Evaluating Partitioned Survival and Markov Decision-Analytic Modeling Approaches for Use in Cost-Effectiveness Analysis: Estimating and Comparing Survival Outcomes

Caitlin Smare1 · Khalid Lakhdari2 · Justin Doan3 · John Posnett1 · Sukhvinder Johal1

Published online: 19 November 2019 © The Author(s) 2019

Abstract

Objective The objective of this study was to assess long-term survival outcomes for nivolumab and everolimus in renal cell carcinoma predicted by three model structures, a partitioned survival model (PSM) and two variations of a semi-Markov model (SMM), for use in cost-effectiveness analyses.

Methods Three economic model structures were developed and populated using parametric curves fitted to patient-level data from the CheckMate 025 trial. Models consisted of three health states: progression-free, progressed disease, and death. The PSM estimated state occupancy using an area under-the-curve approach from overall survival (OS) and progression-free survival (PFS) curves. The SMMs derived transition probabilities to calculate patient flow between health states. One SMM assumed that post-progression survival (PPS) was independent of PFS duration (PPS Markov); the second SMM assumed differences in PPS based on PFS duration (PPS–PFS Markov).

Results All models provide a reasonable fit to the observed OS data at 2 years. For estimating cost effectiveness, however, a more relevant comparison is between estimates of OS over the modeling horizon, because this will likely impact differences in costs and quality-adjusted life-years. Estimates of the incremental mean survival benefit of nivolumab versus everolimus over 20 years were 6.6 months (PSM), 7.6 months (PPS Markov), and 7.4 months (PPS–PFS Markov), reflecting non-trivial differences of +14% and +11%, respectively, compared with PSM.

Conclusions The evidence from this study and previous work highlights the importance of the assumptions underlying any model structure, and the need to validate assumptions regarding survival and the application of treatment effects against what is known about the characteristics of the disease.

1 Introduction

Decision models in healthcare are typically structured around a series of health states descriptive of the disease and characterized by associated costs and health-related quality of life (HRQoL). The expected costs and HRQoL of alternative treatments can be compared by estimating the amount of time spent in each health state over the timeframe of the analysis. In state-transition models (STMs; such as Markov or semi-Markov models [SMMs]), state occupancy is estimated by applying a set of transitions that define the probability that an individual currently in a health state (i) will move to another health state (j) in the following time period. A simple three-state model commonly applied in oncology modeling—progression-free (PF), progressed disease (PD), and death—has three sets of transitions: from PF to PD, from PF to death, and from PD to death (Fig. 1). Transition probabilities are conditional on the starting health state and can be time- and/or treatment-dependent. In a partitioned
survival model (PSM), health state occupancy is estimated directly from the area under the survival curve. In the three-health state example, overall survival (OS) is partitioned into those alive and PF (i.e., progression-free survival [PFS]) and those alive and with PD. The proportion alive at time $t$ is given by the area under the OS curve, and the proportion alive and PF is given by the area under the PFS curve. The proportion alive and with PD is given by the difference between the OS and PFS curves. In this framework there are two implicit transitions: from PF to PD or death, and from PD to death. An inherent assumption of these model structures is that the disease progression is irreversible: patients can move from the PF to the PD health state but they cannot return to the PF health state.

In a recent review of the technology appraisals in oncology conducted by the National Institute for Health and Care Excellence (NICE) between May 2013 and February 2016, 23 of 30 manufacturers' submissions presented a PSM rather than a Markov or semi-Markov approach [1]. Regulatory approval for new oncology interventions is increasingly being granted on the basis of evidence from early-phase, single-arm trials with limited follow-up [2], and, against this background, the PSM may be seen as a relatively straightforward and intuitive approach because state occupancy can be estimated directly from trial-based estimates of OS and PFS. Individual patient data are usually available to estimate survival functions for the target intervention, but not for comparators outside of the clinical trial. An advantage of the PSM approach is that OS and PFS curves can be constructed from Kaplan–Meier curves or from a synthesis of published evidence. However, there are some perceived limitations to this approach [3].

In a within-trial analysis, the two approaches are expected to produce similar results, and both should provide a good fit to the trial data. However, the two approaches employ intrinsically different assumptions that may affect extrapolations of survival. For example, the PSM assumes independence between survival functions: in the three-state example, OS and PFS curves would be extrapolated from the trial data independently. However, depending on the disease, it is possible that the probability of PD is a function of the length of time spent in the PF state, and mortality risk after progression is a function of the time in PD as well as the time spent in PF. Extrapolating PFS and OS independently could also result in the analysis yielding implausible results. For example, the OS and PFS curves could cross, or the extrapolations could produce unrealistic estimated hazards (e.g., the risk of an event in the OS curve being lower than the risk of an event in the PFS curve). In an STM, survival functions are explicitly linked, and extrapolations can be informed by an interaction between underlying disease processes. The extent to which these structural differences affect the ability of either approach to inform decisions about cost effectiveness is uncertain.

