Prediction & Propensity

Charles Miller, PhD
Professor & Associate Vice President
Department of Cardiothoracic and Vascular Surgery
The University of Texas Medical School at Houston
Memorial Hermann Heart & Vascular Institute

Financial Disclosure

- Nothing to Disclose

Individual Prediction - Outcome

\[
\text{Exp}(Y = \beta_0 + \beta_1 X_1)
\]

Population Prediction

Equation solved for each individual patient
Aggregated over entire population

Used for:
- Research
- Generalizing results to populations
- Quality assessment – O/E ratios
- STS hospital comparisons

Limits to Statistical Power

- 10 events per variable required
- Event rate more important than population size
- Rare events hard to model even in large populations
- Complex problems may need many variables to explain

Analytical Framework

\(Y=\text{outcome}\) in conventional models

\(Y=\text{treatment}\) in propensity models

Predictors of treatment – reasons for selection

"Quasi-randomization" to reduce selection bias

Assumption: Treatment possible to randomize
Individual Prediction - Treatment

\[ \text{Exp}(Y=b_0 + b_1 X_1) \]

Population Distribution

Model form still exponential (non-normal)
More common treatment

Population Distribution - Scores

- \( \text{Logit (Log Odds)} \)

- \( \text{Mean} = \text{arithmetic average} \)
- \( \text{Median} = \text{midpoint value} \)
- \( \text{Mode} = \text{most frequent value} \)

- \( \text{Normal Distribution:} \)
  - \( \text{Mean} = \text{Median} = \text{Mode} \)
  - \( 68\% \text{ area } \pm 1\text{SD} \)
  - \( 95\% \text{ area } \pm 2\text{SD} (1.96) \)
  - \( 99.7\% \text{ area } \pm 3\text{SD} \)

Population Distribution - Scores

Treatment probabilities segregate into distributions

Two Main Approaches

- \( \text{Regression-Based Adjustment:} \)
  - Propensity score reduces covariates to single term
  - Can be modeled with other variables
  - Score term may be difficult to interpret

- \( \text{Propensity Matching:} \)
  - Restricts analysis to overlapping score range
  - Simplifies further analysis
  - Approximates randomization

Population Distribution - Scores

Major assumption:
Same overall population

Treatment 1
Treatment 2
Population Distribution - Scores

Requirements:
- Regression can use entire distribution
- Matching only overlapping subset

Regression population

Match population

Treatment 1  Treatment 2

Theory vs Reality

Newer treatment – smaller population

Propensity is for TREATMENT

Population should be eligible for either treatment

Treatment should be randomizable

Not suited for immutable characteristics

Population Distribution - Scores

Theory vs Reality

Newer treatment – smaller population

May be highly selected

May generalize poorly

No TRT

2 1

Fundamental Assumptions

Limitations

Regression Adjustment:
- Preserves sample size
- More flexible
- Doubly robust
- Lose ability to interpret individual risk factors

Propensity Matching:
- Restricts analysis to overlapping score range
- Simplifies further analysis
- Approximates randomization
- Most effective at balancing in large samples
- May sharply reduce sample size in real life
- May generalize poorly
Limitations

Great tool in the proper context
Like any model, may not reflect reality
Patient selection may still be biased
Is not a substitute for randomized trials