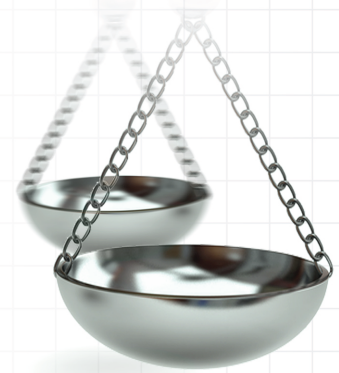


Peter C Wigfield^{*,1} , Bart Heeg¹  & Mario Ouwens²¹Cytel, Weena 316-318, 3012 NJ, Rotterdam, The Netherlands²AstraZeneca, Pepparedsleden 1, 431 50 Mölndal, Sweden*Author for correspondence: Tel.: +31 (0)10 3330 877; peter.wigfield@cytel.com

Aim: National Institute for Health and Care Excellence guidance (Technical Support Document 19) highlights a key challenge of state transition models (STMs) being their difficulty in achieving a satisfactory fit to the observed within-trial endpoints. Fitting poorly to data over the trial period can then have implications for long-term extrapolations. A novel estimation approach is defined in which the predicted overall survival (OS) and progression-free survival (PFS) extrapolations from an STM are optimized to provide closer estimates of the within-trial endpoints. **Materials & methods:** An STM was fitted to the SQUIRE trial data in non-small-cell lung cancer (obtained from Project Data Sphere). Two methods were used: a standard approach whereby the maximum likelihood was utilized for the individual transitions and the best-fitting parametric model selected based on AIC/BIC, and a novel approach in which parameters were optimized by minimizing the area between the STM-predicted OS and PFS curves and the corresponding OS and PFS Kaplan–Meier curves. Sensitivity analyses were conducted to assess uncertainty. **Results:** The novel approach resulted in closer estimations to the OS and PFS Kaplan–Meier for all combinations of parametric distributions analyzed compared with the standard approach. Though the uncertainty associated with the novel approach was slightly larger, it provided better estimates to the restricted mean survival time in 10 of the 12 parametric distributions analyzed. **Conclusion:** A novel approach is defined which provides an alternative STM estimation method enabling improved fits to modeled endpoints, which can easily be extended to more complex model structures.

Plain language summary: A key challenge of state transition models is their difficulty in achieving a satisfactory fit to the observed within-trial endpoints (such as overall survival [OS] and progression-free survival [PFS]). This can have implications for long-term extrapolations and, therefore, for medical decision making as well. A novel approach is defined which provides an alternative estimation method that enables improved fits to modeled endpoints. This approach resulted in closer estimations to the OS and PFS Kaplan–Meier curves for all combinations of parametric distributions analyzed compared with the standard approach. This addresses some of the challenges that are discussed when modeling state transition models, and can easily be extended to more complex model structures.

Tweetable abstract: A key challenge of state transition models is their difficulty in achieving a satisfactory fit to the observed within-trial endpoints, with implications for long-term extrapolations and medical decision making. In this new Methodology article, the authors explore a novel method that helps address some of these challenges.

First draft submitted: 1 March 2023; Accepted for publication: 23 November 2023; Published online: 15 December 2023

Keywords: area between curves • Markov models • optimization • state transition model

Long-term extrapolation is important for health economic modeling to evaluate the long-term cost and benefit of treatment [1–3]. Clinical trial follow-up can sometimes be relatively short, and it is essential that health economic models simulate trial outcomes well to ensure reasonable long-term predictions.

