Survival Data Extrapolation for Cost-effectiveness Modeling

Benjamin Li
June 11th, 2021
Introduction

• Working as a statistician in the Real-World Evidence (RWE) statistics group.
• RWE statistics is part of Quantitative Data Science under the Statistical Research and Data Science Center in Pfizer.
• RWE statistics supports the following areas through the effective deployment of real-world data.
  • Evidence generation
  • Label enhancement
  • Lifecycle management
  • Regulatory filings
  • Publications
  • Enhancement of clinical trials
  • Medical impact of launched products
  • Health economics and value impact
  • Pricing and reimbursement
  • Business development
1. Introduction to cost-effectiveness modeling
2. Estimation of survival benefits due to a new treatment of a disease
3. Example: extrapolation of overall survival data in Tafamidis trial B3461028
4. Introducing finite mixture models as a flexible alternative to standard modeling techniques
5. Questions
Introduction to Cost-effectiveness Modeling
Background Health-Economics / Cost-Effectiveness
Incremental Cost-Effectiveness Ratio (ICER):

Ratio of difference in cost to difference in effectiveness

\[
\text{ICER} = \frac{\text{Costs (new treatment)} - \text{Costs (BSC)}}{\text{Effects (new treatment)} - \text{Effects (BSC)}} = \frac{\text{€/QALY}}{\text{£30,000/QALY - €80,000/QALY}}
\]
Background Health-Economics / Cost-Effectiveness

Quality Adjusted Life Years

Breakthroughs that change patients’ lives
Incremental Cost-Effectiveness Ratio (ICER):

Ratio of difference in cost to difference in effectiveness

\[
\text{ICER} = \frac{\text{Costs (new treatment)} - \text{Costs (BSC)}}{\text{Effects (new treatment)} - \text{Effects (BSC)}} \quad \text{€/QALY}
\]

CE Threshold / willingness to pay
£30,000/QALY - €80,000/QALY
Estimation of Survival Benefits Due to a New Treatment
Extrapolation of Survival Data from Randomized Controlled Trials

- Randomized controlled trials (RCTs) are the main source of survival data associated with a new treatment.

- However, estimates of survival benefits are restricted to that observed directly in the RCTs.

- Extrapolation of survival data collected from RCTs enables estimates of survival benefits for the duration of the economic model.
NICE Technical Support Document on Extrapolation with Patient-level Data

- A review of the Technology Appraisals submitted to the UK National Institute for Health and Clinical Excellence (NICE) demonstrated that a wide range of extrapolation methods had been used, and that a systematic approach was not taken.

- The major issue is sub-optimal justification to the chosen methods.

- In June 2011, the Decision Support Unit (DSU) within NICE issued Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. The report was last updated in March 2013.

- In this report, the DSC recommends a model selection process algorithm.
Survival modeling required for economic evaluation

Patient-level data available

Compare log-cumulative hazard plots, quantile-quantile plots or suitable residual plots to allow initial selection of appropriate models

