Modelling the Natural History of Schizophrenia:
Comparison of Naïve Vs. Advanced Statistical Methods

Aurélie Millier\textsuperscript{a}, Shamil Sadikov\textsuperscript{b}, Santiago G Moreno\textsuperscript{c}, Mondher Toumi\textsuperscript{d}

\textsuperscript{a}Creativ-Ceutical, Paris, France; \textsuperscript{b}Hoffmann-La Roche Ltd, Basel, Switzerland; \textsuperscript{c}University of Lyon, Lyon, France

BACKGROUND

- Schizophrenia is a serious public health problem: it affects approximately 1% of the world’s population\textsuperscript{1} and is a leading cause of disability.\textsuperscript{1}
- In longitudinal studies of schizophrenia patients, we observed correlated outcomes over time that may pose a challenge for modelling.
- Most discrete-time Markov chain models use naïve methods for estimating transition probabilities and resource use from longitudinal data.
- Naïve methods have a number of shortcomings, including not accounting for dependence among observations (due to repeated measures from the same patients); ignoring the non-normality and skewness of data; and not allowing for adjustment of patient characteristics (covariates).
- Advanced statistical methods exist that overcome such limitations, providing more robust results, but are rarely applied.

OBJECTIVE

- This study aims to illustrate how advanced methods can be applied to longitudinal data for populating discrete-time Markov chain models. Results are compared in an empirical data set.

METHODS

- A cohort of 1,208 schizophrenia patients (EuroSC) was interviewed at 6-month intervals over a total of 2 years.
- Patients were classified into 8 health states (HS) using an existing classification\textsuperscript{6} based on the Positive and Negative Syndrome Scale (PANSS).
- HSs represent different symptom profiles, and patients can move forwards and backwards among all of them.
- A cohort Markov model was developed mimicking EuroSC (6-month cycles; 2-year time horizon).
- For the purpose of this case study only two model parameters were investigated (Table 1). They were estimated using the naïve and advanced methods. A different Markov model was populated with estimates derived from each method.

Table 1. Model parameters and method of estimation

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Naïve method</th>
<th>Advanced method (unadjusted)</th>
<th>Advanced method adjusted for patient characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-month transition probabilities</td>
<td>Row proportion of transition matrix, based on raw data (collapsing all time intervals)</td>
<td>Multistate model</td>
<td>Multistate model adjusted for age and gender</td>
</tr>
<tr>
<td>Hospitalisation-days</td>
<td>Arithmetic mean across all over 6 months by HS patients (collapsing all time intervals)</td>
<td>2-part regression model</td>
<td>2-part model adjusted for age and gender</td>
</tr>
</tbody>
</table>

Advanced Method: Multistate Model

- A constant hazard multistate model for panel data was used to estimate the transition rates (MSM package in R statistical software).
- Choice of covariates (age, gender) for illustrative purposes.

Advanced Method: 2-Part Regression Model

- The mean hospitalisation days was computed by multiplying:
  - 1st part: probability of experiencing a hospitalisation over 6 months, given by logistic Generalised Estimating Equations (GEE).
  - 2nd part: mean hospitalisation-days for patients who experienced at least 1 hospital admission, given by non-linear GEE (negative binomial distribution).
- Compound symmetry correlation structure was assumed.

Probabilistic Sensitivity Analysis (PSA)

- The Markov model simulated a cohort of 10,000 patients and was run 1,000 times.
- At each run, input values were generated randomly using the following approach:
  - Transition probabilities generated using cumulative Dirichlet distribution
  - Number of hospitalisation-days generated using a log normal distribution

RESULTS

Estimation of Model Parameters from EuroSC

- Overall, mean transition probabilities were determined to be similar among the three estimation methods, with differences mostly seen beyond the second decimal place (not shown).
- The average 6-month hospitalisation-days (Table 2) ranged:
  - from 4.51 (HS1) to 19.15 (HS8) for the naïve method
  - from 4.17 (HS1) to 14.49 (HS4) for the advanced method
  - from 4.07 (HS7) to 14.67 (HS4) for the advanced adjusted method

Table 2. Hospitalisation-days by method of estimation

<table>
<thead>
<tr>
<th>Health State</th>
<th>Naïve method</th>
<th>Advanced method (unadjusted)</th>
<th>Advanced method adjusted for age &amp; gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1</td>
<td>4.51</td>
<td>4.56</td>
<td>4.27</td>
</tr>
<tr>
<td>HS2</td>
<td>8.31</td>
<td>7.92</td>
<td>7.46</td>
</tr>
<tr>
<td>HS3</td>
<td>11.20</td>
<td>10.12</td>
<td>9.86</td>
</tr>
<tr>
<td>HS4</td>
<td>15.38</td>
<td>14.49</td>
<td>14.67</td>
</tr>
<tr>
<td>HS5</td>
<td>7.53</td>
<td>7.65</td>
<td>7.77</td>
</tr>
<tr>
<td>HS6</td>
<td>12.30</td>
<td>8.74</td>
<td>8.93</td>
</tr>
<tr>
<td>HS7</td>
<td>7.61</td>
<td>4.17</td>
<td>4.07</td>
</tr>
<tr>
<td>HS8</td>
<td>19.15</td>
<td>11.84</td>
<td>11.62</td>
</tr>
</tbody>
</table>

Markov Model Results

- Table 3 presents the Markov model results on the average 6-month hospitalisation-days by patient over all HSs.

Table 3. Markov model results

<table>
<thead>
<tr>
<th>Method of estimation</th>
<th>Mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve method</td>
<td>7.39 (6.70-8.17)</td>
</tr>
<tr>
<td>Advanced method (unadjusted)</td>
<td>6.94 (6.16-7.52)</td>
</tr>
<tr>
<td>Advanced method adjusted for age &amp; gender</td>
<td>6.55 (5.80-7.25)</td>
</tr>
</tbody>
</table>

DISCUSSION

- This case study illustrates the feasibility of applying advanced methods that account for the unique distributional characteristics of longitudinal data, while allowing for adjustment of patient characteristics.
- The assessment of performance of competing statistical methods cannot be achieved with studies of real data alone. Thus, a comprehensive simulation study is underway to properly assess the performance of the different methods in relation to a known truth.

CONCLUSIONS

- This case study demonstrates the feasibility of applying advanced methods for populating Markov models with longitudinal data.
- Ideally, the methods should recognize and account for the unique distributional characteristics of longitudinal data:
  - Multistate model for transition probabilities
  - GEEs for resource utilization
- Research is ongoing to identify the best performing methods under controlled conditions (simulation study).

REFERENCES


ACKNOWLEDGEMENTS

Funding: This research was funded by P. Hoffmann-La Roche, Ltd.
Disclaimer: The authors were responsible for drafting the poster and received editorial, layout, and production support from apothecom, funded by P. Hoffmann-La Roche, Ltd.
Presented at the 16th Annual European Congress of the International Society for Pharmacoeconomics and Outcomes Research, November 2-6, 2013, Dublin, Ireland, PRM105.