Survival extrapolations with and without incorporating external evidence: Evaluating the performance of Bayesian M-spline models vs. parametric models in first-line metastatic castration resistant prostate cancer (mCRPC)

Jack Williams[1], Robert Hettle[1], Iain R. Timmins[2]

1. Health Technology Assessment and Modelling Science, AstraZeneca, Cambridge, UK
2. Statistical Innovation, AstraZeneca, Cambridge, UK

MSR191

Introduction

- Novel survival methods allow for clinical trial data to be combined with external evidence from historical trials, clinical expert opinion, and general population mortality estimates.
- These methods differ to standard approaches commonly used in HTA (e.g. parametric models) by allowing external evidence to directly inform extrapolations, rather than simply informing model selection.

Aim

This study aims to evaluate the performance of the survextrap R package, a Bayesian M-spline approach to survival extrapolation that incorporates both clinical trial and external evidence, versus standard parametric models fitted to clinical trial data only.

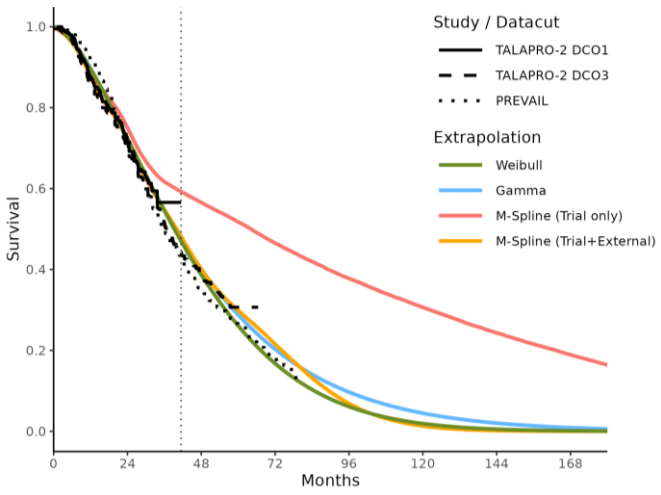
Methods

- Pseudo-patient data were recovered from Overall Survival (OS) Kaplan-Meier (KM) plots for enzalutamide from two mCRPC studies; TALAPRO-2 (NCT03395197) and PREVAIL (NCT01212991).[1,2] The details of each study are shown in **Table 1**.
- Standard parametric models and M-spline models (with and without incorporation of historical PREVAIL data) were fit to the TALAPRO-2 primary analysis data and used to predict survival beyond trial follow-up.[3,4]
- The accuracy of model extrapolations was assessed by comparing model predictions to the final OS analysis of TALAPRO-2.[5]
- Historical PREVAIL data was incorporated into the M-spline model as binomial counts of death from the total number at risk for each 3-month interval, starting from the end of TALAPRO-2 follow-up (41 months) up to the last interval with an event (78 months).
- Scenario analyses evaluated the impact of knot number and knot placement on the extrapolated hazard functions.
- Conditional survival probabilities were also assessed following the end of the available external evidence used in the M-spline model (i.e. survival from 78 months onwards).

Results: Parametric vs. M-spline models

- Using TALAPRO-2 primary analysis the best-fitting parametric models, according to visual fit to TALAPRO-2 and proximity to PREVAIL during extrapolation, were Weibull and Gamma models.
- The M-spline models fit to TALAPRO-2 primary analysis performed well when the external evidence (PREVAIL) was included. When no external evidence was incorporated, extrapolations performed very poorly.
- Overall, the parametric models and M-spline model with external data performed well when predicting the TALAPRO-2 final analysis (**Fig 1**).

Figure 1: Comparison of best-fitting parametric and M-spline survival extrapolations, based on TALAPRO-2 primary analysis (with or without external evidence from PREVAIL study).