Plots are not straight lines
- Consider piecewise or other more flexible models

Plots are not parallel
- Fit individual models

Plots are parallel
- Consider PH/AFT 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
# Parametric Models in Survival Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Probability Density Function</th>
<th>Survival Function</th>
<th>Hazard Function</th>
<th>PH or AFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>$\lambda, \gamma$</td>
<td>$f(t) = \lambda t^{\gamma-1}e^{-\lambda t^\gamma}$</td>
<td>$S(t) = e^{-\lambda t^\gamma}$</td>
<td>$h(t) = \lambda t^{\gamma-1}$</td>
<td>AFT/PH</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\lambda$</td>
<td>$f(t) = \lambda e^{-\lambda t}$</td>
<td>$S(t) = e^{-\lambda t}$</td>
<td>$h(t) = \lambda$</td>
<td>AFT/PH</td>
</tr>
<tr>
<td>Lognormal</td>
<td>$\mu, \sigma$</td>
<td>$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(log t - \mu)^2}{2\sigma^2}}$</td>
<td>$S(t) = 1 - \Phi\left(\frac{log t - \mu}{\sigma}\right)$</td>
<td>$h(t) = \frac{\phi\left(\frac{log t - \mu}{\sigma}\right)}{\sigma t[1 - \Phi\left(\frac{log t - \mu}{\sigma}\right)]}$</td>
<td>AFT</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\alpha, \beta$</td>
<td>$f(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} t^{\alpha-1}e^{-\frac{t}{\beta}}$, where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$</td>
<td>No closed form</td>
<td>No closed form</td>
<td>AFT</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>$\alpha, \beta$</td>
<td>$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \left(1 + \left(\frac{t}{\beta}\right)^\alpha\right)^{-2}$</td>
<td>$S(t) = \frac{1}{1 + \left(\frac{t}{\beta}\right)^\alpha}$</td>
<td>$h(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \left(1 + \left(\frac{t}{\beta}\right)^\alpha\right)^{-2}$</td>
<td>AFT</td>
</tr>
<tr>
<td>Gompertz</td>
<td>$\alpha, \beta$</td>
<td>$f(t) = \beta e^{\frac{\alpha t - \beta}{\alpha(e^{\alpha t} - 1)}}$</td>
<td>$S(t) = e^{-\frac{\beta}{\alpha(e^{\alpha t} - 1)}}$</td>
<td>$h(t) = \beta e^{\alpha t}$</td>
<td>AFT/PH</td>
</tr>
</tbody>
</table>
## Parametric Models in Survival Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Probability Density Function</th>
<th>Survival Function</th>
<th>Hazard Function</th>
<th>PH or AFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Gamma</td>
<td>$\mu, \sigma, Q$</td>
<td>$f(t</td>
<td>\mu, \sigma, Q) = \frac{</td>
<td>Q</td>
<td>(Q^{-2})^{Q^{-2}}}{\sigma t \Gamma(Q^{-2})} e^{(Q^{-2}(Qw-e^{Qw}))}$</td>
</tr>
<tr>
<td>Generalized F</td>
<td>$\sigma, \mu, Q, P$</td>
<td>$f(t) = \frac{\delta^{s_1} e^{s_1w}}{\sigma t \left(1 + s_1 e^{w Q^{-2}} / s_2 \right) \left(s_1 + s_2 \right)} B(s_1, s_2)$ and $B(s_1, s_2) = \frac{\Gamma(s_1) \Gamma(s_2)}{\Gamma(s_1 + s_2)}$ is the beta function.</td>
<td>No closed form</td>
<td>No closed form</td>
<td>AFT</td>
</tr>
</tbody>
</table>
# Parametric Models in Survival Analysis

## Model Assumptions

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard over Time</th>
<th>Log Cumulative Hazard over Log Time</th>
<th>Log Hazard over Time</th>
<th>Logit Survival Probability over Log Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Constant</td>
<td></td>
<td>Linear</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>Monotonic</td>
<td></td>
<td>Linear</td>
<td></td>
</tr>
<tr>
<td>Gompertz</td>
<td>Monotonic</td>
<td></td>
<td></td>
<td>Linear</td>
</tr>
<tr>
<td>Log-logistic</td>
<td></td>
<td></td>
<td></td>
<td>Linear</td>
</tr>
<tr>
<td>Lognormal</td>
<td>Increases from 0 to maximum, then decreases monotonically</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extrapolation of Overall Survival Data in Tafamidis Trial B3461028
Tafamidis Trial B3461028

- A multicenter, international, Phase 3, double-blind, placebo-controlled, randomized study to evaluate the efficacy, safety, and tolerability of daily oral dosing of Tafamidis Meglumine (PF-06291826) 20 mg or 80 mg in comparison to placebo in subjects diagnosed with transthyretin cardiomyopathy (TTR-CM).
- Double blind treatment phase: 30 months
- Primary analysis: all-cause mortality and frequency of cardiovascular-related hospitalizations using the Finkelstein-Schoenfeld method.
- FSFV: December 2013; LSFV: February 2018
- VYNDAQEL® (tafamidis) has been approved by FDA (2019) and EMA (2020).
- Open-label extension study ongoing, with data cuts in 02/2018 and 08/2018
Data Extrapolation

Log Cumulative Hazard Curves

- Placebo
- Pooled Active
- 20 MG
- 80 MG

Log of Time from First Dose (months)
Data Extrapolation

POOLED TREATMENT

Epanechnikov Kernel-Smoothed Hazard Function

Log of Negative Log of Estimated Survivor Function

Bandwidth: 42.29787

Log(Hazard) over Time

Log(Survival Probability) over log(Time)
Data Extrapolation

Fitting of Common Parametric Models vs KM Curves

Placebo

Survival Probability vs Time from First Dose (months)

