ISPOR Europe 2023
12 - 15 November

Delta Hat

Testing a Survival Extrapolation Algorithm for Cancer Immunotherapies: Pass or Fail?

Latimer NR[1,2], Taylor K[1], Hatswell A[1], Ho S[3], Okorogheye G[3], Chen C[3], Kim I[4], Borrill J[3], Bertwistle D[3]

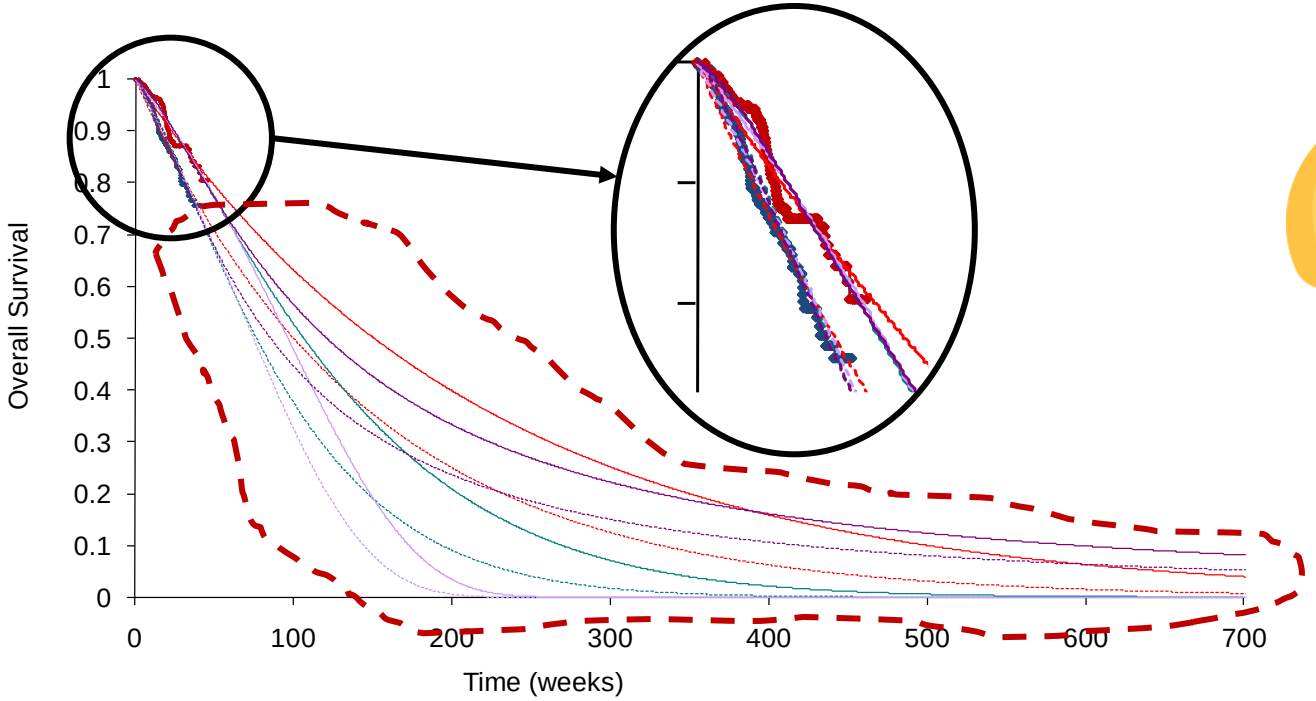
1 Delta Hat Limited, Long Eaton, UK; 2 University of Sheffield, UK; 3 Bristol Myers Squibb, Uxbridge, UK; 4 Bristol Myers Squibb, Lawrenceville, NL, USA

Declaration of interests

- This study was funded by BMS (authors SH, GO, CC, IK, JB, DB report that they are BMS employees and stockholders)
- Delta Hat conducted all analyses independently
- BMS provided helpful insight and comments

