The Impact of Small Sample Size on Selecting Appropriate Distribution for Parametric Survival Extrapolations Models: A Methodological Commentary and Simulation Study

Groff M[1], Tremblay G[1], Daniele P*[1]
1. Cytel, Inc., Waltham, MA, USA

Email: Patrick.daniele@pshta.com

Background and objective

Background: Survival extrapolations form the foundation of partition survival models for cost-effectiveness and cost-utility analysis models. Health economists are increasingly interested in early conceptual modelling of new drugs and technologies for planning a successful health technology assessment (**HTA**) or a submission based on priority. The evidence informing these submissions is often based on phase II clinical studies which are non-randomized, limited in sample, and shorter in duration. Selection of an optimal parametric survival model for long-term estimation of benefits and costs is informed endogenously from the clinical data and is typically based on statistical measures of model fit such as Akaike's information criterion (**AIC**) or Bayesian information criteria (**BIC**). While these measures of model fit may illustrate the best-fitting model based on the likelihood function, they do not consider the underlying disease. An improper choice of distribution used for extrapolation has consequences that will bias the results. The consequence of increased bias can lead to over- or underestimation of the benefits and costs. By understanding the sources of variability that inform health economists' extrapolation decisions, more accurate cost-effectiveness scenarios can be reached.

Objective: This simulation study explored the impact of sample size on parametric survival extrapolation model selection guided using statistical criteria and the potential effect of erroneous extrapolation on economic evaluations.

Methods

We conducted a simulation study wherein:

- Survival data modelling a Weibull distribution were simulated using the 'simsurv' package in R (Brilleman et al., 2021).
- Three permutations of the shape parameter, 0.5, 1.5, or 5.0, using a scale parameter equal to 1.0, were specified to create three separate datasets containing 1000 observations each.
- The simulated datasets were then randomly sampled using 10 sample sizes, starting at n=25 to n=250 by increments of 25, iteratively 1000 times each.
- Each iteration was extrapolated using each of the main seven functional forms requested by HTA organizations in economic evaluation, including exponential, Weibull, Gompertz, simple gamma, generalized gamma, log-normal, and log-logistic.
- The fit of the distributions were evaluated using AIC and BIC.
- The concordance of the actual underlying distribution from which the data were simulated, Weibull, and the AIC/BIC selected parametric distributions were evaluated at each sample size for each simulation dataset.
- Concordance was defined as the percentage agreement between the true Weibull and the simulations best-fitting distributions.
- The impact of improper extrapolation was compared descriptively using the restricted mean survival time (**RMST**).

Figure 1

Frequency of a distribution's selection using BIC. (**Top**) γ = *0.5, α = 1.0;* (**Middle**) γ = *1.5, α = 1.0;* (**Bottom**) γ = *5.0, α = 1.0*

