



**HAL**  
open science

## Assessment of Treatment Effects and Long-term Benefits in Immune Checkpoint Inhibitor Trials Using the Flexible Parametric Cure Model

Thomas Filleron, Marine Bachelier, Julien Mazieres, Maurice Pérol, Nicolas Meyer, Elodie Martin, Fanny Mathevet, Jean-Yves Dauxois, Raphaël Porcher, Jean-Pierre Delord

### ► To cite this version:

Thomas Filleron, Marine Bachelier, Julien Mazieres, Maurice Pérol, Nicolas Meyer, et al.. Assessment of Treatment Effects and Long-term Benefits in Immune Checkpoint Inhibitor Trials Using the Flexible Parametric Cure Model: A Systematic Review. *JAMA Network Open*, 2021, 4 (12), pp.e2139573. 10.1001/jamanetworkopen.2021.39573 . hal-03842894

**HAL Id: hal-03842894**

**<https://hal.science/hal-03842894>**

Submitted on 10 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Assessment of Treatment Effects and Long-term Benefits in Immune Checkpoint Inhibitor Trials Using the Flexible Parametric Cure Model

## A Systematic Review

Thomas Filleron, PhD; Marine Bachelier, MSc; Julien Mazieres, MD, PhD; Maurice Pérol, MD, PhD; Nicolas Meyer, MD, PhD; Elodie Martin, MSc; Fanny Mathevet, MSc; Jean-Yves Dauxois, PhD; Raphael Porcher, PhD; Jean-Pierre Delord, MD, PhD

### Abstract

**IMPORTANCE** Compared with standard cytotoxic therapies, randomized immune checkpoint inhibitor (ICI) phase 3 trials reveal delayed benefits in terms of patient survival and/or long-term response. Such outcomes generally violate the assumption of proportional hazards, and the classical Cox proportional hazards regression model is therefore unsuitable for these types of analyses.

**OBJECTIVE** To evaluate the ability of the flexible parametric cure model (FPCM) to estimate treatment effects and long-term responder fractions (LRFs) independently of prespecified time points.

**EVIDENCE REVIEW** This systematic review used reconstructed individual patient data from ICI advanced or metastatic melanoma and lung cancer phase 3 trials extracted from the literature. Trials published between January 1, 2010, and October 1, 2019, with long-term follow-up periods (maximum follow-up,  $\geq 36$  months in first line and  $\geq 30$  months otherwise) were selected to identify LRFs. Individual patient data for progression-free survival were reconstructed from the published randomized ICI phase 3 trial results. The FPCM was applied to estimate treatment effects on the overall population and on the following components of the population: LRF and progression-free survival in non-long-term responders. Results obtained were compared with treatment effects estimated using the Cox proportional hazards regression model.

**FINDINGS** In this systematic review, among the 23 comparisons studied using the FPCM, a statistically significant association between the time-to-event component and experimental treatment was observed in the main analyses and confirmed in the sensitivity analyses of 18 comparisons. Results were discordant for 4 comparisons that were not significant by the Cox proportional hazards regression model. The LRFs varied from 1.5% to 12.7% for the control arms and from 4.6% to 38.8% for the experimental arms. Differences in LRFs varied from 2% to 29% and were significantly increased in the experimental compared with the control arms, except for 4 comparisons.

**CONCLUSIONS AND RELEVANCE** This systematic review of reconstructed individual patient data found that the FPCM was a complementary approach that provided a comprehensive and pertinent evaluation of benefit and risk by assessing whether ICI treatment was associated with an increased probability of patients being long-term responders or with an improved progression-free survival in patients who were not long-term responders.

JAMA Network Open. 2021;4(12):e2139573. doi:10.1001/jamanetworkopen.2021.39573

### Key Points

**Question** Does the flexible parametric cure model (FPCM) provide additional information compared with the classic Cox proportional hazards regression model in the analysis of randomized immune checkpoint inhibitor (ICI) clinical trials using progression-free survival as an end point?

**Findings** This systematic review of reconstructed individual patient data extracted from ICI advanced or metastatic melanoma and lung cancer phase 3 trials provides empirical evidence that FPCM is a complementary approach to the Cox proportional hazards regression model. The FPCM allows estimation of treatment effects on the overall population and on the following components of the population: long-term responder fraction and progression-free survival in non-long-term responders.

**Meaning** The findings of this review suggest that FPCM is a complementary approach that provides a comprehensive and pertinent evaluation of benefit and risk by assessing whether ICI treatment is associated with an increased probability of patients being long-term responders or with an improved progression-free survival in patients who are not long-term responders.

### + Supplemental content

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

## Introduction

Recent developments in immune checkpoint inhibitors (ICIs) have substantially improved the outcomes of patients with advanced and metastatic cancer across several different tumor types.<sup>1-4</sup> Long-term analysis of the Keynote-001 (Study of Pembrolizumab [MK-3475] in Participants With Progressive Locally Advanced or Metastatic Carcinoma, Melanoma, or Non-Small Cell Lung Carcinoma) study<sup>5</sup> of patients with advanced or metastatic non-small-cell lung cancer (NSCLC) receiving pembrolizumab reported a 5-year overall survival (OS) rate of 15.5%, a clinically meaningful improvement when compared with standard cytotoxic therapies.

Phase 3 trials<sup>1,2</sup> comparing ICIs with standard therapies detect a delayed clinical effect of ICI treatments in terms of progression-free survival (PFS) and OS. Although survival curves for standard therapies and ICI overlap are sometimes even inverted early during follow-up, a clear separation between the 2 curves only becomes apparent several months after starting ICI treatment. The CheckMate-057 (Study of BMS-936558 [Nivolumab] Compared to Docetaxel in Previously Treated Metastatic Non-squamous NSCLC) study, for instance, found that patients with nonsquamous NSCLC treated with chemotherapy had better initial PFS compared with patients receiving ICIs.<sup>2</sup> CheckMate-017 (Study of BMS-936558 [Nivolumab] Compared to Docetaxel in Previously Treated Advanced or Metastatic Squamous Cell NSCLC) similarly found that the PFS of patients with squamous NSCLC treated with nivolumab was identical to the initial 3 months of docetaxel treatment.<sup>2</sup> The ICI treatments may nevertheless still provide durable responses and long-term PFS benefits compared with standard non-ICI agents.<sup>6</sup> The presence of long-term responders is characterized by the appearance of subsequent plateaus in the survival curves as can be observed in patients with melanoma treated with ipilimumab and/or nivolumab<sup>7</sup> and in patients with NSCLC treated with nivolumab.<sup>2</sup> These contrasting observations must be considered when evaluating randomized clinical trials and highlight the challenges of randomized ICI trial analyses. When there is a delayed separation between survival curves and/or the presence of a plateau at the tail end of curves, the assumption of proportional hazards is generally violated, and the classic Cox proportional hazards regression model can no longer adequately quantify the effect size of the treatment.<sup>8-10</sup>

Several alternative approaches have been proposed and discussed to deal with nonproportional hazards,<sup>11</sup> most notably the restricted mean survival time and the weighted log-rank test.<sup>12,13</sup> The restricted mean survival time quantifies the effect of treatment, whereas the weighted log-rank test compares survival curves by allocating different weights to events, depending on the timing of the event. These approaches do not, however, allow for differentiation of whether ICIs increase the duration of the response. To address this question and accurately measure long-term treatment benefits, a previous study<sup>14</sup> focused on analyzing the tail ends of survival curves using alternative approaches, such as milestone survival at prespecified time points. This approach consists of estimating PFS or OS rates and their corresponding CIs at specific time points using the Kaplan-Meier estimator. Milestone analysis has several drawbacks and needs to be interpreted with caution. It does not represent the entire survival curve but only captures information for a single time point. Moreover, PFS or OS estimates and their respective CIs may be misinterpreted when the length of the follow-up is inappropriate or indeed extends too far beyond the last observed event, when the risk set is small, and most patients have already been censored.<sup>15</sup>