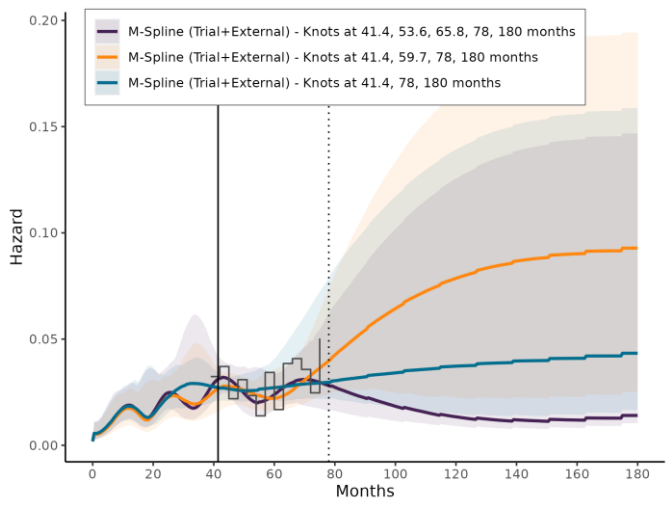
Limitations

- In our example, the survival outcomes of the external evidence (PREVAIL) were closely aligned to the contemporary trial data from TALAPRO-2. Our findings may differ if external evidence does not align with the trial of interest.
- Moreover, our case study had a simple hazard function. The flexibility of M-spline models may perform better when capturing complex hazard functions observed in external evidence, to which standard parametric models are less suited.
- M-spline models can incorporate clinical expert opinion to guide extrapolations, which was not considered in our study, which would likely reduce uncertainty of extrapolations and influence of knot numbers/placement.
- M-spline models can be constrained to use a constant hazard beyond the available evidence, however we did not evaluate this.

Results: M-spline knot number/location

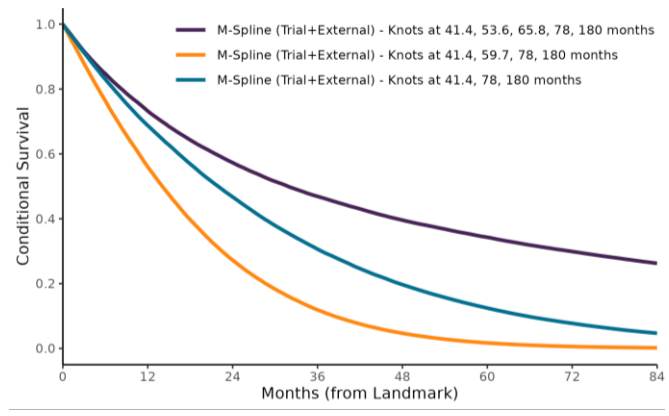
- Even when the M-spline model incorporated external data (PREVAIL, up to 78 months of data), the extrapolated hazards differed significantly based on knot number and knot location used within the period of external trial data (**Fig 2**).
- This differs with other spline methods such as the Royston Parmar cubic splines, which tend to be insensitive to choice of knot location.[6,7]

Figure 2: Hazard functions estimated by M-spline model (including external evidence from PREVAIL) with different knot numbers included in the external evidence period (between 41.4 months and 78 months).



- The knot number and placement for M-spline models has a significant impact on conditional survival beyond the available evidence, in this instance from 78 months onwards (**Fig 3**).

Figure 3: Conditional survival probability from 78 months onwards (contingent on survival up to this landmark).



Conclusions

- The Bayesian M-spline model performed well when high-quality external evidence was available that aligned to the clinical trial data.
- When extrapolating into time periods without trial or external evidence, the M-spline models produced highly uncertain and generally unreliable extrapolations in our case study.
- M-spline models may be a suitable alternative to standard methods where the hazard function is complex and standard extrapolations fail to predict external evidence and/or clinical opinion.
- Where adopted, care should be taken when using M-splines during periods without any external evidence, and extensive sensitivity analysis of knot numbers and placement should be performed since these can significantly alter survival extrapolations.

Table 1: Key study data used for survival analyses

Study	Data cut (Year)	Maximum OS KM (months)	Maturity (n/ N)	Role in analysis
TALAPRO-2 [1]	Primary Analysis (2022)	41.4	32% (129/403)	Clinical trial data used in main survival analysis
TALAPRO-2 [5]	Final Analysis (2024)	66.5	60% (243/403)	Used to validate the trial extrapolations from primary analysis
PREVAIL [2]	Extended overall analysis (2017)	81.1 (last event up to 78 months)	79% (689/872)	Historical trial used as external evidence in M-spline model, and used to support parametric model selection

References

1. Agarwal N, Azad AA, Carles J et al. The Lancet. 2023;402(10398):291-303.
2. Armstrong AJ, Lin P, Tombal B et al. European Urology. 2020;78(3):347-57.
3. Jackson CH. J Stat Softw. 2016;70.
4. Jackson CH. BMC Medical Research Methodology. 2023;23(1):282.
5. Agarwal N, Azad A, Carles J et al. Journal of Clinical Oncology. 2025;43(5_suppl):LBA18-LBA.
6. MJ Rutherford, PC Lambert, MJ Sweeting et al. NICE DSU Technical Support Document 21. 2020.
7. Hettle R, Sackeyfio A, Suri G. Value in Health. 2016;19(7):A347.

Funded by AstraZeneca.

Presented at ISPOR Europe 2025, Glasgow, UK