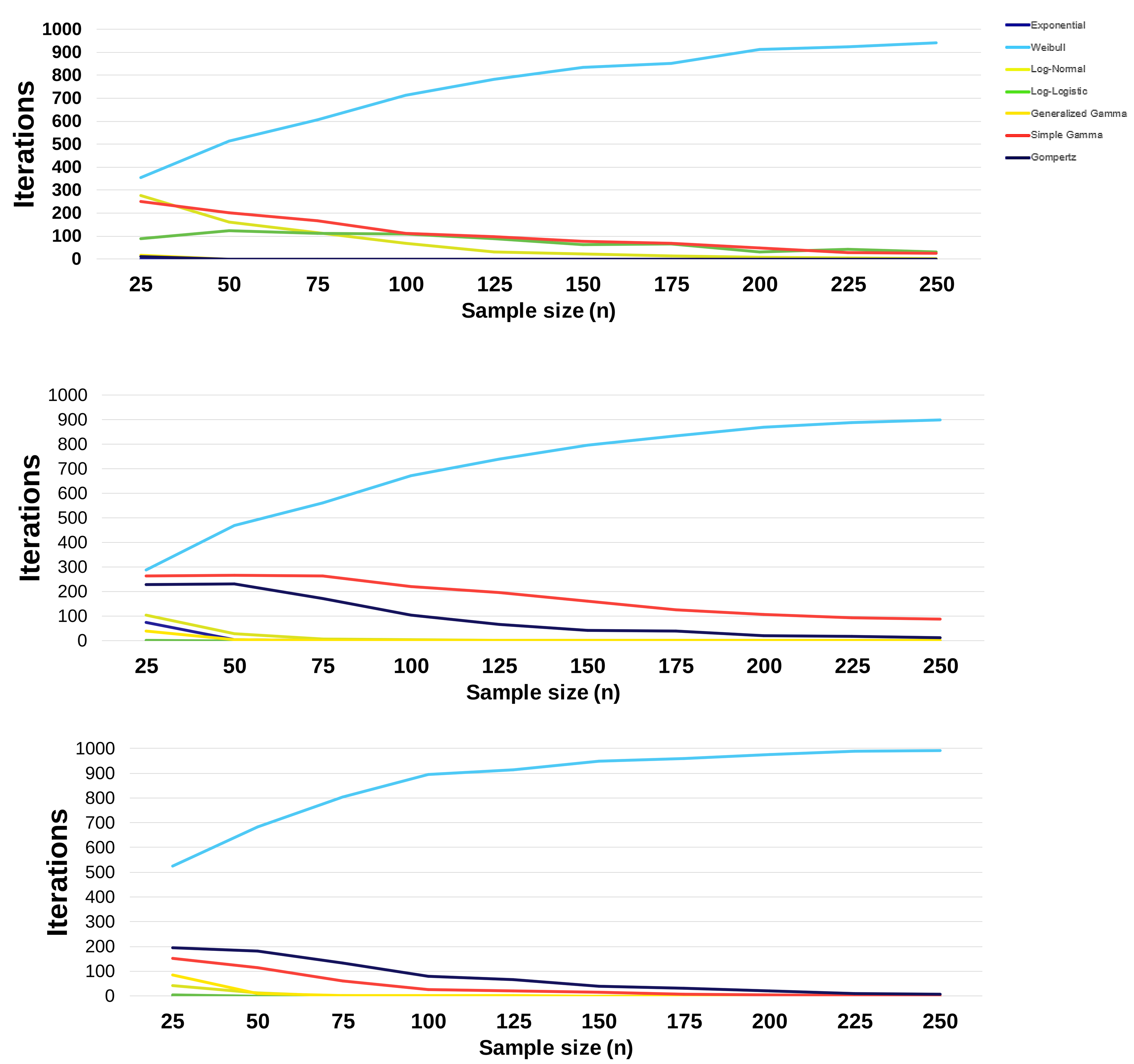
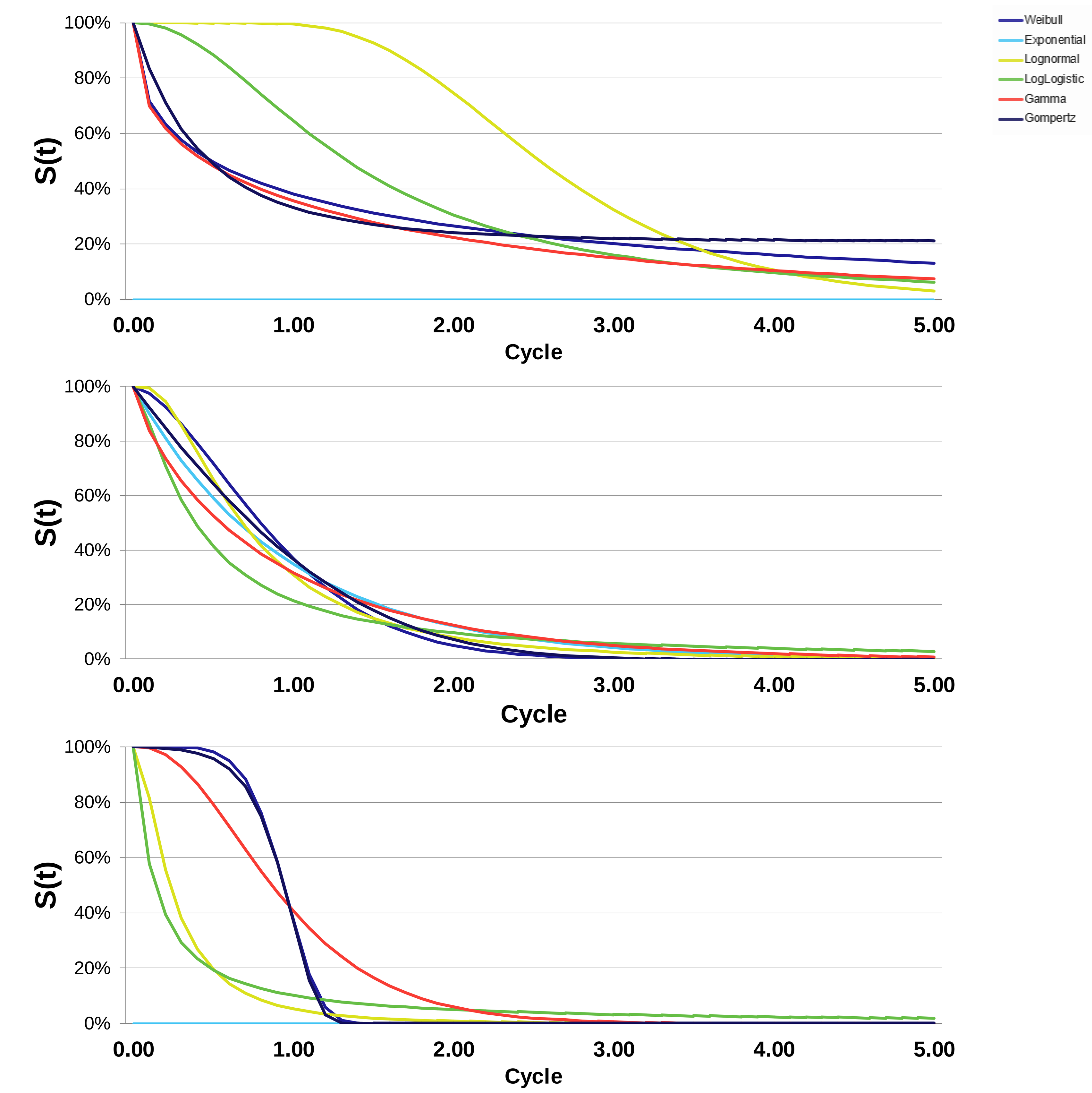


