The Hazards of Applying Hazard Ratios to Accelerated Failure Time Models: A Simulation Study

Hale O[1], Latimer N[1,2]

[1]Delta Hat Limited, Nottingham, UK; [2]Sheffield Centre for Health and Related Research, University of Sheffield, UK

Introduction

- Parametric survival models can be used to extrapolate survival data beyond the observed time period. This is done by assuming that the data follow an underlying distribution – and that once the parameters of the distribution have been established, survival can be estimated at any timepoint
- National Institute for Health and Care Excellence (NICE) technical support document 14 recommends considering six 'standard' distributions, and each of these distributions fall into two categories[1]:
  - Proportional hazard (PH) models – where the ratio of the hazards for each treatment group is constant over time
  - Accelerated failure time (AFT) models – where the treatment multiplicatively influences the time an event will occur
- It is common practice to see hazard ratios applied to these survival models to approximate the survival outcomes of a comparator for which there is no head-to-head data
- Theoretically, the hazard ratio will only provide an accurate measure of relative effectiveness when it is applied to the PH model used to estimate it. However, it is common practice for a hazard ratio to be applied to AFT models, which do not model survival on the hazard scale, in NICE appraisals. This may introduce bias into survival estimates.

Objective

- The aim of our study was to explore whether mismatching the scale of a relative treatment effect and the survival model it is applied to results in bias in long-term extrapolations

Methods

- This simulation study aimed to identify and quantify the bias of survival models that were estimated by applying a hazard ratio to an inferior treatment arm modelled using an AFT
- Survival data were simulated using the R package survsim[2]
- Each simulated data set included 600 patients randomized 1:1 between the intervention and comparator arms
- Several variables that may impact the survival model were included in the simulations (Table 1). These were tested using a factorial design, resulting in 72 scenarios
- The steps followed within the simulation are outlined in Figure 1 and 1,000 iterations were run for each scenario and a mean RMST was calculated
- When applying general population mortality capping, UK age- and sex-specific lifetables were used, assuming a mean age at baseline of 70 years and an all-male population[3]
- Restricted mean survival time (RMST) in the intervention group was used as the primary performance measure. Method estimates were compared to the true RMST restricted to a 30-year time-point

Figure 1. Steps implemented within the simulation study

Simulate survival data	• Data for one of the 72 scenarios were simulated using a jointly fitted parametric model, with the underlying distribution, treatment effect and length of follow-up defined by the scenario
Estimated treatment effect	• A hazard ratio between the two simulated arms was estimated using a Cox proportional hazards model
Estimate the intervention arm	• The "correct" parametric survival model, was fitted to the survival data for the comparator arm • The hazard ratio from the Cox model was then applied to this extrapolation to estimate the survival function for the intervention arm
Assess for bias in the methodology	• The true survival function of the intervention arm was compared with the one estimated using the Cox hazard ratio, using RMST • The percentage bias in RMST was the primary outcome used to assess the method investigational methodology