KM
Weibull
Exponential
Lognormal
Gamma

Logistic
Generalized Gamma
Gompertz
Generalized F
### AIC and BIC Tests - Placebo

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>860.00</td>
<td>863.18</td>
</tr>
<tr>
<td>Weibull</td>
<td>842.93</td>
<td>849.28</td>
</tr>
<tr>
<td>Gamma</td>
<td>844.23</td>
<td>850.58</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>844.77</td>
<td>851.12</td>
</tr>
<tr>
<td>Lognormal</td>
<td>854.89</td>
<td>861.24</td>
</tr>
<tr>
<td>Gompertz</td>
<td>844.28</td>
<td>850.64</td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>844.62</td>
<td>854.15</td>
</tr>
<tr>
<td>Generalized F</td>
<td>846.43</td>
<td>859.14</td>
</tr>
</tbody>
</table>
Data Extrapolation

Fitting of Common Parametric Models vs KM Curves
Pooled Active Treatment

Survival Probability

Time from First Dose (months)

KM
Weibull
Exponential
Lognormal
Gamma

Logistic
Generalized Gamma
Gompertz
Generalized F

Breakthroughs that change patients’ lives

22
## Data Extrapolation

### AIC and BIC Tests – Pooled Active Treatment

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>885.48</td>
<td>889.05</td>
</tr>
<tr>
<td>Weibull</td>
<td>884.37</td>
<td>891.53</td>
</tr>
<tr>
<td>Gamma</td>
<td>884.06</td>
<td>891.21</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>883.60</td>
<td>890.75</td>
</tr>
<tr>
<td>Lognormal</td>
<td>883.50</td>
<td>890.65</td>
</tr>
<tr>
<td>Gompertz</td>
<td>886.27</td>
<td>893.42</td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>885.13</td>
<td>895.85</td>
</tr>
<tr>
<td>Generalized F</td>
<td>887.13</td>
<td>901.43</td>
</tr>
</tbody>
</table>
# Data Extrapolation

## Top Selected Models in each Treatment Arm

<table>
<thead>
<tr>
<th>Model</th>
<th>PLACEBO</th>
<th>POOLED ACTIVE TREATMENT</th>
<th>20 MG</th>
<th>80MG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Log-logistic</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gompertz</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Extrapolation

Gamma Model Fitted to Patient-Level Data for All-cause Mortality

- Pooled Tafamidis: Median survival months: 52.64
- Placebo: Median survival months: 35.16

HR, 0.64 (95% CI, 0.47–0.85); P=0.001
Impact

- Using the updated extrapolations (Feb 2020 cut-off point) the cost-effectiveness improved significantly, with 5%-42%, depending on the assumed distribution and on the type of cost-effectiveness model used. This means that it will be less difficult to show the cost-effectiveness of tafamidis and patients can potentially benefit from faster access to tafamidis.

- Manuscript published in *Cardiology and Therapy*. 
Observations

- Selection of the proper distribution for cost-effectiveness modeling should be based on statistical tests, clinical validity, external data, and biological reasoning.

- In cases when single parametric distribution fail to provide a good fit to the data, finite mixture models should be considered as a viable option.
Finite Mixture Models

- A random variable $Y$ follows a finite mixture distribution if the density $f(y)$ of $Y$ can be written in the form

$$f(y) = \sum_{i=1}^{g} \pi_i f_i(y),$$

where the $f_i(y)$ are densities and the $\pi_i$ satisfy such conditions that

$$0 \leq \pi_i \leq 1 \ (i = 1, \ldots, g)$$

and

$$\sum_{i=1}^{g} \pi_i = 1.$$
Finite Mixture Models – Parameter Estimation

• Suppose that \( y = (y_1, \ldots, y_n)^T \) is a random sample from a population with a \( g \)-component mixture density, that is, for \( j = 1, \ldots, n \),

\[
f(y_j|\Psi) = \sum_{i=1}^{g} \pi_if_i(y_j|\theta_i),
\]

where \( \Psi = (\pi_1, \ldots, \pi_{g-1}, \theta_1, \ldots, \theta_g)^T \) is the vector containing all the unknown parameters. The \( \pi_i \) are the weights, and the \( \theta_i \) contain the respective parameters of each component density.