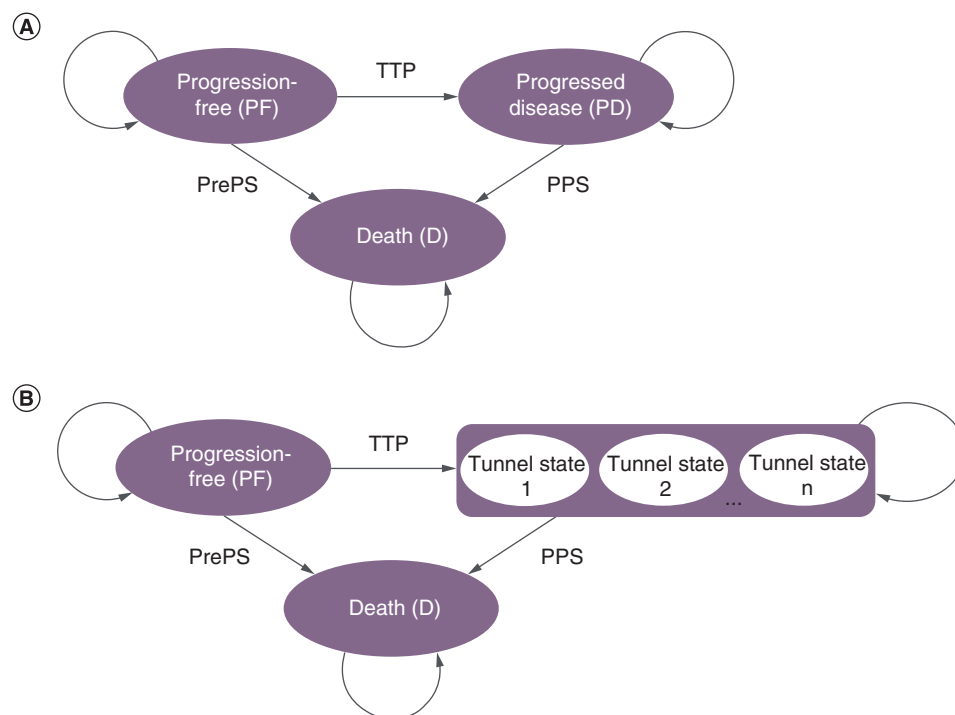


Figure 1. Model structures. (A) Three-state Markov and (B) semi-Markov model structures.
 D: Death; n: Number; PD: Progressed disease; PF: Progression-free; PPS: Post-progression survival; PrePS: Pre-progression survival; TTP: Time to progression.

Partitioned survival analysis (PartSA) is a health economic modeling approach, which is routinely used to inform reimbursement decisions and is outlined in National Institute for Health and Care Excellence Decision Support Unit Technical Support Document (NICE DSU TSD) 19 [2,4]. Within this approach, time to death is partitioned into a series of mutually exclusive health states using trial endpoints (such as overall survival [OS] and progression-free survival [PFS]) directly. Consequently, the post-progression survival (PPS) percentage is defined from randomization as the proportion of patients having progressed but not yet died, and is thus implicitly modeled by PFS and OS.

Compared with PartSA, Markov and semi-Markov models are types of state transition models (STMs) in which the time in each health state instead of the time from randomization is explicitly modeled (including PPS). Three-state Markov (Figure 1A) and semi-Markov (Figure 1B) models usually distinguish the health states progression-free (PF), progressed disease (PD) and death (D). Patients enter the model in PF, and the probability of moving to (or remaining in) health states over time is determined via transition probabilities. From PF, patients can either remain in PF, transition to PD, or transition to D at each timepoint (i.e., model cycle). Parametric functions can be applied to time to progression (TTP), time to death before progression (pre-progression survival or PrePS) and time to death post-progression (PPS) to enable time-dependent transition probabilities.

A semi-Markov model (Figure 1B) differentiates from a Markov model (Figure 1A) in that it also contains one or more transition probabilities which are dependent on the time spent in an intermediate health state (in this case PD), rather than being constant or dependent on calendar time [4]. Figure 1B shows an example whereby this is achieved via the use of tunnel states in PD. Tunnel states are a way of implementing time-dependency by adding 'memory' to a Markov model. This essentially means that the probability of transitioning to the death state (PPS) will differ depending on the time spent in each tunnel state (often called the 'sojourn' time) [5].

NICE DSU TSD 19 highlights a key challenge of STMs being their inability to achieve a satisfactory fit to the observed within-trial OS [4]. This is due to a number of issues. One of which is that OS is no longer determined by a single survival model, as is the case in PartSAs, but by a combination of survival models (e.g., by pre-progression death as well as post-progression death). STMs require time-to-event data on each individual transition probability as well as treatment effect estimates for each. Since most clinical trials only report PFS and OS endpoints, derivation

of time-to-event data required for an STM is not simple. This is because PFS is a combination of progression events and death events from the PF state, and OS is the overall probability of death which is a function of all three transition probabilities. Another challenge with STMs relates to an over-representation of early progressors toward the tail-end of the PPS curve, especially when a naive analysis of PPS is undertaken. This can result in selection bias and lead to an overestimation in the rate of death, causing poor fits to the within-trial OS [6].

The difficulty in achieving a satisfactory fit has also been commented on in at least two prior assessments to NICE (TA632 and TA580), both of which are in oncology and utilized an STM framework [7,8]. At the time of submission, both of these trials reported relatively immature survival (median not reached in either arm). Though the use of immature data is not uncommon in oncology submissions to NICE, the use of an STM approach meant that poor fits to the data over the trial period had potential implications for long-term extrapolations.

Although NICE DSU TSD 14 provides a Survival Model Selection Process Algorithm to guide the fitting of survival models, there is no systematic process to follow when fitting STMs with time-dependent (i.e., time-inhomogeneous) transition probabilities in which endpoints are co-dependent [3]. NICE DSU TSD 21 describes the use of flexible parametric models to implement time-dependent effects, however its application in the context of STMs described here is also limited [1]. Curve fitting in STMs therefore rely on optimizing fits to the individual transitions as described in Williams *et al.* [9].

