Relapse-Free Survival as a Surrogate for Overall Survival in the Evaluation of Stage II–III Melanoma Adjuvant Therapy

Stefan Suciu, Alexander M. M. Eggermont, Paul Lorigan, John M. Kirkwood, Svetomir N. Markovic, Claus Garbe, David Cameron, Srividya Kotapati, Tai-Tsang Chen, Keith Wheatley, Natalie Ives, Gaetan de Schaetzen, Achmad Efendi, Marc Buyse

Affiliations of authors: European Organisation for Research and Treatment of Cancer Headquarters, Brussels, Belgium (SS, GdS, AF); Gustave Roussy Cancer Campus Grand Paris, Villejuif, France (AMME); The Christie NHS Foundation Trust, Manchester, UK (PL); University of Pittsburgh Cancer Institute and School of Medicine, Pittsburgh, PA (JMK); Mayo Clinic Rochester, Rochester, MN (SNM); University of Tubingen, Tubingen, Germany (CG); University of Edinburgh, Western General Hospital, Edinburgh, UK (DC); Bristol-Myers Squibb, Wallingford, CT (SK, T-TC); Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK (KW); Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK (NI); Universitas Brawijaya, Malang, Indonesia (AE); IDDI, Louvain-la-Neuve, Belgium (MB).

Correspondence to: Stefan Suciu, PhD, European Organisation for Research and Treatment of Cancer Headquarters, Avenue E. Mounier 83/11, 1200, Brussels, Belgium (e-mail: stefan.suciu@eortc.be).

Abstract

Background: We assessed whether relapse-free survival (RFS; time until recurrence/death) is a valid surrogate for overall survival (OS) among resected stage II–III melanoma patients through a meta-analysis of randomized controlled trials.

Methods: Individual patient data (IPD) on RFS and OS were collected from 5826 patients enrolled in 11 randomized adjuvant trials comparing interferon (IFN) to observation. In addition, IPD from two studies comparing IFN and vaccination in 989 patients were included. A two-level modeling approach was used for assessing Spearman’s patient-level correlation (\(\rho\)) of RFS and OS and the trial-level coefficient of determination (\(R^2\)) of the treatment effects on RFS and on OS. The results were validated externally in 13 adjuvant studies without available IPD. We then tested the results on the European Organisation for Research and Treatment of Cancer (EORTC) 18071 double-blind trial comparing ipilimumab 10 mg/kg with placebo, which showed a statistically significant impact of the checkpoint inhibitor on RFS and OS. All statistical tests were two-sided.

Results: With a median follow-up of seven years, 12 of 13 trials showed a consistency between the IFN vs No IFN differences regarding RFS (hazard ratio [HR]_{RFS} = 0.88) and OS (HR_{OS} = 0.91), but the small trial, Eastern Cooperative Oncology Group 2696, was an outlier (HR_{RFS} = 0.72 vs HR_{OS} = 1.11). Therefore, even if rho was high, \(R^2\) was low and could not reliably be estimated. Based on the 12 trials, rho remained high (0.89), and the hazard ratios for RFS and OS were strongly correlated (\(R^2 = 0.91\)). The surrogate threshold effect for RFS was estimated to be 0.77. For the EORTC 18071 trial, the hazard ratio for RFS was 0.75, predicting an effect of ipilimumab on OS. This was subsequently confirmed (HR_{OS} = 0.72, 95.1% confidence interval = 0.58 to 0.88, \(P = .001\)).

Conclusions: In high-risk stage II–III melanoma, RFS appeared to be a valid surrogate end point for OS for adjuvant randomized studies assessing interferon or a checkpoint inhibitor. In future similar adjuvant studies, a hazard ratio for RFS of 0.77 or less would predict a treatment impact on OS.
The clinical development of new drugs in oncology is increasingly challenging. With the number of new compounds that are being evaluated and the need to reach conclusions about the efficacy of new treatments as quickly as possible, the search for early clinical end points of biomarkers that can be used as surrogate end points for long-term clinical end points is increasingly important (1). Overall survival (OS) is often used as the primary end point in clinical trials. However, OS, although “simple to measure, easy to interpret, clinically meaningful, and straightforward to explain,” has the disadvantages of requiring extended follow-up and of being confounded by subsequent lines of treatment (2,3). An end point that is reached more rapidly could potentially expedite decisions on efficacy and approval. Any surrogate end point used to substitute for OS must be strongly associated with it (patient-level surrogacy), and the treatment effect on the surrogate end point must be strongly associated with the treatment effect on the true end point. Tumor response is not a surrogate for OS in advanced colorectal cancer, advanced breast cancer, or advanced melanoma (4,5,6). In contrast, progression-free survival has been shown to be an acceptable surrogate for OS in patients with advanced colorectal cancer (7) or advanced melanoma (8,9), but not in patients with advanced breast cancer (5) or prostate cancer (10).

