Survival Analysis – methods often used in HTA

Nick Latimer, ScHARR, University of Sheffield

n.latimer@sheffield.ac.uk
Contents

1. What is survival analysis and why do we need it?
2. Overview of survival analysis methods commonly used in NICE TAs
3. Recent developments
What is survival analysis and why do we need it?
Survival analysis: Intro

The analysis of time-to-event data from a specified time origin (e.g. randomisation) until the occurrence of a particular event or endpoint (e.g. disease progression, death, incidence of complication etc.)

Main problem and distinguishing feature of survival data:

- Sometimes events are not experienced during the study or follow-up period.
- This results in incompletely observed outcomes, called censored observations.
Why survival analysis?

- Uses information from censored patients – standard methods (eg logistic regression) would not
- Estimates how long it takes to experience event
- More informative and sensitive than rates at arbitrary point of time
- Most importantly for economic evaluation – we can use survival analysis to extrapolate survival data and estimate mean survival times
Overview of survival analysis methods commonly used in NICE TAs
Background

- Survival estimates are important parameters in a large % of HTAs
  - Eg, ≈40% of NICE Appraisals are in cancer area. Several others will also include survival data

- **Problem:** Survival data is rarely complete due to limited follow-up
  - Need to extrapolate to estimate total survival effect
  - Key for estimating total QALY gain
  - Need to fit some type of model to extrapolate

- **But several modelling options are available**
  - Potential for inconsistencies in methodology used, evaluation results and subsequent recommendations
Parametric survival models

- Parametric survival models use the assumption that the survival data follows an underlying probability distribution.
- Hence survival can be predicted beyond the end of the trial.
- Several parametric survival models exist, e.g.:
  - Exponential
  - Weibull
  - Gompertz
  - Log-logistic
  - Log normal
  - Generalised Gamma

➔ The key is to pick the most appropriate model based upon the plausibility of the underlying probability distribution.
What is the ‘state of the art’?

- Different survival models will be appropriate in different circumstances

- The use of different models in different HTAs is not necessarily a problem

- However, often chosen methods are not systematically justified

- This could lead to the most appropriate survival model not being chosen
What is the ‘state of the art’?

Reviewed 45 NICE TAs in advanced cancer

<table>
<thead>
<tr>
<th>Method for Estimating Mean</th>
<th>Number of TAs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted Means</td>
<td>17 (38%)</td>
</tr>
<tr>
<td>Parametric Models</td>
<td>32 (71%)</td>
</tr>
<tr>
<td>Weibull</td>
<td>23 (51%) (72%)</td>
</tr>
<tr>
<td>Exponential</td>
<td>20 (44%) (63%)</td>
</tr>
<tr>
<td>Gompertz</td>
<td>6 (13%) (19%)</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>9 (20%) (28%)</td>
</tr>
<tr>
<td>Log normal</td>
<td>6 (13%) (19%)</td>
</tr>
<tr>
<td>Gamma</td>
<td>2 (4%) (6%)</td>
</tr>
<tr>
<td>Piecewise modelling</td>
<td>1 (2%) (3%)</td>
</tr>
<tr>
<td>Other ‘hybrid’ methods</td>
<td>2 (4%)</td>
</tr>
</tbody>
</table>
What is the ‘state of the art’?

<table>
<thead>
<tr>
<th>Method for Justifying Approach</th>
<th>Prevalence in TAs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical tests</strong></td>
<td></td>
</tr>
<tr>
<td>AIC test</td>
<td>Relatively rare and not systematically done in combination with other methods of justification</td>
</tr>
<tr>
<td>BIC test</td>
<td></td>
</tr>
<tr>
<td>Sum of squared deviations</td>
<td></td>
</tr>
<tr>
<td>-2 log likelihood statistic</td>
<td></td>
</tr>
<tr>
<td>Log cumulative hazard plot</td>
<td></td>
</tr>
<tr>
<td>Other tests of the hazard function</td>
<td></td>
</tr>
<tr>
<td><strong>Visual inspection</strong></td>
<td>Common, but often only 1 or a subset of possible models</td>
</tr>
<tr>
<td><strong>External data</strong></td>
<td>Rare</td>
</tr>
<tr>
<td><strong>Clinical validity</strong></td>
<td>Rare</td>
</tr>
</tbody>
</table>
Does it matter?

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean survival (control)</th>
<th>Mean survival (intervention)</th>
<th>Mean survival gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>93.9</td>
<td>130.6</td>
<td>36.8</td>
</tr>
<tr>
<td>Exponential</td>
<td>144.2</td>
<td>217.3</td>
<td>73.1</td>
</tr>
<tr>
<td>Gompertz</td>
<td>78.3</td>
<td>98.2</td>
<td>19.9</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>220.6</td>
<td>305.2</td>
<td>84.6</td>
</tr>
</tbody>
</table>
Does it matter?