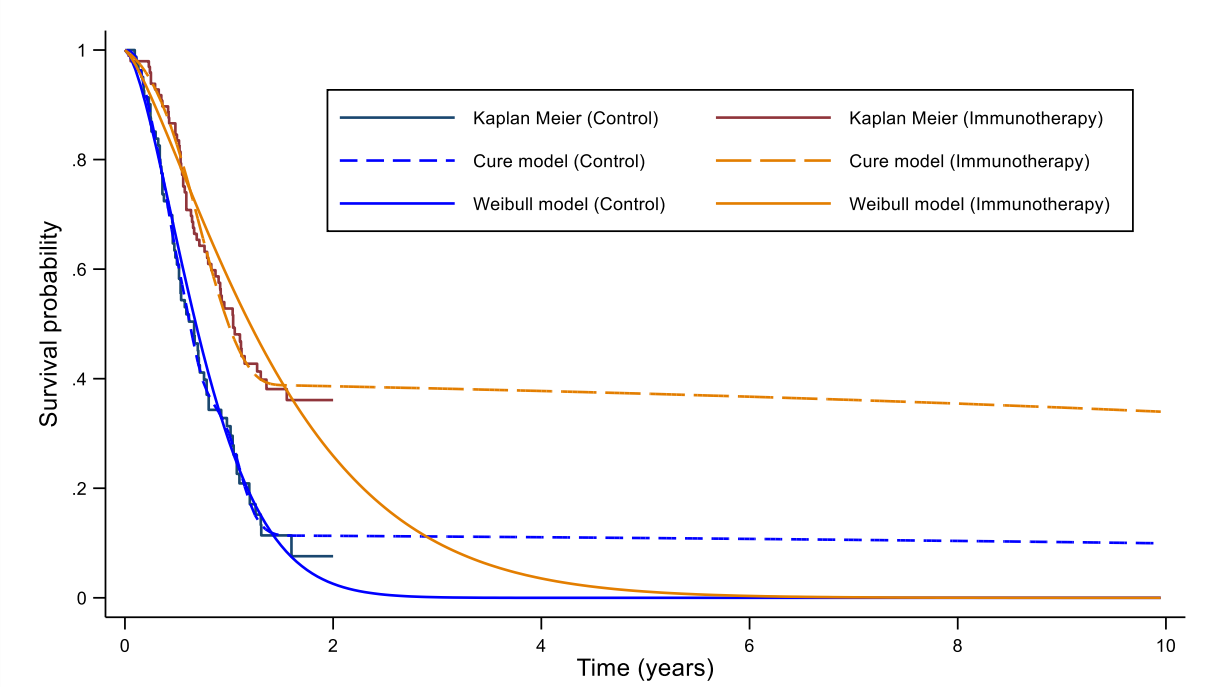
Survival modelling in Health Technology Assessment (HTA)

- Accurately extrapolating survival beyond trial follow-up is essential in HTA
- Different survival models can result in drastically different estimates of treatment benefits



Survival modelling in HTA: Immuno-oncology

- Immuno-oncology is especially affected because survival curves can flatten over time
- “Standard” survival models may not capture this
- “Flexible” survival models (including cure models) may capture the flattening, but are they credible?

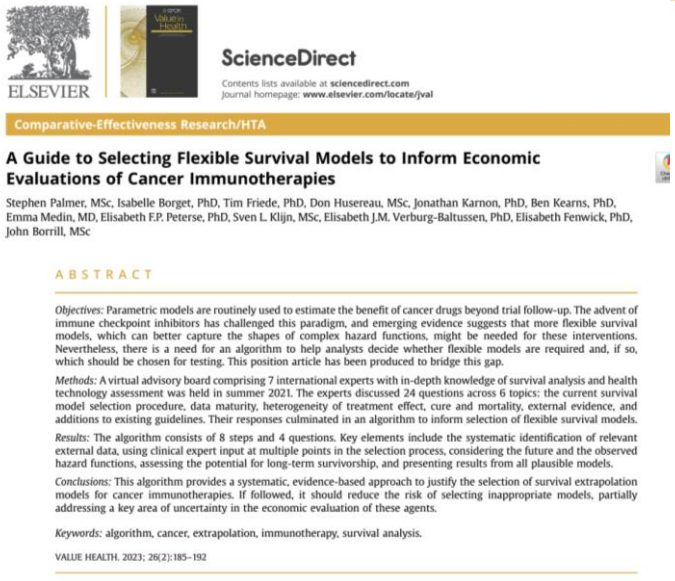


Mean survival (years)	Control	IO	Difference
Weibull model	0.77	1.45	0.67
Cure model	2.93	8.43	5.50

- Different models can result in drastically different estimates of survival benefit
- This has become a crucial discussion point in many appraisals of immuno-oncology treatments

Selecting survival models for immunotherapies

- Palmer *et al*. 2023 present an algorithm for selecting survival models to inform economic evaluations of cancer immunotherapies



ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Comparative-Effectiveness Research/HTA

A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies

Stephen Palmer, MSc, Isabelle Borget, PhD, Tim Friede, PhD, Don Husereau, MSc, Jonathan Karnon, PhD, Ben Kearns, PhD, Emma Medin, MD, Elisabeth F.P. Peterse, PhD, Sven L. Klijn, MSc, Elisabeth J.M. Verburg-Baltussen, PhD, Elisabeth Fenwick, PhD, John Borrill, MSc

ABSTRACT

Objectives: Parametric models are routinely used to estimate the benefit of cancer drugs beyond trial follow-up. The advent of immune checkpoint inhibitors has challenged this paradigm, and emerging evidence suggests that more flexible survival models, which can better capture the shapes of complex hazard functions, might be needed for these interventions. Nevertheless, there is a need for an algorithm to help analysts decide whether flexible models are required and, if so, which should be chosen for testing. This position article has been produced to bridge this gap.

Methods: A virtual advisory board comprising 7 international experts with in-depth knowledge of survival analysis and health technology assessment was held in summer 2021. The experts discussed 24 questions across 6 topics: the current survival model selection procedure, data maturity, heterogeneity of treatment effect, cure and mortality, external evidence, and additions to existing guidelines. Their responses culminated in an algorithm to inform selection of flexible survival models.