For a potential surrogate end point to replace the established end point, it must be formally validated/qualified, a process that has caused considerable controversy in the past two decades (11). A meta-analytic approach using data from several randomized trials has become an accepted method to do this (12,13).

Cutaneous melanoma is an aggressive malignancy accounting for 2.4 deaths per 100,000 inhabitants in the United States (14). The treatment of primary melanoma and loco-regional disease is surgery. The risk of recurrence after definitive surgery can be discriminated using the TNM staging system, which incorporates tumor thickness, tumor ulceration, and nodal involvement (15,16). Stage IIb deep primary melanoma and stage III, with occult micrometastatic or clinically detectable nodal disease at the time of surgery, have a high risk of disease recurrence. Thus, systemic adjuvant therapies, in particular interferon-β (IFN), have been investigated over the last several decades.

Since the first Eastern Cooperative Oncology Group (ECOG) trial that led to the approval of high-dose IFN for resected high-risk stage IIb/III melanoma (17), many studies have investigated various doses, types, and durations of IFN administration (18,19,20,21). A meta-analysis based on individual patient data (IPD) showed that IFN treatment had a statistically significant impact on relapse-free survival (RFS) but a limited impact on OS (22).

More recently, checkpoint inhibitors have revolutionized the treatment of advanced melanoma (23,24) and are being evaluated in the adjuvant setting. The initial results of the European Organisation for Research and Treatment of Cancer (EORTC) 18071 study of adjuvant ipilimumab 10 mg/kg vs placebo in 951 patients showed a statistically significant impact on RFS (25). Based on these results, the US Food and Drug Administration (FDA) licensed ipilimumab 10 mg/kg in stage III resected melanoma.

In our study, using an IPD IFN meta-analysis, we investigated whether RFS is a valid surrogate end point for OS in resected stage II–III melanoma patients. We also aimed to determine the minimum treatment effect required on RFS to predict a meaningful effect on OS in a future adjuvant IFN trial (26). Finally, we examined whether these findings were confirmed for adjuvant ipilimumab (27).

### Methods

#### Data and Trial Characteristics

A systematic literature review was performed to assess the studies to be included in the training set, using the Cochrane Controlled Trials Register, MEDLINE, EMBASE, PubMed, and Web of Science. For this selection, we included randomized adjuvant IFN trials in patients age 18 years or older with resected stage II–III cutaneous melanoma. A total of 13 trials were identified for which IPD were available (Table 1): 11 trials (n = 5826) comparing IFN with observation and two ECOG trials (n = 989), the E1694 trial (n = 882), which compared IFN with the GM2-KLH/QS-21 (GMK) vaccine, and the E2696 trial (n = 107), which compared IFN administered concomitantly or sequentially with GMK vaccine, with GMK vaccine alone.

For the external validation set, we selected randomized adjuvant therapy trials in adult patients with resected stage II–III cutaneous melanoma for which IPD were not available but RFS and OS hazard ratios (HRs) had been published. A total of 13 studies were identified and included in the external validation set (Table 2).

#### Statistical Methods

RFS was defined as the time from random assignment until first recurrence (loco-regional or distant metastasis) or death due to any cause, whichever was observed first. OS was the time from random assignment until death due to any cause. Survival distributions for RFS and OS were estimated using the Kaplan-Meier technique and compared using the log-rank two-tailed test stratified by study. Treatment hazard ratio, 95% or 99% confidence interval (CI), and heterogeneity of the hazard ratios among the studies were evaluated using classical meta-analytical techniques (see the Supplementary Methods, available online) and depicted using forest plots.