To better understand the association of ICIs with treatment response and to provide a comprehensive and pertinent evaluation, the analysis must address 2 specific issues: (1) whether ICIs are associated with an increased probability of being a long-term responder and (2) whether ICIs are associated with an improved PFS in non-long-term responders. To date, various cure models for nonproportional hazards of long-term responders have been developed<sup>16</sup> and applied to ICI trials.<sup>17,18</sup> The current study evaluates the flexible parametric cure model (FPCM) on a representative data set of randomized clinical trials that compared nivolumab with docetaxel in previously treated patients with advanced NSCLC (CheckMate-057). This study also tested FPCM performances on advanced or metastatic melanomas and lung cancer data extracted from several randomized ICI phase 3 trials

## Methods

### Proof of Concept: CheckMate-057

CheckMate-057<sup>19</sup> was a randomized comparative phase 3 trial in patients with previously treated advanced nonsquamous NSCLCs. Patients were randomized to docetaxel (n = 290) and nivolumab (n = 292) treatment arms, and the published analyses<sup>2,20</sup> of the trial report the crossover of PFS curves. The latest follow-up analysis<sup>20</sup> detected no statistical difference in PFS between the 2 treatment arms using a classic Cox proportional hazards regression model (hazard ratio [HR], 0.89; 95% CI, 0.74-1.06). A nonnegligible long-term responder fraction (LRF) was observed in the nivolumab arm, with a 4-year PFS rate estimated at 9.6%. Updated PFS results were used as a representative data set for the PFS analysis (**Figure 1A**). In this study, a meta-analysis was not conducted because the main objective is to present a complementary approach to analyze ICI phase 3 trial data. Individual patient data (IPD) were reconstructed from published Kaplan-Meier curves for each trial arm using the iterative algorithm of Guyot et al.<sup>21</sup>

### Literature Search Strategy and Selection Criteria

A PubMed literature search was conducted in October 2019 to identify phase 3 randomized clinical trial results published between January 1, 2010, and October 31, 2019, which included at least 1 recurrent and/or metastatic melanoma or lung cancer ICI arm and which evaluated phase 3 trials of preselected ICIs. Search strategy and selection criteria are detailed in eAppendix 1 (eTable 1) in the [Supplement](#). As recommended in the statistical literature, trials with clinically insufficient follow-up periods and trials that lacked any clear clinical evidence of an LRF on estimated PFS curves were excluded.<sup>22,23</sup> Among the 643 publications identified (**Figure 2** and eAppendix 1 [eTables 2-4] in the [Supplement](#)), 13 publications corresponding to 11 clinical trials fulfilled the inclusion criteria (melanomas: 8 publications and 6 trials; NSCLCs: 5 publications and 5 trials).

### Data Extraction and IPD Reconstructions

The webplotdigitizer software, version 4.2 was used to extract the time and PFS coordinates from published curves.<sup>24</sup> The number of at-risk patients and the number of events were extracted, if available. These data were then used as input in an iterative algorithm with Stata software, version 16 (StataCorp)<sup>25</sup> that maps digitalized curves back to Kaplan-Meier data by finding numerical solutions to the inverted Kaplan-Meier equations. To validate IPD data reconstructions, we initially evaluated the accuracy of the algorithm (eAppendix 2 and eTable 5 in the [Supplement](#)).

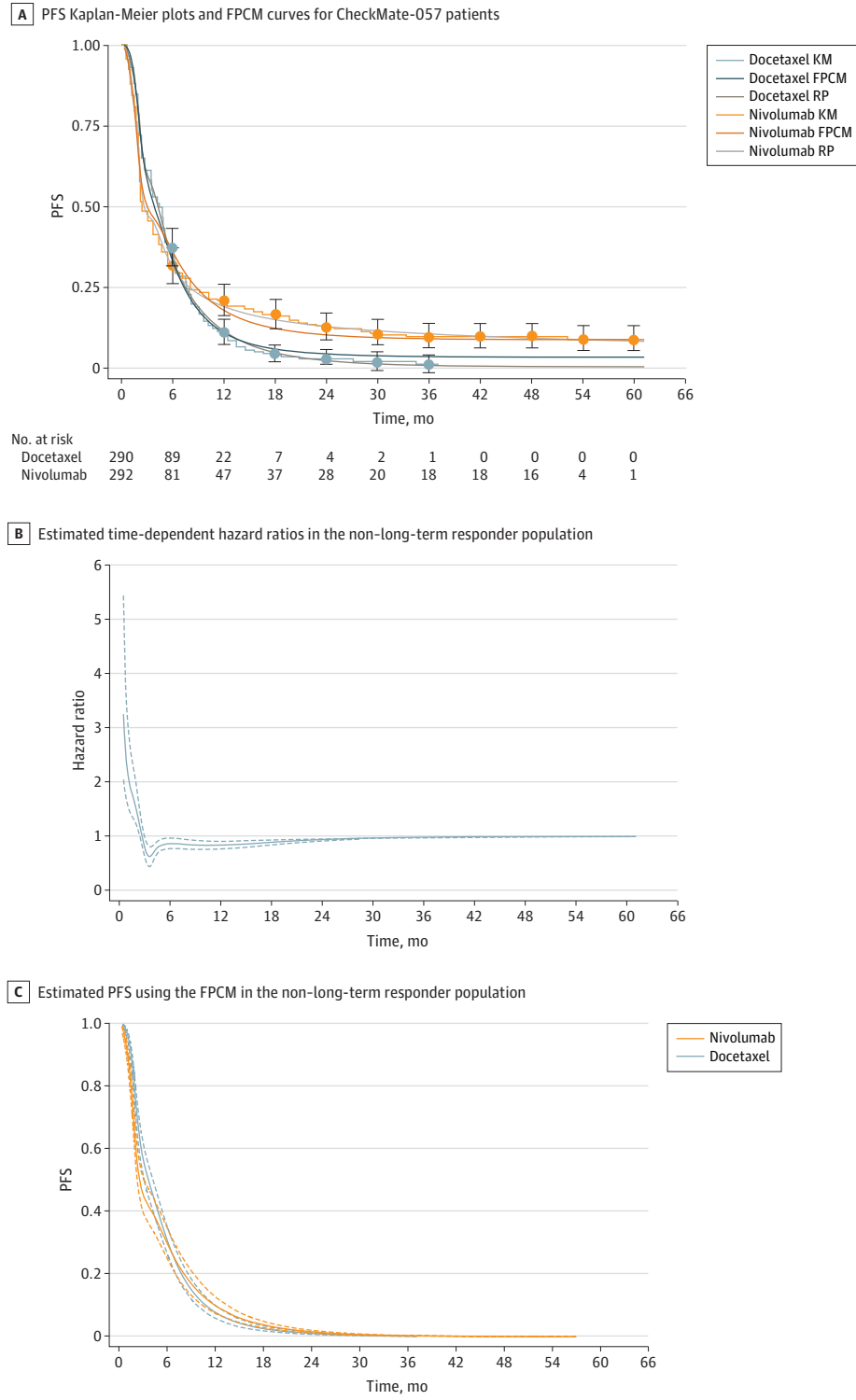
### Statistical Analysis

The FPCMs were used to reanalyze the PFS data. In the flexible parametric survival model, restricted cubic splines with varying spline knots were used to model the log-cumulative hazard function over time, and a time-dependent treatment effect was investigated (eAppendix 3 in the [Supplement](#)). In an FPCM model, such as the Royston-Parmar model (RPM), the log-cumulative hazard function was given as follows: where is a restricted cubic spline function of log time with as the position of the knots and values for the parameters, is a treatment indicator, is the corresponding coefficient, and is a spline function for the time-dependent treatment effect with a vector of knots and values for the parameters. The RPM estimates time-dependent HRs and has been popularized for modeling treatment effects<sup>26</sup> and as a supplementary analysis for randomized clinical trials.<sup>27</sup>

