

ARTICLE

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

Stefan Suci, 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 Schetzen, 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 Tübingen, Tübingen, 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 Suci, PhD, European Organisation for Research and Treatment of Cancer Headquarters, Avenue E. Mounier 83/11, 1200, Brussels, Belgium (e-mail: stefan.suci@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 (ρ) 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 ρ was high, R^2 was low and could not reliably be estimated. Based on the 12 trials, ρ 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.

Received: December 8, 2016; Revised: April 14, 2017; Accepted: May 26, 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

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 or 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- α (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,

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

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

*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*

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

*GMK = GM2-KLH/QS-21; HD = high-dose; ID = intermediate-dose; IFN = interferon; LD = low-dose; observed HR_{RFS} = hazard ratio for IFN vs no IFN comparison regarding relapse-free survival as indicated in each publication; observed HR_{OS} = hazard ratio for IFN vs No IFN comparison regarding overall survival as indicated in each publication; PEG-IFN = pegylated interferon; predicted HR_{OS} = hazard ratio for IFN vs no IFN comparison as predicted by the model included in this paper, given the observed HR_{RFS} as indicated in each publication.

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)

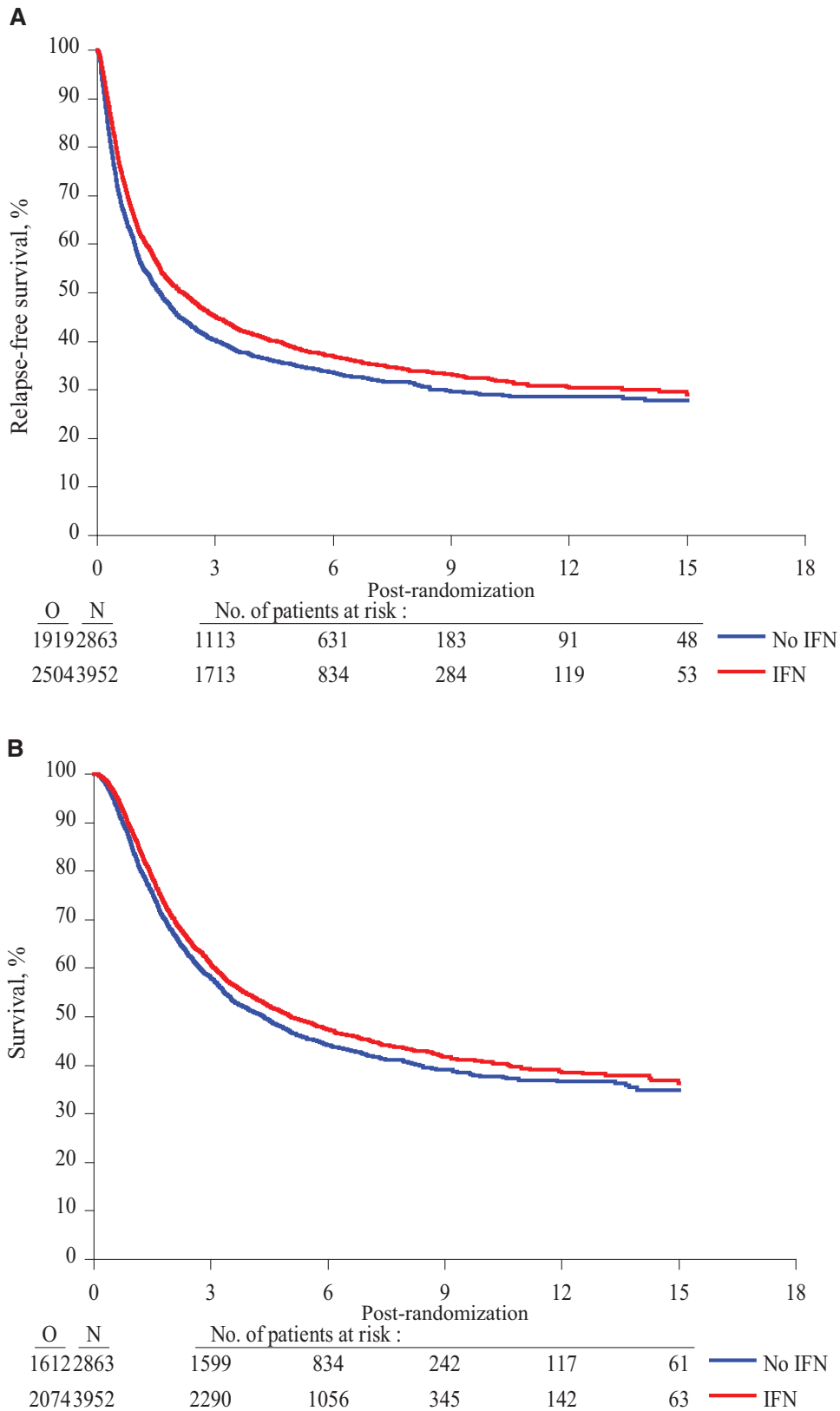


Figure 1. Kaplan-Meier plots by treatment group, according to (A) relapse-free survival and (B) overall survival. N = number of patients; O = number of observed events.