Surrogacy of RFS for OS was assessed through association measures following a two-stage approach estimating the association between the surrogate and the true end point and between the treatment effects on these end points (28). At the first stage, a copula function was used to model the joint distribution of RFS and OS. Three copulas (Clayton, Hougaard, and Plackett) were investigated. The one that fitted the best to the data, using Akaike’s Information Criterion, was chosen. To model the effect of treatment on the marginal distributions of RFS and OS in the joint survivor function, the Cox proportional hazards model was used. The proportional hazards assumption was confirmed by visual inspection of the curves. Spearman’s ρ, as a function of copula parameter, was calculated to measure the patient-level association between RFS and OS. At the second stage, a linear regression model was fitted through the treatment effects on RFS and OS estimated from the first stage with adjustment for the measurement error of these estimates (29). The coefficient of determination of the linear regression, $R^2$, was calculated to measure the trial-level association between RFS and OS. $R^2$ quantifies the proportion of variance in the effects of treatment on OS that is explained by the variance in the treatment effects on RFS. The surrogate threshold effect (STE) was defined as the minimum value of treatment effect on the surrogate end point, for which the predicted effect on the true end point would be different from zero (26). For its computation, see the Supplementary Methods (available online).

An internal validation was carried out using a leave-one-out cross-validation strategy (Supplementary Methods,
available online) (30). An external validation was performed using published hazard ratios of RFS and OS. We used the weighted regression analysis to predict the hazard ratio for OS based on the published hazard ratio for RFS.

All statistical tests were two-sided and a P value of less than .05 was considered statistically significant.

**Results**

**Patient Characteristics**

A total of 6815 patients entered the 13 trials with IPD: 3952 in the IFN arm and 2863 in the No IFN arm (Table 1). The median age was 50 years. Among them, 75% were in disease stage III, median Breslow thickness was 3 mm, and 39% had ulcerated tumors. There were only 55 patients with stage I and 17 patients with resected metastatic disease, of whom 16 were included in the ECOG 2696 trial, which was subsequently excluded from most of the analyses.

The median follow-up was 7.0 vs 6.5 years in the IFN vs no IFN groups and ranged from 4.1 years (Dermatologic Cooperative Oncology Group trial) to 17 years (ECOG 1684) (Table 1).

**Outcomes: RFS, OS, and OS Post-RFS Event**

Among 13 trials, a total of 4423 RFS events were reported: 4250 recurrences and 173 deaths without recurrences. Overall, the median RFS was 1.9 years, ranging from 1.2 to 3.3 years, according to the study (Table 1), and was 2.2 years (95% CI = 1.9 to 2.4 years) in the IFN arm and 1.6 years (95% CI = 1.4 to 1.8 years) in the No IFN arm (Figure 1A). The estimated impact of IFN on the five-year RFS rate was 3.6% (Table 3). The estimated hazard ratio for RFS stratified by study was 0.87 (95% CI = 0.82 to 0.93) (Table 2).

