Regression vs. Propensity Score Matching
A comparison based on the Herniamed registry

Daniela Adolf, Galiya Kaldasheva, Martin Hukauf, Thomas Keller, Ferdinand Köckerling

📞 +49 391 5497013
✉️ daniela.adolf@statconsult.de
@ www.statconsult.de
Outline

- Herniamed registry
  - Objectives and background
- Logistic Regression and Propensity Score Matching
  - Basic principles
  - Regression or matching?
- Example
- Summary
Herniamed registry

- Hernias are generally viewed as routine surgical procedures but
  - average recurrence rate of ten percent
  - more than ten percent of patients develop pain after inguinal hernia operations
- Non-profit organisation Herniamed to improve the results and quality of hernia surgery
- Cornerstone: internet-based quality assurance study, started January 2009
- 577 participating clinics and surgeons in private practice from Germany, Austria, Switzerland and Italy (October 10, 2016) have prospectively registered their patients who had undergone hernia operations
  - 353,271 hernia operations
  - 186,485 1 year follow up entries
  - 12,116 5 years follow up entries
Herniamed registry

- Patient overview
Herniamed registry

- Surgery module

```
<table>
<thead>
<tr>
<th>INGUINAL HERNIA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PATIENT</strong></td>
</tr>
<tr>
<td>Surname: Leistenhernie, Mustermann</td>
</tr>
<tr>
<td>ID Patient: 12345678 (365190)</td>
</tr>
<tr>
<td>Date of Birth: 13.08.1976</td>
</tr>
<tr>
<td>Operation: inguinal hernia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>COMPLICATIONS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RIGHT</strong></td>
</tr>
<tr>
<td>Intraoperative complications: yes</td>
</tr>
<tr>
<td>Hemorrhages: no</td>
</tr>
<tr>
<td>Injuries: yes</td>
</tr>
<tr>
<td>Other:</td>
</tr>
<tr>
<td>blood vessel</td>
</tr>
<tr>
<td>bowel</td>
</tr>
<tr>
<td>bladder</td>
</tr>
<tr>
<td>nerve</td>
</tr>
<tr>
<td>Postoperative complications: no</td>
</tr>
<tr>
<td>Reoperation: no</td>
</tr>
</tbody>
</table>

| **LEFT**          |
| Intraoperative complications: -- Please select -- |
| Hemorrhages:      |
| Injuries:         |
| Other:            |
| blood vessel  |
| bowel    |
| bladder  |
| nerve   |
| Postoperative complications: -- Please select -- |
| Reoperation:      |
```

Next
Herniamed analyses

- Scientific interests mainly focussed on certain aspects, e.g. single treatments
- Physicians had to be convinced that multivariable analysis is necessary
  - Mainly logistic regression
  - Potential predictor variables are prospectively specified (baseline characteristics)
- Necessity to use another method when comparing a regional operation method for incisional hernias to others
  - Propensity score matching
  - Matching variables are prospectively specified
- Currently, we use either regression methods or matching
Basic principles

Logistic regression

- \( y = f(x_1 \ldots x_{k-1}x_k) \) using original data
- Logit: \( \ln \frac{\pi}{1-\pi} \), \( \pi = P(y = 1) \)
- Adjusted odds ratio given for one predictor when the others remain the same
  - For the parameter of interest as well as for all others

Propensity score matching

- \( PS = P(x_k = 1 \mid x_1 \ldots x_{k-1}) \)
- Matching
- \( y = f(x_k) \) using matched samples
  - Paired analysis, e.g. Mc Nemar
  - Adjusted OR for matched pairs
Basic principles

Matching on the propensity score

- One-to-one matching
- Without replacement
- Greedy algorithm:
  - A treated subject is first selected at random
  - The untreated subject (within a specified caliper distance) whose propensity score is closest to that of this randomly selected treated subject is chosen for matching to this treated subject, even if it would better serve as a match for a subsequent treated subject

- Balance assessment using standardized differences
  
  \[ d = \frac{|\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}|}{\sqrt{s^2_{\text{treatment}} + s^2_{\text{control}}/2}} \quad \text{and} \quad d = \frac{|\hat{p}_{\text{treatment}} - \hat{p}_{\text{control}}|}{\sqrt{\hat{p}_{\text{treatment}}(1-\hat{p}_{\text{treatment}}) + \hat{p}_{\text{control}}(1-\hat{p}_{\text{control}})/2}} \], \text{ resp.}

- Balance regulation using different caliper widths
Regression or matching?

- Rosenbaum (1983): In studies with limited resources but large control reservoirs and many confounding variables, the confounding variables can often be controlled by multivariate matching, but the small sample sizes in the final groups do not allow control of all variables by model-based methods [ ] exact matching on a balancing score leads to an unbiased estimate of the average treatment effect.
- Austin (2011): In observational studies with a continuous outcome, without unmeasured confounding and in which the true model is known, logistic regression and propensity score methods should result in similar conclusions
- Martens (2008): Propensity score methods give in general treatment effect estimates that are closer to the true marginal treatment effect than a logistic regression model in which all confounders are modelled
- Shah (2005): Estimates obtained using propensity score methods tend to be modestly closer to the null compared with when regression-based approaches were used for estimating odds ratios or hazard ratios
- Cepeda (2003): The propensity score is a good alternative to control for imbalances between study groups [ ] and produced estimates that were less biased, more robust, and more precise than the logistic regression estimates when there are seven or fewer events per confounder variable
- King (2016): Propensity score matching can increase imbalance, model dependence and stat. bias.
- Austin (2011): It is simpler to determine whether the propensity score model has been adequately specified than to assess whether the regression model has been correctly specified
Example

Are there differences in inguinal surgery between the operation methods

- Lichtenstein - open hernia repair
- TEP - laparoscopic repair: totally extra-peritoneal
- TAPP - laparoscopic repair: transabdominal preperitoneal

cconcerning postoperative complications?

