An exploration of how influential digitisation approaches are in their impact on survival estimates for health technology assessment (HTA)

Charlotte Day[1]*, Ash Bullement[1,2]

[1]Delta Hat, Nottingham, UK
[2]ScHARR, University of Sheffield, UK

Background

Digitisation and patient-level data (PLD) recreation are often used in health technology assessment (HTA) submissions when PLD are unavailable, but the Kaplan-Meier (KM) estimate could be approximated and then used to produce an estimate of life-years (LYs). However, there is currently no published best practice guidance for digitisation when used in HTA submissions and similar work. Additionally, there is often very little or no detailed description of the digitisation method provided in HTA submission materials. This could lead to variable results and consequently an unknown level of uncertainty.

Objectives

This study aims to assess the impact of different digitisation approaches on the approximation and extrapolation of KM estimates when utilised to assess LYs when lacking PLD.

Methods

Two hypothetical case studies with 'true' PLD were generated to produce KM estimates. The first KM exhibited a long tail in survival (i.e., a plateau) and the second had a small initial number of patients at risk. Each KM estimate was digitised using the WebPlotDigitizer software (https://apps.automeris.io/wpd/). Four scenarios utilising different plotting methods were undertaken: manual with minimal plotting (M-), manual with extensive plotting (M+), automated with minimal plotting (A-) and automated with extensive plotting (A+).

In the small number at risk scenario, the M- plotting focused on the corners of the KM whereas in the long tail scenario, the M- plotting was spread evenly as there were no corners. Automated plotting utilises built-in functions to plot the KM. Minimal plotting utilised a smaller number of data points compared to extensive plotting.

After digitisation, pseudo-PLD were generated using the approach defined by Guyot *et al.*, (2012), and parametric curves were fitted. The choice of survival model was informed by statistical goodness-of-fit scores. The KM estimate and extrapolated survival estimates from the scenarios were compared to the values using the 'true' PLD and the differences were expressed as a percentage.

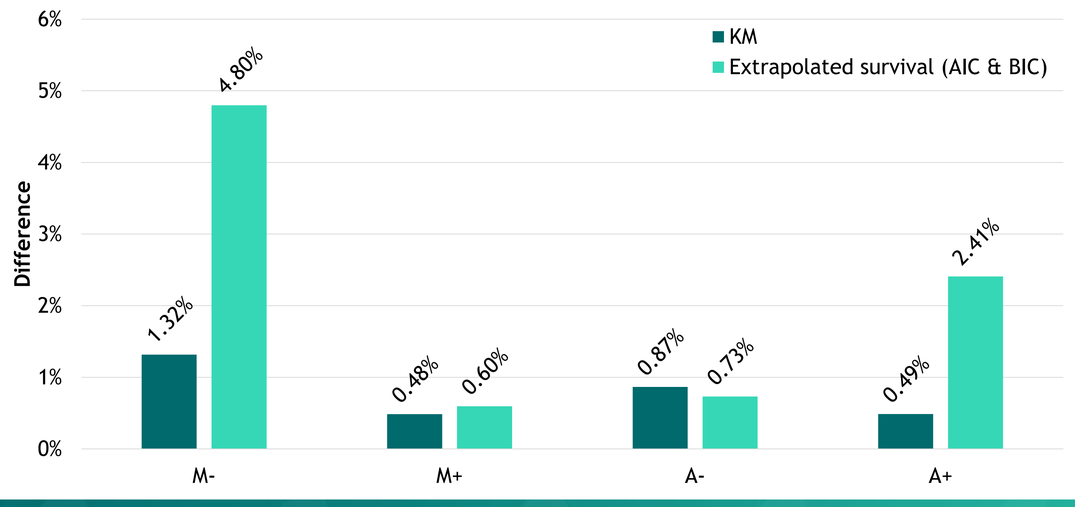
Results

Akaike information criterion (AIC) and Bayesian information criterion (BIC) scores were utilised to select the best-fitting survival model for each scenario (Table 1). The parametric curve with the lowest BIC score in each case was the same as that for the original case study. However, in the scenario with a small number at risk, three scenarios had generalised gamma as the model that had the lowest AIC score, compared to log normal being selected with the 'true' PLD. This could lead to a different survival model being chosen and therefore drastically different results. It should be noted that the differences between the lowest and second lowest scores were consistently less than 3 points in all scenarios of this case study.

Table 1: Summary the survival model selected using lowest AIC and BIC of each scenario

Case-study	Scenario	AIC	BIC
Long tail	'True' PLD	Generalised gamma	Generalised gamma
	M-	✓	✓
	M+	✓	✓
	A-	✓	✓
	A+	✓	✓
Small no. at risk	'True' PLD	Lognormal	Lognormal
	M-	Generalised gamma	✓
	M+	Generalised gamma	✓
	A-	Generalised gamma	✓
	A+	✓	✓

Figure 1: Differences between estimated life-years when utilising the 'true' PLD and each scenario - Long tail case study



In the long tail case study, all the scenarios had the same model chosen by the lowest AIC and BIC scores as when utilising the 'true' PLD. Figure 1 summarises the differences between the estimated life years from the 'true' PLD and each of the scenarios. The largest difference between the 'true' PLD life-years estimate and the scenarios was the M- scenario at 4.80% difference between estimated life-years using the generalised gamma extrapolated parametric curve.