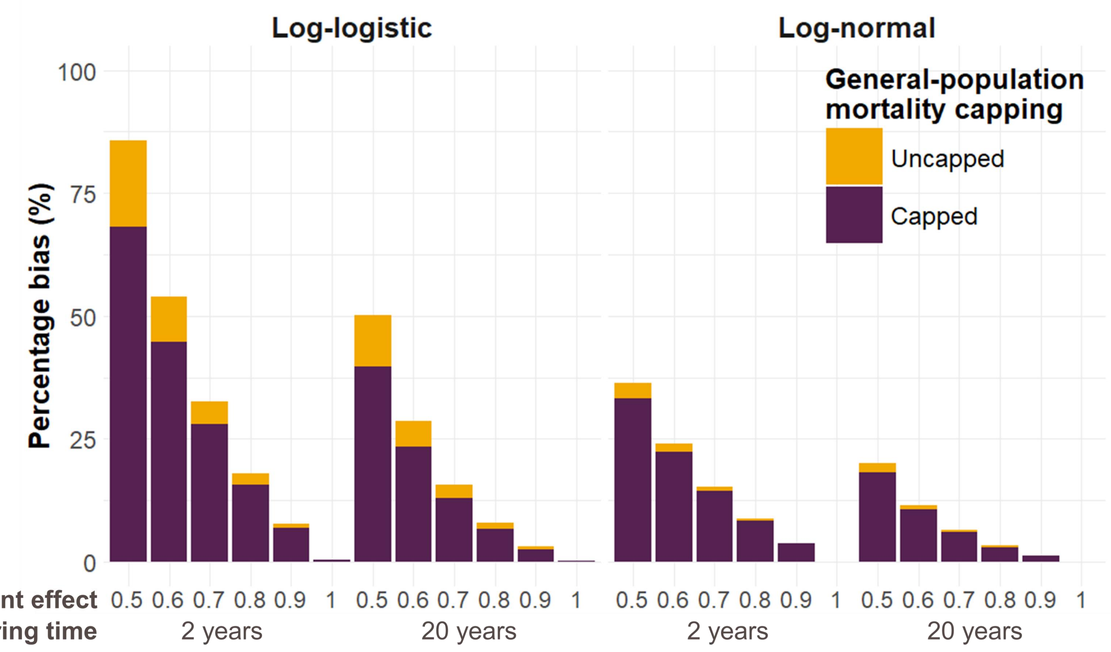
Table 1. Variables tested in the simulation study

Variables	Assessed options
Underlying distribution	• Log-normal • Log-logistic • Weibull
Magnitude of treatment effect between arms (Hazard Ratio)	• 0.5 • 0.6 • 0.7 • 0.8 • 0.9 • 1.0 (No effect)
Censoring	• No censoring • Administrative censoring at 2 years
General-population mortality capping	• No capping • Capped

Results

- The results of the study are presented in figure 2
- The log-logistic model was associated with the greatest deviation in RMST, with bias of up to 86% when modelling a large treatment effect with censored data. The log-normal model was subject to less bias than the log-logistic, but RMST estimates were still overestimated by up to 36%. Applying a hazard ratio to an AFT Weibull resulted in very little bias in RMST estimates and this was consistent across all the modelled scenarios
- 13% of iterations resulted in an underestimate of the intervention RMST, this occurred in scenarios with smaller treatment effects. No iterations resulted in underestimates when the treatment effect was 0.5 or 0.6 between arms
- Capping survival with general-population mortality reduced the bias in RMST, but the log-logistic model still overestimated survival by up to 68%
- Applying administrative censoring increased the bias in intervention arm RMST estimates, this may be related to increased uncertainty when fitting the survival model or when estimating the hazard ratio, rather than increasing the inaccuracy of this methodology

Figure 2. Bias in restricted mean survival estimated by applying a hazard ratio to a comparator arm modelled using an accelerated failure time model



Conclusions

- Applying a hazard ratio to an AFT model fitted to one treatment arm to estimate survival for another treatment arm can lead to extremely biased survival estimates for the latter treatment.
- The magnitude of the bias is dependent upon the distribution of the accelerated failure time model used, the size of the treatment effect, and the length of follow-up in the patient level data
- Some of the bias is mitigated by capping survival with general population mortality, suggesting it is not only due to overestimated survival tails
- The use of this methodology within cost-effectiveness models could bias the results of these models, and in turn impact decision making within health technology assessments

REFERENCES

1. Latimer. NICE DSU Technical Support Document 14. 2013. https://www.sheffield.ac.uk/nice-dsu/tsds **2.** Morina and Navarro. The R Package survsim for the Simulation of Simple and Complex Survival Data. JStatSoft. 2014; 59(2) **3.** Office for National Statistics. National life tables: UK. 2024. Available at https://www.ons.gov.uk/peoplepopulationandcommunity/ birthsdeathsandmarriages/

Presented at ISPOR Europe 2024; 18 – 21 November 2024; Barcelona, Spain.

Poster EE805