Death by a thousand cuts?

A case study how different data cuts impact survival extrapolation

Ash Bullement, (Delta Hat) & Michael Schlichting* (Merck Healthcare KGaA)
Corresponding author: Ash Bullement (abullement@deltahat.com)

EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Glossary of HTA terms: QALY: a quality-adjusted life year. Outcome measure commonly used in HTA. One QALY = one year of life in 'perfect' health = two years of life with a utility of 0.5; ICER: incremental cost-effectiveness ratio. Interpreted as the added total cost (product acquisition, medical resource use, etc.) per QALY gained.
Disclosures: Ash Bullement is an employee of Delta Hat Ltd, which received consultancy fees for the development of this presentation. Michael Schlichting is an employee of Merck Healthcare KGaA.

Health technology assessments (HTAs) of new medicines typically include evidence collected from a registrational clinical trial (or trials). At the time of submission, it is typical that limited data are available in terms of the duration of follow-up, based on a pre-specified interim analysis.

This is commonly seen in appraisals of new cancer drugs, where survival data are available only up until a given time point (usually described in terms of either the minimum or median follow-up of the cohort).

Accordingly, there are consequences for the extrapolation of time-to-event outcomes (e.g., overall survival [OS]) used to inform HTA decision making.

We set out to investigate the challenges presented by sequential data cuts from an HTA perspective (e.g., the National Institute for Health and Care Excellence [NICE]), focusing on a specific case study of JAVELIN Merkel 200 (NCT02155647). Details of other relevant case studies are also provided below, followed by recommendations.

Additional case studies

Tai *et al.*, (2021)

Tai *et al.* updated a cost-effectiveness analysis submitted to NICE using updated data from the pivotal clinical trial (TA381 [later replaced by TA620] of olaparib for maintenance treatment of relapsed platinum-sensitive ovarian, fallopian tube or peritoneal cancer).

The authors found that updating the analysis using the latest data cut caused the incremental QALY gain to change from +0.37 to +0.80, the incremental costs to increase slightly (by £266), and so the ICER essentially **halved**: originally £101,467, decreasing to £45,787.

The difference in the ICER was essentially driven by the fact that the original modelling under-estimated survival for olaparib. In this case study, the updated data **improved** cost-effectiveness.

Bullement *et al.*, (2019)

Bullement *et al.* performed a review of survival extrapolations for cancer immunotherapies based on data available at the time the company submitted to NICE, and compared these to the latest available estimates of survival from the pivotal trial.

In general, the authors found that company projections were reasonable, and that there was no obvious under- or over-estimation on a system level. However, on average, the methods used seemed to slightly **under-estimate** OS.

Later data still only tell us 'part of the story', and so further work is likely needed to understand how well the models perform over a lifetime horizon. However, in this study, the updated data showed broadly **consistent** results with analyses based on previous data.

Vadgama *et al.*, (2022)

Vadgama *et al.*, considered a range of different methods for survival extrapolation in CAR T-cell therapy (axicabtagene ciloleucel [axi-cel]), focusing on models fitted to sequential data cuts. Regulatory approval for axi-cel was based on early data, with objective response rate based on 6 months of minimum follow-up.

Following publication, NICE re-appraised axi-cel with updated survival data available from the pivotal clinical trial. Results were **supportive** of a substantial proportion of long-term survivors, and the model selected by the company was deemed appropriate (a mixture-cure model).

However, the updated data appeared to demonstrate a slightly **lower** 'implied' cure fraction versus the previous data cut. In spite of this, a positive recommendation was still reached concerning cost-effectiveness.

Recommendations

Based on the above case studies, we make the following recommendations for handling multiple sequential data cuts for the same clinical trial for use in an HTA submission:

- Standard parametric models can work well, but may not always be suitable
- Modelling survival for intervention with a novel mechanism of action may require more sophisticated approaches and adequate justification, particularly when new data cuts may influence the selection of the most appropriate model
- External evidence may provide helpful information to address uncertainties associated with trial-based extrapolation methods

Further research is required to investigate different methods for leveraging evidence collected outside of the pivotal trial, and to understand how relevant learnings could be used to in clinical development (e.g., to inform the design of Phase III clinical trials).