In the small initial number at risk case study, the same survival model was chosen utilising the lowest BIC scores. However, the model selected with the lowest AIC score was different in 3 out of 4 scenarios. M-, M+ and A- AIC scores recommended generalised gamma. Figure 2 displays the 'true' KM estimate with the corresponding parametric curve fitted, alongside the plotting method with the largest difference from the original case study. The two different survival models, as chosen by each of the best fit scores, are displayed. There is a clear difference between the lognormal model for the 'true' PLD (AIC and BIC scores) and the generalised gamma model A- scenario (AIC only).

Figure 2: KM estimate and survival model extrapolation(s) for the original case study and A- scenario

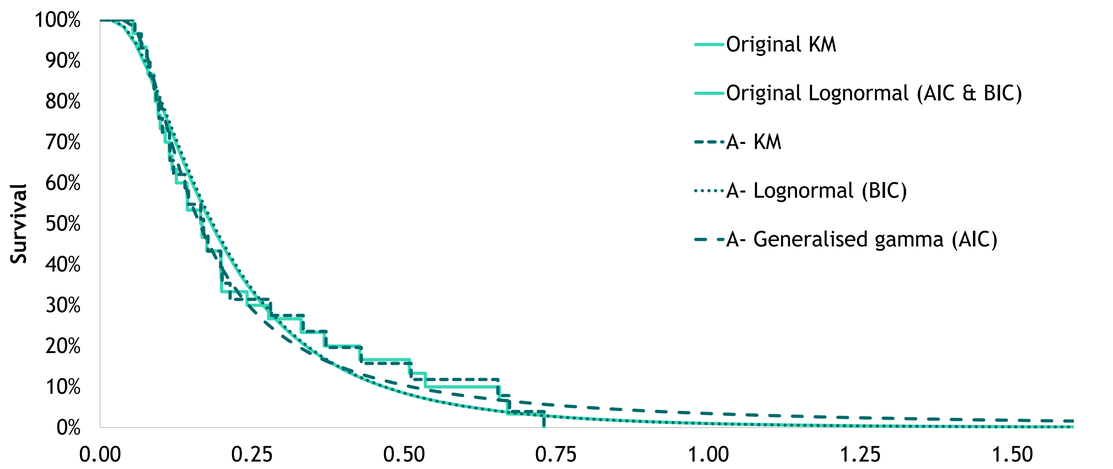
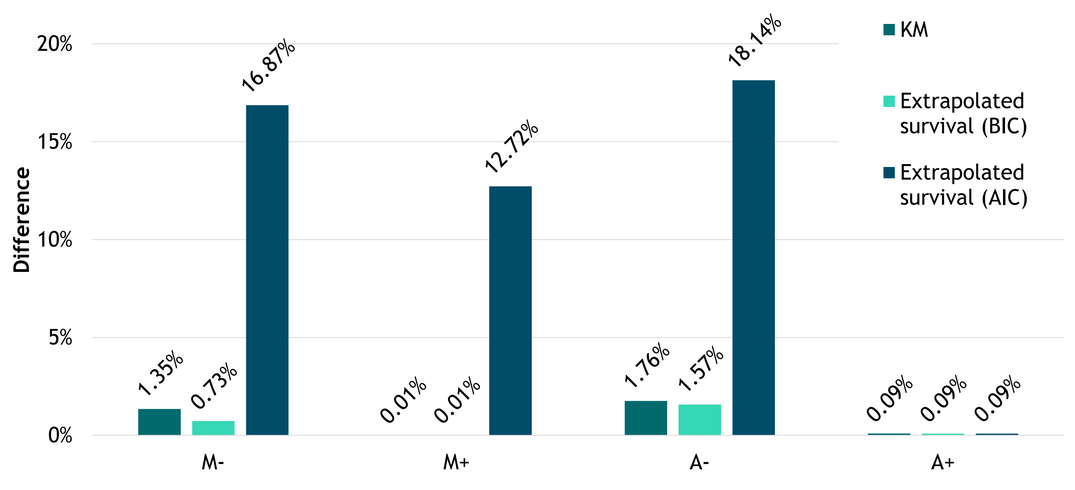


Figure 3 presents the differences between the estimated survival from the 'true' PLD and each of the scenarios using the parametric curves recommended by AIC and BIC scores. The extrapolated estimates of LYs increases greatly with the alternative model selected; the largest difference is seen in the A- scenario with a differences of 18.14%. The largest difference utilising the model selected by BIC scores was also in the A- scenario, with a difference of 1.57%.

Figure 3: Differences between estimated life-years when utilising the 'true' PLD and each scenario - Small no. at risk



Conclusions

A large degree of variation in survival estimates was found, dependent on the digitisation method, especially in cases where the AIC or BIC scores recommended different parametric curves compared to the 'true' PLD. Lack of guidance and detailed explanation of digitisation approaches are expected to have contributed to uncertainty in the accuracy of digitised data and any associated statistical analyses. This has important implications for HTA and future research possibilities, as extrapolations of digitised data are often used to inform comparative efficacy estimates and, by extension, calculations of LYs.

Abbreviations

A-, automated with minimal plotting; A+, automated with extensive plotting; AIC, Akaike information criterion; BIC, Bayesian information criterion; HTA, health technology assessment; LYs, life-years; M-, manual with minimal plotting; M+, manual with extensive plotting; PLD, patient-level data.

References

Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012 Feb 1;12:9. doi: 10.1186/1471-2288-12-9. PMID: 22297116; PMCID: PMC3313891.

Presented at ISPOR Europe 2022, 6-9 November 2022, Vienna, Austria.
Contact: cday@deltahat.com