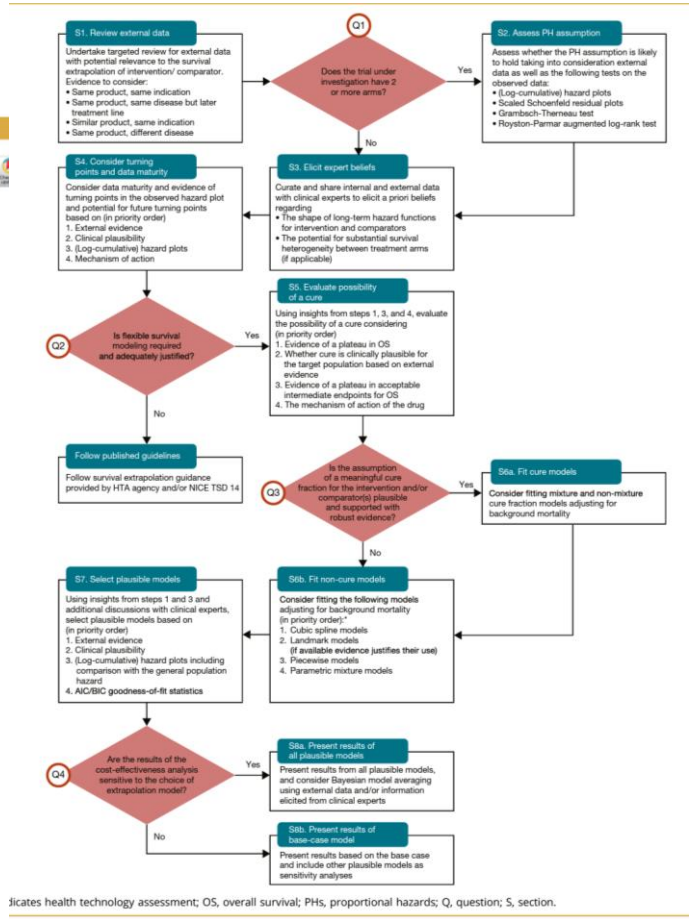
Results: The algorithm consists of 8 steps and 4 questions. Key elements include the systematic identification of relevant external data, using clinical expert input at multiple points in the selection process, considering the future and the observed hazard functions, assessing the potential for long-term survivorship, and presenting results from all plausible models.

Conclusions: This algorithm provides a systematic, evidence-based approach to justify the selection of survival extrapolation models for cancer immunotherapies. If followed, it should reduce the risk of selecting inappropriate models, partially addressing a key area of uncertainty in the economic evaluation of these agents.

Keywords: algorithm, cancer, extrapolation, immunotherapy, survival analysis.

VALUE HEALTH. 2023; 26(2):185–192

Palmer S, Borget I, Friede T, Husereau D, Karnon J, Kearns B, et al. A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies. *Value Health*. 2023 Feb;26(2):185–92



dicates health technology assessment; OS, overall survival; PHs, proportional hazards; Q, question; S, section.

- The algorithm involves 8 steps and 4 questions, involving:

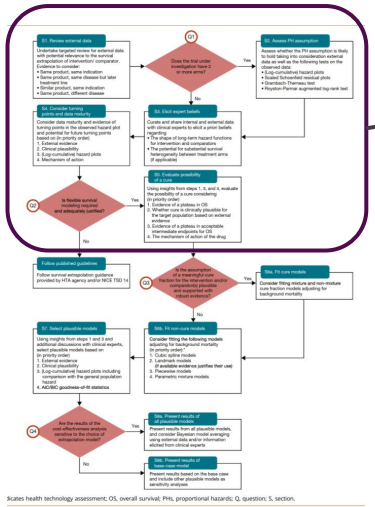
  - Review of external data
  - Assessment of proportional hazards
  - Elicitation of expert beliefs
  - Consideration of turning points in the hazard function and data maturity
  - Evaluation of the possibility of cure
  - Sensitivity analysis