In this analysis we have used individual patient data from a randomized clinical trial of patients with advanced, metastatic renal cell carcinoma (RCC; CheckMate 025) [4] to compare survival outcomes predicted by a PSM and two different STMs. In the PSM approach, parametric survival curves were fitted directly to PFS and OS trial data and extrapolated to a lifetime horizon. The proportion of patients in each health state (PF, PD, and death) at time $t$ was estimated from the areas under the PFS and OS curves. The first STM is described as a post-progression survival (PPS) Markov. In this approach, parametric curves were fitted to PFS, time to progression (TTP), and PPS data from the trial. The proportion of patients in each health state was estimated from a set of time-dependent transition probabilities. The transition probability from PF to PD was estimated from the TTP curve, and is dependent on the time spent in the PF state. The transition from PF to death is based on differences between the PFS and TTP curves. Transitions from PD to death were estimated from the PPS curve. Transitions are dependent on the time spent in PD, but are not dependent on time spent in the PF state. This assumes that the mortality risk after progression does not depend on the time at which the disease progressed. The second STM is described as a PPS–PFS Markov. The difference here is that the transition from PD to death is modeled using tunnel states, which allow the mortality risk to depend on the length of time the patient has been in the PF state. Different PPS curves are generated conditional on the length of PFS. This structure attempts to model the PPS more realistically by using the relationship between a prognostic variable at progression (TTP) and survival. The Markov structures discussed in this study require access to TTP and PPS data, which are not usually published but are available through patient data from a clinical trial. All of the models were implemented in Microsoft Excel® (Microsoft Corp., Redmond, WA, USA).

---

**Fig. 1** Three health state structures used by the partitioned survival model, post-progression survival Markov, and post-progression survival–progression-free survival Markov. **TP** transition probability.
2 Methods

2.1 Data Sources

Patient-level data were available from the CheckMate 025 trial (ClinicalTrials.gov identifier NCT01668784) with 14 and 38 months of minimum follow-up. CheckMate 025 is the pivotal randomized controlled trial for nivolumab compared with everolimus in patients with advanced or metastatic RCC [4]. The dataset with 14 months’ minimum follow-up was used to derive clinical inputs for each model structure in the base case, as these are the data that were available at the time of health technology assessment (HTA) submissions for nivolumab in RCC. The primary endpoint was OS. The median OS was 25.0 months (95% confidence interval [CI] 21.8–not estimable) for patients treated with nivolumab compared with 19.6 months (95% CI 17.6–23.1) for patients treated with everolimus (hazard ratio 0.73; 95% CI 0.57–0.93), indicative of a clinically significant improvement in OS achieved in patients treated with nivolumab. The OS curves from this trial are shown in Fig. 2a. The median PFS as defined by RECIST (Response Evaluation Criteria in Solid Tumors) 1.1 assessment was 4.6 months (95% CI 3.7–5.4) with nivolumab and 4.4 months (95% CI 3.7–5.5) with everolimus (hazard ratio 0.88; 95% CI 0.75–1.03) [4]. The PFS curves from this trial are shown in Fig. 2b [4].

CheckMate 025 was an open-label trial. A protocol amendment after data read-out allowed patients in the everolimus arm to cross over to the nivolumab arm. For the 14-month data cut, 55.4% of patients in the nivolumab arm received subsequent therapy (2.4% of patients received immunotherapy, 53.9% received another approved agent, 3.2% received another investigational agent). In the everolimus arm, 63% of patients received subsequent therapy (5.1% received immunotherapy, 60.8% received other approved agents, 5.4% received other investigational agents including 1% who crossed over to receive nivolumab). Survival was not adjusted for subsequent treatment received and more patients in the everolimus arm went on to receive subsequent treatment. Therefore, the relative benefit of nivolumab may have been underestimated. In CheckMate 025, median PFS was essentially the same for both drugs, which at first glance may suggest that differences in OS only occurred during the post-progression stage and that the difference in OS may be due to subsequent treatment. However, some patients starting immunotherapy have demonstrated an initial increase in tumor lesion size, followed by tumor regression. This could be due to the mechanism of action of immunotherapy, which causes T cells to infiltrate the tumor site. This initial pseudo-progression could be classified inaccurately as disease progression based on RECIST 1.1 assessment [5]. CheckMate 025 allowed treatment beyond progression if there was investigator-assessed clinical benefit and tolerability. Analysis of patients treated beyond first progression in the CheckMate 025 study demonstrated that of 316 patients who progressed in the nivolumab arm, 153 patients were treated ≥ 4 weeks after first progression and 18 patients were treated < 4 weeks beyond progression. Approximately half of patients who were treated with nivolumab beyond first progression had a reduction in tumor burden post-progression and 14% had a ≥ 30% reduction in tumor burden post-progression. Of the 320 patients who progressed in the everolimus arm, 65 were treated ≥ 4 weeks beyond progression and 111 were treated < 4 weeks beyond progression. Approximately one-quarter of patients treated with everolimus beyond first progression had a reduction in tumor burden post-progression and none had a ≥ 30% reduction in tumor burden post-progression. Treatment beyond progression may have contributed to the OS benefit observed with nivolumab, suggesting that RECIST-defined PFS may not be a good surrogate of OS, especially with novel checkpoint inhibitors such as nivolumab [6].

