Evaluating the performance of smoothed hazard functions over study follow-up and by numbers at risk: A simulation study

Jack Williams[1], Michael Sweeting[2], Iain R. Timmins[2], Thomas Haugli-Stephens[3], Robert Hettle[1]

1. Health Technology Assessment and Modelling Science, AstraZeneca, Cambridge, UK,
2. Statistical Innovation, AstraZeneca, Cambridge, UK
3. Health Economics, AstraZeneca, Oslo, Norway

MSR109

Introduction

- Cost-effectiveness analyses routinely require extrapolation of time-to-event data such as overall survival.
- HTA guidelines recommend comparing the within-study empirical hazards to the predicted hazard functions of parametric survival models, to guide the choice of survival extrapolation.[1-3]
- Previous research have shown empirical examples of hazard functions that become unstable at the tail (i.e. near the end of follow-up), when the number of patients at risk is low.[4]

Aim

We aim to evaluate the accuracy of two common smoothed empirical hazard functions in predicting the 'true' hazard rate and assess their performance over the length of follow up and by the number of patients at risk.

Methods

- Smoothed hazards were estimated using *muhaz* and *bshazard* R packages, based on kernel-based and spline smoothing methods, respectively.
- Each function was fit to 1,000 simulated cohorts, each with 1,000 patients, using the default function settings. *Muhaz* fits until 10 patients remain at risk, whilst *bshazard* fits for the duration of study follow-up.
- The time-to-events for simulated cohorts were sampled from a 'true' underlying exponential survival distribution, with a constant hazard rate of 30 cases/100 person-years (Table 1).
- A sensitivity analysis considered a more complex hazard function, with two turning points, using a mixture Weibull distribution. This was created by combining two simulated cohorts (n = 500 each), each with a different underlying Weibull hazard function.
- The accuracy of each hazard function was estimated over the study follow-up by calculating the mean absolute error (MAE) and mean absolute percentage error (MAPE) between the smoothed hazard function fit to each simulation, and the true underlying hazard.
- Results were categorised by the number of patients at risk (intervals of 10) or by follow-up time (i.e. months, intervals of 6) and averaged across all simulated cohorts.

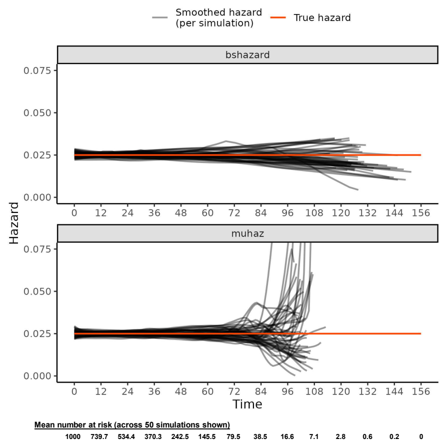
Table 1: Simulation parameters

Cohort sample size (n)	1000
Number of simulations (n)	1000
Event parameters – Exponential (simple hazard)	Rate = 0.025
Censoring parameters – Weibull (simple and complex hazard)	Shape = 3, Scale = 80
Event parameters – Mixture Weibull (complex hazard)	Shape = 1.6, Scale = 40 (50%) Shape = 3.5, Scale = 20 (50%)

Results – Simple hazard

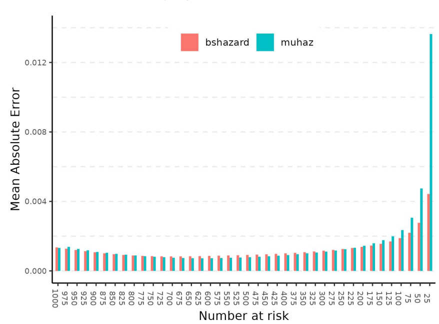
- Using a constant hazard, both functions performed similarly well until the numbers at risk were low, at which point the smoothed hazards became increasingly inaccurate, particularly with the *muhaz* function (**Fig. 1**).
- At the tail (end of study follow-up), the *muhaz* smoothed hazard function was prone to predicting turning points that were not reflective of the true underlying hazard.

