Measuring Balance and Model Selection in Propensity Score Methods

Svetlana V Belitser¹, Edwin P Martens¹,², Wiebe R Pestman¹,², Rolf HH Groenwold¹,², Anthonius de Boer¹ and Olaf H Klungel¹,²

1 Department of Pharmacoepidemiology and Pharmacology, Utrecht Institute of Pharm. Sciences, Utrecht University, Netherlands
2 Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands.

ISPE, 14-17 august 2011
This work was supported by funding from

- the European Community’s Seventh Framework Programme for the Innovative Medicine Initiative
- Pfizer
- GlaxoSmithKline
- Top Institute Pharma
- co-funding from universities, government, and industry, the Dutch Medicines Evaluation Board and the Dutch Ministry of Health
Data simulation

- $x_1, ..., x_9 \sim \mathcal{N}(0, 1)$
- $\text{logit}(p_{i,t}) = \beta_{0,t} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 + \beta_6 x_8 + \beta_7 x_2 x_4 + \beta_8 x_2 x_7 + \beta_9 x_7 x_8 + \beta_{10} x_4 x_5 + \beta_{11} x_1^2 + \beta_{12} x_7^2$

$\implies t_i \sim \text{Bern}(p_{i,t})$

- $\text{logit}(p_{i,y}) = \alpha_{0,y} + 0.89 t + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_2 x_4 + \alpha_8 x_3 x_5 + \alpha_9 x_3 x_6 + \alpha_{10} x_4 x_5 + \alpha_{11} x_1^2 + \alpha_{12} x_6^2$

$\implies y_i \sim \text{Bern}(p_{i,y})$

<table>
<thead>
<tr>
<th>Confounding</th>
<th>variables</th>
<th>terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only treatment-related</td>
<td>$x_1, x_2, x_4, x_5$</td>
<td>$x_1, x_2, x_4, x_5, x_2 x_4, x_4 x_5, x_1^2$</td>
</tr>
<tr>
<td>Only outcome-related</td>
<td>$x_7, x_8$</td>
<td>$x_7, x_8, x_2 x_7, x_7 x_8, x_7^2$</td>
</tr>
<tr>
<td></td>
<td>$x_3, x_6$</td>
<td>$x_3, x_6, x_3 x_5, x_3 x_6, x_6^2$</td>
</tr>
</tbody>
</table>
Data simulation

- $x_1, ..., x_9 \sim \mathcal{N}(0, 1)$
- $\logit(p_{i,t}) = \beta_{0,t} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 + \beta_6 x_8 + \beta_7 x_2 x_4 + \beta_8 x_2 x_7 + \beta_9 x_7 x_8 + \beta_{10} x_4 x_5 + \beta_{11} x_1^2 + \beta_{12} x_7^2$
  $\implies t_i \sim \text{Bern}(p_{i,t})$

- $\logit(p_{i,y}) = \alpha_{0,y} + 0.89 t + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_2 x_4 + \alpha_8 x_3 x_5 + \alpha_9 x_3 x_6 + \alpha_{10} x_4 x_5 + \alpha_{11} x_1^2 + \alpha_{12} x_6^2$
  $\implies y_i \sim \text{Bern}(p_{i,y})$

<table>
<thead>
<tr>
<th>Confounding</th>
<th>variables</th>
<th>terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only treatment-related</td>
<td>$x_1, x_2, x_4, x_5$</td>
<td>$x_1, x_2, x_4, x_5, x_2 x_4, x_4 x_5, x_1^2$</td>
</tr>
<tr>
<td>Only outcome-related</td>
<td>$x_7, x_8$</td>
<td>$x_7, x_8, x_2 x_7, x_7 x_8, x_7^2$</td>
</tr>
<tr>
<td></td>
<td>$x_3, x_6$</td>
<td>$x_3, x_6, x_3 x_5, x_3 x_6, x_6^2$</td>
</tr>
</tbody>
</table>
Marginal effect if $OR_{conditional} = e^{0.89} = 2.45$

Probability of the outcome for each subject in the simulated dataset:

- as if the subject was **exposed to the treatment**

\[
\text{logit}(p_{i,t=1}) = \alpha_{0,y} + 0.89 \times 1 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_2 x_4 + \alpha_8 x_3 x_5 + \alpha_9 x_3 x_6 + \alpha_{10} x_4 x_5 + \alpha_{11} x_1^2 + \alpha_{12} x_6^2
\]
\[
\Rightarrow \bar{p}_{dataset,t=1} = \text{mean}(p_{i,t=1})
\]

- as if the subject was **not exposed to the treatment**

\[
\text{logit}(p_{i,t=0}) = \alpha_{0,y} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_2 x_4 + \alpha_8 x_3 x_5 + \alpha_9 x_3 x_6 + \alpha_{10} x_4 x_5 + \alpha_{11} x_1^2 + \alpha_{12} x_6^2
\]
\[
\Rightarrow \bar{p}_{dataset,t=0} = \text{mean}(p_{i,t=0})
\]

\[
OR_{marginal} = \frac{\bar{p}_{dataset,t=1}/(1 - \bar{p}_{dataset,t=1})}{\bar{p}_{dataset,t=0}/(1 - \bar{p}_{dataset,t=0})} = 2
\]
Propensity score approach to estimate marginal OR

Propensity score: \( PS = P(t = 1|x_1, \ldots, x_p) \)
Marginal OR: \( \text{logit}(p_i) = \beta_0 + \beta_{marg} t + PS \implies OR_{marginal} = e^{\beta_{marg}} \)

We can estimate PS using logistic regression: \( \hat{PS} \)

PS is a balancing score: \( P(x|t = 1, PS) = P(x|t = 0, PS) \)

\( \implies \) sample balance for \( \hat{PS} \): \( P(x|t = 1, \hat{PS}) \approx P(x|t = 0, \hat{PS}) \)

Balance measures
- Overlapping coefficient
- Kolmogorov-Smirnov distance
- Lévy distance
- Absolute \( t \)-statistic
- Absolute mean difference
- Absolute standardized mean difference
Measuring Balance and Model Selection in Propensity Score Methods

Overlapping coefficient

\( \hat{OVL} = \int_{-\infty}^{\infty} \min_t \{ \hat{f}(x|t = 0), \hat{f}(x|t = 1) \} \, dx \)
Kolmogorov-Smirnov distance

\[ \hat{D} = \max_x \{ |\hat{F}(x|t = 0) - \hat{F}(x|t = 1)| \} \]
\[
\hat{L} = \min_{\epsilon > 0} \{ \epsilon : \hat{F}(x-\epsilon|t=0) - \epsilon \leq \hat{F}(x|t=1) \leq \hat{F}(x+\epsilon|t=0) + \epsilon \text{ for all } x \text{ in } \mathbb{R} \}
\]
Measuring Balance and Model Selection in Propensity Score Methods

9 covariates * 5 strata = 45 balance measures.

To get an overall balance measure for every fitted PS model

- **Mean**$(\hat{KS}_{ij})$
- **Median**$(\hat{KS}_{ij})$
- **Weighted average** $\hat{KS}_w = \frac{1}{JJ} \sum_{i=1}^{l} \sum_{j=1}^{J} w_i \hat{KS}_{ij}$
  
  weight 1: $w_i = 1 + \log(\hat{OR}_{x_i,y}) - \frac{1}{l} \sum_{k=1}^{l} \log(\hat{OR}_{x_k,y})$
  
  weight 2: $w_i = 1 + |\log(\hat{OR}_{x_i,y})| - \frac{1}{l} \sum_{k=1}^{l} |\log(\hat{OR}_{x_k,y})|$
40 Propensity Score models