This model was adapted to estimate treatment effects and LRFs by forcing the log cumulative hazard in the flexible parametric survival model to plateau after the last knot. The cumulative hazard function was constrained to have a 0 slope by specifying knots in reverse order, and the last spline parameter was restricted to 0.<sup>28</sup> The FPCMs are a special case of nonmixture cure models in which survival at time  $t$  can be written as (distribution F details are given in eAppendix 3 in the [Supplement](#)).<sup>16</sup> The time-fixed component (constant parameters  $\gamma_{00}$  and  $\beta$ ) is used to model the LRF, and covariates included in the time-dependent component (ie, distribution F) characterize a

short-term effect. Parameters were estimated using maximum likelihood methods. The LRF treatment effect and short-term effect were tested using the Wald test. As proposed by Chen et al,<sup>29</sup> survival of non-long-term responders was modeled as a function of LRF and distribution, which gave

Figure 1. Progression-Free Survival (PFS) of the Study Patients



A, Reconstructed PFS Kaplan-Meier (KM) plots and Flexible Parametric Cure Model (FPCM) curves for patients with non-small cell lung cancer in the CheckMate-057 study. B, Time-dependent hazard ratios with 95% CIs (dashed lines) estimated from the main analysis of the non-long-term responder population (RP). C, PFS and 95% CIs (dashed lines) estimated using FPCM in the non-long-term RP.

a time-dependent HR for non-long-term responders with corresponding 95% CIs (robust bootstrap method with 1000 samples).

The selection of the number of internal spline knots and spline knots for time-dependent effects was assessed using the bayesian information criterion (knot locations are presented in eTable 6 in the Supplement). We performed a sensitivity analysis to test the influence of the number of knots on the LRF estimation (eAppendix 4 in the Supplement). For trials with more than 2 arms (ie, >1 comparison), the FPCM was applied to each comparison. For each comparison, the goodness-of-fit was assessed by comparing the FPCM curves with the Kaplan-Meier estimates and the corresponding RPM curves (knot locations for RPM are presented in eTable 7 in the Supplement). The FPCM and RPM were fitted using the stpm2 module for flexible parametric survival models implemented in Stata.<sup>28</sup> All statistical analyses were performed using Stata software, version 16 (StataCorp).

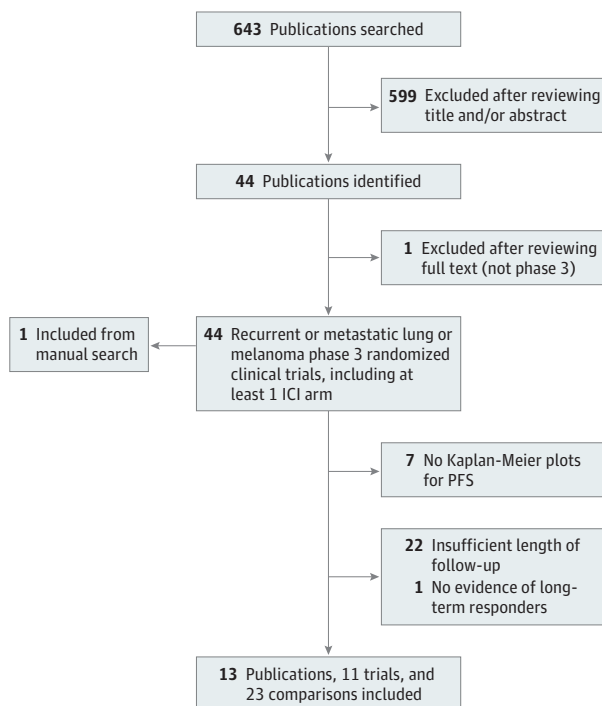
## Results

### Proof of Concept: CheckMate-057 Trial

Figure 1A presents Kaplan-Meier and best-fit FPCM and RPM PFS curves. Best-fit FPCM models comprised 5 internal knots and 1 internal knot for the baseline log-cumulative hazard and time-dependent effect, respectively. Visual inspection of the nivolumab and docetaxel arm FPCMs and Kaplan-Meier plots supported consistency, particularly because FPCM curves were contained within the 95% CIs of the Kaplan-Meier estimates (eFigure 1 in the Supplement). The RPM and FPCM present a similar fit for the tail of the distribution in the ICI arm. The PFS was lower for the RPM compared with the FPCM in the chemotherapy arm.

Although the classic Cox proportional hazards model detected no statistically significant difference between treatment (HR, 0.93; 95% CI, 0.77-1.11), the FPCM identified a treatment effect on both short-term PFS (time-dependent component  $P < .001$ ) and the LRF (time-fixed component

Figure 2. Selection Process of Immune Checkpoint Inhibitor (ICI) Randomized Phase 3 Trials From the Literature



PFS indicates progression-free survival.

$P < .001$ ), with the HR remaining stable (HR, 0.54; 95% CI, 0.42-0.72) at 36 months, after initially decreasing from 5.69 to 0.49 at about 3 months (eFigure 2 in the Supplement). The LRFs were estimated at 3.0% (95% CI, 1.5%-5.3%) for the docetaxel arm and 8.5% (95% CI, 5.6%-12.0%) for the nivolumab arm. Nivolumab was therefore associated with a 5.4% (95% CI, 2.1%-8.8%) increase in the LRF compared with docetaxel. Results from sensitivity analyses were consistent (eAppendix 4 in the Supplement). A difference in the non-long-term responder population between arms over time (Figure 1B) and nivolumab PFS was lower during the first 6 months of treatment compared with docetaxel (Figure 1C).

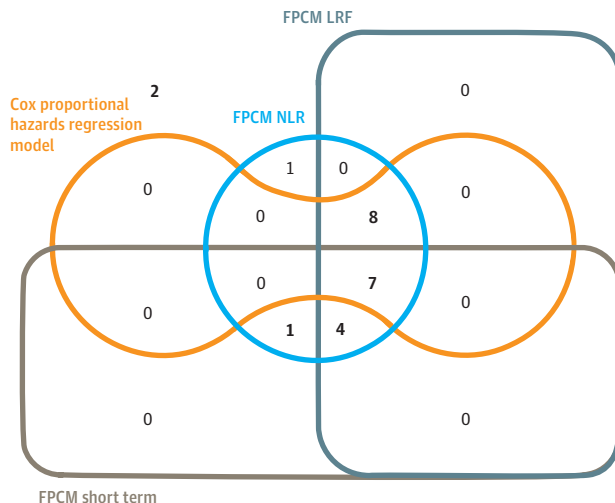
**Characteristics of Selected Randomized Clinical Trials**

Characteristics of the 11 clinical trials are presented in eTables 2 to 4 in the Supplement. The quality of data reconstruction is presented in eTable 5 in the Supplement and was deemed to be good. Data from 13 publications yielded 23 comparisons, with additional follow-up data for another 4 comparisons: 12 melanomas (2 comparisons performed 3 times [CheckMate-067 (Phase 3 Study of Nivolumab or Nivolumab Plus Ipilimumab Versus Ipilimumab Alone in Previously Untreated Advanced Melanoma)]) and 11 NSCLCs (2 comparisons performed twice [CheckMate-017 and CheckMate-057]) (Table).<sup>1,2,7,20,30-38</sup>