• The log likelihood function of the random sample \( y \) is given by

\[
l(\Psi|y) = \sum_{j=1}^{n} \log[f(y_j|\Psi)] = \sum_{j=1}^{n} \log \left[ \sum_{i=1}^{g} \pi_if_i(y_j|\theta_i) \right].
\]
Finite Mixture Models – Parameter Estimation

• The maximum likelihood estimation involves solving \( l(\Psi \mid y) = 0 \) for \( \Psi = (\pi_1, \ldots, \pi_{g-1}, \theta_1, \ldots, \theta_g)^T \).

• To solve \( l(\Psi \mid y) = 0 \) directly is often difficult.

• It can be manipulated so that \( \hat{\Psi} \) satisfies

\[
\hat{\pi}_i = \frac{\sum_{j=1}^{n} \tau_{ij}(y_j \mid \hat{\Psi})}{n} \quad (i = 1, \ldots, g),
\]

and

\[
\sum_{j=1}^{n} \tau_{ij}(y_j \mid \hat{\Psi}) \frac{\partial \log[f_i(y_j \mid \hat{\theta}_i)]}{\partial \theta_i} = 0 \quad (i = 1, \ldots, g),
\]

where

\[
\tau_{ij}(y_j \mid \Psi) = \frac{\pi_if_i(y_j \mid \theta_i)}{\sum_{h=1}^{g} \pi_h f_h(y_j \mid \theta_h)} \quad (i = 1, \ldots, g; j = 1, \ldots, n)
\]
The Expectation and Maximization (EM) Algorithm

• The EM algorithm of Dempser et al. (1977) provides a relatively easier solution to find the MLEs of the parameters in the finite mixture model.

• The EM algorithm is a procedure of iterative computation to calculate the MLEs in cases where the observed data are deemed incomplete.

• In order to apply the EM algorithm, we could turn the observed survival data $y$ into an incomplete data problem.

• Let the random vector $Z = (Z_1, ..., Z_g)^T$ follow a multinomial distribution with $n = 1$ and probability $\pi = (\pi_1, ..., \pi_g)$. That is, $P(Z_i = 1) = \pi_i, (i = 1, ... g)$.

• $Z$ can be viewed as the component label of the mixture distribution. If a random variable $Y$ follows a $g$-component finite mixture distribution, when $Z_i = 1$, the density of $Y$ comes from the $i$th component $f_i(Y)$.

• We don’t observe $z$, the realized value of the random vector $Z$. Therefore $z$ can be viewed as missing data.

• Following the above notation within the EM framework, we observe the incomplete data vector $y = (y_1, ... y_n)^T$, not the complete data vector $x = (y_1, ... y_n, z_1, ..., z_n)^T$. 

Breakthroughs that change patients’ lives
The Expectation and Maximization (EM) Algorithm

- The complete-data log likelihood function is
  \[
  l_c(\Psi|x) = \sum_{j=1}^{n} \sum_{i=1}^{g} z_{ij} \{ \log(\pi_i) + \log[f_i(y_j|\theta_i)] \},
  \]
  where \( \Psi = (\pi_1, ..., \pi_{g-1}, \theta_1, ..., \theta_g)^T \) is the vector containing all the unknown parameters.

- On the \((k + 1)^{th}\) iteration, the E-step requires the computation of the conditional expectation of \(Z_{ij}\) given \(y\) and \(\Psi^{(k)}\), where \(\Psi^{(k)}\) is from the \(k^{th}\) iteration.
  \[
  E_{\Psi^{(k)}}(Z_{ij}|y) = P_{\Psi^{(k)}}(Z_{ij} = 1|y) = \tau_{ij}(y_j|\Psi^{(k)}),
  \]
  where \(\tau_{ij}(y_j|\Psi^{(k)}) = \frac{\pi_{ij}^{(k)} f_i(y_j|\theta_i^{(k)})}{\sum_{h=1}^{g} \pi_{ij}^{(k)} f_h(y_j|\theta_h^{(k)})} \) for \(i = 1, ..., g; j = 1, ..., n\).
The Expectation and Maximization (EM) Algorithm