After this analysis was conducted, a data cut with 38 months’ follow-up subsequently became available. These data were used to assess the predictive accuracy of each model structure. A sensitivity analysis using the dataset with 38 months of follow-up was also conducted.

Longer-term data were also available from the CheckMate 003 trial (ClinicalTrials.gov identifier NCT00730639), which was a phase I study of nivolumab in patients with a range of malignancies, including RCC [7].

2.2 Model Structure

Three economic models were developed in Microsoft Excel®. All have the same structure with three health states (PF, PD, and death [Fig. 1]), weekly cycle length, and lifetime (20-year) horizon.

The PSM uses parametric regressions fitted to the Kaplan–Meier curves for OS and PFS from the CheckMate 025 trial with predictions of survival extrapolated over the time horizon of the analysis. The proportion of patients in each health state at time t was estimated from the areas under the PFS and OS curves.

The PPS Markov and PPS–PFS Markov models use additional information on TTP and PPS from the trial to derive time-dependent transition probabilities (TPs). Three transitions are possible: from PF to PD (TP1), from PF to death (TP2), and from PD to death (TP3). The PPS Markov and PPS–PFS Markov use the same method for calculating TP1 and TP2. The method used to derive TP3 is different, as described later in this section.
TTP data were derived from the clinical trial by censoring PFS on death. Parametric curves were then fitted to the TTP data and extrapolated over a lifetime horizon in order to calculate TP1. Because the shape of the PFS curve reflects both death and PD, differences between PFS and TTP were used to calculate TP2. Changes in the shape of the PFS and TTP curves over time reflect the changing hazards, meaning that TP1 and TP2 are both dependent on the time the patient has spent in the PFS health state. The equations for TP1, TP2, and TP3 to calculate the transition probabilities from one cycle to the next are described as follows, with TP1 and TP2 used as an example below.

\[
TP_1(t, t - u) = 1 - \frac{S_{\text{TTP}}(t)}{S_{\text{TTP}}(t - u)}
\]
\[ TP_2 = 1 - TP_1 - \frac{S(t)}{S(t-u)} \]

where, in general, for \( u > 0 \)

\[ TP(t, t-u) = \Pr(T \leq t|T > t-u) = 1 - \Pr(T > t|T > t-u) = 1 - \frac{\Pr(T > t)}{\Pr(T > t-u)} = 1 - \frac{S(t)}{S(t-u)} \]

where \( S_{\text{TTP}} \) is the predicted survival at time \( t \) from the TTP parametric curve; \( S_{\text{PFS}} \) is the predicted survival at time \( t \) from the PFS parametric curve; \( TP(t, t-u) \) is the transition probability at time \( t \) for cycle length \( u \); and \( \Pr \) is the probability.

PPS was extrapolated by fitting a parametric curve to the PPS Kaplan–Meier curve from the CheckMate 025 trial (Fig. 3). This curve was then used to calculate TP3 for the PPS Markov model. Changes in the hazard of the PPS curve over time mean that TP3 is a function of the time spent in the PPS health state, and hence tunnel states are used to track patients entering the PD health state. A limitation associated with the use of a PPS curve is that data from the patients who are censored before progression or death do not contribute to the estimation of PPS transition probabilities. Data from these patients may still inform the PSM as they are still included in the likelihood function. A further limitation of this method is the application of a single Kaplan–Meier curve for PPS, which assumes the same survival trajectory for all patients entering the PD state irrespective of the time spent in the PF state. TP3 is therefore independent of the

---

**Fig. 3** Post-progression survival curves from the CheckMate 025 trial (14 months’ follow-up) (a), including for patients who progressed in < 8 weeks (b), between 8 and 24 weeks (c), and after 24 weeks (d)
time spent in PF, and this assumption is problematic if there is a correlation between PFS and PPS. A positive relationship between PFS and PPS in individual metastatic RCC patients on first-line treatment has previously been shown [8].

