Survival Analysis in Health Technology Assessment

Is current practice best practice?

John W Stevens, Director, Centre for Bayesian Statistics in Health Economics, University of Sheffield
Martin Hoyle, Senior Research Fellow, Peninsula Medical School, University of Exeter
Noémi Muszbek, Research Scientist, United BioSource Corporation
Edit Remák, Research Scientist, United BioSource Corporation

Scope of Workshop

- Focus on survival analysis with a mortality endpoint
- Issues excluded due to time restrictions:
  - Progression-free survival
  - Switching and cross-over
  - Meta-analysis of multiple sources
Agenda

- Introduction to survival modelling
- Audience poll on criteria for making decisions on methods of survival extrapolation
- Survey of methods used and lessons learned for economic evaluations
- Improved curve fits to summary survival data
- Using external data to validate survival estimates
- Feedback on poll and discussion
Background

“The time horizon of the model should be long enough to capture relevant differences in outcomes across strategies. A lifetime horizon may be required”

ISPOR-SMDM Modelling Good Research Practices Task Force Report 1

Length of follow-up times in clinical trials is limited

Data from Clinical Trials

Observed Survival

Months
Usual Practice(s)

- Fit survival curve to trial data
- Survival distributions (and types)
  - Weibull (PH or AFT)
  - Exponential (PH or AFT)
  - Gompertz* (PH)
  - Lognormal (AFT)
  - Loglogistic (AFT)
  - Generalised gamma (AFT)

PH: Proportional Hazards Model
AFT: Accelerated Failure Time Model
* Not supported in SAS

Sample Curves – fit to data

![Sample Curves](image)
Sample Curves – long-term projections

More than just curve fitting...

- Biological process
- Examine hazards
- Select potential distributions
- Fit curves
- Goodness of fit to original data
- Validate projections
- Clinical/biological plausibility
- External data
- Select distribution(s)
Survey of Methods Used and Lessons Learned For Economic Evaluations

John W Stevens
ScHARR
University of Sheffield
j.w.stevens@sheffield.ac.uk

MRC Extrapolation Project

- Methods of extrapolating evidence from randomised controlled trials (RCTs) for use in economic evaluation models
- Universities of Sheffield, York and Cambridge
Summary of Probability Distributions used to Analyse Patient-level Data in Cost-effectiveness Analyses Alongside RCTs in NICE and non-NICE Submissions

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>12 (52%)</td>
</tr>
<tr>
<td>Weibull</td>
<td>13 (57%)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>3 (13%)</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>3 (13%)</td>
</tr>
<tr>
<td>Gompertz</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>Hybrid Kaplan-Meier</td>
<td>1 (4%)</td>
</tr>
</tbody>
</table>

Modelling Approach

- Many analyses model patient-level time-to-event data according to a limited number of standard probability distributions
- It does not appear to be well-known that most of these standard probability distributions are part of the Generalised F family
Summary of Assumptions Made When Modelling Patient-level Data in Cost-effectiveness Analyses Alongside RCTs in NICE and non-NICE Submissions

<table>
<thead>
<tr>
<th>Model Assumptions</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional hazards assumed</td>
<td>19 (83%)</td>
</tr>
<tr>
<td>Justification for model choice:</td>
<td></td>
</tr>
<tr>
<td>Goodness-of fit</td>
<td>2 (9%)</td>
</tr>
<tr>
<td>Graphical</td>
<td>3 (13%)</td>
</tr>
<tr>
<td>Previous experience</td>
<td>3 (13%)</td>
</tr>
<tr>
<td>Not stated/unclear</td>
<td>15 (65%)</td>
</tr>
<tr>
<td>External validation</td>
<td>9 (39%)</td>
</tr>
</tbody>
</table>
Model Assumptions (1)

- The majority of analyses have assumed proportional hazards
  - This assumes that the treatment effect is constant over the observed and unobserved period
- The majority of analyses provide no justification for the choice of model
- The majority of analyses provide no assessment of external validity