model 1 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_2 x_4 + x_2 x_7 + x_7 x_8 + x_4 x_5 + x_1^2 + x_7^2 \]
model 2 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_2 x_4 + x_2 x_7 + x_7 x_8 + x_4 x_5 + x_1^2 + x_7^2 + x_9 \]
model 3 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_2 x_4 + x_2 x_7 + x_7 x_8 + x_4 x_5 \]
model 4 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_1^2 + x_7^2 \]
model 5 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 \]
model 6 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_9 \]
model 7 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_2 x_4 + x_2 x_7 + x_7 x_8 + x_1^2 + x_7^2 \]
model 8 \[ \text{logit}(t) \sim x_1 + x_2 + x_5 + x_7 + x_8 + x_2 x_7 + x_7 x_8 + x_1^2 + x_7^2 \]
model 9 \[ \text{logit}(t) \sim x_1 + x_2 + x_7 + x_8 + x_2 x_7 + x_7 x_8 + x_1^2 + x_7^2 \]
model 10 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_2 x_4 + x_4 x_5 + x_1^2 \]
model 11 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_2 x_4 + x_4 x_5 + x_1^2 + x_9 \]
model 12 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_2 x_4 + x_4 x_5 \]
model 13 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_1^2 \]
model 14 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 \]
model 15 \[ \text{logit}(t) \sim x_1 + x_4 + x_5 \]
model 16 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_9 \]
model 17 \[ \text{logit}(t) \sim x_1 + x_9 \]
model 18 \[ \text{logit}(t) \sim x_2 + x_7 + x_8 + x_2 x_7 + x_7 x_8 + x_7^2 \]
model 19 \[ \text{logit}(t) \sim x_7 + x_8 + x_2 x_7 + x_7 x_8 \]
model 20 \[ \text{logit}(t) \sim x_7 + x_8 + x_7^2 \]
model 21 \[ \text{logit}(t) \sim x_7 + x_8 \]
model 22 \[ \text{logit}(t) \sim x_7 + x_8 + x_3 + x_6 \]
model 23 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_7 + x_8 + x_2 x_4 + x_2 x_7 + x_7 x_8 + x_4 x_5 + x_1^2 + x_7^2 + x_3 \]
model 24 \[ \text{logit}(t) \sim x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 \]
model 25 \[ \text{logit}(t) \sim x_1 + x_2 + x_4 + x_5 + x_2 x_4 + x_4 x_5 + x_1^2 + x_3 + x_6 + x_3 x_5 + x_3 x_6 + x_6^2 \]
\ldots \]
\ldots
Median(\(\hat{OR}_{ty}\)) and SD(\(\hat{OR}_{ty}\)) when using balance measures for PS model selection.

<table>
<thead>
<tr>
<th>Sample size:</th>
<th>(n=400)</th>
<th>(n=1200)</th>
<th>(n=2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overlapping coefficient</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>3.52 (1.21)</td>
<td>2.59 (0.80)</td>
<td>2.10 (0.38)</td>
</tr>
<tr>
<td>median</td>
<td>3.10 (1.14)</td>
<td>2.30 (0.70)</td>
<td>2.11 (0.54)</td>
</tr>
<tr>
<td>weight1</td>
<td>3.10 (1.15)</td>
<td>2.24 (0.50)</td>
<td>2.08 (0.28)</td>
</tr>
<tr>
<td>weight2</td>
<td>3.51 (1.22)</td>
<td>2.37 (0.70)</td>
<td>2.08 (0.30)</td>
</tr>
<tr>
<td><strong>Kolmogorov-Smirnov distance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>2.54 (0.87)</td>
<td>2.11 (0.35)</td>
<td>2.00 (0.25)</td>
</tr>
<tr>
<td>median</td>
<td>2.54 (0.88)</td>
<td>2.12 (0.60)</td>
<td>2.08 (0.50)</td>
</tr>
<tr>
<td>weight1</td>
<td>2.22 (0.65)</td>
<td>2.06 (0.31)</td>
<td>2.01 (0.24)</td>
</tr>
<tr>
<td>weight2</td>
<td>2.31 (0.76)</td>
<td>2.11 (0.32)</td>
<td>2.01 (0.24)</td>
</tr>
<tr>
<td><strong>Lévy distance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>2.60 (0.90)</td>
<td>2.14 (0.34)</td>
<td>2.00 (0.24)</td>
</tr>
<tr>
<td>median</td>
<td>2.66 (0.93)</td>
<td>2.13 (0.59)</td>
<td>2.06 (0.50)</td>
</tr>
<tr>
<td>weight1</td>
<td>2.29 (0.77)</td>
<td>2.08 (0.32)</td>
<td>2.01 (0.23)</td>
</tr>
<tr>
<td>weight2</td>
<td>2.43 (0.78)</td>
<td>2.12 (0.32)</td>
<td>2.01 (0.24)</td>
</tr>
<tr>
<td><strong>Absolute t-statistic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>2.20 (0.64)</td>
<td>2.02 (0.30)</td>
<td>1.97 (0.22)</td>
</tr>
<tr>
<td>median</td>
<td>2.20 (0.70)</td>
<td>2.06 (0.39)</td>
<td>2.00 (0.35)</td>
</tr>
<tr>
<td>weight1</td>
<td>2.13 (0.64)</td>
<td>2.03 (0.31)</td>
<td>1.98 (0.24)</td>
</tr>
<tr>
<td>weight2</td>
<td>2.21 (0.64)</td>
<td>2.02 (0.30)</td>
<td>1.97 (0.22)</td>
</tr>
<tr>
<td><strong>Absolute standardized mean difference</strong> ((\approx) Absolute mean difference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>2.26 (0.66)</td>
<td>2.05 (0.32)</td>
<td>1.99 (0.24)</td>
</tr>
<tr>
<td>median</td>
<td>2.31 (0.73)</td>
<td>2.11 (0.47)</td>
<td>2.05 (0.36)</td>
</tr>
<tr>
<td>weight1</td>
<td>2.16 (0.65)</td>
<td>2.04 (0.31)</td>
<td>2.00 (0.25)</td>
</tr>
<tr>
<td>weight2</td>
<td>2.22 (0.65)</td>
<td>2.04 (0.33)</td>
<td>1.99 (0.24)</td>
</tr>
</tbody>
</table>