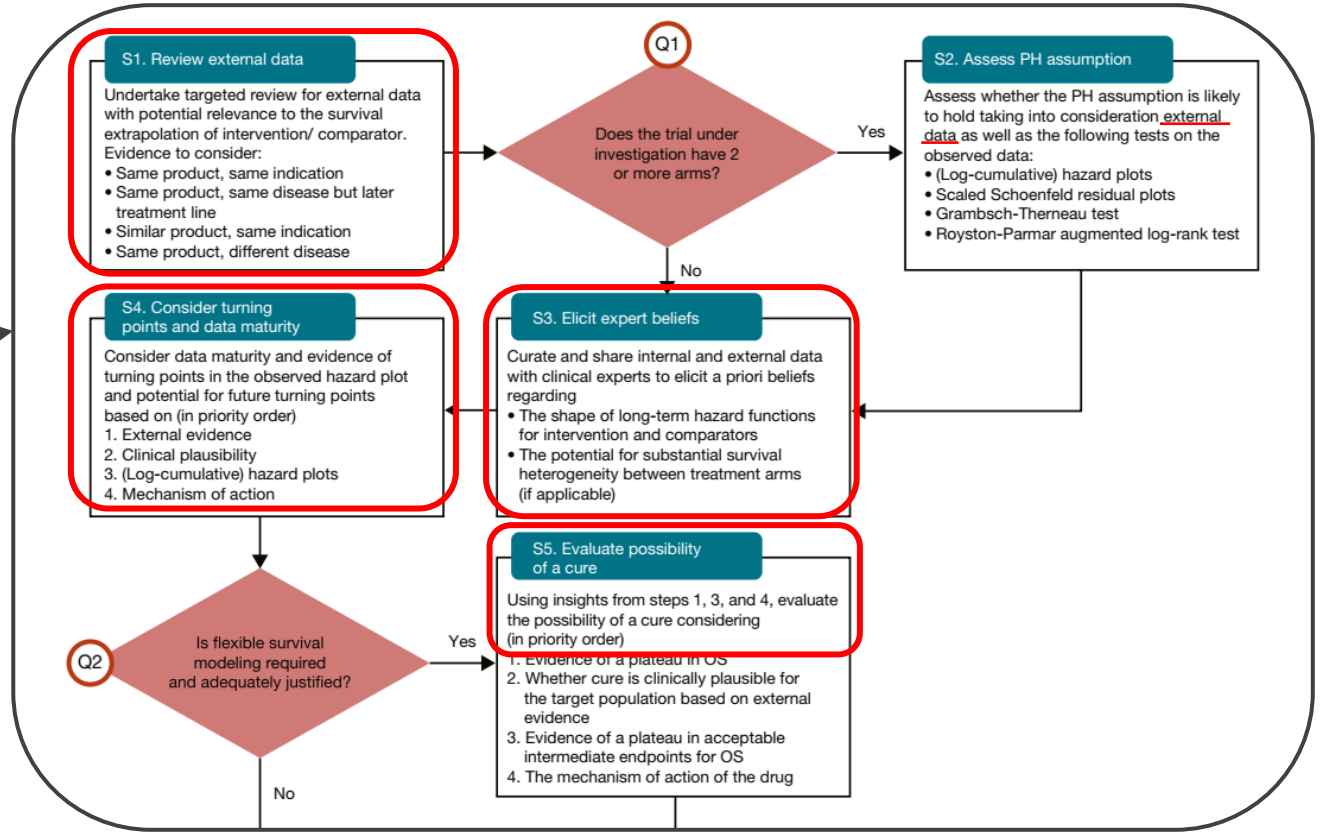
Delta Hat

Selecting survival models for immunotherapies

- Palmer *et al*. 2023 present an algorithm for selecting survival models to inform economic evaluations of cancer immunotherapies

- **A particular novelty of the algorithm is its emphasis on appraising external data / information *before any analyses on the trial in question have been done***



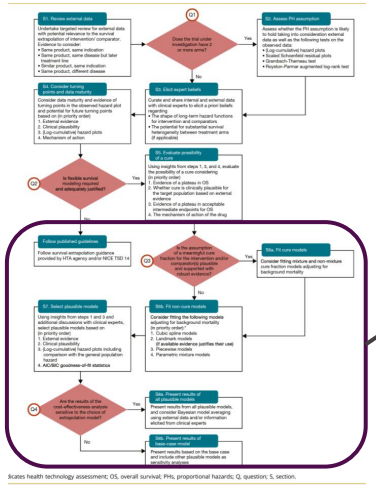


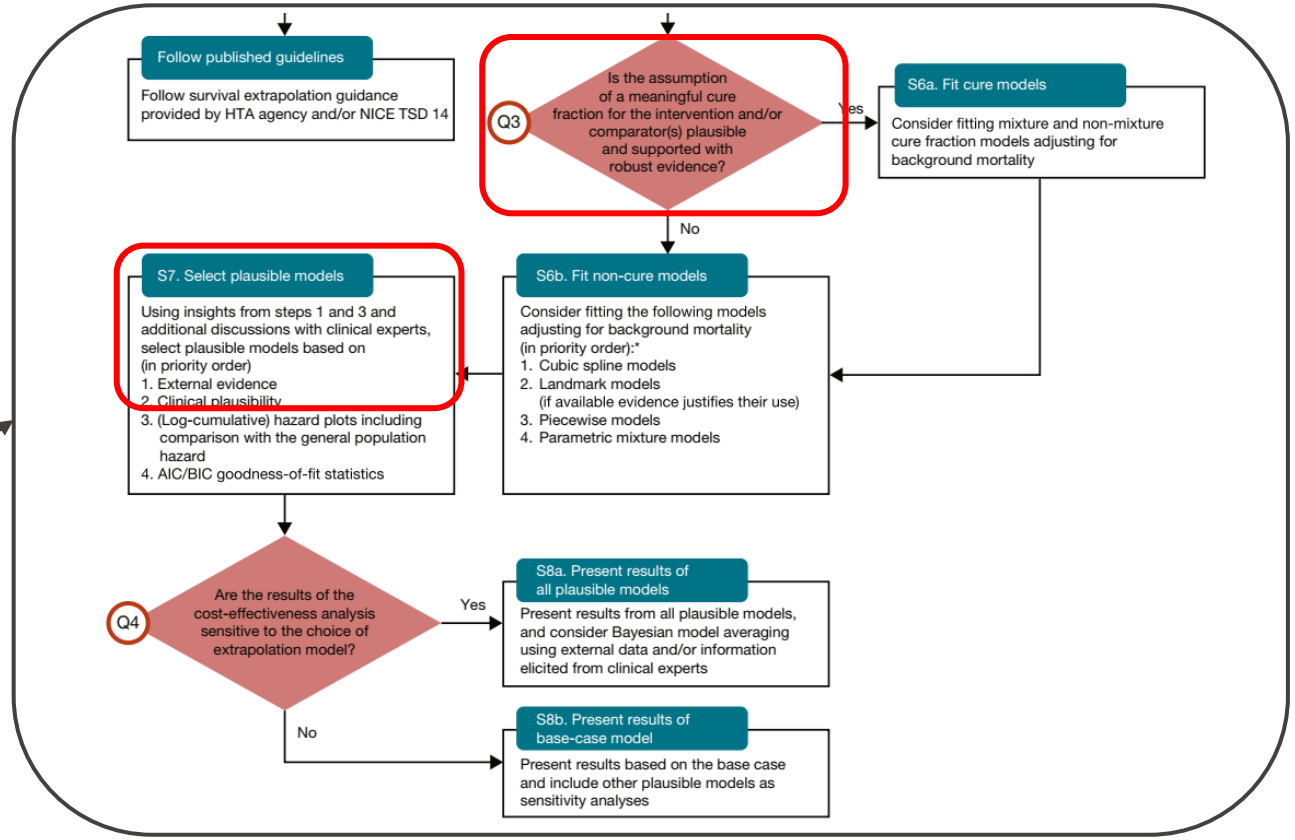
Palmer S, Borget I, Friede T, Husereau D, Karnon J, Kearns B, et al. A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies. Value Health. 2023 Feb;26(2):185-92