Figure 2

Extrapolated survival distributions (n=25). (**Top**) γ = *0.5, α = 1.0;* (**Middle**) γ = *1.5, α = 1.0;* (**Bottom**) γ = *5.0, α = 1.0*



Note: Generalized Gamma not shown.

Results

- The Weibull distribution was the most probable distribution selected during 1000 iterations using BIC as the selection criteria over the three simulated datasets and all sampling sizes (**Figure 1**).
- However, the probability of sampling resulting in a Weibull selection improved with the size of sampling.
- Initially, BIC indicated Weibull was the most appropriate distribution at 31.3%, 51.9%, and 35.6% for shape parameters of 0.5, 1.5, and 5.0, respectively, using an n=25.
- These probabilities increased linearly with each n=25 increment by an average of 6.15% ($p=0.01$), 6.38% ($p<0.01$), and 4.49% ($p=0.01$), for shape parameters of 0.5, 1.5, and 5.0, respectively.
- At a sampling size of n=250, Weibull was appropriately selected at 94.1%, 89.9%, and 99.1% for shape parameters of 0.5, 1.5, and 5.0, respectively.
- AIC demonstrated similar trends to BIC apart from increasing generalized gamma selection at sample sizes greater than n=125 ($p>0.05$).

The RMST was estimated for each of the selected distributions and shown graphically (**Figure 2**).

- The relative differences in the RMST were larger in the shape parameter of 0.5 or 5.0 (top right, bottom right), indicating greater differences to the RMST as the shape parameter deviated further from 1.0.
- These differences were greatest in the extrapolated log-normal and log-logistic distributions, which were selected more frequently when the shape parameter was less than 1.0 and the sample size was less than n=100.
- In these selections, the RMST was overestimated by as large as 151% and 6% in simulation samples of 0.5, and 5.0, respectively, and underestimated by as small as 9% and 64% in samples of 0.5, and 5.0, respectively.

Conclusion

- Time to event analyses are sensitive to sample size. The frequency of selecting the appropriate distribution to model survival outcomes when selected by AIC/BIC and the underlying distribution is Weibull, should be approached with caution.
- An incorrect extrapolation of the sampled data can lead to under- and over-estimation of the RMST, thus directly impacting the calculation of the benefits and costs in an economic evaluation.
- The likelihood of selecting the correct distribution improves significantly with each n=25 increment up to n=250 and as the shape parameter becomes more positive.
- The magnitude of the absolute differences observed between the true Weibull RMST and the statistically best fitting distributions, as ascertained in the simulation, were dependent on the shape parameter and not the sample size as expected.
- The comparative difference in the simulated RMST to the underlying sampled deviated further when the shape parameter was set to 0.5 or 5.0, instead of 1.0.
- The AIC and BIC selected differently with regards to the generalized gamma distribution; one possible explanation is due to the additional degree of freedom required to estimate a three parameter distribution.
- The generalized gamma failed to converge at small sample sizes but was selected by the AIC at higher sample sizes as the most appropriate extrapolation.
- In these cases, the difference in RMST was small (1%) and demonstrates the flexibility of the distribution with a larger sample size.

Limitations

- This study was limited to the Weibull distribution and therefore, these findings can not be extrapolated to other commonly used distributions.
- Future research should repeat currently used methodology to the other six distributions.
- The current research was conducted on simulated data and does not necessarily represent the heterogeneity found within real-world evidence.
- Nonetheless, findings demonstrate a key point that analysts should take care to select their extrapolation distributions based on prior research of the survival outcome under investigation and visual inspection aided by statistical criteria.

References

1. Brilleman, SL, Wolfe, R, Moreno-Betancur, M, & Crowther, MJ (2021). Simulating Survival Data Using the simsurv R Package. *Journal of Statistical Software*, *97*(1), 1-27.

Abbreviations: AIC, Akaike's information criteria; BIC, Bayesian information criteria; HTA, health technology assessment; RMST, restricted mean survival time