# CHARACTERISING ERRORS INCURRED BY MODEL SELECTION BASED ON AIC FOR EXTRAPOLATION OF SURVIVAL IN HEALTH TECHNOLOGY ASSESSMENT – A SIMULATION STUDY

Young R[1], Jacob I[1], McEwan P[1]

[1] Health Economics and Outcomes Research Ltd, Cardiff, UK

PCN59

## Introduction

- Analysts are frequently required to extrapolate time-to-event outcomes from clinical trials beyond the observed follow-up period to provide expected survival times for health technology assessment (HTA).
- Increasingly, decision-makers are asked to assess the performance and validity of multiple models of survival.
- Appropriate statistical methods for ranking model adequacy for models from differing statistical families do not exist, and current guidance (e.g. NICE DSU Technical Support Document 14 [1]) indicates the use of likelihood-based information criteria such as Akaike information criterion (AIC) to comment on model goodness of fit.
- These statistics are the most regularly reported evidence used in HTA[2] and are often used to give absolute ranking of competing model types, resulting in pressure to select models based on their relative criteria.
- A simulation study was performed to demonstrate the impact of allowing the AIC to unduly influence model selection on estimations of expected survival.

## Methods

- Six parametric distributions frequently used for modelling survival time distribution and suggested by TSD14 were chosen:
  - Exponential, Weibull, Gompertz, lognomal, log-logistic and generalised gamma
- For each two parameter distribution, a range of parameters (n=101) were generated, with the distribution constrained to have a median survival time of 1 arbitrary unit. For the exponential distribution, the rate parameter was fixed and the number of repetitions was increased 10 times versus the two-parameter models.
- A fixed grid of censoring times (n=20) varying from 0.1 of median to 2 times median was defined.
- For each resultant parameter set and each censoring time, 100 sets of random survival times were drawn for simulated single arm studies of N=100 and N=300 patients. These survival times were then censored at the specified time.
- The observed survival times of each simulated study were fitted using the maximum-likelihood fitting package *flexsurv* [3] in **R** to each of the six specified distributions.
- The fitted model using the true generating distribution and the model with the lowest AIC were tested for two errors:
  1. The 95% confidence interval (CI) of the mean of the distribution obtained by 1000 bootstrap samples of the asymptotic multivariate normal distribution of the fitted parameters did not contain the mean of the true distribution
  2. The mean survival associated with the maximum likelihood parameters of the distribution deviated by more than 10% from the mean of the true distribution