True \(OR_{ty} = 2\).

200 simulations per sample size.
\( \hat{OR}_{ty} \) and \( \text{Median}(\hat{OR}_{ty}) \) for the PS models only with confounders:

- **with all 4 confounders, all 7 terms**
  - Sample size = 400 800 1200 1600 2000
  - Median = 2.10 1.99 2.04 2.00 2.00

- **with all 4 confounders, 6 terms without \( x_1^2 \)**
  - Sample size = 400 800 1200 1600 2000
  - Median = 2.35 2.24 2.27 2.23 2.24

- **with all 4 confounders, 5 terms without \( x_2x_4, x_4x_5 \)**
  - Sample size = 400 800 1200 1600 2000
  - Median = 2.15 2.07 2.10 2.06 2.07

- **with all 4 confounders, 4 terms**
  - Sample size = 400 800 1200 1600 2000
  - Median = 2.43 2.32 2.34 2.29 2.30

- **with 3 confounders**
  - Sample size = 400 800 1200 1600 2000
  - Median = 2.64 2.50 2.57 2.51 2.53

- **with 1 confounder**
  - Sample size = 400 800 1200 1600 2000
  - Median = 3.13 3.02 3.05 3.00 3.03
\( \hat{OR}_{ty} \) and Median(\( \hat{OR}_{ty} \)) for the PS models with all 7 confounding terms

- **Only all 7 confounding terms**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.10, 1.99, 2.04, 2.00, 2.00

- **All 7 confounding terms and one independent variable**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.10, 1.99, 2.04, 2.00, 2.00

- **All 7 confounding terms and all only outcome-related terms**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.06, 1.97, 2.04, 1.99, 2.01

- **All 7 confounding terms and all only treatment-related terms**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.07, 1.96, 2.00, 1.94, 1.97

- **All 7 confounding terms and all only treatment- and outcome-related terms**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.07, 1.96, 2.00, 1.94, 1.97

- **Only 4 confounding terms**
  - Sample size: 400, 800, 1200, 1600, 2000
  - Median: 2.43, 2.32, 2.34, 2.29, 2.30

---

Measuring Balance and Model Selection in Propensity Score Methods
## Conclusions

### Balance and Balance measures for PS model selection
- more attention for creating, measuring, and reporting the balance
- balance measures can help to select a PS model among a variety of possible models

### PS models
- not only all confounding variables but also all confounding terms (squares, interaction, etc.) in PS model to get marginal effect

---

This work is published <online> in

*Pharmacoepidemiology and Drug Safety 2011*

”Measuring balance and model selection in propensity score methods”