**Treatment Effect and Estimation of LRFs**

The main FPCM analysis is presented in eTables 8 and 9 and eFigures 3 and 4 in the Supplement. The HRs estimated using the Cox proportional hazards regression model and FPCM are reported in the Table. Figure 3 shows overlaps of the significant results obtained with the Cox proportional hazards regression model and with individual components of the FPCM. Using the FPCM, a time-dependent model was retained for 12 comparisons (melanoma: n = 5; NSCLC: n = 7). A statistically significant effect was identified on short-term PFS in all comparisons; for 11 comparisons (melanoma: n = 5; NSCLC: n = 6), we also identified an improvement in LRF. Among the 11 remaining comparisons, a treatment effect on LRF was identified by the FPCM for 8 comparisons (melanoma: n = 5; NSCLC: n = 3), with 3 comparisons showing no statistical differences between arms. Overall, results indicated a short-term effect on PFS for 12 comparisons (melanoma: n = 5; NSCLC: n = 7) and an LRF increase for 19 comparisons (melanoma: n = 10; NSCLC: n = 9). The Cox proportional hazards regression model did not find any statistically significant treatment effects for 8 comparisons (melanoma: n = 3; NSCLC: n = 5), whereas the FPCM retained a time-dependent treatment effect in

Figure 3. Overlaps of Significant Results Obtained With the Cox Proportional Hazards Model and With Individual Components of the Flexible Parametric Cure Model (FPCM)



LRF indicates long-term responder fraction; NLR, non-long-term responders.



Table. Hazard Ratios Estimated Using the FPCM and the Classic Cox Proportional Hazards Regression Model

| Trial                                 | Experimental vs standard comparison                      | Hazard ratio (95% CI)               |                   | FPCM P value <sup>b</sup> |                   | PFS of non-long-term responders in FPCM | Source                                 |
|---------------------------------------|--|-------------------------------------|-------------------|---------------------------|-------------------|---|--|
|                                       |  | Cox proportional hazards regression | FPCM <sup>a</sup> | LRF effect                | Short-term effect |   |  |
| <b>Melanoma: first-line treatment</b> |  |                                     |                   |                           |                   |   |  |
| CheckMate-066                         | Nivolumab vs dacarbazine                                 | 0.41 (0.32-0.52)                    | 0.40 (0.32-0.51)  | <.001                     | NA                | SFT                                     | Ascierto et al, 2019 <sup>30</sup>     |
| CheckMate-067                         | Nivolumab vs ipilimumab                                  | 0.59 (0.49-0.71)                    | Time varying      | <.001                     | <.001             | SDE                                     | Wolchok et al, 2017 <sup>31</sup>      |
|                                       |  | 0.56 (0.46-0.67)                    | Time varying      | <.001                     | <.001             | SDE                                     | Hodi et al, 2018 <sup>7</sup>          |
|                                       |  | 0.61 (0.51-0.74)                    | Time varying      | <.001                     | <.001             | SDE                                     | Larkin et al, 2019 <sup>32</sup>       |
|                                       | Nivolumab plus ipilimumab vs ipilimumab alone            | 0.43 (0.35-0.53)                    | Time varying      | <.001                     | <.001             | SDE                                     | Wolchok et al, 2017 <sup>31</sup>      |
|                                       |  | 0.40 (0.33-0.49)                    | 0.40 (0.33-0.49)  | <.001                     | NA                | SFT                                     | Hodi et al, 2018 <sup>7</sup>          |
|                                       |  | 0.41 (0.33-0.49)                    | 0.41 (0.33-0.49)  | <.001                     | NA                | SFT                                     | Larkin et al, 2019 <sup>32</sup>       |
| <b>Melanoma: first line or later</b>  |  |                                     |                   |                           |                   |   |  |
| Intergroup trial E1690                | 10 mg/kg vs 3 mg/kg of ipilimumab                        | 0.86 (0.74-1.01)                    | 0.87 (0.75-1.02)  | .09                       | NA                | NS                                      | Ascierto et al, 2017 <sup>33</sup>     |
| Keynote-006                           | Pembrolizumab every 2 weeks vs ipilimumab                | 0.57 (0.46-0.69)                    | 0.57 (0.47-0.69)  | <.001                     | NA                | SFT                                     | Robert et al, 2019 <sup>1</sup>        |
|                                       | Pembrolizumab every 3 weeks vs ipilimumab                | 0.57 (0.47-0.70)                    | 0.57 (0.47-0.70)  | <.001                     | NA                | SFT                                     |  |
| <b>Melanoma: second line or later</b> |  |                                     |                   |                           |                   |   |  |
| CheckMate-037                         | Nivolumab vs ICC   | 0.78 (0.59-1.02)                    | Time varying      | .03                       | <.001             | SDE                                     | Larkin et al, 2018 <sup>34</sup>       |
| CA184-002                             | Ipilimumab plus GP100 vs GP100                           | 0.85 (0.69-1.03)                    | 0.84 (0.68-1.02)  | .08                       | NA                | SFT                                     | Hodi et al, 2010 <sup>35</sup>         |
| <b>NSCLC: first-line treatment</b>    |  |                                     |                   |                           |                   |   |  |
| CA184-104                             | Ipilimumab plus chemotherapy vs chemotherapy             | 0.90 (0.77-1.05)                    | 0.90 (0.77-1.05)  | .18                       | NA                | NS                                      | Govindan et al, 2017 <sup>36</sup>     |
| CheckMate-227<br>PDL1 ≥ 1%            | Nivolumab plus ipilimumab vs chemotherapy                | 0.82 (0.70-0.98)                    | Time varying      | <.001                     | <.001             | SDE                                     | Hellman et al, 2019 <sup>37</sup>      |
|                                       |  | 0.83 (0.71-0.98)                    | 0.83 (0.71-0.98)  | .02                       | NA                | SFT                                     |  |
| CheckMate-227                         | Nivolumab plus ipilimumab vs chemotherapy                | 0.78 (0.61-0.99)                    | Time varying      | <.001                     | <.001             | SDE                                     |  |
| PDL1 < 1%                             | Nivolumab plus ipilimumab vs nivolumab plus chemotherapy | 1.00 (0.79-1.27)                    | Time varying      | .06                       | <.001             | SDE                                     |  |
|                                       |  | 0.71 (0.56-0.90)                    | 0.72 (0.57-0.91)  | .007                      | NA                | SFT                                     |  |
| <b>NSCLC: second line</b>             |  |                                     |                   |                           |                   |   |  |
| CheckMate-017                         | Nivolumab vs docetaxel                                   | 0.64 (0.49-0.84)                    | 0.64 (0.49-0.83)  | <.001                     | NA                | SFT                                     | Horn et al, 2017 <sup>2</sup>          |
|                                       |  | 0.65 (0.50-0.85)                    | Time varying      | <.001                     | .009              | SDE                                     | Antonia et al, 2019 <sup>20</sup>      |
| CheckMate-057                         | Nivolumab vs docetaxel                                   | 0.92 (0.77-1.11)                    | Time varying      | .004                      | <.001             | SDE                                     | Horn et al, 2017 <sup>2</sup>          |
|                                       |  | 0.93 (0.77-1.11)                    | Time varying      | <.001                     | <.001             | SDE                                     | Antonia et al, 2019 <sup>20</sup>      |
| <b>NSCLC: second line or later</b>    |  |                                     |                   |                           |                   |   |  |
| OAK                                   | Atezolizumab vs docetaxel                                | 0.98 (0.87-1.11)                    | Time varying      | <.001                     | <.001             | SDE                                     | Fehrenbacher et al, 2018 <sup>38</sup> |

