Accurate Predictions of Life Year Gains for Immuno-Oncology Therapies in the Long Term? An Analysis Based on Published Checkmate 057 Nivolumab Data

A. Porteous,[1] K. Herbert,[2] C. Painter[1]

[1]Costello Medical, London, UK; [2]Costello Medical, Cambridge, UK

PCN20

OBJECTIVES

- This research aimed to retrospectively analyse the accuracy of different extrapolations of overall survival performed at early data cuts in predicting realised long-term life years based on the CheckMate 057 trial of nivolumab in advanced non-squamous non-small-cell lung cancer, to provide an indication as to whether certain parametric models might be more appropriate for immuno-oncology therapies.

BACKGROUND

- Immuno-oncology (I-O) therapies have emerged in the last few years as treatments for a variety of cancers. Due to their novel mechanism of action, the survival profiles for I-O therapies may be associated with a plateau, as well as evidence of a delayed effect, leading to complex hazard functions.
- Therefore, using best statistical fit to determine the extrapolation of short-term trial data at the time of Health Technology Assessment (HTA) might be expected to lead to inaccurate predictions of long-term survival.

METHODS

- Published Kaplan-Meier (KM) overall survival data for nivolumab from successive interim data cuts of the CheckMate 057 trial (NCT01673867) in advanced non-squamous non-small-cell lung cancer (NSCLC) (minimum follow-ups of 13.2, 17.2 and 24.2 months)[1–4] were digitised using DigitizeIt software.
- The digitised data points were converted to pseudo individual patient time-to-event data using the algorithm published in Guyot *et al.* 2012.[5]
- The generated time-to-event data were parameterised to estimate Exponential, Weibull, LogNormal, LogLogistic, Gompertz and Generalised Gamma parametric survival models and spline models (1–4 knots) as per the guidance from the National Institute for Health and Care Excellence Decision Support Unit in Technical Support Document 14.[6]
- The resulting models were tested for statistical fit using Akaike information criterion (AIC) and Bayesian information criterion (BIC).
- For each model, cumulative life years (LYs) were estimated over a 51.8-month time horizon (corresponding to the longest available duration of observed KM data, which comes from a data cut with minimum follow-up of 40.3 months).[7]
- Estimated LYs from extrapolated models were compared to realised LYs calculated from the published long-term KM data observed over 51.8 months to determine which interim models provided the most accurate predictions vs the long-term KM data.
- This methodology is displayed conceptually in **Figure 1**.

