# Parametric Model Selection: Beyond AIC/BIC in Health Economic Contexts

MSR53

Beatrice Suero, MSc[1]; Roya Gavanji, MSc[1]; Aidan Dineen[1] • [1]EVERSANA, Burlington, ON, Canada

## Background and Objectives

- Health Technology Assessment (HTA) submissions rely on time-to-event analyses (e.g., overall survival, progression-free survival) to inform cost-effectiveness.[1]
- Trial data are often immature, with many events yet to occur, requiring extrapolation beyond observed follow-up.
- Parametric survival models are the standard for estimating outcomes over a patient's lifetime.
- Model selection typically involves assessing statistical criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).
- However, AIC/BIC emphasize in-sample fit and may not adequately reflect extrapolation performance—especially in the tail, where data are sparse.[2]
- Visual fit to Kaplan-Meier curves, residual diagnostics, and assumption testing (e.g., proportional hazards [PH], accelerated failure time) can provide additional insight.
- Clinical expert input supports plausibility of long-term projections and helps justify model choice.[3]
- The objective of this study was to evaluate and illustrate a comprehensive approach for assessing the suitability of parametric survival models in HTA submissions, highlighting the limitations of relying solely on AIC/BIC and the value of incorporating visual inspection, diagnostic testing, and clinical plausibility—especially when trial data are immature and long-term extrapolation is required.

## Methods

- We conducted a simulation study to evaluate the suitability of various parametric survival models. The steps are summarized in **Figure 1**.

**Figure 1: Simulation Study Methods**

**1 Simulate Survival Data**
- Survival data were simulated from a Weibull distribution with a shape parameter of 0.8 and a scale parameter that varied by treatment group to reflect a modest treatment effect.
- Censoring was introduced using exponential censoring times to reflect approximately 40% censoring in the sample.
- For each simulation, we generated a sample of 200 patients and applied right censoring to produce observed survival times.

**2 Fit Parametric Models**
- Standard parametric survival models—Exponential, Weibull, Log-normal, Log-logistic, Gompertz, Gamma, and Generalized Gamma—were fit to each simulated dataset.*

**3 Model Performance**
- Model performance was evaluated based on the Akaike Information Criterion (AIC), and the best-fitting model for each iteration was identified as the one with the lowest AIC.

**4 Iterations**
- This process was repeated over 500 simulation iterations.

**5 Summarize Results**
- We summarized how frequently each model was selected as the best fit to assess how often AIC correctly identifies the true underlying distribution (Weibull) and how often alternative models were incorrectly favored—especially when extrapolation to the tails is required.

* All models included treatment as a covariate, although no formal hypothesis testing was conducted on treatment effects.

### Model Fit Diagnostics

- Visual goodness of fit was assessed by overlaying extrapolated parametric survival curves onto the Kaplan-Meier (KM) estimate to assess both in-sample and extrapolated performance.
- The PH assumption was tested statistically and visually using Schoenfeld residuals plot. Deviations from horizontal residual trends indicated potential PH violation. Models parametrized as PH were Exponential, Weibull, and Gompertz.
- The accelerated failure time (AFT) assumption was tested using quantile-quantile (Q-Q) plots survival times. If the quantile pairs approximately lie on a straight line, this suggests that the AFT assumption is met. Models parametrized as AFT were Log-normal, Log-logistic, Gamma, and Generalized Gamma.

## Results

**Table 1: Frequency of Best Fitting Model According to AIC**

| | Frequency | |
|---|---|---|
| **Model** | **N** | **%** |
| Weibull | 169 | 33.8% |
| Gamma | 146 | 29.2% |
| Log-Logistic | 69 | 13.8% |
| Exponential | 36 | 7.2% |
| Gompertz | 35 | 7.0% |
| Generalized Gamma | 23 | 4.6% |
| Log-Normal | 22 | 4.4% |

## Results

### Model Fit Statistics

- The frequency with which each parametric model achieved the lowest AIC across 500 iterations is summarized in **Table 1**.
- Notably, the Weibull model had the lowest AIC in only about one-third of simulations. This suggests that, in many cases, alternative models may be selected based on AIC alone, even if they are not the most appropriate fit.

### Additional Model Fit Diagnostics

- Visual inspection of the extrapolated curves presented in **Figure 2** suggests that the Weibull model provided the best fit within the observed data and yielded a plausible extrapolation for the tail.
- Model assumption diagnostics indicated that both the PH (Figure 3a) and AFT (Figure 3b) assumptions were reasonably satisfied. This is consistent with the properties of the Weibull distribution, which can meet both assumptions under appropriate conditions.