## Results

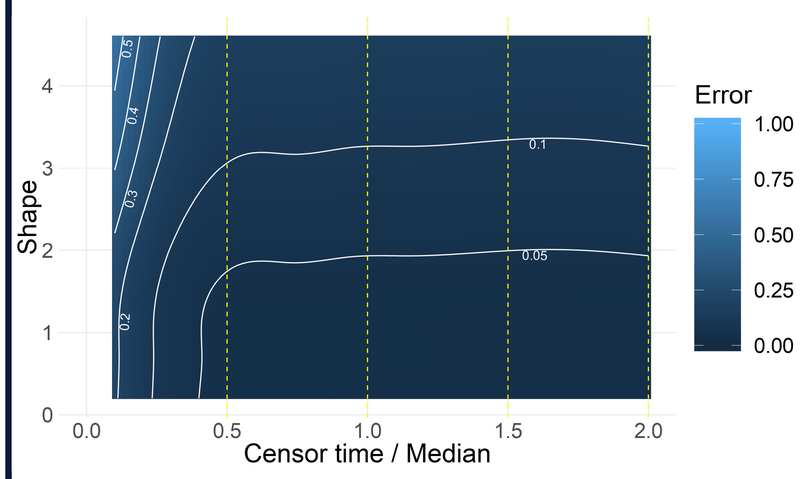


**Figure 1:** Reference "true mean out of 95% confidence interval" errors, Weibull distribution

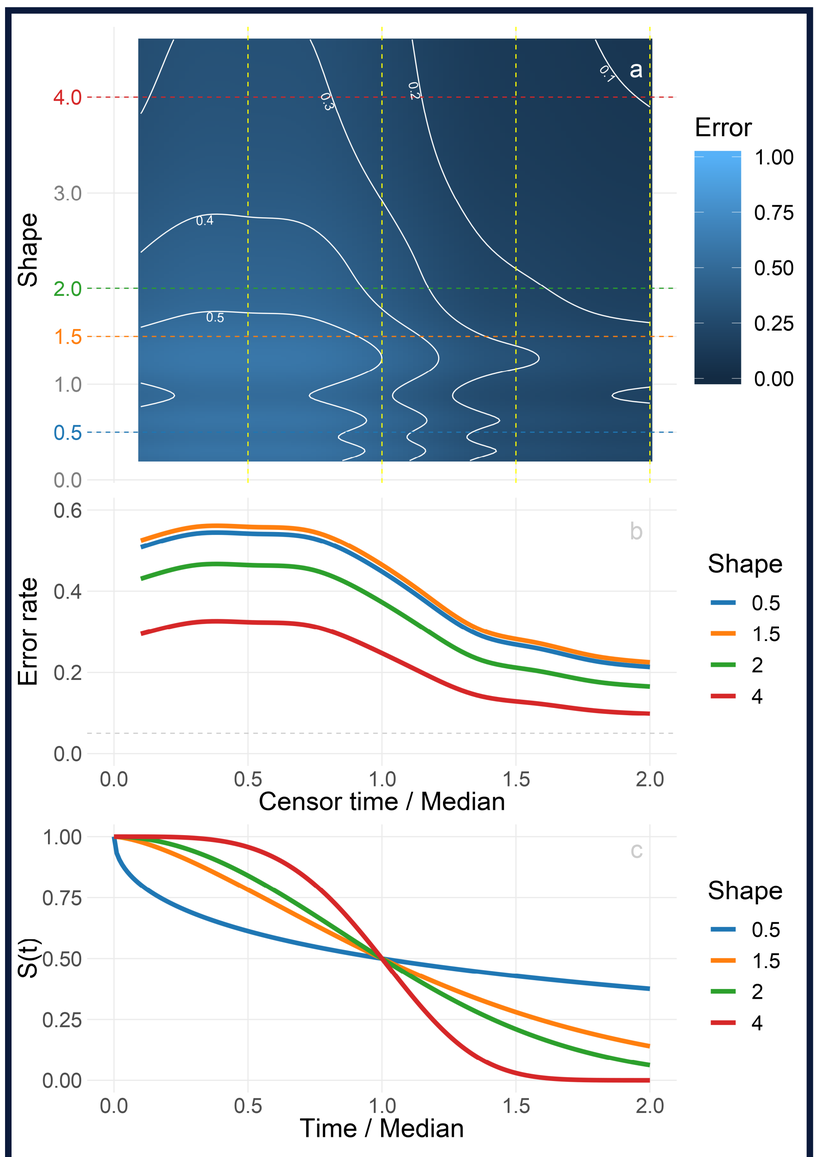


**Figure 2:** Mean out of 95% confidence interval errors for fitted model with lowest AIC, true distribution Weibull. (a) Contour plot of error rate using generalised additive model with logistic link conditional upon smoothing splines of shape and censoring time; (b) cross-sectional error rates for various Weibull shapes; (c) True survival time distribution for the illustrative shapes.