Abbreviations: CA184-002, MDX-010 Antibody, MDX-1379 Melanoma Vaccine, or MDX-010/MDX-1379 Combination Treatment for Patients With Unresectable or Metastatic Melanoma; CA184-104, Phase 3 Trial in Squamous Non Small Cell Lung Cancer Subjects Comparing Ipilimumab Vs Placebo in Addition to Paclitaxel and Carboplatin; CheckMate-017, Study of BMS-936558 [Nivolumab] Compared to Docetaxel in Previously Treated Advanced or Metastatic Squamous Cell NSCLC; CheckMate-037, A Study to Compare BMS-936558 to the Physician's Choice of Either Dacarbazine or Carboplatin and Paclitaxel in Advanced Melanoma Patients That Have Progressed Following Anti-CTLA-4 Therapy; CheckMate-057, Study of BMS-936558 [Nivolumab] Compared to Docetaxel in Previously Treated Metastatic Non-squamous NSCLC; CheckMate-066, Study of Nivolumab (BMS-936558) Compared With Dacarbazine in Untreated, Unresectable, or Metastatic Melanoma; CheckMate-067, Phase 3 Study of Nivolumab or Nivolumab Plus Ipilimumab Versus Ipilimumab Alone in Previously Untreated Advanced Melanoma; E1690, Phase 3 Trial in Subjects With Metastatic Melanoma Comparing 3 mg/kg Ipilimumab Versus 10 mg/kg Ipilimumab; FPCM, flexible

parametric cure model; GP100, glycoprotein 100; ICC, investigator choice chemotherapy; Keynote-006, Study of Pembrolizumab [MK-3475] in Participants With Progressive Locally Advanced or Metastatic Carcinoma, Melanoma, or Non-Small Cell Lung Carcinoma; LRF, long-term responder fraction; NA, not applicable; NS, nonsignificant; NSCLC, non-small cell lung cancer; OAK, Study of Atezolizumab Compared With Docetaxel in Participants With Locally Advanced or Metastatic Non-Small Cell Lung Cancer Who Have Failed Platinum-Containing Therapy; PDL1, programmed cell death ligand 1; PFS, progression-free survival; SDE, significantly deleterious effect of the experimental treatment during early time points followed by a significant beneficial effect (direction of the effect varies over time); SFT, significantly in favor of the experimental treatment.

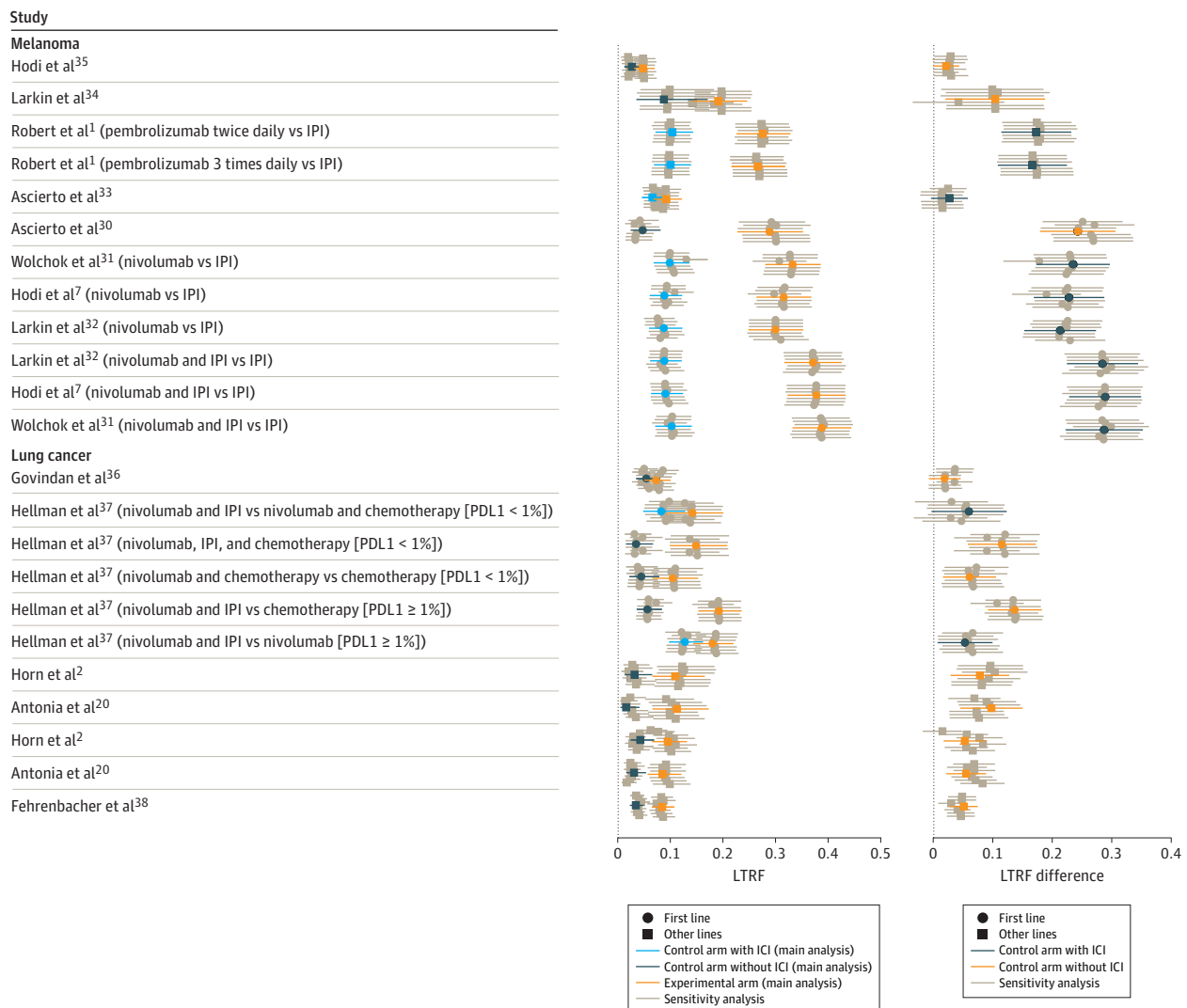
<sup>a</sup> For models with time-dependent effects, a single hazard ratio may not provide a relevant measure of the treatment effect.

<sup>b</sup> Treatment effect on the LRF was only tested for models with time-varying effects.



5 comparisons (melanoma: n = 1; NSCLC: n = 4). Statistically significant effects were observed on short-term PFS for all 5 comparisons and an increase in the LRF for 4 comparisons (melanoma: n = 1; NSCLC: n = 3). In the sensitivity analyses for knot location and number of knots, concordant results were observed for all comparisons except for 4, which yielded inconclusive results for short-term effect for 2 of them (eTable 10 and eFigure 5 in the Supplement). **Figure 4A** presents LRF estimates and their corresponding 95% CIs for standard and experimental arms. The rate of long-term responders varied from 1.5% (CheckMate-017<sup>20</sup>) to 12.7% (CheckMate-227 [An Investigational Immuno-therapy Trial of Nivolumab, or Nivolumab Plus Ipilimumab, or Nivolumab Plus Platinum-doublet Chemotherapy, Compared to Platinum Doublet Chemotherapy in Patients With Stage IV Non-Small Cell Lung Cancer]<sup>37</sup>) for the standard treatment arm and from 4.6% (CA184-002<sup>35</sup> [MDX-010 Antibody, MDX-1379 Melanoma Vaccine, or MDX-010/MDX-1379 Combination Treatment for Patients With Unresectable or Metastatic Melanoma]) to 38.8% (CheckMate-067<sup>31</sup>) for the experimental arm. Figure 4B shows LRF differences with 95% CIs for each of the 23 comparisons. Differences in LRFs varied from 1.8% to 28.8% and were larger in melanoma (median, 22.0%, range,

Figure 4. Long-term Responder Fraction (LTRF) and Rate of Long-term Response for the Standard and Experimental Arms



In the center, comparison of LTRFs (95% CIs) in the different treatment arms. At right is the estimation of LTRF differences between the trial arms. IPI indicates immune checkpoint inhibitor; PDL1, programmed cell death ligand 1.