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean Survival (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>22.9</td>
</tr>
<tr>
<td>Exponential</td>
<td>28.3</td>
</tr>
<tr>
<td>Gompertz</td>
<td>23.1</td>
</tr>
<tr>
<td>Log Normal</td>
<td>28.9</td>
</tr>
</tbody>
</table>
‘State of the art’ summary

- A wide variety of models are used

- **Chosen models often not systematically justified**

- Standard models were usually used – very little use of more flexible models (Generalised F, Generalised Gamma, Piecewise models, spline-based models)

- Too often there was a reliance on justification by visual inspection of a small number of models

- External validity / clinical plausibility was rarely addressed
A familiar story?

http://www.youtube.com/watch?v=MyaVq-UPD2A
Recent developments
Choosing a model

- We need to take a systematic approach to survival modelling

→ NICE DSU have published a technical support document

(DSU Technical Support Document and Latimer MDM paper (2013))

- Construct plots to examine PH and AF assumptions
- Consider proportional treatment effect and fit appropriate models
- Assess internal validity of models (stats tests, monotonicity of hazards over time)
- Analyse external validity (external data, clinical plausibility)
- Present sensitivity analysis using alternative models
- Don’t just pick a Weibull!

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.

- Plots are not straight lines
- Compare log-cumulative hazard plots, q-q plots (or suitable residual plots) to allow initial selection of appropriate models
- Plots are parallel
- Consider PH/AF models
- Plots are not parallel
- Fit individual models
- Plots are not straight lines
- Consider piecewise or other more flexible models

Compare model fits to select the most appropriate model taking into account the completeness of the survival data:

Complete survival data:
- AIC
- BIC
- Log-cumulative hazard plots
- Other suitable statistical tests of internal validity

Incomplete survival data:
- Visual inspection
- External data
- Clinical validity
- AIC
- BIC
- Log-cumulative hazard plots
- Other suitable tests of internal and external validity
- Consider duration of treatment effect

Choose most suitable model based on above analysis.

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.

We need to take a systematic approach to survival modelling.

→ NICE DSU have published a technical support document.

(DSU Technical Support Document and Latimer MDM paper (2013))

Choose a model

- Construct plots to examine PH and AF assumptions.
- Consider proportional treatment effect and fit appropriate models.
- Assess internal validity of models (stats tests, monotonicity of hazards over time).
- Analyse external validity (external data, clinical plausibility).
- Present sensitivity analysis using alternative models.
- Don’t just pick a Weibull!

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.

We need to take a systematic approach to survival modelling.

→ NICE DSU have published a technical support document.

(DSU Technical Support Document and Latimer MDM paper (2013))

Choose a model

- Construct plots to examine PH and AF assumptions.
- Consider proportional treatment effect and fit appropriate models.
- Assess internal validity of models (stats tests, monotonicity of hazards over time).
- Analyse external validity (external data, clinical plausibility).
- Present sensitivity analysis using alternative models.
- Don’t just pick a Weibull!

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.

We need to take a systematic approach to survival modelling.

→ NICE DSU have published a technical support document.

(DSU Technical Support Document and Latimer MDM paper (2013))

Choose a model

- Construct plots to examine PH and AF assumptions.
- Consider proportional treatment effect and fit appropriate models.
- Assess internal validity of models (stats tests, monotonicity of hazards over time).
- Analyse external validity (external data, clinical plausibility).
- Present sensitivity analysis using alternative models.
- Don’t just pick a Weibull!

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.

We need to take a systematic approach to survival modelling.

→ NICE DSU have published a technical support document.

(DSU Technical Support Document and Latimer MDM paper (2013))

Choose a model

- Construct plots to examine PH and AF assumptions.
- Consider proportional treatment effect and fit appropriate models.
- Assess internal validity of models (stats tests, monotonicity of hazards over time).
- Analyse external validity (external data, clinical plausibility).
- Present sensitivity analysis using alternative models.
- Don’t just pick a Weibull!

Complete sensitivity analysis using alternative plausible survival models, and taking into account uncertainty in model parameter estimates.
More complex methods

- Sometimes the standard parametric models won’t be appropriate
  - Hazard plots will have kinks in them
  - Predicted survival times will not match up with external data
  - Models will not fit the data well

- In these circumstances other approaches are required:
  - Piecewise models
  - Royston and Parmar’s flexible spline-based parametric models
  - Bayesian methods and explicit use of external information/data

- Note: extrapolation not based on fact (by definition) – opinions on how to extrapolate may differ. Impossible to ascertain “best” answer
Conclusions

- We need to take a systematic approach to survival modelling
- Survival analyses often a key focus in HTA appraisals
- Some lack of consensus remains
  - Should we start with ‘standard’ models and go through the DSU TSD process
  - Or disregard these and move straight to other methods
- Further research is highly desirable
  - E.g. how to define and measure “valid” and “plausible”, how best to use external information/data
Further reading

• Latimer NR. Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data: Inconsistencies, limitations and a practical guide. Med Decis Making. published online 22 January 2013


• Collett, D. Modelling survival data in medical research (2nd ed.), Boca Raton: Chapman & Hall/CRC, 2003


• Jackson, C.H., Sharples, L.D., Thompson, S.G. Survival models in health economic evaluations: Balancing fit and parsimony to improve prediction. The International Journal of Biostatistics 2010; 6;1;34