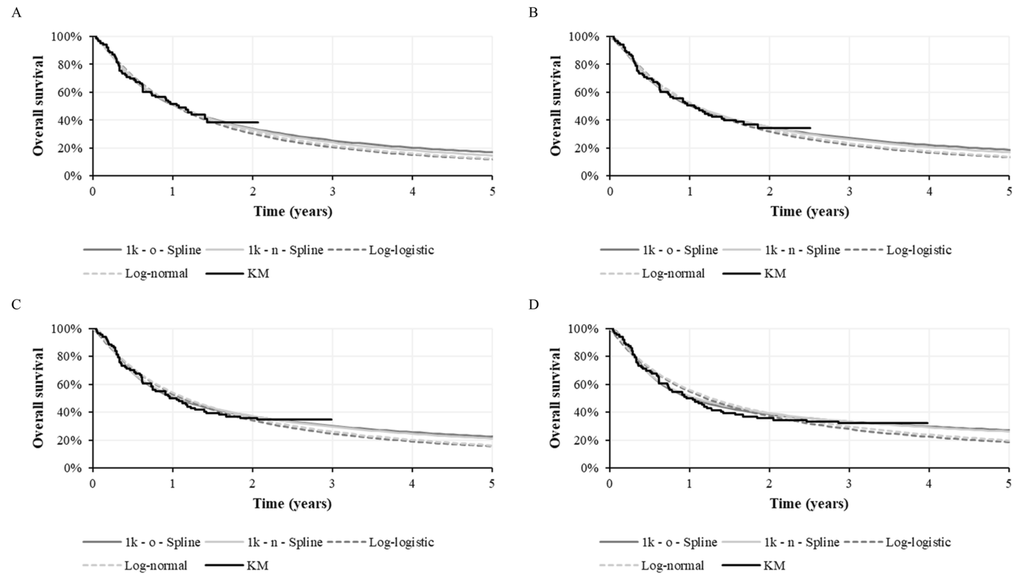
References: Tai TA, Latimer NR, Benedict Á, *et al.* Prevalence of Immature Survival Data for Anti-Cancer Drugs Presented to the National Institute for Health and Care Excellence and Impact on Decision Making. Value Health. 2021 Apr;24(4):505-512; Bullement A, Meng Y, Cooper M, *et al.* A review and validation of overall survival extrapolation in health technology assessments of cancer immunotherapy by the National Institute for Health and Care Excellence: how did the initial best estimate compare to trial data subsequently made available? J Med Econ. 2019 Mar;22(3):205-214; Vadgama S, Mann J, Bashir Z, *et al.* Predicting Survival for Chimeric Antigen Receptor T-Cell Therapy: A Validation of Survival Models Using Follow-Up Data From ZUMA-1. Value Health. 2022 Jun;25(6):1010-1017.

JAVELIN Merkel 200 case study

In this study, we investigated how standard and spline-based parametric survival model estimates changed as we informed the model with four sequential data cuts from the JAVELIN Merkel 200 study [NCT02155647]. The data cuts represented minimum follow-up periods of 12, 18, 24, and 36 months. We found that spline-based models provided the best fit to the trial data across each of the four data cuts.

However, all of the extrapolation methods fitted to the 12-month data appeared to underestimate the 'true' long-term survival. Based on the difference in restricted mean survival time (RMST) between the fitted models and the Kaplan-Meier estimate, the models underestimated survival by between 0.5 to 1.1 months. While these values may appear small in magnitude, such differences extrapolated over a lifetime horizon can have important implications for cost-effectiveness analysis.

Fitted models from Part A of the JAVELIN Merkel 200 clinical trial



Notes: A, 12-month data cut; B, 18-month data cut; C, 24-month data cut; D, 36-month data cut.
Key: k, knot(s); KM, Kaplan-Meier; n, normal; o, odds.
Source: Image taken from: Bullement, A., Willis, A., Amin, A. et al. Evaluation of survival extrapolation in immuno-oncology using multiple pre-planned data cuts: learnings to aid in model selection. BMC Med Res Methodol 20, 103 (2020). Image re-produced here in accordance with the terms set out in the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

Get involved!

Join us

Want to get more involved in this discussion and help impact the future of EU HTA? Become a member of the HTA ESIG today - scan the QR code or email htasig@psiweb.org