Delta Hat

Selecting survival models for immunotherapies

- Palmer *et al*. 2023 present an algorithm for selecting survival models to inform economic evaluations of cancer immunotherapies
- **A particular novelty of the algorithm is its emphasis on appraising external data / information *before any analyses on the trial in question have been done***





Palmer S, Borget I, Friede T, Husereau D, Karnon J, Kearns B, et al. A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies. Value Health. 2023 Feb;26(2):185-92

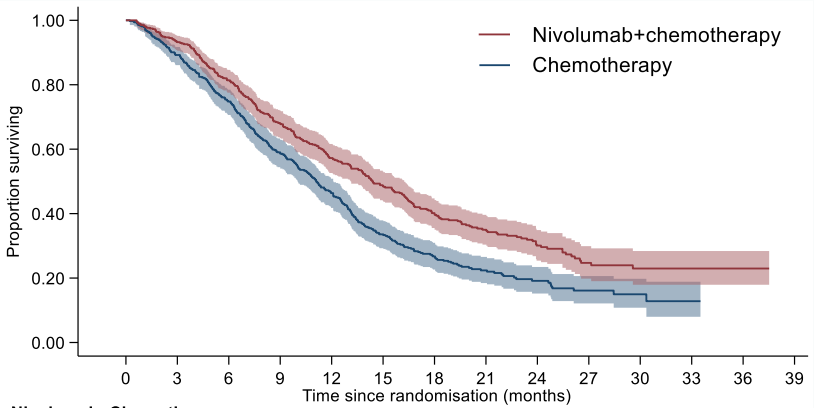
Delta Hat

Testing the Palmer *et al*. algorithm

CheckMate 649 compared first line nivolumab plus chemotherapy versus chemotherapy alone, in patients with advanced gastric, gastro-oesophageal junction, or oesophageal adenocarcinoma

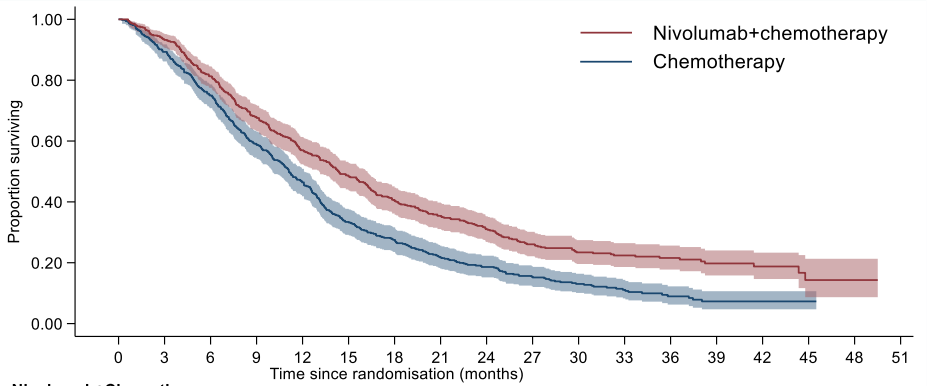
- 12-, 24- and 36-month data-cuts are available
- Pivotal trial in NICE TA 857

12-month data-cut



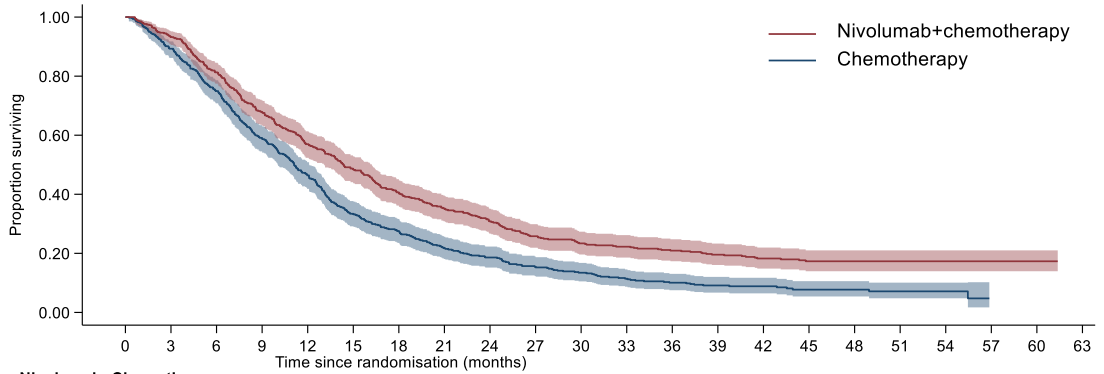
Nivolumab+Chemotherapy														
At-risk	473	438	377	313	261	198	149	96	65	33	22	9	1	0
Censored	0	3	9	11	14	39	55	91	110	133	142	155	163	164
Died	0	32	87	149	198	236	269	286	298	307	309	309	309	309
Chemotherapy														
At-risk	482	421	350	271	211	138	98	56	34	19	8	2	0	0
Censored	0	10	13	19	21	37	50	78	93	103	113	118	120	120
Died	0	51	119	192	250	307	334	348	355	360	361	362	362	362