Figure 1: Smoothed hazards across 50 simulated cohorts.



- The MAE increased with lower numbers of patients at risk, from 0.0022 (8.9% MAPE) and 0.0018 (7.3%) with between 91 and 100 patients at risk, up to 0.0065 (26%) and 0.0032 (13%) with between 21 and 30 patients at risk, for *muhaz* and *bshazard* respectively (**Fig. 2**).

Figure 2: Mean absolute error based on number of patients at risk, where the underlying true hazard is constant.



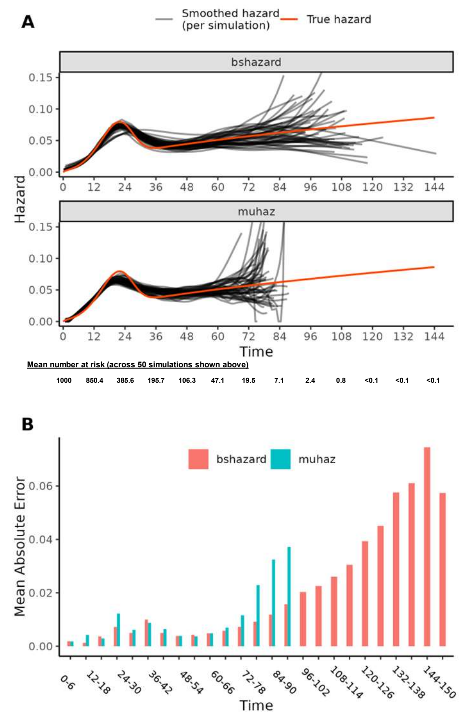
Limitations

- Our analysis did not consider how inaccuracies in the smoothed hazard functions may influence the choice of extrapolation model (or the consequences of this).
- We did not consider smaller sample sizes or alternative censoring patterns
- Identifying an exact 'threshold' at which smoothed hazard functions become unreliable is challenging, and further research is required.

Results – Complex hazard

- Using a complex hazard function (two turning points), both smoothed hazard functions performed well when the sample size was large (**Fig. 3A**).
- The largest error in the smoothed hazard functions was observed at the tail (**Fig. 3B**), with a small increase in the error around the turning points in the hazard.
- Overall, few simulations captured the tail of the hazard function well.

Figure 3: A complex hazard function (mixture Weibull) with A) smoothed hazard functions vs. the 'true' underlying hazard (n=50 simulations shown) and B) the Mean Absolute Error across time intervals for the two smoothed hazard functions (n=1000 simulations).



Conclusions

- With a large sample size, both the *muhaz* and *bshazard* smoothed hazard functions perform well until the numbers at risk were low, for both simple and complex hazard shapes.
- At the tail, when the numbers at risk are low, observed turning points in the hazard function may not reflect the true underlying hazard.
- To aid interpretation, we recommend that both smoothed hazard methods are presented, alongside confidence intervals (where possible), and with the numbers of patients at risk and numbers of events displayed.

Existing smoothed hazard functions accurately predict the 'true' hazard until numbers at risk become small. The hazard at the tail is often misrepresented and should not be relied on as evidence of a turning point in model extrapolations.

References

1. Medicinrådet [Danish Medicines Council]. Ansøgningsskema til vurdering af nye lægemidler Copenhagen: Medicinrådet2024
2. Norwegian Medicinal Products Agency. Submission guidelines for single technology assessment of medicinal products. Oslo: Direktoratet for medisinke produkter. 2024
3. Rutherford MJ, Lambert PC, Sweeting MJ et al. NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis. 2020.
4. Williams J, Hettle R, Haugli-Stephens T. PT16 Impact of Data Maturity on the Estimation of the within-Trial Hazard Function: An Example from Metastatic Castration Resistant Prostate Cancer. Value in Health. 2023;26(12):S425-S6.

Funded by AstraZeneca. All authors are employees of AstraZeneca.

Presented at ISPOR Europe 2024, Barcelona, Spain