- Figure 1 shows the reference error rate for the true mean being out of the 95% confidence interval of the estimate when the true generating distribution is known and used for fitting. Errors rates greater than 5% are assumed to be due to breakdown in the multivariate normal assumption for the parameter distribution.
- Figure 2 shows the equivalent error rate when the generating distribution is unknown and the model with minimum AIC is chosen. Error rates increase with increasing true mean / decreasing shape, and are seen to peak around a shape value of 1, suggesting that misclassification of models as exponential causes large errors in mean estimation here. Error rates with follow-up less than the true median of the distribution are high, with shapes in the range of 0.5 to 1.5 experiencing a greater than 50% error rate. Error rates decrease sharply with observation around and after median follow-up, but error rates remain above expected until follow-up is sufficient for >99% observation of the survival time distribution.

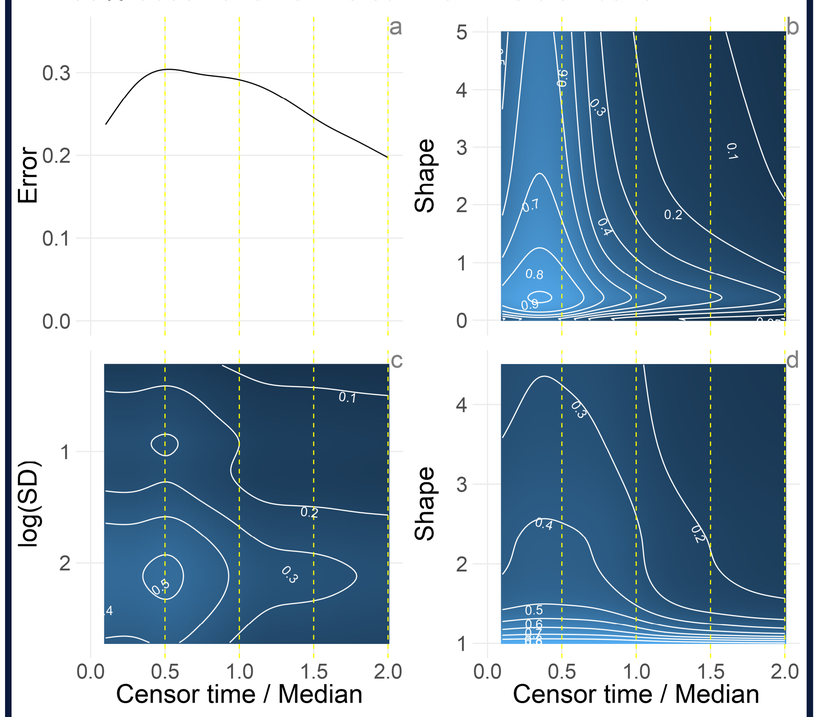


**Figure 3:** Mean out of CI error for (a) exponential distribution, (b) Gompertz distribution, (c) lognormal distribution, (d) log-logistic distribution

## Point estimates

- Figure 3 shows the same error rates for the other generating distributions. The error rate for Gompertz distributions was particularly high with follow-up below median.
- Figure 4 shows the rate of point estimate of mean from maximum likelihood parameters deviating by > 10% from the true mean. This point estimate, often used for "base case" cost effectiveness evaluation, is affected both by uncertainty due to lack of follow-up and by error in model selection. As a result, for the Weibull model, follow-up below 0.5 times median results in very poor estimation if the true distribution is known; for follow-ups of twice median survival more than 20% of estimated means from distributions with decreasing hazards deviate by more than 10%. When the distribution is unknown and selected by AIC ranking, this error is further compounded and more follow-up is required to reduce the error in estimation.
- Figure 5 demonstrates this error for other generating distributions. Note that the area of the lognormal plot is weighted more towards distributions with higher mean (higher SD) than the other plots.