24-month data-cut



Nivolumab+Chemotherapy																								
At-risk	473	440	380	315	263	223	187	161	141	107	81	61	43	26	19	6	2	0						
Censored	0	1	5	7	9	10	10	12	12	24	40	57	73	87	93	104	108	110						
Died	0	32	88	151	201	240	276	300	320	342	352	355	357	360	361	363	363	363						
Chemotherapy																								
At-risk	482	424	353	275	215	154	125	97	83	62	46	31	18	11	6	1	0	0						
Censored	0	7	10	14	15	16	18	20	20	27	34	43	51	55	60	65	66	66						
Died	0	51	119	193	252	312	339	365	379	393	402	408	413	416	416	416	416	416						

36-month data-cut



Nivolumab+Chemotherapy																												
At-risk	473	440	380	315	263	223	187	161	141	118	105	100	94	81	66	53	37	24	17	6	2	0						
Censored	0	1	5	7	9	10	10	11	11	11	13	13	13	20	30	40	56	69	76	87	91	93						
Died	0	32	88	151	201	240	276	301	321	344	355	360	366	372	377	380	380	380	380	380	380	380						
Chemotherapy																												
At-risk	482	424	353	275	215	154	125	97	83	69	60	51	44	35	28	18	14	10	5	0	0	0						
Censored	0	7	10	14	15	16	18	20	20	20	20	20	21	26	32	39	43	46	51	55	0	55						
Died	0	51	119	193	252	312	339	365	379	393	402	411	417	421	422	425	425	426	426	427	427	427						

NICE: National Institute for Health and Care Excellence; TA: Technology Appraisal

Delta Hat

Testing the Palmer *et al*. algorithm

Methods

- Applied Palmer *et al*. algorithm to 12- and 24-month data-cuts from CheckMate 649
- Selected preferred survival models based on our application of the algorithm
- Compared predictions to data observed in the 36-month data-cut
  (concentrating on the Combined Positive Score (CPS) ≥ 5 subgroup)

Aims

- **Applicability.** How straightforward is the algorithm to apply?
- **Model selection.** Does the algorithm result in survival models that extrapolate survival accurately?
- **Implications for HTA.** What impact might the algorithm have on the HTA process and decision making?
- **Development.** Would the algorithm benefit from any amendments?

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

1. **Review external data** [Evidence from a range of RCTs, real world data from the UK, US, Canada and the Netherlands]
2. **Assess proportional hazards** [Hazard plots from CheckMate 649, external information on treatment effects of IOs]
3. **Elicit expert beliefs** [Informed by all expert beliefs documented in the appraisal documents for NICE TA857]
4. **Consider turning points in the hazard function**
5. **Consider the need for flexible survival modelling**
6. **Evaluate the possibility of a cure**
7. **Consider the plausibility and justifiability of a cure assumption**

Informed by Steps 1-3

8. **Selection of potentially plausible models…**

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

8. Selection of potentially plausible models…

- Based on responses to all previous steps

Findings: Plausibility criteria

Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%
Expectations around hazards	Initial increase followed by decrease. In both treatment groups hazards will converge towards background rates in the long-term (5+ years), but will always remain above background mortality levels. Age-related increase in the very long-term in longest-term survivors.	
Treatment effect waning	Uncertain, but in the small percent who live beyond a few years, hazards will equalise	

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

8. Selection of potentially plausible models…

- Based on responses to all previous steps

Findings: Plausibility criteria

Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%
Expectations around hazards	Initial increase followed by decrease. In both treatment groups hazards will converge towards background rates in the long-term (5+ years), but will always remain above background mortality levels. Age-related increase in the very long-term in longest-term survivors.	
Treatment effect waning	Uncertain, but in the small percent who live beyond a few years, hazards will equalise	

Delta Hat

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

8. Selection of potentially plausible models…

- Based on responses to all previous steps

Findings: Plausibility criteria

Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%
Expectations around hazards	**Initial increase followed by decrease.** **In both treatment groups hazards will converge towards background rates in the long-term (5+ years), but will always remain above background mortality levels.** **Age-related increase in the very long-term in longest-term survivors.**	
Treatment effect waning	Uncertain, but in the small percent who live beyond a few years, hazards will equalise	