**Figure 2a: Visual Inspection of Extrapolations vs KM Curve – Treatment**

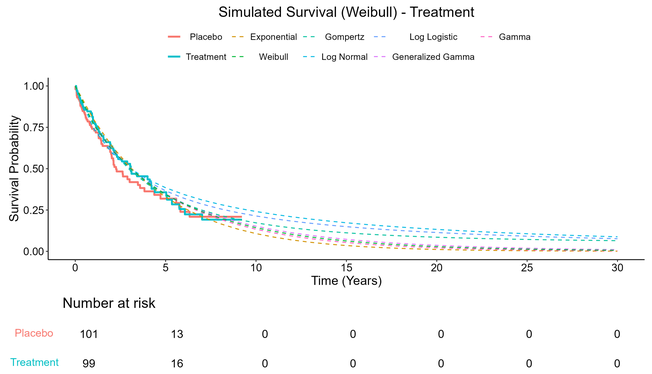



**Figure 2b: Visual Inspection of Extrapolations vs KM Curve – Placebo**

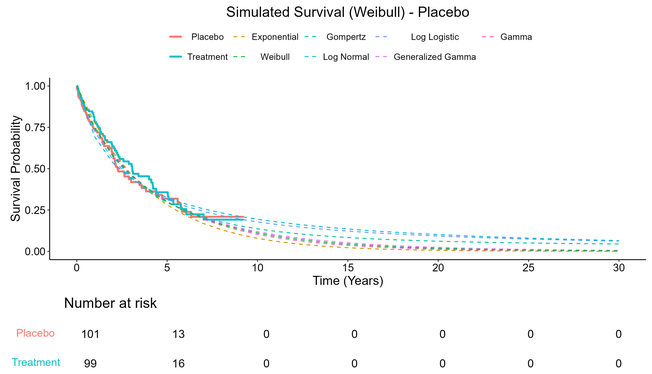



**Figure 3a: PH Test – Schoenfeld Residuals Plot**

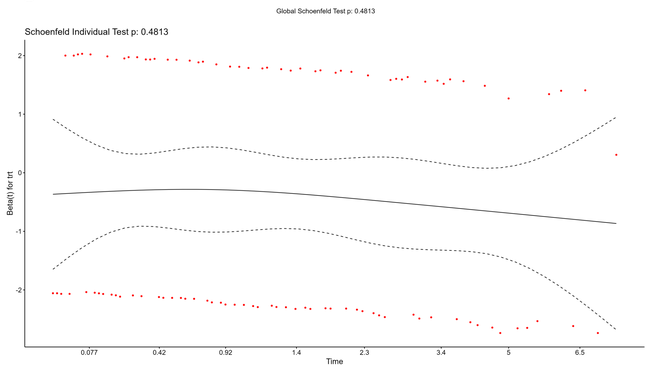
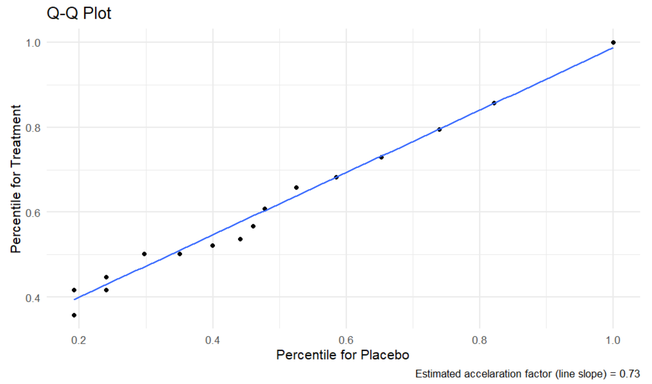



**Figure 3b: AFT Test – Q-Q Plot**




## Discussion

**When selecting the best parametric model, there are various criteria that should be considered in addition to AIC and BIC. The evidence herein highlights that by leveraging approaches including visual inspection, model performance, and model assumption tests, the most appropriate, reliable and interpretable model is chosen for the data. Supplemental feedback from clinical experts can improve the clinical plausibility of survival extrapolations based on the selected model.**

**Abbreviations** AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; HTA = Health Technology Assessment; KM = Kaplan-Meier; PH = proportional hazards.

**References**

**1.** Latimer, N. R. (2013). Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Medical Decision Making, 33(6), 743-754. **2.** Gallacher, D., Kimani, P., & Stallard, N. (2021). Extrapolating parametric survival models in health technology assessment: a simulation study. Medical Decision Making, 41(1), 37-50. **3.** Soares, M. O., Dias, S., & Ades, A. E. (2018). Expert elicitation for use in health technology assessment: protocol for a multiple stakeholder workshop (DSU Technical Support Document 26). Decision Support Unit, ScHARR, University of Sheffield. https://www.sheffield.ac.uk/nice-dsu/tsds/expert-elicitation-tsd

**Presented at the**
**2025 ISPOR Conference**
**Montreal, QC, Canada**
**May 13 to 16,2025**

EVERSANA®