Instead of assessing fits to the individual transitions independently from one another, this paper focuses on an optimization approach whereby the predicted OS and PFS extrapolations from an STM are optimized to provide closer estimates of the within-trial endpoints. The methodology will be illustrated using a three-state time-inhomogeneous Markov model using data from the SQUIRE trial [10].

Methods

The SQUIRE trial (Thatcher *et al.*) is a multinational, randomized, multicenter, open-label, phase III study [10]. Necitumumab plus gemcitabine and cisplatin (NecGemCis; $n = 545$) is compared with gemcitabine and cisplatin alone (GemCis; $n = 548$) as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (NSCLC). OS (the primary endpoint) was significantly longer in the NecGemCis arm (median 11.5 months; 95% confidence interval [CI] 10.4–12.6) compared with the GemCis arm (median 9.9 months; 95% CI 8.9–11.1), with a stratified hazard ratio of 0.84 (95% CI 0.74–0.96; $p = 0.01$). NecGemCis also resulted in a significant improvement in PFS compared with GemCis (median 5.7 months [95% CI 5.6–6.0] vs 5.5 months [95% CI 4.8–5.6]; stratified hazard ratio 0.85 [95% CI 0.74–0.98; $p = 0.02$]).

Project Data Sphere (an open-access data sharing platform) was used to obtain patient-level data [11]. Due to the assumption of dependency between endpoints, STMs are useful alongside trials that show a significant PFS with a benefit expected to be translated to OS. Considering the significant benefit in PFS as well as OS, the SQUIRE trial was therefore deemed to be an appropriate choice from the options available. Only the control arm (GemCis) was publicly accessible from Project Data Sphere, and therefore assessing whether both treatment arms resulted in clinically implausible outcomes was not possible.

Data from the SQUIRE trial were fitted by a typical illness-death model which contained the ability to apply time-independent or time-dependent transition probabilities, as presented in Figure 2 (top). In the model, the three health states distinguished were PF, PD and D. The corresponding transitions were TTP with hazard $\lambda_1(t)$, PrePS with hazard $\lambda_2(t)$, and PPS with hazard $\lambda_3(t)$.

Following the multi-state approach outlined by Williams *et al.* [12], a parametric survival distribution was assigned to each transition. From the individual transitions, PFS and OS can then be derived (Figure 2 [bottom]). The PFS hazard rate is equal to the sum of the hazard $\lambda_1(t)$ of TTP and the hazard $\lambda_2(t)$ of PrePS:

$$\lambda_{PFS}(t) = \lambda_1(t) + \lambda_2(t)$$

In the case where exponential distributions are applied for all three transitions (i.e., TTP, PrePS, PPS): $TTP \sim \text{Exp}(\lambda_1)$ and $PrePS \sim \text{Exp}(\lambda_2)$ imply that $PFS = \min(TTP, PrePS) \sim \text{Exp}(\lambda_1 + \lambda_2)$. The probability ω that death occurs before progression is then equal to $\omega = \lambda_2/(\lambda_1 + \lambda_2)$ [13].