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

8. Selection of potentially plausible models…

- Based on responses to all previous steps

Findings: Plausibility criteria

Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%
Expectations around hazards	Initial increase followed by decrease. In both treatment groups hazards will converge towards background rates in the long-term (5+ years), but will always remain above background mortality levels. Age-related increase in the very long-term in longest-term survivors.	
Treatment effect waning	**Uncertain, but in the small percent who live beyond a few years, hazards will equalise**	

Application of the Palmer *et al*. algorithm

Steps 1-8. Determining candidate survival models

8. Selection of potentially plausible models…

- Based on responses to all previous steps

Findings: Plausibility criteria

Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%
Expectations around hazards	Initial increase followed by decrease. In both treatment groups hazards will converge towards background rates in the long-term (5+ years), but will always remain above background mortality levels. Age-related increase in the very long-term in longest-term survivors.	
Treatment effect waning	**Uncertain, but in the small percent who live beyond a few years, hazards will equalise**	

Findings: Candidate models

- Log normal models
- Log-logistic models
- Generalised gamma models
- Flexible parametric models
- Mixture cure models
- Non-mixture cure models

→All models to include background mortality and SMR uplift

→All models to be fit independently to treatment groups

SMR: Standardised Mortality Ratio

Results: Models fit to 12-month data-cut

Compared to Plausibility Criteria

- Only log-logistic and non-mixture cure models with a 15-year cure time-point met our plausibility criteria
- FPMs under-predicted survival at 10 years, and mixture cure models and non-mixture cure models with earlier cure time-points (7-10 years) over-predicted survival compared to our plausibility criteria

Model		Mean survival (50 years)	Survival %			AIC	Cure %
			Year 4	Year 10	Year 20		
12-month data-cut, CPS≥5							
1. Log-logistic, SMR=2.5	Chemo	1.44	6.1%	1.0%	0.1%	829	N/A
	Nivo+chemo	1.99	11.1%	2.0%	0.3%	906	N/A
2. Non-mixture cure (FPM, 4df), SMR=2.5, 15-year boundary knot	Chemo	1.97	7.4%	3.6%	1.8%	829	5.7%
	Nivo+chemo	3.28	15.3%	8.3%	4.1%	909	13.1%

- Log-logistic models predict survival at low end of plausible range
- Non-mixture cure models predict survival at top end of plausible range

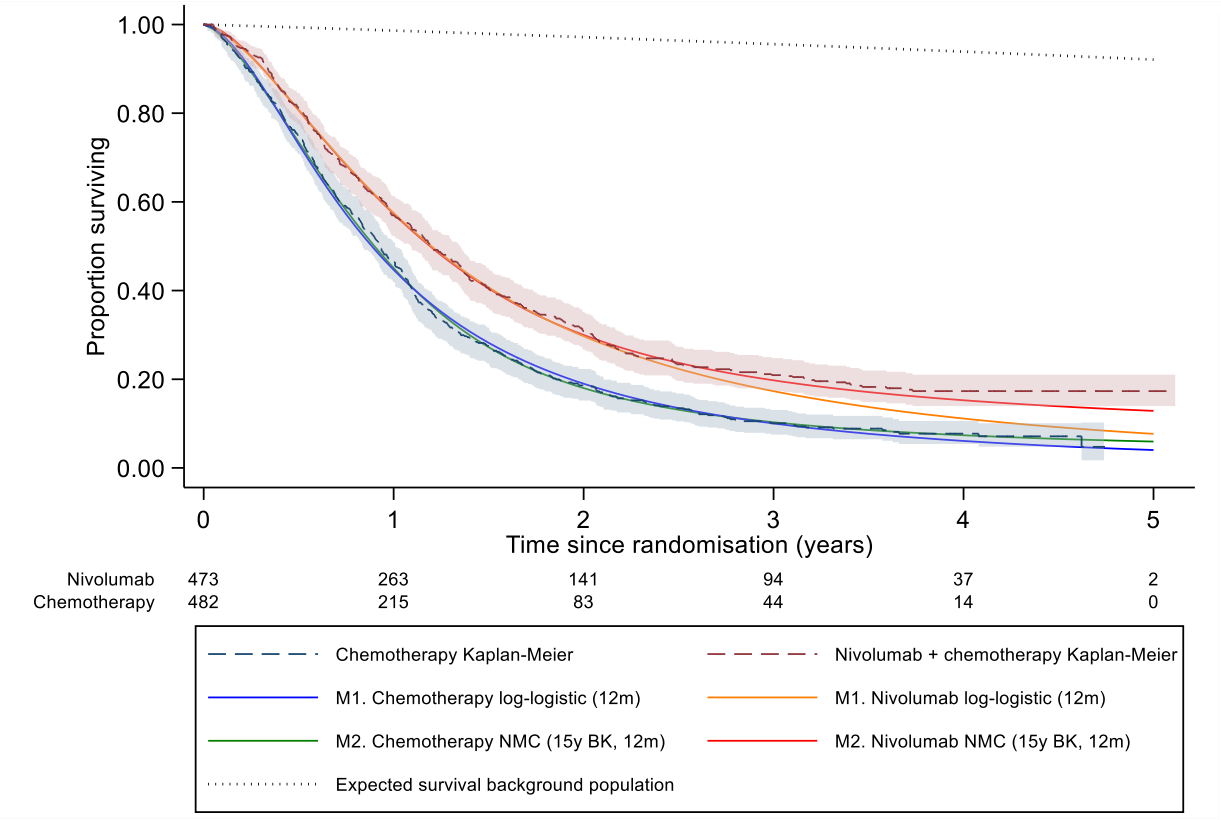
Issue	Chemotherapy	Nivolumab + chemotherapy
Plausible expectations for survival proportions	4 years: 3-7% 10 years: 1-4%	May approximately double survival that is observed with chemotherapy. 20 years: approximately 3%