The following baseline characteristics have also to be accounted for:

- Age
- BMI
- Sex
- ASA score (I / II / III-IV)
- Defect size (I / II / III)
- EHS localizations (medial, femoral, lateral, scrotal)
- Preoperative pain (yes / no / unknown)
- Coumarin
- Antiplatelet therapy
Example – data selection

All hernias based on the exports at October 10th, 2016 (n=353,271)

Inguinal hernias (n=233,834)

Primary inguinal hernias via Lichtenstein, TEP or TAPP (n=182,213)

Completely documented unilateral primary inguinal hernias via Lichtenstein, TEP or TAPP, patient’s age ≥ 16 (n=141,140)

Completely documented elective unilateral primary inguinal hernias via Lichtenstein, TEP or TAPP, patient’s age ≥ 16, operated before September 1st, 2015 (n=102,476)

Completely documented elective unilateral primary inguinal hernias via Lichtenstein, TEP or TAPP, patient’s age ≥ 16, operated before September 1st, 2015 and having a 1 year follow-up (n=71,809)

Lichtenstein (n=27,359)  TEP (n=17,905)  TAPP (n=26,545)
Example – implementation

• Logistic regression models

```plaintext
PROC LOGISTIC data=hernia_fu(where=(vgl_LichtTAPP ne .));
  class sex asa preop_pain defect_size EHS_medial EHS_lateral EHS_femoral EHS_scrotal
coumarin thrombo vgl_LichtTAPP /order=internal param=glm;
  model post_compl = vgl_LichtTAPP sex asa preop_pain defect_size EHS_medial
 EHS_lateral EHS_femoral EHS_scrotal coumarin thrombo age_10 BMI_5;
RUN;
```

• Propensity Score Matching
  – Logistic regression on pairwise operation method comparisons
  – % gmatch (Bergstrahl E, Kosanke J (2003). Division of Biomedical Statistics and Informatics, Mayo Clinic)
  – McNemar and adjusted OR for matched pairs

<table>
<thead>
<tr>
<th>Caliper widths</th>
<th>0.1 std</th>
<th>0.25 std</th>
<th>0.5 std</th>
<th>0.75 std</th>
<th>1.0 std</th>
<th>1.25 std</th>
<th>1.5 std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairs n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lichtenstein vs. TEP</td>
<td>14760 (82.4)</td>
<td>14933 (83.4)</td>
<td>15418 (86.1)</td>
<td>15988 (89.3)</td>
<td>16512 (92.2)</td>
<td>16973 (94.8)</td>
<td>17314 (96.7)</td>
</tr>
<tr>
<td>Lichtenstein vs. TAPP</td>
<td>19577 (73.8)</td>
<td>19789 (74.5)</td>
<td>20337 (76.6)</td>
<td>21084 (79.4)</td>
<td>21832 (82.2)</td>
<td>22639 (85.3)</td>
<td>23406 (88.2)</td>
</tr>
<tr>
<td>TEP vs. TAPP</td>
<td>17580 (98.2)</td>
<td>17631 (98.5)</td>
<td>17730 (99.0)</td>
<td>17811 (99.5)</td>
<td>17844 (99.7)</td>
<td>17848 (99.7)</td>
<td>17849 (99.7)</td>
</tr>
</tbody>
</table>
Example – balance assessment

- **Lichtenstein vs. TAPP**

```latex
\begin{center}
\begin{tabular}{l|c|c}
 & Lichtenstein & TAPP & Stand. Diff. \\
\hline
ASA score I & 6144 & 30.21 & 6278 & 30.87 & 0.014 & 0.175 \\
ASA score II & 11125 & 54.70 & 11028 & 54.23 & 0.010 & 0.115 \\
ASA score III-IV & 3068 & 15.09 & 3031 & 14.90 & 0.005 & 0.356 \\
\end{tabular}
\end{center}
```

- **Caliper = 0.5 std**
Example – Results

Postoperative complications

- Logistic Regression
- Propensity Score Matching
Example – Results

• Presentation of the matched pair results

<table>
<thead>
<tr>
<th></th>
<th>Lichtenstein</th>
<th>TAPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Postoperative complications</td>
<td>yes</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>602</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Disadvantage</th>
<th>OR for matched pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lichtenstein</td>
<td>TAPP</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Postoperative complications</td>
<td>762</td>
<td>3.75</td>
</tr>
</tbody>
</table>
Summary

✓ Logistic regression and propensity score matching yield similar results on the Herniamed registry
✓ Similar uncertainty of estimates
✓ Slightly smaller effects using propensity score matching
✓ Quite robust results using propensity score matching within a wide range of caliper widths

+ Propensity score matching is suitable for rare events (but also in general cases)
  – Choice of caliper widths
  – No information about effects of baseline characteristics
References


Thank you for your attention.