When the tail wags the dog:
a transparency framework for survival extrapolation in oncology HTA

Andrew H. Briggs · Ziyi Lin · Alex Hill · Elisabeth Fenwick*
London School of Hygiene & Tropical Medicine, London, UK *also OPEN Health, Oxford, UK

The problem
How much of the answer comes from beyond the data?

Survival extrapolation is unavoidable in oncology HTA. Trials rarely follow patients long enough for the decision-relevant time horizon, so observed Kaplan–Meier data must be projected forward with parametric models. NICE TSD 14 and TSD 21 give the toolkit for fitting and selecting these models, but no current standard tells the committee how much of the estimated treatment benefit comes from beyond the observed data. Models with near-identical AIC can produce wildly different ICERs.

Transparency framework
Decompose the survival benefit. Three statistics. One diagnostic.

We decompose the unrestricted mean survival time (UMST) of each arm into a restricted mean survival time (RMST) up to a cutoff t* and an extrapolated mean survival time (EMST) beyond it:

$$UMST(0, H) = RMST(0, t^*) + EMST(t^*, H)$$

Applied to both arms, the decomposition partitions the incremental survival benefit ΔUMST into an observed-period component ΔRMST and an extrapolated-period component ΔEMST.

Arm	RMST (0 to t*)	EMST (t* to H)	UMST (0 to H)
Treatment (T)	RMST_T	EMST_T	UMST_T
Comparator (C)	RMST_C	EMST_C	UMST_C
Benefit (T − C)	**ΔRMST**	**ΔEMST**	**ΔUMST**

Three transparency statistics fall out:

1. **EBS** – Extrapolation Benefit Share

   $$\Delta EMST \div \Delta UMST$$

   The share of modelled benefit that lies beyond t*
2. **SRS** – Survival Ratio Share

   $$(EMST\_T / EMST\_C) \div (RMST\_T / RMST\_C)$$

   Whether the relative benefit is tail loaded vs observation loaded
3. **HRT** – Whether the model assumes the treatment effect is amplifying, stable, or attenuating

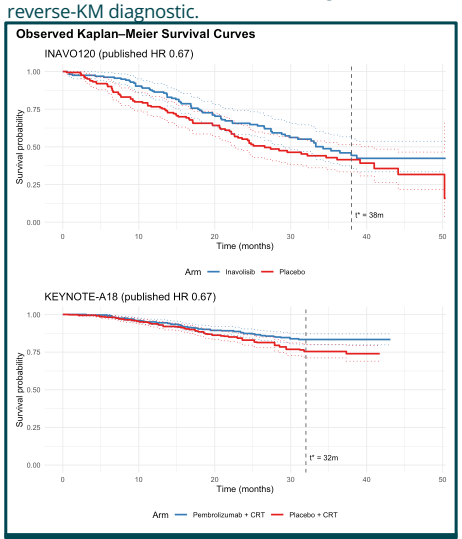
Alongside these we propose the implied HR(t) trajectory as the primary diagnostic for model plausibility, with a formal reversal test (does HR(t) cross 1.0 between t* and H?)

Two case studies
Identical headline effect. Very different curves

Two oncology overall-survival datasets, both with a published hazard ratio of 0.67, but on opposite ends of the maturity spectrum.

INAVO120 (Jhaveri et al., NEJM 2025). Inavolisib + palbociclib + fulvestrant vs placebo + palbociclib + fulvestrant in PIK3CA-mutated HR+/HER2− advanced breast cancer. n = 325. Mature overall survival data, medians reached in both arms: 34.0 months (Inavolisib) vs 27.0 months (Placebo). Reconstructed Cox HR 0.74. t* = 38 months.

KEYNOTE-A18 (Lorusso et al., Lancet 2024). Pembrolizumab + chemoradiotherapy vs placebo + chemoradiotherapy in locally advanced cervical cancer. n = 1,060. Immature data, neither median reached; 36-month OS 82.6 % (Pembrolizumab) vs 74.8 % (Placebo). Reconstructed Cox HR 0.69. t* = 32 months, anchored on the censoring-informed reverse-KM diagnostic.



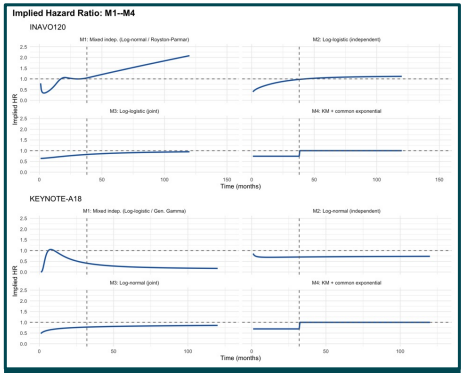
Observed Kaplan–Meier survival curves for overall survival in each case study. Dotted lines show 95% confidence intervals. The dashed vertical line marks the selected t*, beyond which survival is extrapolated

The M1–M4 model ladder
Four extrapolation strategies, increasing structural constraint on the treatment effect and KM + tail

Any survival extrapolation makes an implicit assumption about how the treatment effect evolves beyond the data. The transparency statistics are most informative when applied across a ladder of approaches that varies that assumption deliberately, from no constraint to complete constraint.