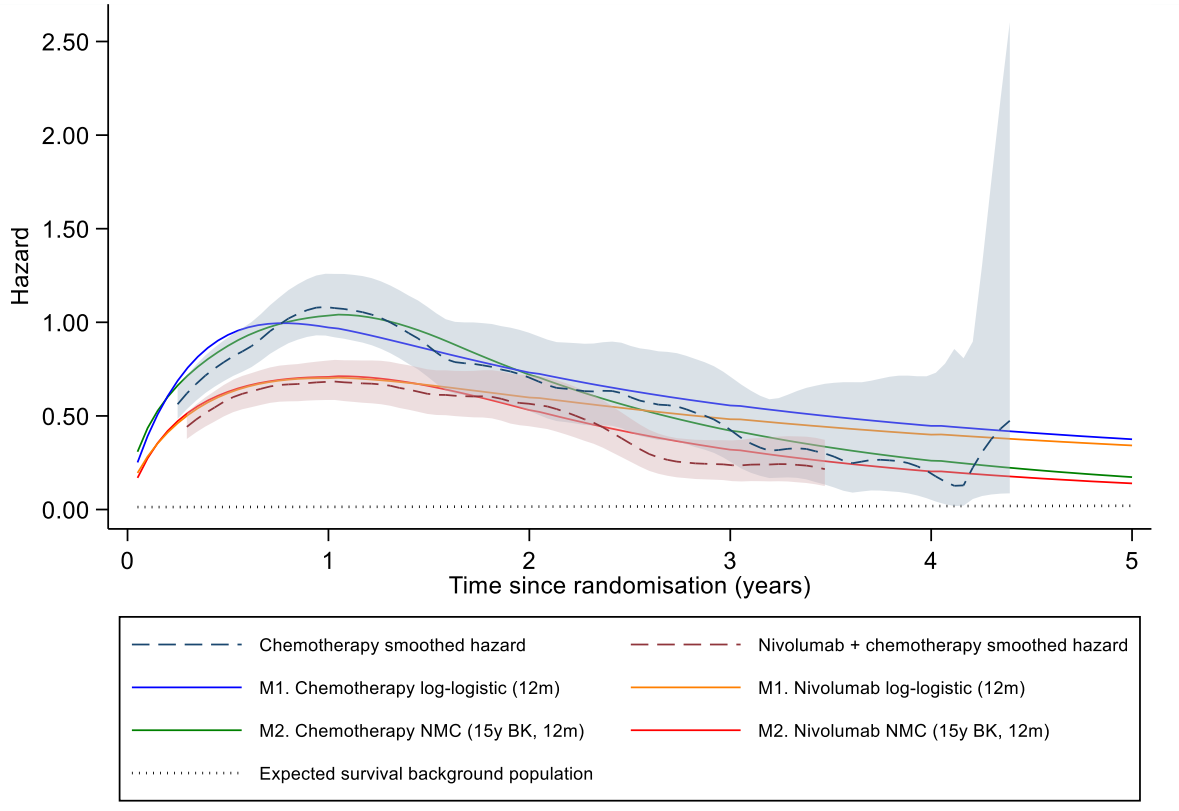
AIC: Akaike Information Criterion; CPS: Combined Positive Score; SMR: Standardised Mortality Ratio; FPM: Flexible Parametric Model; df: degrees of freedom

Results: Models fit to 12-month data-cut

Compared to 36-month data

- Models seem pessimistic: especially the log-logistic models
  - For nivolumab + chemo neither model produced survival predictions within the confidence intervals of the observed data at year 5





NMC: Non-mixture cure; BK: Boundary Knot

Delta Hat

Results: Models fit to 24-month data-cut

- Did not change much!
- The same models produced extrapolations that fell within the plausible range (log-logistic and non-mixture cure)
- FPMs still under-predicted survival and mixture cure models still seemingly over-predicted survival

- The log-logistic models seemed very pessimistic compared to the 36-month data-cut
- The non-mixture cure model seemed pessimistic, but did produce survival predictions that were within the confidence intervals of the observed data at year 5

→ **So similar results to those for models fitted to the 12-month data-cut, just with slightly improved extrapolations (as we would expect)**

FPM: Flexible Parametric Model

Conclusions

Applicability

- The Palmer *et al.* algorithm was simple to use and offered a systematic procedure for model selection

Model selection

- Allowed us to successfully narrow down the list of plausible models
- Resulted in models that provided credible extrapolations (though possibly pessimistic?)

Implications for HTA

- The algorithm “front-loads” the work, before models are fitted
- Should reduce disagreement around model choice; reduced need for additional modelling during appraisals

Development

- The Palmer *et al.* algorithm may benefit from some modifications
- The algorithm does not require that plausibility criteria are explicitly defined
- This provides an additional mechanism to ensure preferred models are selected in an unbiased manner

Thank you