### Table 1. Training set of 13 trials in adjuvant melanoma with individual patient data available

<table>
<thead>
<tr>
<th>Trial identification</th>
<th>Trial question</th>
<th>Sample size, No.</th>
<th>Median* RFS, y</th>
<th>Median* OS, y</th>
<th>Median* follow-up, y</th>
<th>Estimated HR&lt;sub&gt;RFS&lt;/sub&gt;</th>
<th>Estimated HR&lt;sub&gt;OS&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECOG 1684 (17)</td>
<td>High-dose IFN vs observation</td>
<td>287</td>
<td>1.4</td>
<td>3.3</td>
<td>17.0</td>
<td>0.76</td>
<td>0.84</td>
</tr>
<tr>
<td>ECOG 1690 (34)</td>
<td>High-dose vs low-dose IFN vs observation</td>
<td>642</td>
<td>2.4</td>
<td>7.0</td>
<td>10.7</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>NCCTG 83-7052 (35)</td>
<td>High-dose IFN vs observation</td>
<td>264</td>
<td>2.1</td>
<td>5.5</td>
<td>15.1</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>EORTC 18952 (36)</td>
<td>Intermediate-dose IFN (1 or 2 y) vs observation</td>
<td>1388</td>
<td>1.8</td>
<td>4.5</td>
<td>4.7</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>WHO16 (37)</td>
<td>Low-dose IFN vs observation</td>
<td>444</td>
<td>1.2</td>
<td>2.7</td>
<td>6.3</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>UKCCCR AIM-High (38)</td>
<td>Low-dose IFN vs observation</td>
<td>674</td>
<td>1.3</td>
<td>3.8</td>
<td>5.7</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>DeCOG (39)</td>
<td>Low-dose IFN vs observation</td>
<td>293</td>
<td>1.3</td>
<td>4.0</td>
<td>4.1</td>
<td>0.72</td>
<td>0.63</td>
</tr>
<tr>
<td>Scottish MG (40)</td>
<td>Low-dose IFN vs observation</td>
<td>94</td>
<td>1.4</td>
<td>2.4</td>
<td>6.6</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>EORTC 18871 (41)</td>
<td>Very low-dose IFN vs observation</td>
<td>281</td>
<td>1.3</td>
<td>3.5</td>
<td>8.0</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>DKG 80-1 (41)</td>
<td>Very low-dose IFN vs observation</td>
<td>203</td>
<td>2.0</td>
<td>5.3</td>
<td>8.0</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td>EORTC 18991 (20)</td>
<td>PEG-IFN (5 y) vs observation</td>
<td>1256</td>
<td>2.5</td>
<td>5.8</td>
<td>7.6</td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>ECOG 1694 (42)</td>
<td>High-dose IFN vs GMK vaccine</td>
<td>882</td>
<td>3.3</td>
<td>6.8</td>
<td>6.1</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>ECOG 2696 (43)</td>
<td>High-dose IFN + GMK vaccine vs GMK vaccine</td>
<td>107</td>
<td>2.4</td>
<td>7.3</td>
<td>7.1</td>
<td>0.72</td>
<td>1.11</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>6815</td>
<td>1.9</td>
<td>4.7</td>
<td>6.8</td>
<td>0.87</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*All treatment groups pooled together. GMK — GM2-KLH/QS-21; HR OS = hazard ratio IFN vs no IFN for overall survival; HR RFS = hazard ratio IFN vs no IFN for relapse-free survival; IFN = interferon; PEG-IFN = pegylated interferon.

### Table 2. External validation set of 3 trials in adjuvant melanoma without individual patient data*

<table>
<thead>
<tr>
<th>No.</th>
<th>Trial identification</th>
<th>Trial question</th>
<th>Sample size</th>
<th>Observed HR&lt;sub&gt;RFS&lt;/sub&gt;</th>
<th>Observed HR&lt;sub&gt;OS&lt;/sub&gt;</th>
<th>Predicted HR&lt;sub&gt;OS&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Middleton (44)</td>
<td>1-y vs 1-mo HD-IFN</td>
<td>194</td>
<td>0.59</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>Austrian MMCG (45)</td>
<td>ID-IFN vs observation</td>
<td>311</td>
<td>0.62</td>
<td>0.83</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>French CGM (19)</td>
<td>ID-IFN vs observation</td>
<td>487</td>
<td>0.75</td>
<td>0.72</td>
<td>0.83</td>
</tr>
<tr>
<td>4</td>
<td>Nordic IFN 1 y (46)</td>
<td>1-y ID-IFN vs observation</td>
<td>569</td>
<td>0.77</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>5</td>
<td>China (47)</td>
<td>1-y vs 1-mo HD-IFN</td>
<td>158</td>
<td>0.81</td>
<td>0.69</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>Sunbelt Trial (48)</td>
<td>HD-IFN vs observation</td>
<td>218</td>
<td>0.82</td>
<td>1.07</td>
<td>0.88</td>
</tr>
<tr>
<td>7</td>
<td>Nordic IFN 2 y (46)</td>
<td>2-y ID-IFN vs observation</td>
<td>570</td>
<td>0.83</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>EADO 2001/CMII Trial (49)</td>
<td>3-y PEG IFN vs 1.5-y LD-IFN</td>
<td>896</td>
<td>0.91</td>
<td>1.09</td>
<td>0.94</td>
</tr>
<tr>
<td>9</td>
<td>HeCOG (50)</td>
<td>1-y vs 1-mo HD-IFN</td>
<td>364</td>
<td>0.94</td>
<td>1.1</td>
<td>0.96</td>
</tr>
<tr>
<td>10</td>
<td>DeCOG-MM-ADJ-0 (51)</td>
<td>IFN vs IL-2</td>
<td>223</td>
<td>1.0</td>
<td>1.0</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>DeCOG-MM-ADJ-3 (52)</td>
<td>LD-IFN with or without a modified</td>
<td>650</td>
<td>1.0</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>DeCOG-MM-ADJ-2 (39)</td>
<td>LD-IFN + DTIC vs observation</td>
<td>294</td>
<td>1.01</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>13</td>
<td>DeCOG-MM-ADJ-4 (53)</td>
<td>5-y LD IFN vs 1.5-y LD-IFN</td>
<td>840</td>
<td>1.05</td>
<td>1.03</td>
<td>1.03</td>
</tr>
</tbody>
</table>