The hazard for OS(t) is equal to

$$\lambda_{OS}(t) = PF(t) \times \lambda_2(t) + PD(t) \times \lambda_3(t)$$

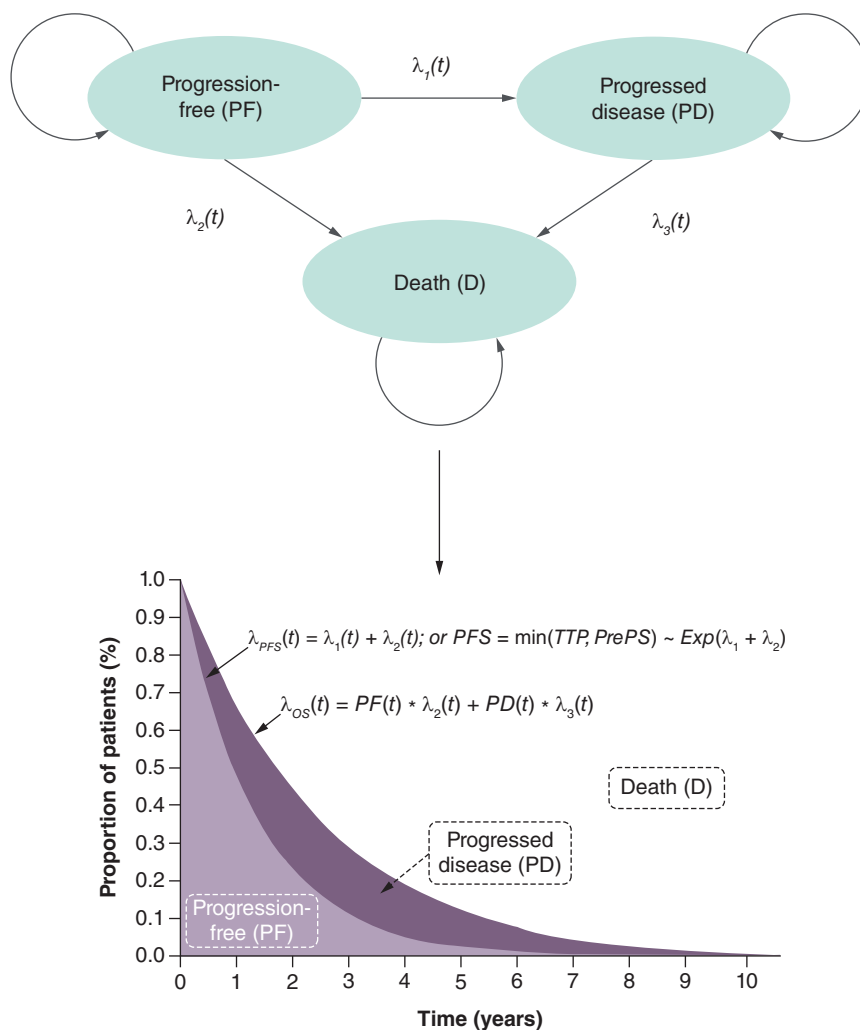


Figure 2. Typical illness-death model (top) and multistate approach (bottom).

D: Death; OS: Overall survival; PD: Progressed disease; PF: Progression-free; PFS: Progression-free survival; PrePS: Pre-progression survival; t: Time; TTP: Time to progression.

where $PF(t)$ and $PD(t)$ represents the percentage of patients in the progression-free and progressed disease states, respectively.

The cumulative distribution function for OS is then comprised as a function of PrePS and PPS, but is also indirectly reliant on TTP [14]:

$$F_{OS}(t) = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2 - \lambda_3} \exp^{-\lambda_3 t} - \frac{\lambda_2 - \lambda_3}{\lambda_1 + \lambda_2 - \lambda_3} \exp^{-(\lambda_1 + \lambda_2)t}$$

where F_{OS} represents the distribution function of OS.

Two approaches were used to estimate the parameters of the survival distributions for the individual transitions. In the first (i.e., 'standard') approach, the maximum likelihood was used and Akaike information criterion (AIC) / Bayesian information criterion (BIC) values were obtained (Supplementary Table 1) to select the best-fitting parametric model for the individual transitions (i.e., for TTP, PrePS and PPS). This is broadly in line with the NICE DSU TSD 14 Survival Model Selection Process Algorithm when complete survival data is available, noting that visual fits and clinical plausibility of the extrapolations were not assessed in this instance [3]. The predicted PFS and OS outcomes from the STM were then derived based on the chosen distributions selected for each of

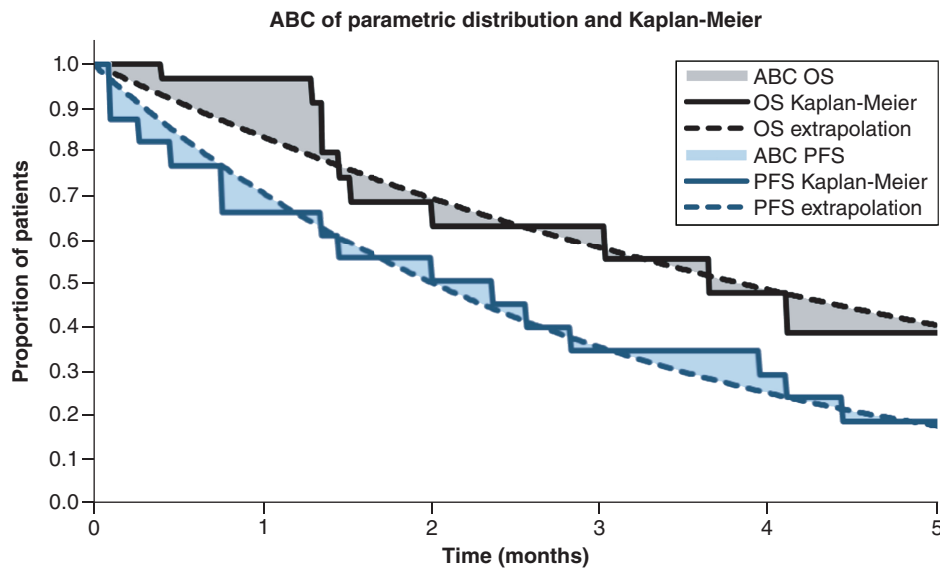


Figure 3. Example: estimation of area between curves (predicted parametric distribution vs Kaplan–Meier).
 NB. This is purely an illustrative example to highlight the approach being used (not showing data used in the analysis).
 ABC: Area between curves; OS: Overall survival; PFS: Progression-free survival.

the individual transitions. This approach is reflective of the methodology commonly being utilized to obtain best model fits in time-dependent STMs [8,15,16].