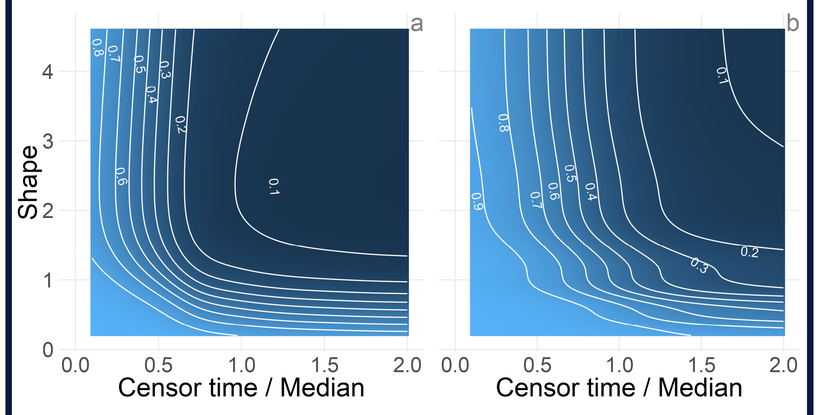


**Figure 4:** Estimated mean >10% deviation from true mean for Weibull distribution, contour plots of error rate using generalised additive model with logistic link conditional upon smoothing splines of shape and censoring time (a) for fitted Weibull distributions (b) for distributions selected by minimum AIC.

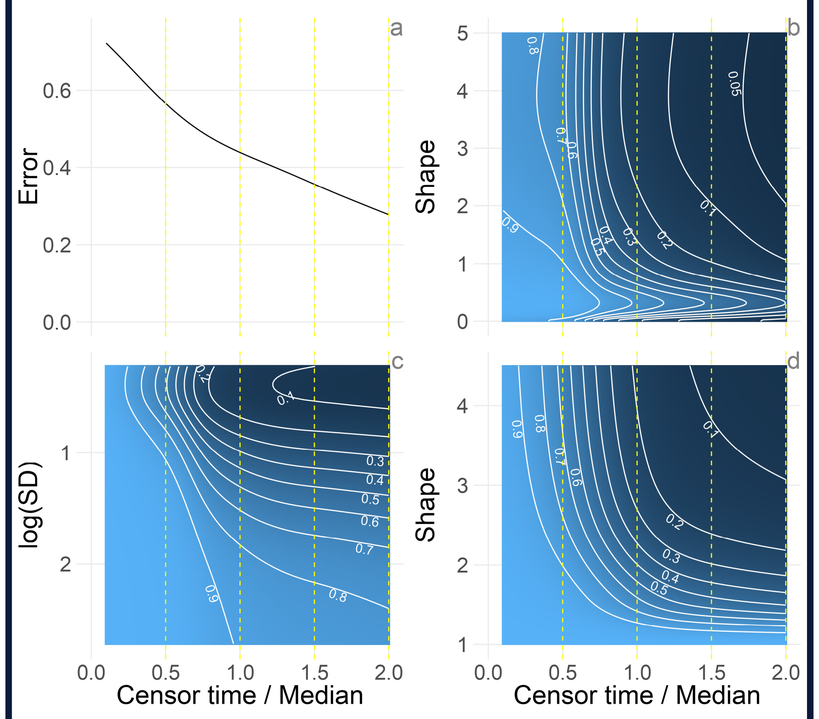


**Figure 5** Estimated mean >10% deviation from true mean for (a) exponential distribution, (b) Gompertz distribution, (c) lognormal distribution, (d) log-logistic distribution

## Conclusions

- Cost-effectiveness affecting error rates associated with model selection by AIC rank are very high, particularly below median follow-up.
- Point estimation of mean survival even with known distributions given the sample sizes in common studies frequently results in errors of greater than 10% when follow-up is below median.
- AIC ranking should never be used to perform unsupervised model selection.

## References

1. Latimer, N (2011) NICE DSU Technical support document 14. *Available from http://www.nicedsu.org.uk*
2. Bell Gorrod, H et al (Accepted 2019) A review of survival analysis methods used in NICE technology appraisals of cancer treatments. *Medical Decision Making*
3. Jackson, C (2016) flexsurv: A platform for parametric survival modelling in R. *Journal of Statistical Software, 70 (8)*

Presented at ISPOR Europe in Copenhagen, Denmark; 2-6 November, 2019