*GMK — GM2-KLH/QS-21; HD — high-dose; ID — intermediate-dose; IFN — interferon; LD — low-dose; observed HR<sub>RFS</sub> = hazard ratio for IFN vs no IFN comparison regarding for relapse-free survival as indicated in each publication; observed HR<sub>OS</sub> = hazard ratio for IFN vs No IFN comparison regarding overall survival as indicated in each publication; PEG-IFN — pegylated interferon; predicted HR<sub>OS</sub> = hazard ratio for IFN vs no IFN comparison as predicted by the model included in this paper, given the observed HR<sub>RFS</sub> as indicated in each publication.
Figure 1. Kaplan-Meier plots by treatment group, according to (A) relapse-free survival and (B) overall survival. N = number of patients; O = number of observed events.
### Table 3. Summary of results for the training set*

<table>
<thead>
<tr>
<th>Result</th>
<th>All patients (n = 6815)</th>
<th>All but ECOG 2696 (n = 6708)</th>
<th>No vaccine trials (n = 5826)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No IFN (95% CI), %</td>
<td>IFN (95% CI), %</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td></td>
<td>No IFN (95% CI), %</td>
<td>IFN (95% CI), %</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td></td>
<td>No IFN (95% CI), %</td>
<td>IFN (95% CI), %</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>5-y RFS</td>
<td>35.2 (33.4 to 37.0)</td>
<td>38.8 (37.3 to 40.4)</td>
<td>0.87 (0.82 to 0.93)</td>
</tr>
<tr>
<td>5-y OS</td>
<td>47.2 (45.3 to 49.0)</td>
<td>50.3 (48.6 to 51.9)</td>
<td>0.91 (0.85 to 0.97)</td>
</tr>
<tr>
<td>$p$ (95% CI)</td>
<td>0.90 (0.90 to 0.90)</td>
<td>0.89 (0.88 to 0.89)</td>
<td>0.91 (0.81 to 1.01)</td>
</tr>
<tr>
<td>$R^2$ (95% CI)</td>
<td>0.24† (−0.16 to 0.65)</td>
<td>0.91 (0.81 to 1.01)</td>
<td>0.77</td>
</tr>
<tr>
<td>Surrogate threshold effect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Hazard ratio for IFN vs no IFN comparison was estimated using meta-analytical methods (see the Supplementary Methods, available online). Best copula used: Clayton (first analysis, based on all patients) or Plackett copula (two remaining analyses). $p$ = individual-level correlation estimate; CI = confidence interval; ECOG: Eastern Cooperative Oncology Group; HR = hazard ratio; IFN = interferon; OS = overall survival; $R^2$ = trial-level correlation estimate; RFS = relapse-free survival.

†Unreliable estimate (large standard errors).
‡A surrogate end point is considered valid at the trial level if $R^2$ is sufficiently close to 1 and is considered valid at a patient level if Spearman’s $p$ is sufficiently close to 1 (13). The SAS macro of Burzykowski et al. was used (http://ibis.stat.be/software/surrogate).

---

In order to determine whether the model is robust, we cross-validated it using the leave-one-out method. Figure 3 shows that the estimated trial-level association between RFS and OS was extremely high, indicating a large patient-level association between RFS treatment and OS (Table 3). In contrast, the estimated trial-level association $R^2$ between RFS and OS treatment differences was very low. The latter estimate was extremely unreliable, with a large standard error (SE), and hence a wide 95% confidence interval.