In order to account for possible differences in survival between progressed patients based on time spent in the PF state, three different Kaplan–Meier PPS curves were generated from the CheckMate 025 trial data for the PPS–PFS Markov model (Fig. 3). Patients were divided into three subgroups (tertiles) dependent on the time at which they progressed (the resulting PFS intervals are based on the selected percentiles). PPS curves were generated for patients who progressed within 8 weeks of randomization (< 33.3rd percentile: n = 224), patients who progressed between 8 and 24 weeks of randomization (33.3rd–66.6th percentile, n = 195), and beyond 24 weeks of randomization (> 66.6th percentile, n = 198). The initial 8-week timepoint was chosen to correspond with the first tumor assessment point in CheckMate 025 and the portioning of data into the aforementioned periods gave roughly equal sample sizes in each group. The discrepancy in patient numbers in each tertile arose due to the fact that progression was not a continuous measure, but rather was measured at discrete 4-week intervals in CheckMate 025. Each PPS curve was extrapolated by fitting a parametric curve to the Kaplan–Meier data. These curves were then used to derive transition TP3 for the PPS–PFS Markov. On entering the PD health state, tunnel states are used to track patients, and the model then assigns different transition probabilities based on PPS estimates conditional on TTP. In the PPS–PFS Markov, transition probabilities are dependent on the time spent in the PF state and the time after PD.

2.3 Data Analysis

For the initial analysis parametric curves were fitted to the OS, PFS, TTP, and PPS Kaplan–Meier data from CheckMate 025 (14-month analysis, with minimum follow-up period of 14 months) and extrapolated over a lifetime horizon. For the base-case analysis, the best-fitting models were chosen based on Akaike’s information criterion (AIC), Bayesian information criterion (BIC), and visual inspection. For OS for everolimus, the three best-fitting distributions according to AIC and BIC criteria (log-normal, log-logistic, and generalized gamma) provided OS curves that plateaued after 3 years, meaning that OS in the everolimus arm was higher than in the nivolumab arm. This is not likely to be clinically plausible, given the OS curves for nivolumab demonstrate higher OS throughout the CheckMate 025 study period. The fourth best-fitting distribution (gamma) was therefore selected.

In addition, a number of scenario analyses were run in order to determine how changes to the parametric curve choices affected the results. The scenarios included exponential distributions fitted to all endpoints, Weibull distributions fitted to all endpoints, generalized gamma distributions, gamma distributions fitted to all endpoints, and log-normal distributions fitted to all endpoints. These distributions were chosen as they provided the best fit to the CheckMate 025 trial data for at least one endpoint.

Additional trial data have since become available with ~38 months’ minimum follow-up. These data have been used to evaluate the long-term survival predictions of the analysis. The parametric survival modeling was conducted using the flexsurv package in the statistical software, R 3.3.0 [9]. Table 1 shows the final functional choices for the trial data in the base case and for the scenario analyses.

