The Effects of Treatment Use when Externally Validating a Prediction Model that Did not Include Treatment as a Predictor

Romin Pajouheshnia

Rolf H H Groenwold, Karel G M Moons, Johannes B Reitsma, Linda M Peelen

MEMTAB 2016
Introduction

Prognostic prediction models:

- Individualized risk prediction
- Guidance over decision to initiate treatment
  - Treatments that effectively reduce the risk of developing the prognostic outcome

E.g. QRISK2, Framingham Risk Score

**Predictions:**
Estimates of the untreated prognosis of an individual
Introduction

External validation set

- Individuals who remain untreated (with a certain treatment) throughout follow up
Introduction

External validation set

• Individuals who remain untreated (with a certain treatment) throughout follow up

Individuals may receive treatment before or during follow up

![Diagram showing follow up timeline with events and censoring points.]

- Event
- Censoring
- Treatment initiation

Follow up:

Start of follow up

End of follow up
Introduction

- Effective treatment will modify risk
- The observed model performance could be affected

Untreated risk

E.g. Treated if risk > 20%

Untreated Receive treatment during the study

**Aim**

To evaluate the effect of treatment use in a validation set on model performance when validating a prediction model intended for use in untreated individuals.

**Method**

Simulation study

- Compare the model performance observed in treated and untreated validation sets
  - Across different treatment settings
  - Using different analytical approaches
Simulation design

Step 1: Simulate untreated development set

Development data set (1 million individuals)

Two predictors: $X_1$, $X_2$

\[
\text{logit}(Y) = B_0 + B_1X_1 + B_2X_2 \quad \rightarrow \quad \text{Outcome } Y
\]

Step 2: Fit an optimal model

Develop a prediction model:

\[
\text{logit}(Y^*) = b_0 + b_1X_1 + b_2X_2
\]

Average risk of outcome $= 20\%$

True risk
Simulation design

Step 1: Simulate untreated development set

Development data set (1 million individuals)

Two predictors: X1, X2

\[ \logit(Y) = B0 + B1*X1 + B2*X2 \rightarrow \text{Outcome } Y \]

Step 2: Fit an optimal model

Develop a prediction model:

\[ \logit(Y*) = b0 + b1*X1 + b*X2 \]

Step 3: Simulate untreated external validation set

Draw a validation set from the same background population

• New individuals

• “Ideal” validation set (case-mix issues aside)

Average risk of outcome = 20%
Simulation design

Step 4: Simulate treatment in the validation set

After baseline measurements, allocate treatment to part of the validation set.

- Proportion treated: \(~50\%\)
- Average treatment effect: **Odds Ratio = 0.5** (*Highly effective*)

<table>
<thead>
<tr>
<th>Validation set</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trial</td>
<td>Treatment is randomly allocated</td>
</tr>
<tr>
<td>Observational: no threshold</td>
<td>Probability of treatment is higher in high-risk individuals</td>
</tr>
<tr>
<td>Observational: threshold</td>
<td>Treatment is only allocated to individuals with a baseline risk &gt; 20 %</td>
</tr>
</tbody>
</table>
Simulation design

Step 5: Assess the effect of treatment on the findings of the validation

A. Measure the performance of the model in the full treated validation sets

B. Exclude treated individuals from each set and measure the performance of the model in only the untreated individuals

Calibration

• Observed/expected ratio (O:E)
  • O:E within the highest and lowest risk deciles
  • Calibration plots

Discrimination

• C-index
### Results

#### A. Ignore treatment

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O:E overall</td>
<td>O:E lowest risk decile</td>
</tr>
<tr>
<td>Development set</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>Validation set Untreated</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td>Validation set Randomized trial</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>Validation set Observational: no threshold</td>
<td>0.74</td>
<td>0.99</td>
</tr>
<tr>
<td>Validation set Observational: threshold</td>
<td>0.75</td>
<td>1.02</td>
</tr>
</tbody>
</table>
### Results

#### A. Ignore treatment

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O:E overall</td>
<td>O:E lowest risk decile</td>
</tr>
<tr>
<td><strong>Development set</strong></td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untreated</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized trial</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational: no threshold</td>
<td>0.74</td>
<td>0.99</td>
</tr>
<tr>
<td>Observational: threshold</td>
<td>0.75</td>
<td>1.02</td>
</tr>
</tbody>
</table>
## Results

### A. Ignore treatment

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O:E overall</td>
<td>O:E lowest risk decile</td>
</tr>
<tr>
<td><strong>Development set</strong></td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untreated</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized trial</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational: no threshold</td>
<td>0.74</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational: threshold</td>
<td>0.75</td>
<td>1.02</td>
</tr>
</tbody>
</table>
Results

B. Exclusion of treated individuals

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O:E overall</td>
<td>O:E lowest risk decile</td>
</tr>
<tr>
<td>Full validation set</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Untreated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized trial</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Untreated only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational: no threshold</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Untreated only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational: threshold</td>
<td>1.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>
# Results

## B. Exclusion of treated individuals

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O:E overall</td>
<td>O:E lowest risk decile</td>
</tr>
<tr>
<td><strong>Full validation set</strong></td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Untreated only (Untreated)</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Untreated only (Randomized trial)</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Untreated only (Observational: no threshold)</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Untreated only (Observational: threshold)</td>
<td>1.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Results

B. Exclusion of treated individuals

- Exclusion of treated individuals appears to give correct estimate of calibration
- Limited interpretability when treatment use is non-random
Summary

• If use of an effective treatment in a validation set is ignored, model performance may be underestimated

• Where treatment allocation is random
  • Exclusion of treated individuals during the analysis resulted in correct estimates of model performance

• Where treatment allocation is non-random
  • Exclusion of treated individuals is not an appropriate method to obtain correct estimates of model performance
Reality: more complicated

Follow up

Start of follow up  End of follow up

Event
Censoring
Treatment initiation
Treatment discontinued
Reality: more complicated

Additional challenges

• Heterogeneous treatment use
  • Time on-treatment
  • Treatment switching

• Heterogeneous treatment effects
  • Differential effectiveness across study participants
  • Changes in effect on risk over time

• Multiple treatments/interventions

• Not an isolated problem
Conclusions

When validating a model in a (partially) treated set of individuals:

- Caution required when interpreting findings
- Exclusion of individuals treated after baseline is suboptimal
- Further investigation is needed into methods to account for treatment use
Thank you for your attention.

Are there any questions?

R.pajouheshnia@umcutrecht.nl
References


Proportion treated and effectiveness of treatment

Treated validation set: Randomized trial

Model performance

- Proportion treated
- Treatment effect odds ratio

Legend:
- O:E overall
- O:E lowest risk decile
- OE: highest risk decile
- c-index
Proportion treated and effectiveness of treatment

Treated validation set:
Observational: no threshold

Model performance:
Proportion treated vs Treatment effect odds ratio

Graphs showing:
- O:E overall
- O:E lowest risk decile
- OE: highest risk decile

C-index graph:
- c-index

Proportion treated vs Treatment effect odds ratio
Proportion treated and effectiveness of treatment

Treated validation set:
Observational: threshold

Model performance

Proportion treated

Treatment effect odds ratio
B. Exclusion of treated individuals
Calibration

- Effect of excluding treated individuals from the analysis on the observed calibration