In the two-stage procedure, the Clayton copula provided the best fit (Supplementary Table 1, available online). After excluding the ECOG 2696 trial, the Plackett copula provided the best fit (Supplementary Table 2, available online). This also indicated that there was an important association between the RFS and OS treatment differences. The STE was computed from the intersection of the horizontal line representing a hazard ratio of 1 for OS (null hypothesis of no treatment difference) with the upper prediction limit; the vertical line representing a hazard ratio of 1 for RFS (null hypothesis of no treatment difference) with the upper prediction limit. This line represents the natural log transformation of the hazard ratio for RFS and OS and $\exp(0.0106 + 0.9874 \ln(HRRFS))$, where $HRRFS = \exp(0.0106 + 0.9874 \ln(HRRFS))$. The STE corresponds to a hazard ratio for OS as a function of the hazard ratio for RFS: $HRRFS = \exp(0.0106 + 0.9874 \ln(HRRFS))$. $HROS = \exp(0.0106 + 0.9874 \ln(HRRFS))$, where $HROS$ represents the natural log transformation of the hazard ratio for OS and $HRRFS$ represents the natural log transformation of the hazard ratio for RFS. This model indicates that the risk reduction due to IFN is 27% in the No IFN arm (Figure 1B). In contrast, the estimated trial-level association $R^2$ between RFS and OS treatment differences was extremely low, with a very high $SE$, and hence a very wide 95% confidence interval. The latter estimate was extremely unreliable, with a large standard error (SE), and hence a wide 95% confidence interval.

---

*S. Suciu et al.*
Figure 2. Treatment effects on overall survival (OS) vs treatment effects on relapse-free survival (RFS; 12 trials, $n = 6708$, Plackett copula) and prediction of hazard ratio for OS for the European Organisation for Research and Treatment of Cancer study 18071 based on the initial estimate of a hazard ratio for RFS of 0.75. Model equation: $HR_{\text{OS}} = \exp (0.0106 + 0.9874 \times \ln (HR_{\text{RFS}}))$. $HR_{\text{OS}}$ = hazard ratio for IFN vs no IFN comparison regarding overall survival; $HR_{\text{RFS}}$ = hazard ratio for IFN vs no IFN comparison regarding relapse-free survival; OS = overall survival; RFS = relapse-free survival.

Figure 3. Internal validation through leave-one-out analysis: observed hazard ratio (HR) for OS for left-out trial (True HR) vs predicted hazard ratio for OS (Pred HR) and 95% prediction interval (Pred interval) for predicted hazard ratio for OS. To assess model accuracy, a leave-one-out cross-validation strategy was used: each unit of analysis was left out once, and the linear model was then constructed from scratch using the remaining data (30). This model was then re-applied to the left-out study in order to compare the predicted and observed treatment effect on OS. Based on the linear regression models, a 95% prediction interval was calculated for each study. $HR_{\text{OS}}$ = hazard ratio for IFN vs no IFN comparison regarding overall survival; $HR_{\text{RFS}}$ = hazard ratio for IFN vs no IFN comparison regarding relapse-free survival; IFN = interferon; OS = overall survival.
the observed hazard ratios for OS of the 12 trials were within the 95% confidence interval of the predicted hazard ratio for OS.

Out of 13 trials without IPD available, all but one hazard ratio for OS were included in the corresponding 95% prediction intervals, which could be constructed based on the model (Figure 4).

Prediction of $H_{ROS}$ Based on $H_{RFS}$ for the EORTC 18071 Trial

For the EORTC 18071 trial, with a median follow-up of 2.7 years at the time of the final analysis of RFS, the estimated hazard ratio for RFS was 0.75 (95% CI = 0.64 to 0.90) and the three-year RFS rate improvement was 11.7% (25). Based on the model indicated above, linking the hazard ratio for RFS and hazard ratio for OS, the predicted hazard ratio for OS was 0.76. With a median follow-up of 5.3 years, the estimated hazard ratio for OS was 0.72 (95.1% CI = 0.58 to 0.88), and the five-year OS rate improvement was 11% (P = .001) (27).