3 Results

The OS estimates produced by each model were initially compared in terms of goodness of fit to the CheckMate 025 trial data (Table 2). All of the models provide a reasonable fit to the observed data at 1 and 2 years, particularly for nivolumab, and this is evident from visual inspection of the respective survival curves (Fig. 3). Three-year OS rates from a later cut of the CheckMate 025 data were used to evaluate the long-term estimates of each model structure. The 3-year OS rate for the nivolumab arm of CheckMate 025 with 38 months’ follow-up was 39.1%. The PPS–PFS Markov was able to predict this most closely at 36.2%, followed by the PSM at 33.1%. The PPS Markov under-predicted the 3-year OS at 27.6% (Table 2). The 3-year OS rate in the everolimus arm from the CheckMate 025 data with 38 months’ follow-up was 29.5%. The PPS–PFS Markov was able to most closely predict this at 26.5%, closely followed by the PSM at 25.6%. The PPS Markov under-predicted the 3-year OS at 19.2% (Table 2). When comparing the results from the 14-month and 38-month analyses from the everolimus arm, it is important to consider that 47 patients in the trial crossed over from everolimus to nivolumab by 2 years, and this could have implications for the survival observed in the everolimus arm. At 3, 4, and 5 years there are meaningful differences in predictions of survival between the models (Figs. 4, 5). The differences between the PPS–PFS Markov and PPS Markov are the most marked, and the PPS Markov consistently produces the lowest survival estimates for both nivolumab and everolimus. Additional assessment of the nivolumab arms of each model was conducted using the conditional survival from 1 to 5 years for patients who received nivolumab in the CheckMate 003 trial (Table 2) [7]. Estimates of 5-year conditional survival years for the PSM, PPS Markov, and PPS–PFS Markov are 18.0%, 9.9%, and 22.1%, respectively, for nivolumab, versus 48.6% in the CheckMate 003 trial. The PPS–PFS Markov produces estimates that are
| Curve choice                  | Nivolumab             | Everolimus            |
|------------------------------|-----------------------|-----------------------|
| **Base-case analysis**       |                       |                       |
| OS                           | Gamma                 | Gamma                 |
| PFS                          | Generalized gamma     | Log-normal            |
| TTP                          | Generalized gamma     | Log-normal            |
| PPS (overall)                | Generalized gamma     | Exponential           |
| PPS (PFS < 8 weeks)          | Exponential           | Log-normal            |
| PPS (PFS 8–24 weeks)         | Exponential           | Exponential           |
| PPS (> 24 weeks)             | Weibull               | Exponential           |
| **Sensitivity analysis 1 (exponentials only)** |                       |                       |
| OS                           | Exponential           | Exponential           |
| PFS                          | Exponential           | Exponential           |
| TTP                          | Exponential           | Exponential           |
| PPS (overall)                | Exponential           | Exponential           |
| PPS (PFS < 8 weeks)          | Exponential           | Exponential           |
| PPS (PFS 8–24 weeks)         | Exponential           | Exponential           |
| PPS (> 24 weeks)             | Exponential           | Exponential           |
| **Sensitivity analysis 2 (Weibulls only)** |                       |                       |
| OS                           | Weibull               | Weibull               |
| PFS                          | Weibull               | Weibull               |
| TTP                          | Weibull               | Weibull               |
| PPS (overall)                | Weibull               | Weibull               |
| PPS (PFS < 8 weeks)          | Weibull               | Weibull               |
| PPS (PFS 8–24 weeks)         | Weibull               | Weibull               |
| PPS (> 24 weeks)             | Weibull               | Weibull               |
| **Sensitivity analysis 3 (generalized gammas only)** |                       |                       |
| OS                           | Generalized gamma     | Generalized gamma     |
| PFS                          | Generalized gamma     | Generalized gamma     |
| TTP                          | Generalized gamma     | Generalized gamma     |
| PPS (overall)                | Generalized gamma     | Generalized gamma     |
| PPS (PFS < 8 weeks)          | Generalized gamma     | Generalized gamma     |
| PPS (PFS 8–24 weeks)         | Generalized gamma     | Generalized gamma     |
| **Sensitivity analysis 4 (log-normals only)** |                       |                       |
| OS                           | Log-normal            | Log-normal            |
| PFS                          | Log-normal            | Log-normal            |
| TTP                          | Log-normal            | Log-normal            |
| PPS (overall)                | Log-normal            | Log-normal            |
| PPS (PFS < 8 weeks)          | Log-normal            | Log-normal            |
| PPS (PFS 8–24 weeks)         | Log-normal            | Log-normal            |
| PPS (> 24 weeks)             | Log-normal            | Log-normal            |
| **Sensitivity analysis 5 (38-month data)** |                       |                       |
| Overall survival             | Log-logistic          | Log-logistic          |
| PFS                          | Generalized gamma     | Log-normal            |
| Time to progression          | Generalized gamma     | Generalized gamma     |
| PPS (overall)                | Exponential           | Generalized gamma     |
| PPS (PFS < 8 weeks)          | Exponential           | Generalized gamma     |
| PPS (PFS 8–24 weeks)         | Log-logistic          | Log-normal            |
| PPS (> 24 weeks)             | Weibull               | Log-logistic          |

OS overall survival, PFS progression-free survival, PPS post-progression survival, TTP time to progression
Table 2  Comparison of nivolumab and everolimus overall survival as predicted by each model type with data from CheckMate 025 and CheckMate 003