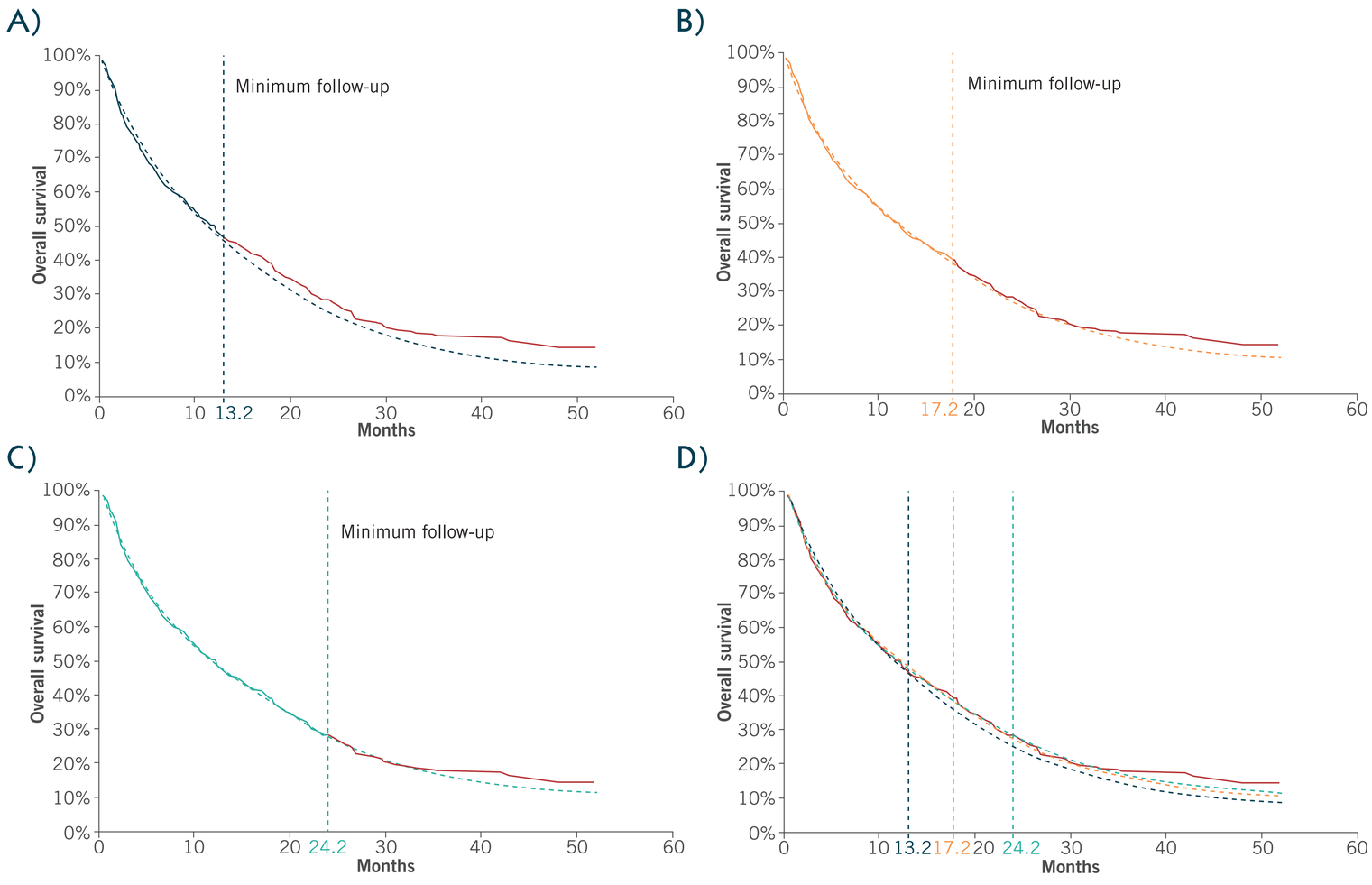
RESULTS

- Goodness-of-fit statistics (AIC and BIC) for the parametric models are presented in **Table 1**, and differences between estimated and realised cumulative LYs over 51.8 months are presented in **Table 2**.
- Of the standard parametric models, the LogNormal model provided the best statistical fit (i.e. the lowest AIC and BIC) at all interim data cuts. In contrast, the LogLogistic model provided the most accurate estimates of the LYs accumulated over 51.8 months at all interim data cuts.
- Overall, models that reflected a decrease in mortality rate over time (LogLogistic, Lognormal and Generalised Gamma) provided better statistical fit and more accurate estimates of the LYs accumulated over 51.8 months.
- Of the standard parametric models, the Exponential and Weibull models were seen to most substantially underestimate the realised LYs accumulated over 51.8 months.
- Due to their increased flexibility, the spline models offered better statistical fit than the standard parametric models at all interim data cuts. However, as the spline models increased in complexity with additional knots, the accuracy of LY predictions decreased, which could suggest that these models were overfitting the data.

CONCLUSIONS

- This study provides empirical evidence that the model of best statistical fit for short-term survival data may not provide the most accurate estimates of LYs for I-O therapies over the long-term. Prioritisation of best statistical fit may therefore lead to modelled cost-effectiveness estimates that are, in hindsight, inaccurate, with resultant implications for access where cost-effectiveness results inform decision making.
- One limitation of this approach is that 'accuracy' is measured against realised LYs calculated from long-term observed KM data that are themselves associated with uncertainty (e.g. due to small patient numbers informing the tail of the curve).
- Furthermore, this represents a single analysis. Further research is required to determine whether the results of this study are generalisable across I-O therapies and indications, and can hence provide meaningful learnings that might support the use of particular model choices at early data cuts; for example, those that allow for long-term survivors.

