	
Hispanic or Latino	94 (33.6%)	199 (28.1%)	
Not reported	4 (1.4%)	25 (3.5%)	
Body mass index (kg/m^2)			-0.0677
Mean (SD)	30.1 (6.33)	30.5 (6.69)	
Median	30.0	29.8	
U.S. geographic region, n (%)			0.2203
Midwest	116 (41.4%)	287 (40.6%)	
South	80 (28.6%)	146 (20.7%)	
West	61 (21.8%)	199 (28.1%)	
Northeast	23 (8.2%)	75 (10.6%)	

Note: Std diff is calculated as the difference in means or proportions between cohorts (new – SoC), divided by the standard deviation overall. A std diff less than 20% (small effect size) indicates good balance between cohorts

Table 2. Pre-match Clinical Characteristics

Characteristic	New Procedure (N=280)	SoC Procedure (N=707)	Std Diff
Chronic pain condition,[a] n (%)	120 (42.9%)	329 (46.5%)	-0.0740
Pre-existing autoimmune disease, n (%)	31 (11.1%)	80 (11.3%)	-0.0077
Pre-existing diabetes, n (%)	12 (4.3%)	30 (4.2%)	0.0021
Pre-existing depression, n (%)	83 (29.6%)	267 (37.8%)	-0.1725
History of allergic/hypersensitivity reactions, n (%)	126 (45.0%)	344 (48.7%)	-0.0733
History of prior abdominal surgery,[b] n (%)	150 (53.6%)	357 (50.5%)	0.0616
Pre-existing abnormal uterine bleeding,[c] n (%)	40 (14.3%)	70 (9.9%)	0.1348
Disorders associated with hysterectomy,[d] n (%)	21 (7.5%)	48 (6.8%)	0.0276
Hormonal contraception use, n (%)			0.5055
Injectable hormonal contraception	58 (20.7%)	58 (8.2%)	
Other hormonal contraception	132 (47.1%)	267 (37.8%)	
No hormonal contraception	90 (32.1%)	382 (54.0%)	

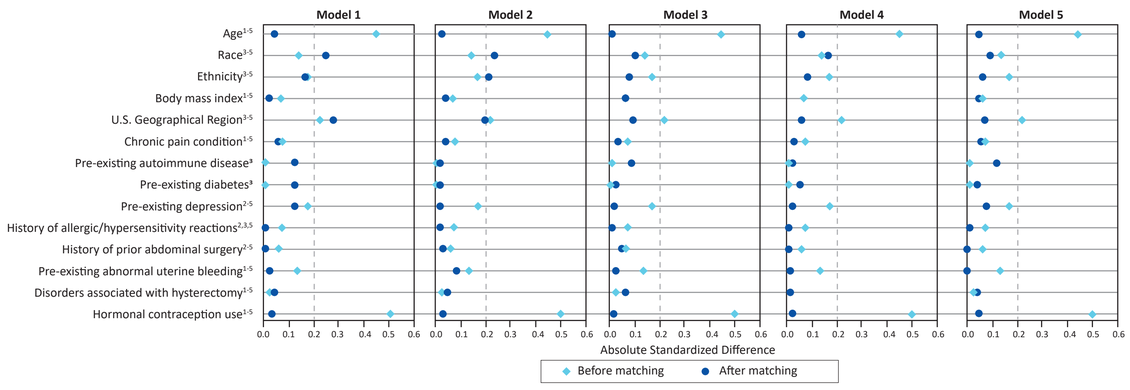
a Chronic pain condition includes chronic/pelvic/abdominal pain, chronic headache, migraine, fibromyalgia, dysmenorrhea (ever) and dyspareunia (ever)
b History of prior abdominal surgery includes caesarean births and any other abdominal surgery
c Pre-existing abnormal uterine bleeding includes intracyclic bleeding (during last 6 months) and prescribed medication to control bleeding in the past 2 years
d Disorders associated with hysterectomy includes endometriosis, fibroids and ovarian cancer
Note: Std diff is calculated as the difference in means or proportions between cohorts (new – SoC), divided by the standard deviation overall. A std diff less than 20% (small effect size) indicates good balance between cohorts

- All 5 models yielded good match results (Table 3)
- The proportion matched between cohorts ranged from 92.1% (Model 2) to 95.0% (Model 4) across the 5 models
- Models 1 and 2 were removed from consideration due to poor balance (Model 1) and the worst match rate (Model 2)
- Models 3, 4 and 5 were then considered:
  - Models 3, 4, and 5 achieved balance across all 14 confounders, as mean std diff ranged from 0.0383 (Model 4) to 0.0530 (Model 5), and maximum std diff from 0.1023 (Model 3) to 0.1616 (Model 4)
  - Among Tier 1 confounders, the maximum std diff ranged from 0.0608 (Model 4) to 0.0686 (Model 5)

Table 3. Propensity Score Model Selection Criteria Evaluation

Model	Mean absolute std diff among all variables		Maximum absolute std diff among all variables		Maximum absolute std diff among Tier 1 variables		New cohort patients matched (%)	
	Value	Rank	Value	Rank	Value (covariate)	Rank	Value	Rank
Model 1	0.0931	5	0.2799	5	0.0607 (chronic pain)	1	94.6%	2
Model 2	0.0718	4	0.2403	4	0.0853 (preexisting abnormal uterine bleeding)	5	92.1%	5
Model 3	0.0481	2	0.1023	1	0.0671 (body mass index)	3	93.9%	4
Model 4	0.0383	1	0.1616	3	0.0608 (age)	2	95.0%	1
Model 5	0.0530	3	0.1177	2	0.0686 (chronic pain)	4	94.3%	3

Note: Tier 1 covariates include age, BMI, chronic pain condition, pre-existing abnormal uterine bleeding, disorders associated with hysterectomy and hormonal contraception use.
Tier 2 covariates include pre-existing depression, history of prior abdominal surgery and history of allergic/hypersensitivity reactions. Tier 3 covariates include race, ethnicity, U.S. geographical region, pre-existing autoimmune disease and pre-existing diabetes.



Note: Superscript numbers following covariates on y-axis represent inclusion in the respective model

Figure 1. Covariate Absolute Standardized Differences Before and After Matching

Recommended Model

- None of the models achieved the best ranking in all matching consideration criteria, so the model selection could not be determined on data alone
- After clinical and analytical considerations, Model 4 was selected as the final model for the following reasons:
  - Highest cohort match percent
  - Balance across all confounders (mean std diff and maximum std diff <0.20)
  - After eliminating Models 1 and 2 from consideration:
    - Best mean and maximum absolute std diff among Tier 1 variables
    - Lowest mean std diff and lowest maximum std diff for Tier 1 confounders
  - Only negative is an elevated std diff on race, a Tier 3 confounder, which was determined not to be of clinical importance

Conclusions

- This analysis provides further evidence to the added value of clinical expertise in PS matching model development and selection for controlling for bias in observational studies
- Clinical expertise is important to instilling confidence in a largely data-driven approach, for the interpretation of the study's results in the medical community

Abbreviations: PS, propensity score; SD, standard deviation; SoC, standard of care; std diff, standardized difference; U.S., United States

Presented at ISPOR Europe 2022

labcorp Drug Development