ASSESSING THE IMPACT OF THE CHOICE OF MODELLING STRATEGY FOR QUANTITATIVE COVARIATES ON RISK ADJUSTMENT

Lynne Moore PhD, James Hanley PhD, André Lavoie PhD, Alexis Turgeon MD MSc, FRCPC

Affiliations:
Department of Biostatistics and Epidemiology, McGill University
Unité de Traumatologie-Urgence-Soins Intensifs du CHA, Université Laval

Funding:
Canadian Institutes of Health Research
Fonds de la Recherche en Santé du Québec
Context

- Observational study
- Exposure (E) – Disease (D) association
- Adjustment for Risk Factors (RF)
- E and D are binary, RF are quantitative
- RF have non-linear association with response
Methods for modelling RF

1. Single linear term?
2. Dummy variables on categories?
3. Fractional polynomials?
4. Generalized Additive Models?
OBJECTIVE

Evaluate whether GAM can offer better risk adjustment than single linear terms, dummy variables on categories, and FP
METHODS
Study population

• Quebec Trauma Registry 1999-2006
• Regionalized inclusive trauma system
• 59 trauma centers (level I, II, III, and IV)
• Mandatory participation of all trauma centers
• Uniform inclusion criteria
• Standardized data collection protocols
Analysis

EXPOSURE: Trauma center

DISEASE: Hospital mortality

RISK FACTORS:
1. Age
2. Injury Severity Score
3. Glasgow Coma Score
4. Respiratory rate
5. Systolic blood pressure
1. Randomly select 2 hospitals
2. Derive an adjusted OR of hospital mortality
3. Compare OR obtained with GAM to that obtained with 3 other modelling strategies
4. Repeat 1-3 100 times by resampling with replacement
Logit $\pi_i = \alpha + \beta_1 \text{AGE}_i + \beta_2 \text{ISS}_i + \beta_3 \text{GCS}_i + \beta_4 \text{RR}_i + \beta_5 \text{SBP}_i + \beta_6 \text{HOSP}_i$
Dummy variables on categories

\[
\text{Logit} \pi_i = \alpha + \beta_1 \text{AGE}_{55-64_i} + \beta_2 \text{AGE}_{65-74_i} + \beta_3 \text{AGE}_{75-84_i} + \beta_4 \text{AGE}_{>84_i} \\
+ \beta_5 \text{ISS}_{9-15_i} + \beta_6 \text{ISS}_{16-24_i} + \beta_7 \text{ISS}_{25-40_i} + \beta_8 \text{ISS}_{>40_i} \\
+ \beta_9 \text{GCS}_{9-12_i} + \beta_{10} \text{GCS}_{6-8_i} + \beta_{11} \text{GCS}_{4-5_i} + \beta_{12} \text{GCS}_{3_i} \\
+ \beta_{13} \text{RR}_{0_i} + \beta_{14} \text{RR}_{1-5_i} + \beta_{15} \text{RR}_{6-9_i} + \beta_{16} \text{RR}_{>29_i} \\
+ \beta_{17} \text{SBP}_{0_i} + \beta_{18} \text{SBP}_{1-49_i} + \beta_{19} \text{SBP}_{50-75_i} + \beta_{20} \text{SBP}_{76-89_i} + \beta_{21} \text{HOSP}_{i}
\]
Fractional Polynomials

\[ \text{Logit } \pi_i = \alpha + \beta_1 \text{AGE}_i^{p1} + \beta_2 \text{AGE}_i^{p2} + \beta_3 \text{ISS}_i^{p1} + \beta_4 \text{ISS}_i^{p2} + \beta_5 \text{GCS}_i^{p1} + \beta_6 \text{GCS}_i^{p2} + \beta_7 \text{RR}_i^{p1} + \beta_8 \text{RR}_i^{p2} + \beta_9 \text{HOSP}_i \]
Logit $\pi_i = \alpha + \beta_1 \text{HOSP}_i + s(\text{AGE}_i) + s(\text{ISS}_i) + s(\text{GCS}_i) + s(\text{RR}_i) + s(\text{SBP}_i)$
RESULTS

- \( N = 88,235 \)
- 4731 (5.4%) hospital deaths
- Crude mortality: 1.3% to 14.3%
Predictive accuracy

Discrimination and model fit in the whole study sample (n=123 732) according to the four modeling strategies

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single linear term</td>
<td>0.886 (0.882-0.889)*</td>
<td>37944</td>
</tr>
<tr>
<td>Dummy variables on categories</td>
<td>0.894 (0.891-0.897)*</td>
<td>35364</td>
</tr>
<tr>
<td>FP</td>
<td>0.901 (0.898-0.904)</td>
<td>35054</td>
</tr>
<tr>
<td>GAM</td>
<td>0.901 (0.898-0.905)</td>
<td>34908</td>
</tr>
</tbody>
</table>

*p<0.0001 when compared to the generalized additive model
Predictive accuracy: Age

![Graph showing predictive accuracy for different methods based on age.](image-url)
### Risk adjustment

Median changes in Odds Ratio (OR), Standard Error (SE), and statistical significance (p<0.05) over 100 risk-adjusted hospital comparisons.

<table>
<thead>
<tr>
<th>Model</th>
<th>Median absolute % change in OR* (q1;q3)</th>
<th>Median % change in SE*</th>
<th>% changed statistical significance*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No adjustment</td>
<td>30.60 (12.13 ; 46.56)</td>
<td>-25.66 (-30.76 ; -21.12)</td>
<td>48</td>
</tr>
<tr>
<td>Single linear term</td>
<td>9.94 (4.26 ; 28.84)</td>
<td>-4.62 (-7.73 ; -3.29)</td>
<td>13</td>
</tr>
<tr>
<td>Categories</td>
<td>3.74 (1.93 ; 6.75)</td>
<td>0.48 (-0.89 ; 1.46)</td>
<td>15</td>
</tr>
<tr>
<td>FP</td>
<td>3.59 (1.30 ; 6.18)</td>
<td>0.83 (0.31 ; 2.40)</td>
<td>2</td>
</tr>
<tr>
<td>GAM</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Compared to the Generalized Additive Model
In summary

Results suggest that:

• Single linear term:
  – most parsimonious model
  – can lead to residual confounding

• Dummy variables on categories:
  – Similar effect estimates
  – Change in statistical significance
  – Clinically implausible step functions

• FP :
  – equivalent risk adjustment to GAM
Limitations

- One clinical context
- GAM gold standard?
- Care needed in fitting FP and GAM
CONCLUSION

GAM and FP can both offer superior risk adjustment than the other modelling strategies tested but should be implemented with care!
Thank You!
Systolic Blood Pressure
Injury Severity Score

![Histogram of Injury Severity Score](image-url)
Glasgow Coma Scale score
Injury Severity Score
Glasgow Coma Score

![Graph showing mortality rates vs. Glasgow Coma Scale]

- No transformation
- Dummy variables on categories
- Fractional polynomials
- Generalized additive model
- Observed
Respiratory Rate

![Graph showing the relationship between respiratory rate and mortality.](image-url)
Systolic Blood Pressure

The graph shows the relationship between systolic blood pressure and mortality. The x-axis represents systolic blood pressure, while the y-axis represents mortality (%). Different lines and markers represent various models, including:

- No transformation
- Dummy variables on categories
- Fractional polynomials
- Generalized additive model
- Observed

The observed data points are shown as black squares, and the lines represent different models of the relationship between systolic blood pressure and mortality.