1 Conceptual depiction of study methodology



Long-term observed KM data compared with extrapolations based on successive interim data cuts

KM Data	Extrapolation	
——		Long-term observed KM data (minimum follow-up 40.3 months)
——	- - -	A) First interim data cut (minimum follow-up 13.2 months)
——	- - -	B) Second interim data cut (minimum follow-up 17.2 months)
——	- - -	C) Third interim data cut (minimum follow-up 24.2 months)
		D) All data cuts

KM: Kaplan-Meier.

1 Summary of goodness-of-fit data for nivolumab overall survival at interim data cuts

Model	Data Cut (minimum follow-up)					
	13.2 months		17.2 months		24.2 months	
	AIC	BIC	AIC	BIC	AIC	BIC
Exponential	1463.9	1466.7	1600.6	1603.3	1768.4	1771.2
Weibull	1465.7	1471.2	1602.2	1607.8	1769.9	1775.5
LogNormal	**1455.5**	**1461.1**	**1592.8**	**1598.4**	**1758.9**	**1764.4**
LogLogistic	1461.9	1467.4	1600.6	1606.1	1767.5	1773.0
Gompertz	1464.6	1470.1	1602.6	1608.1	1768.8	1774.3
Gen. Gamma	1457.5	1465.8	1594.8	1603.2	1760.9	1769.2
Spline 1 knot	1458.2	1466.5	1596.1	1604.5	1764.0	1772.3
Spline 2 knot	**1451.2**	**1462.3**	1585.6	**1596.8**	1752.4	**1763.5**
Spline 3 knot	1452.1	1466.0	1584.5	1598.4	1751.3	1765.2
Spline 4 knot	1452.8	1469.5	**1581.0**	1597.7	**1751.2**	1767.8

A smaller AIC or BIC value represents a better goodness of fit. Green highlighting indicates the standard parametric distribution with the lowest AIC or BIC. Orange highlighting indicates the spline model with the lowest AIC or BIC. AIC: Akaike information criterion; BIC: Bayesian information criterion; Gen.: generalised.

2 Accuracy of predicted LYs over 51.8 months compared with published long-term KM data

Model	Difference from realised LYs over 51.8 months		
	Data cut (minimum follow-up)		
	13.2 months	17.2 months	24.2 months
Exponential	-7.4%	-6.5%	-4.3%
Weibull	-6.6%	-7.1%	-4.1%
LogNormal	2.1%	-0.5%	-1.2%
LogLogistic	**1.7%**	**0.1%**	**-0.3%**
Gompertz	-2.3%	-6.5%	-2.8%
Gen. Gamma	2.6%	-0.5%	-1.3%
Spline 1 knot	**0.5%**	-2.4%	-1.6%
Spline 2 knot	-5.7%	-7.4%	-3.9%
Spline 3 knot	-6.4%	-7.6%	-3.6%
Spline 4 knot	-6.5%	-10.7%	-4.3%

Green highlighting indicates the standard parametric model with the most accurate prediction of realised LYs. Orange highlighting indicates if a spline model more accurately predicts realised LYs. Gen.: generalised; KM: Kaplan-Meier; LYs: life years.

References

1. Borghaei H *et al.* New Engl J Med 2015;373:1627–39; **2.** Horn L *et al.* Phase 3, randomized trial (CheckMate057) of nivolumab vs docetaxel in advanced non-squamous (Non-SQ) nonsmall cell lung cancer (NSCLC): subgroup analyses and patient-reported outcomes (PROs). Presented at the European Cancer Congress 2015; **3.** Horn L *et al.* J Clin Oncol 2019;35:3924; **4.** National Institute for Health and Care Excellence. TA484: Nivolumab for previously treated non-squamous non-small-cell lung cancer. Committee Papers. 2017. Available at: https://www.nice.org.uk/guidance/ta484/documents/committee-papers [Last accessed: 16 September 2019]; **5.** Guyot P *et al.* BMC Med Res Methodol 2012;12:9; **6.** National Institute for Health and Care Excellence Decision Support Unit. Technical Support Document 14. 2013. Available at: http://nicedsu.org.uk/technical-support-documents/survival-analysis-tsd/ [Last accessed: 16 September 2019]; **7.** Vokes *et al.* Ann Oncol 2018;29:959–65.

Acknowledgements

The authors thank Charlotte Bright, Costello Medical, for graphic design assistance.

Presented at the ISPOR 22nd Annual European Congress | Copenhagen, Denmark | 2–6 November 2019