• In the M-step, we maximize

\[ Q(\Psi; \Psi^{(k)}) = \sum_{j=1}^{n} \sum_{i=1}^{g} \tau_{ij}(y_j | \Psi^{(k)}) \{ \log(\pi_i) + \log[f_i(y_j | \theta_i)] \} \]

with respect to \( \Psi \) and get

\[ \pi_i^{(k+1)} = \frac{\sum_{j=1}^{n} \tau_{ij}(y_j | \Psi^{(k)})}{n} \quad (i = 1, \ldots, g) \]

and

\[ \sum_{j=1}^{n} \sum_{i=1}^{g} \tau_{ij}(y_j | \Psi^{(k)}) \left\{ \frac{\partial}{\partial \theta_i} \log[f_i(y_j | \theta_i)] \right\} = 0 \]

• The iteration continues until \( l(\Psi^{(k+1)} | y) - l(\Psi^{(k)} | y) \) is less than an arbitrarily small amount.

• \( l(\Psi | y) \) is a non-decreasing function and the sequence will converge to a stable point.
Enhancing the EM Algorithm with Censored Quantile Regression

• The EM algorithm will converge to a stable point, but not necessarily to the global maximum.

• Need to find a good initial value to avoid convergence to local maxima.

• Current methods are more suitable for uncensored data that are normally distributed.

• We suggest to do a rough exhaustive search using censored quantile regression (QR) that will ideally produce a solution that is in the general area of the global maximum.
Hodi et al. (2010) published results from a double-blind Phase III clinical trial that investigated the efficacy and safety of ipilimumab in patients with previously treated metastatic melanoma.

Three treatment arms:
1. ipilimumab plus a glycoprotein 100 peptide vaccine (IPI+GP100, 403 patients)
2. ipilimumab alone (IPI, 137)
3. glycoprotein 100 alone (GP100, 136)
Applying the Mixture Models

- Digitized data from Hodi et al. (2010) are used to demonstrate the model selection process.
## Applying the Mixture Models

- **Model fitting**

<table>
<thead>
<tr>
<th>Models</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>Restricted Mean (95% CI) T=51</th>
<th>Extrapolated Mean (95% CI) Lifetime</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM</td>
<td></td>
<td></td>
<td>6.47</td>
<td></td>
</tr>
<tr>
<td>Mixture of 3 Weibull</td>
<td>-615.95</td>
<td>1247.90</td>
<td>6.51 (5.55 – 7.61)</td>
<td>6.84 (5.7 – 8.16)</td>
</tr>
<tr>
<td>Single Weibull</td>
<td>-988.70</td>
<td>1981.40</td>
<td>5.86 (5.36 – 6.37)</td>
<td>5.86 (5.37 – 6.37)</td>
</tr>
<tr>
<td>Gamma</td>
<td>-955.92</td>
<td>1915.84</td>
<td>5.67 (5.25 – 6.09)</td>
<td>5.67 (5.30 – 6.10)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>-861.76</td>
<td>1727.51</td>
<td>5.22 (4.89 – 5.57)</td>
<td>5.22 (4.88 – 5.60)</td>
</tr>
<tr>
<td>Loglogistic</td>
<td>-816.97</td>
<td>1637.93</td>
<td>4.41 (4.18 – 4.69)</td>
<td>4.42 (4.18 – 4.66)</td>
</tr>
<tr>
<td>Generalized F</td>
<td>-707.59</td>
<td>1423.18</td>
<td>5.25 (4.64 – 5.72)</td>
<td>5.39 (4.68 – 6.01)</td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>-736.31</td>
<td>1478.63</td>
<td>5.35 (4.91 – 5.91)</td>
<td>5.52 (4.97 – 6.28)</td>
</tr>
<tr>
<td>Gompertz</td>
<td>-996.38</td>
<td>1996.75</td>
<td>6.12 (5.70 – 7.00)</td>
<td></td>
</tr>
</tbody>
</table>

- **Manuscript accepted by *Value in Health*.**
Applying the Mixture Models

Progression-free Survival – IPI+GP100

- KM
- Mixture of 3 Weibull
- Single Weibull
- Gamma
- Lognormal
- Loglogistic

Breakthroughs that change patients’ lives
Questions?