In the second (i.e., ‘novel’) approach, instead of optimizing over the individual transitions, the area between the curves (ABC) of the estimated PFS and OS distributions from the STM, and the PFS and OS Kaplan–Meier curves was directly optimized rather than relying on the maximum likelihood. Kaplan–Meier curves are non-parametric statistics, which are used to visualize the probability of an event (e.g., death) occurring over time [17]. This was achieved by optimizing the survival parameters of the individual transitions (e.g., TTP Weibull: shape ($\lambda \in (0, +\infty)$), scale ($\kappa \in (0, +\infty)$)), using the PFS and OS Kaplan–Meier curves as optimization criteria. The ABC was restricted until the end of trial follow-up, and provides an estimate of the absolute difference in restricted mean survival time (RMST) between the Kaplan–Meier and the predicted extrapolation in months. Compared with the standard approach, this approach treats all prediction error as equally valued throughout the trial period (e.g., a 2% prediction error of survival at 12 months would be valued just as much as a 2% prediction error of survival at 24 months).

Figure 3 shows a visual example of how the ABC of the Kaplan–Meier and predicted distributions were obtained. Note that this serves as merely an illustration of the approach, and not the actual data being used for the analyses. The method functions by minimizing the absolute difference between the Kaplan–Meier and predicted extrapolations obtained from the STM for OS and PFS, by simultaneously optimizing TTP, PrePS and PPS. In both approaches, the ABCs were obtained so that a comparison between the methods could be made.

Sensitivity analyses were conducted to characterize uncertainty between the two approaches. In line with Williams *et al.* (2017) [12], the standard approach involved varying coefficients of the parametric distributions over each of the individual transitions using 1000 Monte Carlo iterations. For the novel approach, since parametric uncertainty could not be obtained from the already optimized parameters, bootstrapping of the OS and PFS time-to-event data was performed. After each bootstrap, the survival parameters were optimized to the resampled Kaplan–Meier estimates for each iteration, and the results recorded (1000 iterations).

Six commonly used parametric distributions (Weibull, exponential, log-normal, log-logistic, Gompertz, generalized gamma) were used [3]. Though it is possible to apply different distributions for each of the three transitions (e.g., Weibull TTP, exponential PrePS, log-normal PPS), the same distributions were tested for all transitions (e.g., Weibull TTP/PrePS/PPS) to limit the possible combinations of distributions that were assessed.

The STM was developed in Microsoft Excel, and the Solver Add-in in combination with Visual Basics for Applications was utilized to perform the optimization approach. Survival analyses using the patient-level data were undertaken in R.

Table 1. Overview of ABC values (standard approach vs novel approach).

Parametric distribution (TTP, PrePS and PPS)	ABC _{TTP} (months)		ABC _{PrePS} (months)		ABC _{PPS} (months)		ABC _{PFS} (months)		ABC _{OS} (months)	
	Standard approach	Novel approach	Standard approach	Novel approach	Standard approach	Novel approach	Standard approach	Novel approach	Standard approach	Novel approach
log-normal	1.46	2.45	4.49	6.50	0.67	0.98	0.80	0.57	0.30	0.26
Weibull	2.25	2.59	1.49	1.35	0.63	2.11	1.04	0.90	1.10	0.26
exponential	1.68	2.08	3.18	2.33	0.55	1.95	1.37	1.04	0.69	0.56
log-logistic	1.52	2.29	3.05	6.01	0.55	0.96	0.59	0.49	0.57	0.29
Gompertz	1.81	3.71	0.81	11.31	0.52	0.91	1.27	1.03	0.89	0.27
generalized gamma	0.92	2.61	0.84	7.17	0.32	1.17	0.81	0.40	0.99	0.24

ABC: Area between curves; OS: Overall survival; PFS: Progression-free survival; PPS: Post-progression survival; PrePS: Pre-progression survival; TTP: Time to progression.

Table 2. Sensitivity analyses (1000 Monte Carlo simulations).

Endpoint	Distribution	Standard approach			Novel approach		
		Restricted mean LY (95% CI)	% deviation from RMST		Restricted mean LY (95% CI)	% deviation from RMST	
PFS	Log-normal	0.55	(0.51–0.59)	10%	0.50	(0.37–0.62)	0%
	Weibull	0.54	(0.51–0.58)	8%	0.45	(0.31–0.57)	-10%
	Exponential	0.55	(0.50–0.59)	10%	0.48	(0.37–0.65)	-4%
	Log-logistic	0.54	(0.50–0.58)	8%	0.51	(0.41–0.62)	2%
	Gompertz	0.55	(0.48–0.64)	10%	0.48	(0.35–0.65)	-4%
	Generalized gamma	0.56	(0.51–0.65)	12%	0.51	(0.39–0.66)	2%
OS	Log-normal	1.00	(0.93–1.07)	-1%	1.00	(0.82–1.18)	-1%
	Weibull	0.97	(0.91–1.03)	-4%	1.03	(0.91–1.17)	2%
	Exponential	0.97	(0.90–1.04)	-4%	0.99	(0.88–1.15)	-2%
	Log-logistic	1.00	(0.93–1.06)	-1%	0.99	(0.83–1.12)	-2%
	Gompertz	0.98	(0.86–1.14)	-3%	0.98	(0.85–1.11)	-3%
	Generalized gamma	0.99	(0.90–1.15)	-2%	0.99	(0.84–1.13)	-2%

CI: Confidence interval; LY: Life years; OS: Overall survival; PFS: Progression-free survival; RMST: Restricted mean survival time.