Model Assumptions (2)

- The choice of probability distribution is a subjective decision that should be based on clinical plausibility as well as statistical criteria
- It can also be informative to plot the hazard function, and the cumulative hazard function against time (or their logarithms against log time)
  - Sufficient conditions for the shape of a hazard function are given by Glaser (1980)
- Patients are typically heterogeneous
  - No single standard distribution may adequately represent the data
  - In the absence of prognostic data, mixture models may better represent the data
Uncertainty

- Parameters in time-to-event models may be estimated using Classical or Bayesian methods
- There is more to uncertainty than sampling variation
  - Structural uncertainty (e.g. model averaging)
  - Parameter uncertainty (e.g. scope, study quality)
  - Calibration (i.e. formal use of external evidence)
- In PSA we are putting probability distributions on unknown parameters, which makes it unequivocally a Bayesian analysis
- The joint distribution of parameters is not necessarily (even approximately) multivariate normal

Conclusions

- Model choice is inadequately justified
- Model choice tends to be limited to a few standard distributions
- Structural and parameter uncertainty is dealt with inadequately
Introduction

- Total costs and QALYs are often calculated from time to event data such as overall survival and progression-free survival for economic evaluations.
- Ideally, parameters are estimated using Classical or Bayesian methods applied to individual patient data. However, such data are often not available.
  - Who has regular access to IPD?
- Instead, curves are commonly fit to summary Kaplan-Meier graphs (who has done this?) by:
  - Regression; or
  - least squares;
  - any others?
Introduction

- Two recent more accurate methods:
  - Hoyle & Henley (2011) (BMC Medical Research Methodology)
  - Guyot et al. (2012) (BMC Medical Research Methodology)

- Anyone aware of, or used methods?

Step A: Estimate Underlying Individual Patient Data

Hoyle & Henley (2011)
Step A: Estimate Underlying Individual Patient Data

One example simulated trial
(500 patients with decreasing underlying hazard)
(Hoyle & Henley 2011)

Step B: Fit Curve

- Next, fit a curve to estimated underlying individual patient data:
  - Interval censor times of events

Hoyle & Henley (2011)
Step B: Fit Curve

- Simulation results:
  - Method gives more accurate curve fits than traditional methods (regression and least squares)
  - Fits often very similar to those derived directly from the underlying individual patient data
  (Hoyle & Henley 2011)

Example Application

ICER £62,000 per QALY in regression method vs. £43,000 per QALY using Hoyle & Henley (2011)
Guyot et al. (2012)

- Similarities with Hoyle & Henley (2011)
  - Both use;
    - Kaplan-Meier,
    - Numbers at risk at each time interval
  - Both assume;
    - Censoring occurs at constant rate in each time interval.

- Differences with Hoyle & Henley (2011)
  - Guyot et al estimates precise times of deaths, Hoyle & Henley interval-censored times.
  - Only Guyot et al uses total number of events, if available.
  - Guyot et al uses iterative algorithm compared to closed-form analysis of Hoyle & Henley.
  - Guyot et al all done in R, Hoyle & Henley mostly in Excel, some in R.
  - Only Hoyle & Henley simulated accuracy of method, and compared accuracy with basic methods.

Conclusions

- Methods are recommended for cost-effectiveness analysis when only summary survival data is available.
- Accurately estimate underlying IPD
- Accurately estimate uncertainty in curve fits, not available using traditional methods.
- Cost-effectiveness can be substantially changed using the methods.
- Papers & easy-to-use Excel spreadsheets & R code to implement the methods available;
Using External Data to Validate Survival Estimates

Example in hepatocellular carcinoma
Noémi Muszbek
United BioSource Corporation, UK
noemi.muszbek@unitedbiosource.com

Credible Distribution?