Discussion

In this study, we evaluated whether RFS is a valid surrogate end point for OS in adjuvant interferon melanoma studies. Over several decades, the regulatory bodies (FDA and European Medicines Agency) assessed the value of IFN as an adjuvant treatment for melanoma based on efficacy evidence, mainly OS and RFS, and toxicity. Many randomized IFN trials, differing in interferon type and dose, treatment duration, design, and follow-up, have been reported. High-dose IFN was approved in 1996 based on the results of the ECOG E1684 trial in stage Ib/III melanoma patients, which showed a statistically significant impact on both RFS and OS (17). With longer follow-up (median = 12.6 years), updated results showed that the RFS benefit was maintained (HR = 0.72, P = .02) but the impact on OS decreased (HR increased from 0.67 to 0.82) due, in part, to competing causes of death (18). In Europe, low-dose IFN was approved based on a study in stage II patients that showed a statistically significant impact on RFS and a borderline significant impact on OS (19). In 2011, the FDA approved pegylated IFN for stage III melanoma patients based on the EORTC 18991 trial (20). With a median follow-up of 3.8 years, pegylated IFN treatment statistically significantly prolonged RFS (HR = 0.82, 95% CI = 0.71 to 0.96, P = .011) but not OS. Updated results with a median follow-up of 7.6 years showed that the impact on RFS had decreased (HR = 0.87, 95% CI = 0.76 to 1.00) (21).

Based on collected IPD for patients randomized in 13 adjuvant IFN trials, we showed that RFS was highly predictive of OS at the patient level. Such strong correlation cannot solely be explained by 173 patients (3.9% of the RFS events) who died without documented recurrence, but rather by the very poor outcome of advanced melanoma at that time, with a median survival of less than one year. Overall, the impact of IFN across the aggregated regimens was modest regarding RFS (HR$_{RFS}$ = 0.87, 95% CI = 0.82 to 0.93) and OS (HR$_{OS}$ = 0.91, 95% CI = 0.85 to 0.97). At the trial level, a good concordance between the magnitude of the hazard ratio for RFS and hazard ratio for OS was observed in all but one study. Indeed, the ECOG 2696 phase II trial enrolled both metastatic resectable stage III and IV disease and reported discrepant treatment differences between RFS and OS. Keeping this study in the model resulted into a low and an unreliable $R^2$ trial-level correlation between hazard ratio for RFS and hazard ratio for OS.

Based on the remaining 12 trials, the individual-level association between RFS and OS remained high (0.89). In addition, there was a high trial-level association between the effects of adjuvant IFN on RFS and on OS. The estimated $R^2$ was high.
ARTICLE

mab (200 mg flat dose) to high-dose IFN or ipilimumab, has

The SWOG S1404 study (NCT02506153), comparing pembrolizu-

mined. A number of studies have already completed accrual,

This was accepted by both European and US authorities.

were a secondary end point in the same respective populations.

A grant was provided by Bristol-Myers Squibb.

of RFS. The estimated STE was 0.77, suggest-

was the primary or coprimary end point. In the EORTC 18071 trial, which compared high-dose ipilimumab (10 mg/kg) with placebo, the primary end point was RFS. In 2015, the FDA approved high-dose ipilimumab based on its statistically significant impact on RFS (P = .0013); the OS results were still immature at the time (25). The observed hazard ratio for RFS was 0.75, lower than the estimated hazard ratio for RFS of 0.88 provided by the IFN vs observation IPD meta-analysis (Table 3) and lower than the STE (0.77). The recent publication of the OS results confirmed the hazard ratio for RFS (0.76, 95% CI = 0.64 to 0.89), and the estimated hazard ratio for OS was 0.72 (95.1% CI = 0.58 to 0.88, P = .001) (27), lower than 0.76, as pre-

dicted by the model. This suggests that the strong correlation between RFS and OS observed in the IFN trials is also correct for ipilimumab, a CTLA-4 inhibitor.

Postrecurrence treatment is a potential confounder for OS for all treatment modalities. In the last five years, the advent of new therapies has resulted in an improvement of median OS, from nine to 25–30 months in metastatic melanoma. The ongo-
ing evolution of treatment algorithms, including sequential application of these modalities, could weaken the surrogacy of RFS for OS, both at the patient and trial levels, as has happened in breast cancer and myeloma (5,31,32,33). However, this was not seen in the EORTC 18071 ipilimumab trial, where 23.5% of patients who had an RFS event on placebo received subsequent ipilimumab, 9.7% received a PD-1 inhibitor, and 27.2% received BRAF-directed therapy. Furthermore, the two-year OS rate post-

RFS event was approximately 50% in both arms.