Results

Table 1 presents an overview of the ABC values that were obtained using the standard approach versus the novel approach for all tested parametric distributions (note that a lower ABC indicates a better fit of the tested parametric distribution to the corresponding Kaplan–Meier data). When comparing the ABC for PFS (i.e., ABC_{PFS}) and OS (i.e., ABC_{OS}), the novel approach resulted, as expected, in closer estimations to the Kaplan–Meier for all combinations of distributions analyzed compared with the standard approach. In the standard approach, the log-logistic showed the lowest ABC_{PFS} (0.59 months), while the log-normal showed the lowest ABC_{OS} (0.30 months). This is in contrast to the novel approach, in which the generalized gamma presented both the lowest ABC_{PFS} (0.40 months) and ABC_{OS} (0.24 months).

An overview of the parametric distributions estimated over PFS and OS using the standard approach compared with the novel approach is provided in Figure 4. Survival curves for all included parametric distributions are plotted, together with the associated ABC_{PFS} and ABC_{OS} between the predictions and observed Kaplan–Meier curves. From the figure, it is clear that utilizing the novel approach provides better visual fits to the underlying data (especially for OS) compared with use of the standard approach across the entire trial follow-up.

Although the novel approach resulted in closer predictions to the PFS and OS Kaplan–Meiers, this was not the case for TTP, PrePS and PPS. For TTP and PPS, the standard approach resulted in lower ABCs for all parametric distributions tested compared with the novel approach. While for PrePS, only the Weibull distribution (standard approach 1.49 months, novel approach 1.35 months) and the exponential distribution (standard approach 3.18 months, novel approach 2.33 months) provided better fits to the data when using the novel method. Parametric distributions plotted for all endpoints are shown in the Supplementary Material (Supplementary Figure 1 for the standard approach and Supplementary Figure 2 for the novel approach). Sensitivity analyses were conducted using Monte Carlo simulations. The RMST over the trial follow-up was 0.50 years for PFS, and 1.01 years for OS. Table 2 shows the restricted mean life years for both approaches. In all but two cases (PFS Weibull, OS log-logistic) did the

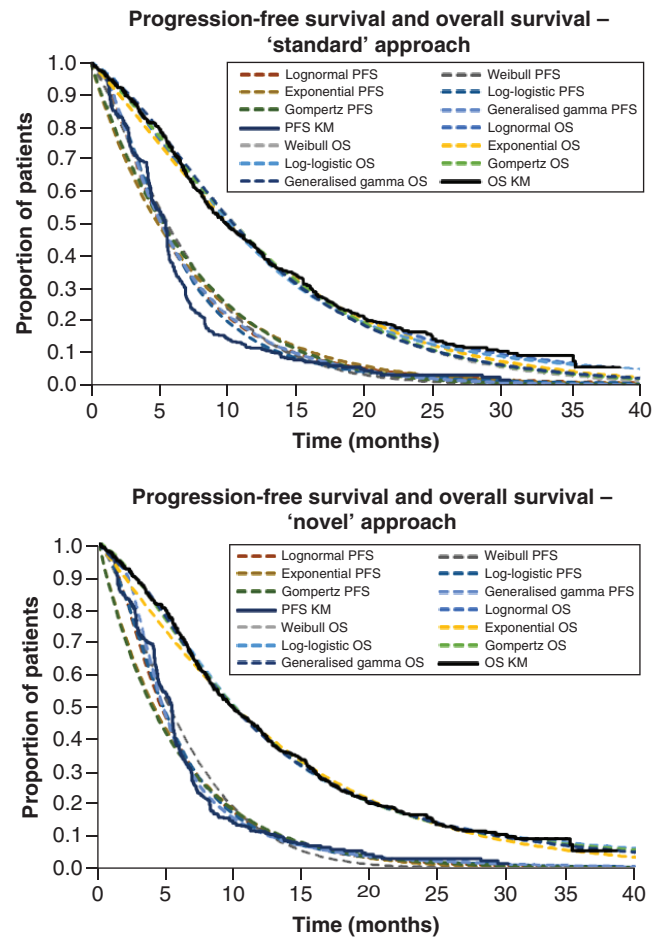


Figure 4. Parametric distributions for overall survival and progression-free survival (standard vs novel approach).