| Model type                          | Overall survival (%) | Conditional survival from 1 to 5 years (%) | Restricted mean survival analyses over a 20-year time horizon (months) |
|-------------------------------------|----------------------|------------------------------------------|---------------------------------------------------------------|
|                                     | 1 year | 2 years | 3 years | 4 years | 5 years |                                           |
| Nivolumab (observed)                |        |         |         |         |         |                                           |
| CheckMate 025 (14 months’ follow-up)| 76.0   | 51.7    | NA      | NA      | NA      | NA                                         |
| CheckMate 025 (38 months’ follow-up)| 76.0   | 52.2    | 39.1    | 29.4    | NA      | NA                                         |
| CheckMate 003                       | 70.6   | 48.1    | 41.2    | 37.8    | 34.3    | 48.6                                       |
| Everolimus (observed)               |        |         |         |         |         |                                           |
| CheckMate 025 (14 months’ follow-up)| 66.7   | 44.3    | NA      | NA      | NA      | NA                                         |
| CheckMate 025 (38 months’ follow-up)| 67.0   | 42.1    | 29.5    | 23.2    | NA      | NA                                         |
| Nivolumab (base-case analysis)      |        |         |         |         |         |                                           |
| PSM                                 | 74.7   | 50.7    | 33.1    | 21.2    | 13.4    | 18.0                                       |
| PPS Markov                          | 76.7   | 49.5    | 27.6    | 14.0    | 7.6     | 9.9                                        |
| PPS–PFS Markov                      | 73.7   | 52.1    | 36.2    | 24.5    | 16.3    | 22.1                                       |
| Everolimus (base-case analysis)     |        |         |         |         |         |                                           |
| PSM                                 | 65.1   | 41.2    | 25.6    | 15.8    | 9.7     | 14.9                                       |
| PPS Markov                          | 64.9   | 35.9    | 19.2    | 10.2    | 5.4     | 8.4                                        |
| PPS–PFS Markov                      | 60.5   | 39.4    | 26.5    | 18.1    | 12.5    | 20.7                                       |
| Nivolumab (sensitivity analysis 1: exponential curves fit to all endpoints) |        |         |         |         |         |                                           |
| PSM                                 | 72.1   | 52.7    | 38.4    | 28.1    | 20.5    | 28.4                                       |
| PPS Markov                          | 76.7   | 51.0    | 32.2    | 20.0    | 12.3    | 16.1                                       |
| PPS–PFS Markov                      | 75.8   | 55.0    | 40.6    | 30.6    | 23.5    | 31.0                                       |
| Everolimus (sensitivity analysis 1: exponential curves fit to all endpoints) |        |         |         |         |         |                                           |
| PSM                                 | 64.1   | 42.4    | 28.0    | 18.5    | 12.3    | 19.1                                       |
| PPS Markov                          | 65.9   | 35.7    | 18.2    | 9.1     | 4.5     | 6.8                                        |
| PPS–PFS Markov                      | 62.3   | 38.7    | 25.1    | 16.7    | 11.3    | 18.1                                       |
| Nivolumab (sensitivity analysis 2: Weibull curves fit to all endpoints)        |        |         |         |         |         |                                           |
| PSM                                 | 75.1   | 50.5    | 31.9    | 19.2    | 11.1    | 14.8                                       |
| PPS Markov                          | 77.5   | 50.2    | 29.9    | 16.9    | 9.2     | 11.9                                       |
| PPS–PFS Markov                      | 76.1   | 53.7    | 35.8    | 22.5    | 13.2    | 17.4                                       |
| Everolimus (sensitivity analysis 2: Weibull curves fit to all endpoints)       |        |         |         |         |         |                                           |
| PSM                                 | 65.2   | 41.2    | 25.4    | 15.4    | 9.2     | 14.1                                       |
| PPS Markov                          | 65.4   | 35.6    | 18.9    | 10.1    | 5.5     | 8.4                                        |
| PPS–PFS Markov                      | 62.8   | 38.8    | 23.8    | 14.4    | 8.6     | 13.7                                       |
| Nivolumab (sensitivity analysis 3: generalized gamma curves fit to all endpoints) |        |         |         |         |         |                                           |
| PSM                                 | 74.7   | 50.7    | 33.2    | 21.4    | 13.6    | 18.2                                       |
| PPS Markov                          | 76.7   | 49.5    | 27.6    | 14.0    | 7.6     | 9.9                                        |
| PPS–PFS Markov                      | 74.6   | 51.3    | 32.9    | 19.2    | 10.8    | 14.4                                       |
| Everolimus (sensitivity analysis 3: generalized gamma curves fit to all endpoints) |        |         |         |         |         |                                           |
| PSM                                 | 63.3   | 42.5    | 30.2    | 22.3    | 17.0    | 26.8                                       |
| PPS Markov                          | 63.5   | 37.2    | 23.1    | 15.1    | 10.2    | 16.1                                       |
| PPS–PFS Markov                      | 60.1   | 38.8    | 25.7    | 17.1    | 11.3    | 18.9                                       |
| Nivolumab (sensitivity analysis 4: log-normal curves fit to all endpoints)      |        |         |         |         |         |                                           |
| PSM                                 | 72.6   | 52.7    | 40.1    | 31.7    | 25.8    | 35.5                                       |
| PPS Markov                          | 74.5   | 52.2    | 38.6    | 29.8    | 23.8    | 31.9                                       |
| PPS–PFS Markov                      | 72.7   | 54.5    | 42.9    | 34.8    | 29.0    | 39.8                                       |

△ Adis
closest to the CheckMate 003 cohort [7], although the discrepancy is large. The discrepancy may be due to different baseline characteristics of patients entering CheckMate 025 and CheckMate 003 and to dosing differences between the two trials. The 5-year conditional survival estimates for everolimus patients, predicted by the PSM, PPS Markov, and PPS–PFS Markov model structures were comparable to those predicted by each model structure in the nivolumab arm at 14.9%, 8.4%, 20.7%, respectively (Table 2).

From the viewpoint of modeling cost effectiveness, the most relevant comparison between the different model structures is in estimates of incremental mean survival over the modeling horizon because this drives differences in quality-adjusted life-years (QALYs). Estimates of the difference in mean survival between nivolumab and everolimus over 20 years are 6.62 months (PSM), 7.6 months (PPS Markov), and 7.4 months (PPS–PFS Markov). The latter two reflect non-trivial differences of +14% and +11%, respectively, compared with PSM.

