Comparison of Bayesian Network and Decision Tree Methods for Predicting Access to the Renal Transplant Waiting List

Sahar BAYAT, Marc CUGGIA, Delphine ROSSILLE, Michèle KESSLER, Luc FRIMAT

INSERM U936, Université Rennes 1, IFR 140, Rennes, France
EA 4003, Nancy Université, France
Introduction

- Renal replacement therapy (RRT):
  - Hemodialysis
  - Peritoneal dialysis
  - Kidney transplantation

- Kidney transplantation:
  - Longer survival
  - Lower long-term cost
  - Graft shortage

- Selection criteria diverges from one center to another
Introduction

• Ideally, selection based on medical factors:
  - Women
  - Elderly
  - Distance from transplantation department
  - Private ownership of dialysis facilities

• NEPHROOLOR healthcare network:
  - French region: Lorraine
  - Access to the renal transplant waiting list:
    - Age
    - Medical factors
  - Conventional statistical methods and Bayesian networks: similar results
• Compare the performance of Bayesian networks and decision trees for predicting registration on the renal transplant waiting list in NEPHROCOLOR network
Material and method

- **NEPHROOLOR healthcare network:**
  - Combines public and private for-profit dialysis facilities
  - Only one transplant centre at university hospital of Nancy

- **Study population:**
  - Adult patients
  - Living in Lorraine
  - Starting RRT in NEPHROOLOR network facilities (incident patients)
  - Between July 1, 1997 and June 30, 2003
Material and method

• Data collection:

- **Social and demographic data**: age, sex and distance between the patient's residence and the department performing transplantation

- **Clinical and biological data at first RRT**: existence of diabetes, cardiovascular disease, respiratory disease, hepatic disease, psychiatric disorder, past history of malignancy, physical impairment of ambulation, Body Mass Index (<20, 20-24.99, ≥25), hemoglobin in (<11 g/dl, ≥11) and serum albumin (<3 g/dl, 3-3.49, ≥3.5)

- **Data related to medical follow up in the NEPHROCOLOR network**: ownership of nephrology facility where the first RRT was performed (public or private), medical follow-up in the department performing transplantation versus 12 other facilities without transplantation
Material and method

- Statistical analysis:
  - Data set:
    1. Training set: 90%
    2. Validation set: 10%
      - Comparison of the two sets: $\chi^2$
  - Training set: Modelling registration on the waiting list by Bayesian network and decision tree
  - Validation set: predictive performances of both models (sensitivity, specificity and positive predictive values)
    - Difference between the two models: McNemar test
Material and method

- **Bayesian network:**
  - Conditional dependences between the variables
  - Probabilistic relationships: diseases and symptoms
  - Directed acyclic graph:
    - Nodes: variables
    - Arcs: relationship between variables
    - *not necessarily a cause-effect relationship*
Material and method

• **Decision tree approach:**
  
  - Tree-structured classifier
  - Built by partitioning data into homogenous classes

  - Root node split into child nodes:
    - Selecting the variable that best classifies the samples according to a split criterion

• **CART** method
Results

• Patients’ characteristics:

- 809 patients included
- mean age: 62.1 ± 14.2 years
- 59.6% male
- 34.5% diabetes
- 44.2% cardiovascular disease
- 11.1% respiratory disease
- 14.1% past history of malignancy
- 19.5% physical impairment
- 5.9% psychiatric disorder

- 212 (26.2%) registered on the transplant waiting list
Results

• Training set:
  729 patients

• Validation set:
  80 patients

• No significant difference between the characteristics of the two sets
Results – Bayesian network

Ownership of facility → Follow up in transplantation center → Albumin

Distance from transplantation center → Registration on the waiting list → Respiratory disease

BMI → Diabetes → Cardiovascular disease → Age

Physical impairment → Sex → Hemoglobin → Psychiatric disorder

Past history of malignancy → Hepatic disease
Results – Bayesian network

• Predictive performances on validation set:

  ▶ Sensitivity: 90.0 % (95%CI: 76.8–100)

  ▶ Specificity: 96.7% (95%CI: 92.2–100)

  ▶ Positive predictive value: 90.0% (95%CI: 76.8–100)

• Correct predictions:

  ▶ 18 out of 20 registrations

  ▶ 58 out of 60 non registrations
Results – Decision tree

NR: Non Registered, R: Registered, CVD: CardioVascular Disease
Bayat – MIE 2009
Results – Decision tree

• Predictive performances on validation set:
  - Sensitivity: 90.0 % (95%CI: 76.8–100)
  - Specificity: 96.7% (95%CI: 92.2–100)
  - Positive predictive value: 90.0% (95%CI: 76.8–100)

• Correct predictions:
  - 18 out of 20 registrations
  - 58 out of 60 non registrations
Results – Bayesian network and Decision tree

- High predictive performances on validation set
- McNemar: No significant difference between the models
- Predictions discordant for 2 patients
- Kappa of concordance: 0.93
Discussion

- Decision tree and the Bayesian methods showed:

  - High performances for predicting access to renal transplant waiting list in NEPHROOLOR network
  - Models highly concordant
  - Age the most important variable for both models
## Discussion

<table>
<thead>
<tr>
<th>Bayesian network</th>
<th>Decision tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Albumin</td>
<td>Albumin</td>
</tr>
</tbody>
</table>
## Discussion

<table>
<thead>
<tr>
<th>Bayesian network</th>
<th>Decision tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Albumin</td>
<td>Albumin</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>BMI</td>
</tr>
<tr>
<td>Follow-up in transplantation center</td>
<td>Distance from transplantation center</td>
</tr>
</tbody>
</table>
## Discussion

<table>
<thead>
<tr>
<th>Bayesian network</th>
<th>Decision tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Albumin</td>
<td>Albumin</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>BMI</td>
</tr>
<tr>
<td>Follow-up in transplantation center</td>
<td>Distance from transplantation center</td>
</tr>
</tbody>
</table>

Visualizes other relationships:

[Diagarm showing relationships between various factors related to health outcomes.]
Discussion

Bayesian network

Cardiovascular disease
Diabetes
Albumin

Respiratory disease
Follow-up in transplantation center

Visualizes other relationships:

Decision tree

Cardiovascular disease
Diabetes
Albumin

BMI
Distance from transplantation center

Bayat – MIE 2009
## Discussion

<table>
<thead>
<tr>
<th>Bayesian network</th>
<th>Decision tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Albumin</td>
<td>Albumin</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>BMI</td>
</tr>
<tr>
<td>Follow-up in transplantation center</td>
<td>Distance from transplantation center</td>
</tr>
<tr>
<td>Visualizes other relationships</td>
<td></td>
</tr>
<tr>
<td>Links variables:</td>
<td>Decision rules :</td>
</tr>
<tr>
<td>complex, direct and indirect ways</td>
<td>Easily derived from decision tree</td>
</tr>
<tr>
<td>interpretation more problematic</td>
<td>Simpler interpretation tool for physicians</td>
</tr>
</tbody>
</table>

Bayat – MIE 2009
Conclusion

• Bayesian network and decision tree predict access to renal transplant waiting list in NEPHROCOLOR with high accuracy

• Models are complementary:
  - Bayesian network: global view of associations
  - Decision tree: more easily interpretable

• Formalizing and optimizing the health care process