KM: Kaplan–Meier; OS: Overall survival; PFS: Progression-free survival.

novel approach result in closer point estimates to the RMST compared with the standard approach (i.e., 10 of 12 distributions). The standard approach included the PFS RMST within the 95% CI for only the log-logistic and Gompertz distributions, while the novel approach included the PFS RMST within the 95% CI for all parametric distributions. The uncertainty associated with the novel approach, however, was slightly larger than the standard approach.

Discussion

This paper proposes a novel optimization approach in which the predicted OS and PFS extrapolations from an STM are optimized to provide closer estimates to the within-trial endpoints. This method optimizes the average ABC of the estimated OS and PFS distributions and their respective Kaplan–Meier curves to minimize the prediction error, rather than relying on the maximum likelihood. This methodology is applied to a case study in NSCLC utilizing a three-state time-inhomogeneous Markov model.

As expected due to the choice of optimization criterion, the novel approach resulted in closer estimations (in terms of ABC) to OS and PFS compared with the standard approach for all parametric distributions tested. Meanwhile, the predictions for TTP, PrePS and PPS were oftentimes worse when using the novel method. This is to be anticipated based on how the analyses were conducted. In the example shown, the parametric distributions for TTP, PrePS and PPS were all optimized to ensure the best fits to OS and PFS alone, without applying any weights to the other endpoints. In reality, it is important that all measured endpoints are accurately modeled. For future research, varying the distribution of weights for these endpoints based on importance is required so that a middle-ground can be established. This would then result in better fits to the multistate transitions, but at the expense of fitting less well to the OS and PFS Kaplan–Meiers. It should also be noted that finetuning of the modeling by, for example, evaluating whether a three-state model is appropriate, and whether different combinations of survival distributions would fit better, are some aspects that would improve the estimations.

The key assumption utilized in the novel approach is that the Kaplan–Meier estimate is effectively assumed to be the objective truth, and that by consequence, parametric models are fitted to enable as close estimates as possible. Since Kaplan–Meier is a non-parametric estimator and relies on data from a finite sample of patients, it can therefore be greatly affected by small numbers of patients at risk, as well as censoring. Compared with using the maximum likelihood, this approach effectively treats all prediction error as equally valued. This is an important limitation, and for future research this method needs to adequately take this into account by defining a minimal number of at-risk patients for its use.

Bootstrapping of the Kaplan–Meier curves for each transition (i.e., TTP, PrePS and PPS) was required for the novel approach in order for uncertainty to be assessed, since it could not be obtained from the already optimized parameters. This is in contrast to the standard approach whereby uncertainty could be provided through sampling of random values using a Cholesky decomposition. In sensitivity analyses, the uncertainty associated with the novel approach was higher than the standard approach, however use of the maximum likelihood was not able to capture the PFS RMST within the 95% CIs in four out of the six distributions. It is therefore not surprising that uncertainty was greater with the novel approach, since the underlying data used to produce model fits were changed. Though the novel method improves the fits to OS and PFS, it requires bootstrapping and subsequent optimizations which in itself can take a considerable amount of time to conduct. This is a crucial limitation (particularly with NICE's movement toward a probabilistic base case) [18], and future research is planned in order to run these analyses in statistical packages which should help to reduce the computation time. A comparison between the two methods in terms of uncertainty should therefore be interpreted with caution.

Though this paper has outlined an approach using a three-state STM, it is reasonable to assume that this method could also be extended to more complex models with multiple health states as well as different model structures. A key strength of this approach is that it essentially does not require direct estimation of PPS in order to produce closer estimations to OS. This is a key area of debate within STMs, and a likely reason why these models are unable to achieve a satisfactory fit to the observed within-trial endpoints.

An important limitation of this research is that it was only conducted using data from one treatment arm (GemCis) from one clinical trial (SQUIRE) [10]. This was primarily due to the limited options available from the dataset being used. Validation is therefore required to test whether similar outcomes will be shown in different trials, utilizing different treatment arms, and with different levels of data maturity.

In conclusion, this paper outlines an alternative STM estimation approach enabling improved fits to OS and PFS and potentially addressing some of the challenges around PPS estimation discussed in NICE DSU TSD 19 [4].

Summary points

- Long-term extrapolation is important for health economic modeling to evaluate the long-term cost and benefit of treatment.
- Markov and semi-Markov models are types of state transition models (STMs) which enable transitions between health states (including post-progression survival) to be explicitly characterized.
- National Institute for Health and Care Excellence guidance (Technical Support Document 19 or NICE TSD 19) highlights a key challenge of STMs being their difficulty in achieving a satisfactory fit to the observed within-trial endpoints.
- Instead of assessing fits to the individual health state transitions independently from one another, this paper focuses on an optimization approach whereby the predicted overall survival (OS) and progression-free survival (PFS) extrapolations from the STM are optimized to provide closer estimates of the within-trial endpoints.
- The methodology is illustrated using a three-state time-inhomogeneous Markov model using data from the SQUIRE trial (data obtained from Project Data Sphere; an open-access data sharing platform).
- The novel approach resulted in closer estimations to the OS and PFS Kaplan–Meier for all combinations of parametric distributions analyzed compared with the standard approach.
- Though the uncertainty associated with the novel approach was slightly larger, it provided better estimates to the restricted mean survival time in 10 of the 12 parametric distributions analyzed.
- A novel approach is defined which provides an alternative STM estimation method enabling improved fits to modeled endpoints, which can easily be extended to more complex model structures.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://bpl-prod.literatumonline.com/doi/10.57264/cer-2023-0031>