2.1%-28.8%) compared with NSCLC (median, 5.9%; range, 1.8%-13.6%). In 18 (melanoma: n = 10; NSCLC: n = 8) of the 23 comparisons, LRFs of the experimental arm were at least twice that of the control arm. In sensitivity analyses, results were consistent for all comparisons (Figure 4A and B). Treatment arm PFS curves and time-dependent HRs in non-long-term responders are detailed in eFigures 6 and 7 in the Supplement. Visual assessment of graphs identified favorable experimental arm treatment effects for 9 comparisons (melanoma: n = 6; NSCLC: n = 3) (Table). Detrimental followed by beneficial experimental treatment effects were demonstrated for 12 comparisons (melanoma: n = 5; NSCLC: n = 7). No differences were observed over time for the remaining comparisons.

## Discussion

This systematic review supports that the FPCM may improve the analysis of randomized ICI clinical trials using PFS as an end point. When compared with the classic Cox proportional hazards regression model, the FPCM provides additional information for trials with complex survival patterns by incorporating a time-dependent HR estimation and/or taking into account LRFs. By extending the evaluation of treatment benefits to include a time scale and an LRF, the FPCM offers a direct clinically meaningful interpretation of treatment effects. The LRF component of the analysis, which refers to functional cure, is particularly relevant for patients.

Our findings specifically indicate that FPCM is a suitable strategy to fit randomized ICI phase 3 trial results and confirm that ICI treatments increase LRFs when compared with standard therapies. Considerable variations in the improvements of the LRFs, associated with the heterogeneity of clinical trial characteristics, were observed across trials. Melanomas have greater LRF increases compared with NSCLCs, irrespective of treatment lines. In melanomas, no significant differences were observed in the trial that included a population of highly pretreated patients<sup>35</sup> and the phase 3 trial that compared the benefit-risk profile of 10 mg/kg vs 3 mg/kg of ipilimumab.<sup>33</sup> The increase in the LRF was similar for the CheckMate-066 (Study of Nivolumab [BMS-936558] Compared With Dacarbazine in Untreated, Unresectable, or Metastatic Melanoma) trial (nivolumab vs dacarbazine) and for the ipilimumab vs nivolumab comparison in CheckMate-067. Analysis of 3 trials with at least 1 follow-up identified a nonnegligible fraction of long-term responders in the initial analysis. Notably, analysis of subsequent follow-up data indicated that these LRFs persisted over time and that the differences in LRFs between the different groups also remained constant. Indeed, accounting for follow-up data for the corresponding trials sometimes yielded different estimates of the drug's benefit. Our results are consistent with reports from a previous study<sup>39</sup> that introduced the concept of the functional curative potential of immunotherapy. Because our analysis focuses on the PFS end point, which is defined as disease progression or death, the *cure* term in FPCM is somewhat inappropriate because no one can actually be *cured* of death. However, in the interest of consistency with the literature, we will keep referring to our approach as FPCM.

The current study not only quantifies differences in the number of long-term responders in various treatment arms but also assesses the magnitude of long-term treatment benefits, which is particularly useful in the context of drug approvals where minimal clinically important differences between treatment arms have been predefined and where an experimental treatment may be considered clinically relevant if the observed difference is greater than this predefined threshold. Our FPCM approach also complements other systems that measure long-term benefits,<sup>40,41</sup> specifically the American Society of Clinical Oncology Value Framework version 2 bonus and the European Society for Medical Oncology Value Framework, which both incorporate bonuses and adjustments that capture data from the tail ends of survival curves.<sup>42,43</sup>

The FPCMs have numerous advantages compared with other approaches that identify LRFs. Compared with milestone estimations, the FPCM examines the entire survival curve and does not require a previously specified meaningful survival milestone. However, use of the FPCM requires careful justification and mature data. Applying FPCM to estimate LRFs is warranted when there is

enough supportive evidence from follow-up data to substantiate the identification of long-term responders and when survival plots exhibit tail-end plateaus. For any other cases, we recommend estimating treatment effects using a classic flexible parametric model without cure, instead of the FPCM.<sup>12</sup> Because the FPCM is a classic flexible parametric model with a restriction on 1 of the parameters, it is comparable to a standard RPM for testing the assumption of cure. Because the formal test compares the fit over the whole time scale and not only the tail end that is used to estimate the cure proportion,<sup>28</sup> it is recommended that the assumption of cure and the fit of the model be assessed visually from the graphs.<sup>44</sup> Because no patient can be *cured* of death, analyses presented using the FPCM focus on PFS rather than OS, even though PFS is a secondary end point. Classic flexible parametric models without cure, such as the RPM,<sup>12</sup> are better suited to evaluating treatment effects for OS end points.

### Limitations

The current study has a number of limitations. The study was not performed on original IPD but on reconstructed IPD. This type of approach has nevertheless been previously used by others<sup>45,46</sup> and the accuracy of the reconstruction algorithm itself has been validated both in the literature<sup>47</sup> and in our current analysis. When we compare results for trials with at least 1 follow-up, some discordant results may be associated with the quality of reconstructions. Our HRs were also estimated without adjusting or stratifying for randomization of stratification factors. Stratified analyses reduce bias when estimating treatment effects that violate the assumption of proportionality. Our study was also restricted to 23 FPCM comparisons of mature data because the length of follow-up of many of the trials retrieved from our initial PubMed searches was insufficient to identify long-term responders. Several comparisons involved late follow-ups of only a few patients, which may not allow sensible conclusions to be drawn because the interpretation of the Kaplan-Meier estimates may suffer from representativeness bias.<sup>48</sup> For these comparisons, the RPM and FPCM have discrepancies at the tail ends of the distributions. A recent simulation study<sup>49</sup> highlights that the FPCM may allow for the extrapolation of data and the corresponding LRF estimates may be accurate if the follow-up is sufficiently long. Finally, the FPCM needs to address several issues associated with model selection<sup>50</sup> and assessment schedule.<sup>51</sup> The degree of complexity, dictated by the number and location of knots, needs to be balanced between goodness of fit and the risk of overfitting. For this reason, an automated process was used rather than a data-driven approach, and a sensitivity analysis was performed to evaluate the influence of model specification. The sensitivity analysis yielded concordant results for the short-term component of all but 3 of the comparisons. For these comparisons, the inconclusive results observed were in part related to the model selection process and tended to yield nonsignificant results. These inconclusive results concern only short-term PFS treatment effects and did not influence LRFs. As found in the sensitivity analysis, the number of knots and locations have little effect on the LRF result, which is pertinent for the patients. The oscillation observed in the time-dependent HR during the first few months after randomization may be associated not only with the assessment schedule but also with the hyperprogressive nature of the disease in ICI-treated patients.<sup>5</sup>

### Conclusions

The FPCM approach described in the current systematic review may have utility for both clinicians and health authorities to better describe treatment effects and estimate LRFs. This approach does not replace the classic Cox proportional hazards regression model for the primary analysis of randomized clinical trials but complements classic methods to evaluate treatment benefits. Because issues of delayed treatment effects and LRFs are not specific to ICI,<sup>52</sup> an FPCM approach may be more widely applicable to clinical research to improve estimates of treatment benefits for other treatment strategies, such as targeted therapies. We suggest that clinicians and statisticians include an FPCM analysis in randomized phase 3 trial assessment to help regulatory agencies and clinicians

evaluate the benefit-risk ratio of different therapies and to guide the selection of an optimal treatment strategy for individual patients. This FPCM approach may also help answer the fundamental question that all patients ask, "What is my probability of being functionally cured?"

## ARTICLE INFORMATION

**Accepted for Publication:** October 25, 2021.

**Published:** December 21, 2021. doi:[10.1001/jamanetworkopen.2021.39573](https://doi.org/10.1001/jamanetworkopen.2021.39573)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2021 Filleron T et al. *JAMA Network Open*.