In order to determine how changes to the parametric curve choice might affect the results, a scenario was run with exponential distributions fitted to all endpoints. Although the exponential curve was not the best choice for goodness of fit for most endpoints according to AIC/BIC criteria, each

![Image of graphs showing overall survival for nivolumab and everolimus over 20 years.](Image)

**Fig. 4** Extrapolations of overall survival for nivolumab (a) and everolimus (b) over a lifetime horizon. Kaplan–Meier data from the CheckMate 025 trial with 14 months’ follow-up. *PFS* progression-free survival, *PPS* post-progression survival, *PSM* partitioned survival model

### Table 2 (continued)

| Model type | Overall survival (%) | Conditional survival from 1 to 5 years (%) | Restricted mean survival analyses over a 20-year time horizon (months) |
|------------|---------------------|------------------------------------------|---------------------------------|
|            | 1 year | 2 years | 3 years | 4 years | 5 years | PSM           | PPS Markov | PPS–PFS Markov | PSM           | PPS Markov | PPS–PFS Markov |
| Everolimus (sensitivity analysis 4: log-normal curves fit to all endpoints) | | | | | | | | | | |
| PSM       | 62.6   | 43.3   | 32.2   | 25.1   | 20.2   | 32.3 | 43.30 |
| PPS Markov | 62.5   | 38.5   | 26.3   | 19.3   | 14.9   | 23.9 | 36.23 |
| PPS–PFS Markov | 59.7   | 40.0   | 29.1   | 22.4   | 18.0   | 30.1 | 40.78 |
| Nivolumab (sensitivity analysis 5: re-analysis with 38-month dataset) | | | | | | | | | | |
| PSM       | 74.4   | 52.6   | 38.6   | 29.5   | 23.4   | 31.4 | 50.78 |
| PPS Markov | 76.1   | 53.8   | 37.7   | 26.4   | 18.6   | 24.5 | 38.83 |
| PPS–PFS Markov | 73.5   | 53.6   | 39.1   | 27.9   | 19.6   | 26.6 | 41.11 |
| Everolimus (sensitivity analysis 5: re-analysis with 38-month dataset) | | | | | | | | | | |
| PSM       | 63.3   | 41.9   | 29.9   | 22.7   | 17.9   | 28.3 | 41.72 |
| PPS Markov | 61.5   | 37.4   | 25.0   | 17.8   | 13.2   | 21.5 | 31.80 |
| PPS–PFS Markov | 56.6   | 36.9   | 25.7   | 18.7   | 14.2   | 25.1 | 33.60 |

*NA* not available, *PFS* progression-free survival, *PPS* post-progression survival, *PSM* partitioned survival model
model still provided reasonable fit to the trial data for the nivolumab arm. A reduction in goodness of fit was observed for the everolimus arm. The dataset with 38 months’ follow-up was used to assess the long-term survival predictions of each model. Both the PSM and the PPS–PFS Markov provided a very close fit to the 3-year OS rate for nivolumab from CheckMate 025 at 3 years (39.1%), with estimates of 38.4% and 40.6%, respectively (Table 2). The PPS Markov underpredicted the 3-year OS rate, with an estimate of 32.2%. The PSM provided the closest fit to the 3-year OS rate for everolimus from CheckMate 025 (29.5%), with an estimate of 28.0%. The PPS–PFS Markov and PPS Markov estimated 3-year OS rates of 25.1% and 18.2%, respectively (Table 2). The incremental OS between nivolumab and everolimus estimated by the PSM, PPS Markov, and PPS–PFS Markov was 9.5, 8.9, and 16.7 months, respectively. As expected, the results of these scenarios demonstrate that the outcomes predicted by each model are sensitive to the choice in parametric curves, which reiterates the importance of validating each choice (Table 2).

4 Discussion

PSM and Markov decision-analytic models were developed to compare the differences in estimated survival outcomes between these modeling approaches. Two different STM analytic approaches were investigated, one in which PPS transition probabilities are independent of time spent in the PF state, and the other in which PPS transition probabilities were dependent on time spent in the PF state, based on a Kaplan–Meier analysis to estimate PPS for each of the three PFS groups (PFS ≤ 8 weeks, PFS > 8 and ≤ 24 weeks, and PFS > 24 weeks). The three approaches were easily implementable in Microsoft Excel®, making them accessible approaches for most analysts and appropriate for most HTA submission processes across the world.

The main conclusion from this analysis is that differences in the assumptions underlying the choice of model structure are likely to produce different estimates of QALY gains, which could have a significant impact on estimates of cost effectiveness. The difference in the estimated longer-term OS probabilities of each model was significantly decreased when a sensitivity analysis was conducted using more mature data, which highlights the additional uncertainty associated with the use of datasets with a short minimum follow-up. When using the earlier 14-month data cut, the PPS–PFS Markov was able to predict most closely the observed 3-year data for nivolumab. This may have implications in a HTA landscape where drug manufacturers are increasingly seeking reimbursement of new drugs based on early data cuts from randomized clinical trials.

There is evidence to suggest that some patients in the CheckMate 025 trial may have experienced pseudo-progression. It is worth noting that pseudo-progression has implications for any economic model that includes progression as a key determinant of costs and/or outcomes. A possible method to overcome the issue of...
Models for Cost-Effectiveness Analysis

5 Conclusion

The evidence from our study and previous work highlights the importance of the assumptions underlying any model structure, and the need to validate assumptions regarding survival and the application of treatment effects against the known disease characteristics.
Acknowledgements Medical writing and editorial assistance were funded by Bristol-Myers Squibb and provided by PAREXEL International.