Author contributions

PC Wigfield, B Heeg and M Ouwers were involved in the methods conception and design. PCW was responsible for data analyses. All authors were involved in drafting and revising the manuscript.

Financial disclosure

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Competing interests disclosure

PC Wigfield and B Heeg work at Cytel Inc. as external consultants for numerous pharmaceutical companies outside the submitted work. M Ouwers is an employee of AstraZeneca. The authors have no other competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Data sharing statement

Patient-level data used in this analysis was obtained from Project Data Sphere; an open-access data sharing platform.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Rutherford MJ, Lambert PC, Sweeting MJ *et al.* NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis (2020). <https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress>.
2. Woods BS, Sideris E, Palmer S, Latimer N, Soares M. Partitioned survival and state transition models for healthcare decision making in oncology: where are we now? *Value Health* 23(12), 1613–1621 (2020).
3. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med. Decis. Making* 33(6), 743–754 (2013).
- **Highlights the importance of long-term extrapolation of clinical trial data for use in healthcare decision making, and the need for a systematic approach to reduce the possibility of bias in cost-effectiveness results.**
4. Woods BS, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19: Partitioned Survival Analysis For Decision modeling In Health Care: A Critical Review (2017). <https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress>
- **Highlights the key challenge of state transition models with regards to their inability to achieve a satisfactory fit to the observed within-trial overall survival, and the need for an alternative approach to help solve this problem.**
5. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Tech. Assess.* 3(10), 1–152 (1999).
6. García-Albéniz X, Maurel J, Hernán MA. Why post-progression survival and post-relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. *Int. J. Cancer* 136(10), 2444–2447 (2015).
- **Highlights the issues of selection bias in relation to modeling of post-progression survival, which impacts the modeling of overall survival endpoints in state transition models.**
7. National Institute for Health and Care Excellence (NICE). Trastuzumab emtansine for adjuvant treatment of HER2-positive early breast cancer [TA632] (2020). www.nice.org.uk/guidance/ta632/evidence
- **Provides a good visualisation of the issue; i.e., difficulty in achieving a satisfactory fit to within-trial endpoints when using a state transition model.**
8. National Institute for Health and Care Excellence (NICE). Enzalutamide for hormone-relapsed non-metastatic prostate cancer [TA580] (2019). www.nice.org.uk/guidance/ta580/evidence
- **Compares use cases of multi-state models and partitioned survival models, and the differences in assumptions required.**
9. Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. *Med. Decis. Making* 37(4), 427–439 (2017).

10. Thatcher N, Hirsch FR, Luft AV *et al.* Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. *Lancet Oncol.* 16(7), 763–774 (2015).
11. Green AK, Reeder-Hayes KE, Corty RW *et al.* The Project Data Sphere Initiative: accelerating cancer research by sharing data. *Oncologist* 20(5), 464–e20 (2015). <https://data.projectdatasphere.org/projectdatasphere/html/home>
- **Provides a framework for the modeling of a multi-state model, and is compared in the paper as a standard approach to compare against.**
12. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness analysis in R using a multi-state modeling survival analysis framework: a tutorial. *Med. Decis. Making* 37(4), 340–352 (2017).
13. Xia F, George SL, Wang X. A multi-state model for designing clinical trials for testing overall survival allowing for crossover after progression. *Stat. Biopharm. Res.* 8(1), 12–21 (2016).
14. Fleischer F, Gaschler-Markefski B, Bluhmki E. A statistical model for the dependence between progression-free survival and overall survival. *Stat. Med.* 28(21), 2669–2686 (2009).
15. National Institute for Health and Care Excellence (NICE). Durvalumab for maintenance treatment of unresectable non-small-cell lung cancer after platinum-based chemoradiation [TA578] (2019). www.nice.org.uk/guidance/ta578/evidence.
16. National Institute for Health and Care Excellence (NICE). Blinatumomab for treating acute lymphoblastic leukaemia in remission with minimal residual disease activity [TA589] (2019). www.nice.org.uk/guidance/ta589/evidence
17. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53(282), 457 (1958).
18. Thom H. Deterministic and probabilistic analysis of a simple Markov model: how different could they be? *App. Health Econ. Health Pol.* 20(3), 447–449 (2022).