Our study indicates that a positive impact of the initial adju-
vant treatment on the RFS as compared with observation/pla-

celbo should represent sufficient evidence to the regulatory bodies to approve an adjuvant melanoma treatment, especially if the safety profile is acceptable. In the ECOG-ACRIN US Intergroup study (NCT01274338), which compared ipilimumab 10 mg/kg with 3 mg/kg and high-dose IFN, the trial adopted coprimary end points of RFS and OS. In the new EORTC 1325 (KEYNOTE-054) phase III double-blind trial (NCT02362594), which compared pembrolizumab (200 mg flat dose) with pla-

celbo after complete resection of high-risk stage III melanoma, the coprimary end points were RFS in the overall population; in the subgroup of patients with PD-L1-positive expression, OS was a secondary end point in the same respective populations. This was accepted by both European and US authorities.

Whether RFS will correlate with OS for the newer checkpoint inhibitors compared with an active drug remains to be determined. A number of studies have already completed accrual, and the initial results will be available in the next one to two years. Checkmate-238 (NCT02388906), which comparing nivol-

mab 3 mg/kg i.v. q. two weeks to ipilimumab 10 mg/kg, com-

completed accrual in September 2015; its primary end point is RFS. The SWOG S1404 study (NCT02506153), comparing pembrolizu-

mab (200 mg flat dose) to high-dose IFN or ipilimumab, has coprimary end points of RFS and OS. Our data would suggest that a statistically significant impact of newer checkpoint inhibitors on RFS, with a hazard ratio for RFS of 0.77 or less, may pre-
dict an OS benefit. This could be used to support early approval based on RFS.

A number of studies (eg, BRIM-8 [NCT01667419], Combi-AD [NCT01682083]) comparing BRAF-directed therapy with placebo in the adjuvant setting have completed accrual, and their RFS treatment results are still pending. While there is good evidence of a correlation between RFS and OS for targeted therapy in the metastatic disease setting (9), it remains to be seen if this is the case in the adjuvant setting; however, it is not an unreasonable suggestion.

The meta-analytic approach used in this paper has informed claims of surrogacy, or lack thereof, in many cancer settings (1,4,5,7,9–13,26,28,30,32). This approach makes use of all rele-

vant patient-level data. A theoretical limitation of this approach is that it establishes correlation rather than causation. A practi-
cal limitation is that a range of treatment effects must be ob-

served for the regression analysis to be possible, which was the case in the present situation (Figure 2).

In conclusion, using the IPD database collected from resected stage II–III patients entered in 13 adjuvant IFN trials, we demonstrated that RFS is a surrogate end point for OS only at the patient level. In 11 trials comparing IFN with observation and in a large phase III trial comparing IFN with vaccination, RFS appeared to be a surrogate end point for OS at the trial level as well. This was validated internally and externally and also confirmed successfully in a large adjuvant ipilimumab trial. Further improvement of postrecurrence treatment outcome might weaken this surrogacy. The evaluation of efficacy results in terms of RFS, along with the safety profile, should continue to allow the regulatory authorities and medical community to properly appraise the value of new adjuvant melanoma treat-

ments. Overall survival evaluation, performed later, should sup-

port the initial RFS findings.

Funding

A grant was provided by Bristol-Myers Squibb.

Notes

The protocol was written by the members of the European Organisation for Research and Treatment of Cancer (EORTC) Headquarters team and a Bristol-Myers Squibb representative. IPD were collected and computerized at International Melanoma Meta-Analysis Collaborative Group (IMMACG) secre-
tariat and transferred to the EORTC Headquarters. Data were analyzed at the EORTC Headquarters. An initial draft of the manuscript was prepared by the EORTC Headquarters team. All the authors participated in the revision and finalization of the manuscript and made the decision to submit the manuscript for publication.

The following groups (listed alphabetically, with the names of investigators/statisticians nominated by each group) supplied individual patient data for this project: Eastern Cooperative Oncology Group, United States: John Kirkwood, Sandra Lee; EORTC: Alexander Eggermont, Stefan Suciu; Dermatologic Cooperative Oncology Group: Claus Garbe, Michael Kressig; North Central Cancer Treatment Group: Svetomir N. Markovic, Vera Suman; Scottish MG: David Cameron, Valerie Doherty, Rona Mackie; United Kingdom Co-ordinating
Committee on Cancer Research AIM-High: Paul Lorigan, Barry Hancock, Lesley Turner; World Health Organization 16 Melanoma Group: Natalic Cascaric (deceased), Rosaria Bufalino.

References


