Propensity Score Calibration

HUG, October 27, 2009

Enhancing the effectiveness of health care for Ontarians through research
We’ve Got Data – Observational Data

• Data collected to administer the health care system
• Not collected for research
• Generally concerned with payments
  ➤ Payments to physicians
  ➤ Payments to pharmacists
  ➤ Hospital budgets based on patient load

Would like to use this data to answer research questions
Advantages of Admin Data

- Low cost relative to collecting primary data
- (but definitely not low-cost)
- Encompasses (almost) the full range of recipients and providers
  - not just physicians and patients in academic centres, or subjects who meet strict enrollment criteria
- Can follow people across the continuum of care
- Can include people who do not receive care, as well as those who do
Disadvantages of Admin Data

• People are not randomly assigned to their treatments
  ▶ E.g. People receiving drug A were older and sicker than people receiving drug B.
  ▶ Differences in outcome may be due to differences the patients, rather than differences due to their treatment

• Clinical information (severity of disease) lacking
  ▶ May be supplemented by chart reviews

• Long delay in (some) data availability
Propensity Score Analysis

• One method for analyzing observational data
• Propensity score = probability of being in one of the two treatment groups
• Calculated using logistic regression

proc logistic;
  model Drug_A = age sex comorbidities ...;
  output out = results predicted = phat;
proc logistic;
  model Drug_A = age sex comorbidities ...;
  output out = results predicted = phat;

Propensity score =
  logit(probability(Drug_A))
Once the Propensity Score is Calculated...

- Match people from treatment A with people from treatment B
  - Match on the propensity score (+/- caliper)
  - Can additionally match on “important” variables (e.g. age, sex, specific comorbidities)
  - 1:1 or 1:many
- Use paired analysis to compare outcomes of matched sets (e.g. conditional logistic regression)
Enhancing the effectiveness of health care for Ontarians through research
But ...
Missing Key Clinical Information

- Observational data are not collected in order to promote your research question
- Missing key clinical information
  - Smoking status
  - Immigration status
  - BMI
  - Self-reported health status
Calibrated Propensity Score

• You must have a source of additional data for a subset of your sample
  ➤ Survey data (e.g. CCHS)
  ➤ Chart review
For each person in the calibrated data set

- Original propensity score
  \[ PS_{\text{original}} = \alpha_0 + \alpha_1(\text{sex}) + \alpha_1(\text{age}) \]

- Calibrated propensity score
  \[ PS_{\text{calibrated}} = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{sex}) + \beta_3(\text{smoking status}) \]

- Relationship between original and calibrated PS
  \[ PS_{\text{calibrated}} = \delta_0 + \delta_1(\text{treatment}) + \delta_2(PS_{\text{original}}) \]
• $PS_{\text{calibrated}} = \delta_0 + \delta_2(PS_{\text{original}})$
  ▶ No need to re-run the matching

• $PS_{\text{calibrated}} = \delta_0 + \delta_1(\text{treatment}) + \delta_2(PS_{\text{original}})$
  ▶ Use this relationship to calculate the (estimated) $PS_{\text{calibrated}}$ for everyone
  ▶ Re-create the matched pairs
Stürmer’s paper

- Original analysis used the PS as a covariate rather than matching on the PS
- Logistic regression used to estimate the effect of treatment on a binary outcome

```
proc logistic;
model outcome = treatment propensity_score;
```
Original analysis

\[ \text{PS\_original} = \logit(\alpha_0 + \alpha_1(\text{sex}) + \alpha_1(\text{age})) \]

\[ \text{P(outcome)} = \logit(\beta_0 + \beta_1(\text{treatment}) + \beta_2(\text{PS\_orig})) \]

What is wanted

\[ \text{PS\_calibrated} = \logit (\gamma_0 + \gamma_1(\text{age}) + \gamma_2(\text{sex}) + \gamma_3(\text{smoking status}) ) \]

\[ \text{P(outcome)} = \logit(\eta_0 + \eta_1(\text{treatment}) + \eta_2(\text{PS\_calibrated})) \]

How to get there

\[ \text{PS\_calibrated} = \delta_0 + \delta_1(\text{treatment}) + \delta_2(\text{PS\_original}) \]

\[ \eta_1 = \beta_1 - \frac{\delta_1 \beta_2}{\delta_2} \]
References

PS Calibration:

Propensity score methods:
Biom J. 2009; 51(1): 171-84