We compare four strategies, each fitted to both arms over the observed period and projected to a 240-month horizon:

- **M1** Independent distribution by arm
- **M2** Common distribution, independent fit
- **M3** Common distribution, joint fit
- **M4** K-M to t* then common exponential tail



Implied hazard ratio HR(t) for the four selected models in each case study. Dashed horizontal line: HR = 1 (no treatment effect). Dashed vertical line: t*. Models where HR(t) crosses 1.0 imply treatment effect reversal.

Observed vs Extrapolated
Two trials, four model classes

The table in the next column shows the restricted means in each arm (RMST) followed by the EMST summing to the UMST. The differences between the arms are key quantities for the transparency statistics that follow.

Case study	Model	RMST (0 to t*) T	RMST (0 to t*) C	RMST (0 to t*) Δ	EMST (t* to H) T	EMST (t* to H) C	EMST (t* to H) Δ	UMST (0 to H) T	UMST (0 to H) C	UMST (0 to H) Δ
INAVO120	M1	28.0	24.9	3.1	20.4	26.1	-5.7	48.4	51.0	-2.6
INAVO120	M2	28.1	25.0	3.1	28.0	24.5	3.5	56.1	49.5	6.6
INAVO120	M3	28.0	25.0	3.0	31.7	22.1	9.6	59.7	47.1	12.7
INAVO120	M4	28.1	25.0	3.0	18.4	17.2	1.2	46.4	42.2	4.2
KEYNOTE-A18	M1	29.4	28.7	0.8	204.6	72.6	132.0	234.1	101.3	132.8
KEYNOTE-A18	M2	29.5	28.6	0.9	133.1	97.2	35.9	162.6	125.8	36.8
KEYNOTE-A18	M3	29.5	28.6	1.0	125.0	102.3	22.7	154.5	130.9	23.7
KEYNOTE-A18	M4	29.4	28.6	0.8	221.9	200.7	21.2	251.3	229.3	22.0

Transparency statistics
Same trial, different model = wildly different transparency profile

The transparency statistics quantify what the trajectory diagnoses. Models that pass the reversal test are not all alike – and models that fail it can be flagged numerically even without inspecting the curve.

Case study	Model	Reversal?	ΔUMST (m)	EBS	SRS	HRT
INAVO120	M1	Yes	-2.6	220%	0.69	1.48
INAVO120	M2	Yes	6.6	54%	1.02	1.11
INAVO120	M3	No	12.7	76%	1.28	1.11
INAVO120	M4	No	4.2	28%	0.95	1.00
KEYNOTE-A18	M1	No	132.8	99%	2.74	0.57
KEYNOTE-A18	M2	No	36.8	97%	1.33	1.02
KEYNOTE-A18	M3	No	23.7	96%	1.18	1.06
KEYNOTE-A18	M4	No	22.0	96%	1.08	1.00

Recommendations
Two things HTA committees can do

1. **Lead with the implied HR trajectory**
   Examine the HR trajectory ahead of fit statistics, and apply the reversal test to exclude implausible models from the base case.
2. **Report EBS, SRS, and HRT alongside ICERs**
   Routine reporting makes extrapolation dependence visible: an EBS of 70 % tells the committee immediately that the majority of claimed benefit arises from beyond the observed data.

References

Latimer NR. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. Sheffield: NICE Decision Support Unit; 2011 (updated 2013). • Rutherford MJ, Lambert PC, Sweeting MJ, Pennington B, Crowther MJ, Abrams KR, et al. NICE DSU Technical Support Document 21: Flexible methods for survival analysis. Sheffield: NICE Decision Support Unit; November 2020 (updated March 2022). • Jhaveri KL, Im SA, Saura C, Loibl S, Kalinsky K, Schmid P, et al. Overall survival with inavolisib in PIK3CA-mutated advanced breast cancer. N Engl J Med 2025;393(2):151–61. • Lorusso D, Xiang Y, Hasegawa K, Scambia G, Leiva M, Ramos-Elias P, et al. Pembrolizumab or placebo with chemoradiotherapy followed by pembrolizumab or placebo for newly diagnosed, high-risk, locally advanced cervical cancer (ENGOT-cx11/GOG-3047/KEYNOTE-A18): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2024;404(10460):1321–32

andrew.briggs@lshtm.ac.uk · Manuscript in preparation