- Proportionality
- Monotonicity
- Fits observational data well graphically
- Statistical tests (AIC, BIC)
- Diagnostic plots
- Plausibility with what is seen in clinical practice
- Biological/clinical explanation

Source: Latimer (2011); Connock (2011)
NOTE: Not in order of importance or timing
Background

- Unresectable hepatocellular carcinoma*
  - Almost two co-existent diseases (carcinoma and the underlying liver disease)
  - Patients unsuitable for curative therapy
  - Major risk factors: hepatitis C and B, cirrhosis
- Two RCTs comparing sorafenib and placebo**
  - SHARP and Asia Pacific trials
  - At the end of follow-up (72-74 weeks) 0-34% of patients alive depending on trial and treatment arm
- Analysis: Lognormal distribution fitted best
- However: Long tail of lognormal distribution often criticised for offering survival benefit for years after the clinical trials

* Llovet et al. (2003); ** Llovet et al. (2008), Cheng et al. (2009)

Credible Distribution?

- Behaviour of the hazard
  - No proportionality
  - Not monotonic
- Fit
  - Smallest AIC/BIC
  - Graphically good fit except at very end, where only 5 events
  - Diagnostic plots show best fit
- Plausibility
  - Consistent with what was seen in clinical practice
- Biologic / clinical explanation
  - Different growth rate of the tumour remnants?
  - Genetic variability regulating response and toxicity?
  - Absence of major risk factors?
  - Different survival for the underlying liver disease?
Credible Distribution?

- Proportionality
- Monotonicity
- Fits observational data well graphically
- Statistical tests (AIC, BIC)
- Diagnostic plots
- Plausibility with what is seen in clinical practice
- Biological/clinical explanation
- External data

Source: Latimer (2011); Connock (2011)

External Data

- Literature in oncology regarding distributions for overall survival
  - Stream of research initiated by Boag explores the use of lognormal distribution* (‘cure model’)

- Published Kaplan-Meier curves in unresectable hepatocellular carcinoma
  - 99 longer term (>72 weeks) Kaplan-Meier curves in various patient populations

- Registry data from the New South Wales Cancer Registry
  - Long-term (32 years) follow-up of 3,280 patients with hepatocellular carcinoma (local, regional, distant HCC and with unknown localisation)

### Results from the Literature Review

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Best fit according to AIC</th>
<th>Best fit according to BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Kaplan-Meier Curves</td>
<td>Percentage of Kaplan-Meier Curves</td>
</tr>
<tr>
<td>Lognormal</td>
<td>62</td>
<td>62.6%</td>
</tr>
<tr>
<td>Weibull</td>
<td>6</td>
<td>6.1%</td>
</tr>
<tr>
<td>Loglogistic</td>
<td>22</td>
<td>22.2%</td>
</tr>
<tr>
<td>Gompertz</td>
<td>8</td>
<td>8.1%</td>
</tr>
<tr>
<td>Other*</td>
<td>1</td>
<td>1.0%</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

*Curve could not be read in for analysis

### Statistical Criteria from the NSW Cancer Registry

<table>
<thead>
<tr>
<th>All Patient Population with Hepatocellular Carcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>Exponential</td>
</tr>
<tr>
<td>Weibull</td>
</tr>
<tr>
<td>Gompertz</td>
</tr>
<tr>
<td><strong>Lognormal</strong></td>
</tr>
<tr>
<td>Loglogistic</td>
</tr>
</tbody>
</table>

Lognormal and loglogistic distributions provided the best fit also for patients with local, regional and distant HCC or for those with unknown localization of the disease.
Diagnostic Plots from the NSW Cancer Registry

Issues to Address

- Functional form of the curve due to disease or treatment?
  - How closely does the patient population have to match trial population?
  - What would influence the functional form?
    - Initial or subsequent treatment options?
    - Where patients start on the curve?
    - Patient/disease characteristics?

- What if external data shows different functional form?
- Lack of understanding of the disease to explain functional forms
- Potential future steps: use external data for calibration
Thank You For Your Attention