Author contributions All authors contributed to study conception and design and drafted the manuscript. CS and SJ performed the analysis.

Funding The work was sponsored by Bristol-Myers Squibb, which was involved in the design, data collection, data analysis, manuscript preparation, and manuscript review.

Data availability statement The analysis reported in this study uses patient-level data from the CheckMate 025 trial. The patient-level data are not publicly available, but the results of the trial have been reported in a number of publications. The trial results supporting the findings of this analysis are presented graphically within the article. The survival analysis was implemented in R using the flexsurv package.

Compliance with Ethical Standards

Conflict of interest This study was funded by Bristol-Myers Squibb. PAREXEL international received a consultancy fee from Bristol-Myers Squibb to support this analysis and develop the manuscript. Caitlin Smare, John Posnett, and Sukhinder Johal are employees of PAREXEL International. Justin Doan and Khalid Lakhdari are employees of Bristol-Myers Squibb. Justin Doan is a shareholder in Bristol-Myers Squibb.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Woods B, Soares M, Sideris E, Palmer S. Partitioned survival analysis: a critical review of the approach and application to decision modelling in health care [abstract no. 4K-1]. Society for Medical Decision Making: 6th Biennial European Conference; 12–14 Jun 2016; London.

2. Hettle R, Posnett J, Borrill J. Challenges in economic modeling of anticancer therapies: an example of modeling the survival benefit of olaparib maintenance therapy for patients with BRCA-mutated platinum-sensitive relapsed ovarian cancer. J Med Econ. 2015;18:516–24.

3. NICE Decision Support Unit. Partitioned survival analysis for decision modelling in health care: a critical review. DSU technical support document. Sheffield: National Institute for Health and Care Excellence; 2017.

4. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. N Engl J Med. 2015;373:1803–13.

5. Dranitsaris G, Cohen RB, Acton G, et al. Statistical considerations in clinical trial design of immunotherapeutic cancer agents. J Immunother. 2015;38:259–66.

6. Escudier B, Motzer RJ, Sharma P, et al. Treatment beyond progression in patients with advanced renal cell carcinoma treated with nivolumab in CheckMate 025. Eur Urol. 2017;72:368–76.

7. McDermott DF, Motzer RJ, Atkins MB, et al. Long-term overall survival (OS) with nivolumab in previously treated patients with advanced renal cell carcinoma (aRCC) from phase I and II studies [abstract no. 4507]. J Clin Oncol. 2016;34:4507.

8. Negrier S, Bushmakin AG, Cappelleri JC, et al. Assessment of progression-free survival as a surrogate end-point for overall survival in patients with metastatic renal cell carcinoma. Eur J Cancer. 2014;50:1766–71.

9. Development Core Team R. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015.

10. Johal S, Santi I, Doan J, George S. Is RECIST-defined progression-free survival a meaningful endpoint in the era of immunotherapy? [poster no. 488]. 2017 Genitourinary Cancers Symposium; 16–18 Feb 2017; Orlando.

11. McEwan P, Gordon J, Ward T, Pentrod JR, Yuan Y. Empirical assessment of the impact of model choice (Markov state transition versus partitioned survival) in modelling small cell lung cancer [poster no. PRM83]. International Society for Pharmacoconomics and Outcomes Research 19th Annual European Congress 2016; 29 Oct–2 Nov 2016; Vienna.

12. Goeree R, Villeneuve J, Goeree J, Pentrod JR, Orsini L, Tahami Monfareed AA. Economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC in Canada: a comparison of modeling approaches to estimate and extrapolate survival outcomes. J Med Econ. 2016;19:630–44.

13. Briggs A, Baker TM, Gilloteau I, Orsini L, Wagner S, Palv V. Partitioned survival versus state transition modeling in oncology: a case study with nivolumab in advanced melanoma [abstract no. A338]. International Society for Pharmacoconomics and Outcomes Research 18th Annual Congress; 7–11 Nov 2015; Milan.

14. Coyle D, Coyle K. The inherent bias from using partitioned survival models in economic evaluation [abstract no. PRM74]. Value Health. 2014;17:A194.

15. 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. 2017;37:427–39.

16. Othus M, Barlogie B, LeBlanc ML, Crowley J. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res. 2012;18(14):3731–6.

17. Gibson E, Koblubaier I, Begum N, Dranitsaris G, Liew D, McEwan P, et al. Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: a systematic approach to data analysis and extrapolation. Pharmacoeconomics. 2017;35:1257–70.

18. Ouweens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating lifetime benefits associated with immunooncology therapies: challenges and approaches for overall survival extrapolations. Pharmacoeconomics. 2019;37:1129–38. https://doi.org/10.1007/s40273-019-00806-4.