**Corresponding Author:** Thomas Filleron, PhD, Institut Claudius Regaud, Institut Universitaire du Cancer de Toulouse, 1 avenue Irène Joliot Curie, 31059 Toulouse Cedex, France ([filleron.thomas@iuct-oncopole.fr](mailto:filleron.thomas@iuct-oncopole.fr)).

**Author Affiliations:** Department of Biostatistics, Institut Claudius Regaud, Institut Universitaire du Cancer Toulouse, Toulouse, France (Filleron, Bachelier, Martin, Mathevet); Department of Pneumology, Centre Hospitalier Universitaire de Toulouse Larrey, Toulouse, France (Mazieres); Department of Medical Oncology, Léon Bérard Cancer Center, Lyon, France (Pérol); Institut Universitaire du Cancer Toulouse Oncopôle, Toulouse, France (Meyer); Institut de Mathématiques de Toulouse, Université de Toulouse, Centre National de la Recherche Scientifique, Institut National des Sciences Appliquées de Toulouse, Toulouse, France (Dauxois); Assistance Publique des Hôpitaux de Paris, Hôpital Hôtel Dieu, Centre d'Épidémiologie Clinique, INSERM U1153, Paris, France (Porcher); Department of Medical Oncology, Institut Claudius Regaud, Institut Universitaire du Cancer Toulouse, Toulouse, France (Delord).

**Author Contributions:** Dr Filleron had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Filleron, Dauxois, Delord.

*Acquisition, analysis, or interpretation of data:* All authors.

*Drafting of the manuscript:* Filleron, Bachelier, Martin, Dauxois.

*Critical revision of the manuscript for important intellectual content:* Filleron, Mazieres, Pérol, Meyer, Mathevet, Dauxois, Porcher, Delord.

*Statistical analysis:* Filleron, Bachelier, Martin, Mathevet, Dauxois.

*Obtained funding:* Filleron.

*Supervision:* Filleron, Mazieres, Meyer, Delord.

**Conflict of Interest Disclosures:** Dr Filleron reported receiving grants from Fondation BMS for Immuno-Oncology during the conduct of the study and working as a consultant for Cellectis outside the submitted work. Dr Pérol reported receiving personal fees from Roche, AstraZeneca, Bristol Myers Squibb, Merck Sharp Dohme, and Sanofi outside the submitted work. Dr Meyer reported receiving personal fees from BMS, MSD, Novartis, Pierre Fabre, Sanofi, and Sun Pharma outside the submitted work. No other disclosures were reported.

**Funding/Support:** This study was conducted in part with support from the Bristol Myers Squibb Foundation for Research in Immuno-Oncology.

**Role of the Funder/Sponsor:** The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** T Petra Neufing, PhD, assisted with the English proofreading. She was compensated for her work.

## REFERENCES

1. Robert C, Ribas A, Schachter J, et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol*. 2019;20(9):1239-1251. doi:[10.1016/S1470-2045\(19\)30388-2](https://doi.org/10.1016/S1470-2045(19)30388-2)
2. Horn L, Spigel DR, Vokes EE, et al. Nivolumab versus docetaxel in previously treated patients with advanced non-small-cell lung cancer: two-year outcomes from two randomized, open-label, phase iii trials (CheckMate 017 and CheckMate 057). *J Clin Oncol*. 2017;35(35):3924-3933. doi:[10.1200/JCO.2017.74.3062](https://doi.org/10.1200/JCO.2017.74.3062)
3. Motzer RJ, Tannir NM, McDermott DF, et al; CheckMate 214 Investigators. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med*. 2018;378(14):1277-1290. doi:[10.1056/NEJMoa1712126](https://doi.org/10.1056/NEJMoa1712126)

4. Socinski MA, Jotte RM, Cappuzzo F, et al; IMpower150 Study Group. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378(24):2288-2301. doi:10.1056/NEJMoa1716948
5. Garon EB, Hellmann MD, Rizvi NA, et al. Five-year overall survival for patients with advanced non-small-cell lung cancer treated with pembrolizumab: results from the Phase I KEYNOTE-001 study. *J Clin Oncol*. 2019;37(28):2518-2527. doi:10.1200/JCO.19.00934
6. Borcoman E, Kanjanapan Y, Champiat S, et al. Novel patterns of response under immunotherapy. *Ann Oncol*. 2019;30(3):385-396. doi:10.1093/annonc/mdz003
7. Hodi FS, Chiarion-Sileni V, Gonzalez R, et al. Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *Lancet Oncol*. 2018;19(11):1480-1492. doi:10.1016/S1470-2045(18)30700-9
8. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13-15. doi:10.1097/EDE.Ob013e3181c1ea43
9. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014;32(22):2380-2385. doi:10.1200/JCO.2014.55.2208
10. Saad ED, Zalcberg JR, Péron J, Coart E, Burzykowski T, Buyse M. Understanding and communicating measures of treatment effect on survival: can we do better? *J Natl Cancer Inst*. 2018;110(3):232-240. doi:10.1093/jnci/djx179
11. Freidlin B, Korn EL. Methods for accommodating nonproportional hazards in clinical trials: ready for the primary analysis? *J Clin Oncol*. 2019;37(35):3455-3459. doi:10.1200/JCO.19.01681
12. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175-2197. doi:10.1002/sim.1203
13. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika*. 1982;69(3):553-566. doi:10.1093/biomet/69.3.553
14. Chen T-T. Milestone survival: a potential intermediate endpoint for immune checkpoint inhibitors. *J Natl Cancer Inst*. 2015;107(9):djv156. doi:10.1093/jnci/djv156
15. Carter RE, Huang P. Cautionary note regarding the use of CIs obtained from Kaplan-Meier survival curves. *J Clin Oncol*. 2009;27(2):174-175. doi:10.1200/JCO.2008.18.8011
16. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. *Clin Cancer Res*. 2012;18(14):3731-3736. doi:10.1158/1078-0432.CCR-11-2859
17. Ouwens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. *Pharmacoeconomics*. 2019;37(9):1129-1138. doi:10.1007/s40273-019-00806-4
18. Gibson E, Koblbauer I, Begum N, et al. Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: a systematic approach to data analysis and extrapolation. *Pharmacoeconomics*. 2017;35(12):1257-1270. doi:10.1007/s40273-017-0558-5
19. ClinicalTrials.gov. Study of BMS-936558 (Nivolumab) Compared to Docetaxel in Previously Treated Metastatic Non-squamous NSCLC (CheckMate057). NCT01673867. Accessed November 4, 2021. <https://clinicaltrials.gov/ct2/show/NCT01673867>
20. Antonia SJ, Borghaei H, Ramalingam SS, et al. Four-year survival with nivolumab in patients with previously treated advanced non-small-cell lung cancer: a pooled analysis. *Lancet Oncol*. 2019;20(10):1395-1408. doi:10.1016/S1470-2045(19)30407-3
21. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12(1):9. doi:10.1186/1471-2288-12-9
22. Legrand C, Bertrand A. Cure models in oncology clinical trials. In: *Textbook of Clinical Trials in Oncology: A Statistical Perspective*. Chapman & Hall/CRC Press I Taylor & Francis Group; 2019:465-492.
23. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000;56(1):227-236. doi:10.1111/j.0006-341X.2000.00227.x
24. Rohatgi A. WebPlotDigitizer—extract data from plots, images, and maps. Accessed September 1, 2020. <https://automeris.io/WebPlotDigitizer/>
25. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan-Meier curves. *Stata J*. 2017;17(4):786-802. doi:10.1177/1536867X1801700402
26. Wärnberg F, Garmo H, Emdin S, et al. Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS Trial. *J Clin Oncol*. 2014;32(32):3613-3618. doi:10.1200/JCO.2014.56.2595

27. Perren TJ, Swart AM, Pfisterer J, et al; ICON7 Investigators. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med*. 2011;365(26):2484-2496. doi:10.1056/NEJMoa1103799
28. Andersson TM-L, Lambert PC. Fitting and modeling cure in population-based cancer studies within the framework of flexible parametric survival models. *Stata J*. 2012;12(4):623-638. doi:10.1177/1536867X1201200404
29. Chen M-H, Ibrahim JG, Sinha D. A new bayesian model for survival data with a surviving fraction. *J Am Stat Assoc*. 1999;94(447):909-919. doi:10.1080/01621459.1999.10474196
30. Ascierto PA, Long GV, Robert C, et al. Survival outcomes in patients with previously untreated *BRAF* wild-type advanced melanoma treated with nivolumab therapy: three-year follow-up of a randomized phase 3 trial. *JAMA Oncol*. 2019;5(2):187-194. doi:10.1001/jamaoncol.2018.4514
31. Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*. 2017;377(14):1345-1356. doi:10.1056/NEJMoa1709684
32. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*. 2019;381(16):1535-1546. doi:10.1056/NEJMoa1910836
33. Ascierto PA, Del Vecchio M, Robert C, et al. Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol*. 2017;18(5):611-622. doi:10.1016/S1470-2045(17)30231-0
34. Larkin J, Minor D, D'Angelo S, et al. Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: a randomized, controlled, open-label phase III trial. *J Clin Oncol*. 2018;36(4):383-390. doi:10.1200/JCO.2016.71.8023
35. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363(8):711-723. doi:10.1056/NEJMoa1003466
36. Govindan R, Szczesna A, Ahn MJ, et al. Phase III trial of ipilimumab combined with paclitaxel and carboplatin in advanced squamous non-small-cell lung cancer. *J Clin Oncol*. 2017;35(30):3449-3457. doi:10.1200/JCO.2016.71.7629
37. Hellmann MD, Paz-Ares L, Bernabe Caro R, et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med*. 2019;381(21):2020-2031. doi:10.1056/NEJMoa1910231
38. Fehrenbacher L, von Pawel J, Park K, et al. Updated efficacy analysis including secondary population results for OAK: a randomized phase III study of atezolizumab versus docetaxel in patients with previously treated advanced non-small cell lung cancer. *J Thorac Oncol*. 2018;13(8):1156-1170. doi:10.1016/j.jtho.2018.04.039
39. Zimmermann S, Peters S. Appraising the tail of the survival curve in the era of PD-1/PD-L1 checkpoint blockade. *JAMA Oncol*. 2019;5(10):1403-1405. doi:10.1001/jamaoncol.2019.2186
40. Everest L, Shah M, Chan KKW. Comparison of long-term survival benefits in trials of immune checkpoint inhibitor vs non-immune checkpoint inhibitor anticancer agents using ASCO Value Framework and ESMO Magnitude of Clinical Benefit Scale. *JAMA Netw Open*. 2019;2(7):e196803. doi:10.1001/jamanetworkopen.2019.6803
41. Vivot A, Créquit P, Porcher R. Use of late-life expectancy for assessing the long-term benefit of immune checkpoint inhibitors. *J Natl Cancer Inst*. 2019;111(5):519-521. doi:10.1093/jnci/djy211
42. Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology Value Framework: revisions and reflections in response to comments received. *J Clin Oncol*. 2016;34(24):2925-2934. doi:10.1200/JCO.2016.68.2518
43. Cherny NI, Dafni U, Bogaerts J, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol*. 2017;28(10):2340-2366. doi:10.1093/annonc/mdx310
44. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol*. 2011;11(1):96. doi:10.1186/1471-2288-11-96
45. Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol*. 2016;34(15):1813-1819. doi:10.1200/JCO.2015.64.2488
46. Alexander BM, Schoenfeld JD, Trippa L. Hazards of hazard ratios—deviations from model assumptions in immunotherapy. *N Engl J Med*. 2018;378(12):1158-1159. doi:10.1056/NEJMc1716612
47. Satagopan JM, Iasonos A, Kanik JG. A reconstructed melanoma data set for evaluating differential treatment benefit according to biomarker subgroups. *Data Brief*. 2017;12:667-675. doi:10.1016/j.dib.2017.05.005



48. Gebski V, Garès V, Gibbs E, Byth K. Data maturity and follow-up in time-to-event analyses. *Int J Epidemiol*. 2018;47(3):850-859. doi:10.1093/ije/dyy013
49. Kearns B, Stevenson MD, Triantafyllopoulos K, Manca A. The extrapolation performance of survival models for data with a cure fraction: a simulation study. *Value Health*. 2021;24(11):1634-1642. doi:10.1016/j.jval.2021.05.009
50. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73. doi:10.7326/M14-0698
51. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst*. 2007;99(6):428-432. doi:10.1093/jnci/djk091
52. Moore K, Colombo N, Scambia G, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med*. 2018;379(26):2495-2505. doi:10.1056/NEJMoa1810858

**SUPPLEMENT.****eAppendix 1.** Literature Searches**eAppendix 2.** Data Extraction and Accuracy of IPD Reconstruction**eAppendix 3.** Splines and Flexible Parametric Cure Model**eAppendix 4.** Royston and Parmar Model – Overall Survival**eAppendix 5.** Model Selection and Sensitivity Analysis**eTable 1.** Detailed Search Strategy**eTable 2.** Randomized Phase III Trials of Advanced/Metastatic Melanoma Identified From the Literature**eTable 3.** Randomized Phase III Trials of Advanced/Metastatic NSCLC Identified From the Literature**eTable 4.** Randomized Phase III Trials Which Included Patients With Advanced/Metastatic SCLC Identified From Reviewing the Literature**eTable 5.** Reconstructed Data Quality After Digitalization (PFS)**eTable 6.** Internal Knot Locations for Baseline Hazard and Time-Dependent Treatment Effects (FPCM)**eTable 7.** Internal Knot Locations for Baseline Hazard and Time-Dependent Treatment Effects (Royston Parmar)**eTable 8.** Number of Internal Knots for Baseline Hazard ( $K_0$ ) and Time-Dependent Treatment Effects ( $K_1$ ) for the Main Analysis – Advanced/Metastatic Melanoma**eTable 9.** Number of Internal Knots for Baseline Hazard ( $K_0$ ) and Time-Dependent Treatment Effects ( $K_1$ ) for the Main Analysis – Advanced/Metastatic NSCLC**eTable 10.** Estimation of Treatment Effects Using the FPCM and the Cox Proportional Model for the Main Analysis and Treatment Effects Obtained in the Sensitivity Analysis (6 Models)**eFigure 1.** Reconstructed PFS Kaplan-Meier Plots With 95% Confidence Intervals and FPCM Curves (Main Model) for Patients With Non-Small Cell Lung Cancer In Checkmate-057**eFigure 2.** Time-Dependent Hazard Ratios With 95% Confidence Intervals Estimated From the Main Analysis Overall Population**eFigure 3.** Reconstructed PFS Kaplan-Meier Plots With 95% Confidence Intervals and FPCM Curves (Main Model) for Patients With Melanoma**eFigure 4.** Reconstructed PFS Kaplan-Meier Plots With 95% Confidence Intervals and FPCM Curves (Main Model) for Patients With Non-Small-Cell Lung Cancer**eFigure 5.** Checkmate 227 PDL1<1% Nivo+Ipi vs Nivo+ Chemo: Time-Dependent HRs With 95% Confidence Intervals Estimated From the Main and Sensitivity Analyses**eFigure 6.** Non-Long-Term Responder Population for patients with Melanoma**eFigure 7.** Non-Long-Term Responder Population for patients with NSCLC