Table 3. Summary of results for the training set*

Result	All patients (n = 6815)			All but ECOG 2696 (n = 6708)			No vaccine trials (n = 5826)		
	No IFN (95% CI), %	IFN (95% CI), %	HR (95% CI)	No IFN (95% CI), %	IFN (95% CI), %	HR (95% CI)	No IFN (95% CI), %	IFN (95% CI), %	HR (95% CI)
5-y RFS	35.2 (33.4 to 37.0)	38.8 (37.3 to 40.4)	0.87 (0.82 to 0.93)	35.2 (33.4 to 37.0)	38.8 (37.2 to 40.4)	0.88 (0.82 to 0.93)	34.0 (32.1 to 35.9)	38.0 (36.3 to 39.6)	0.88 (0.83 to 0.94)
5-y OS	47.2 (45.3 to 49.0)	50.3 (48.6 to 51.9)	0.91 (0.85 to 0.97)	47.0 (45.1 to 48.9)	50.2 (48.5 to 51.8)	0.91 (0.85 to 0.97)	45.4 (43.4 to 47.5)	49.0 (47.3 to 50.7)	0.91 (0.85 to 0.98)
ρ (95% CI)	0.90 (0.90 to 0.90)			0.89 (0.88 to 0.89)			0.89 (0.88 to 0.89)		
R^2 (95% CI)	0.24† (-0.16 to 0.65)			0.91 (0.81 to 1.01)			0.92 (0.84 to 1.01)		
Surrogate threshold effect for HR_{RFS}^\ddagger				0.77					0.79

*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). ρ = 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 ρ is sufficiently close to 1 (13). The SAS macro of Burzykowski et al. was used (<http://ibio.stat.be/software/surrogate>).

(Table 3), and no statistically significant heterogeneity among the studies was detected (Supplementary Figure 1, available online).

A total of 3686 patients were followed until death. Overall, the median OS was 4.7 years, ranging from 2.7 to 7.3 years among the 13 trials (Table 1), and the median OS rates were 5.0 years (95% CI = 4.7 to 5.7) in the IFN arm and 4.4 years (95% CI = 3.9 to 4.8) in the No IFN arm (Figure 1B). The improvement in the five-year OS rate was 3.1% (Table 3). The hazard ratio for OS stratified by study was 0.91 (95% CI = 0.85 to 0.97), and no statistically significant heterogeneity among the studies was detected (Supplementary Figure 2, available online). However, for the small vaccination trial (ECOG 2696), the largest difference between the overall median RFS and median OS was observed: 2.4 vs 7.3 years (Table 1). Also, this was the only trial where discrepant results were observed between treatment differences regarding RFS ($HR_{RFS} = 0.72$) and OS ($HR_{OS} = 1.11$).

After an RFS event, the median OS was 0.9 years in both treatment groups and the two-year OS rate was 26.4% in the IFN arm vs 27.1% in the No IFN arm (Supplementary Figure 3, available online). This also indicated that there was an important association between the RFS and OS at the patient level because the vast majority of patients who experienced an RFS event ultimately died.

Surrogacy Results

In the two-stage procedure, the Clayton copula provided the best fit to data (Supplementary Table 1, available online). Spearman's ρ was extremely high, indicating a large patient-level association between RFS and OS (Table 3). In contrast, the trial-level association R^2 between the 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.

After excluding the ECOG 2696 trial, the Plackett copula provided the best fit to data (Supplementary Table 2, available online). In this setting, the Spearman's ρ remained very high and the estimated trial-level correlation R^2 was also high (Table 3). Hence, based on the 12-trial training set, a regression model could be developed for the hazard ratio for OS as a function of the hazard ratio for RFS:

$HR_{OS} = \exp(0.0106 + 0.9874 \times \ln(HR_{RFS}))$, where $\ln(HR_{RFS})$ represents the natural log transformation of the hazard ratio for RFS and "exp" is the exponential power, inverse of the ln function. This model indicates that the risk reduction due to IFN treatment was approximately the same for OS and for RFS.

Figure 2 shows the linear regression corresponding to the model, which may be used to predict treatment effects of future interferon trials on OS, from the one observed on RFS, and the 95% prediction limits of the regression line.

The STE was computed from the intersection of the horizontal line representing a hazard ratio for 1 for OS (null hypothesis of no treatment difference) with the upper prediction limit; from that intersection, the vertical line cuts the x-axis at the STE. The STE corresponded to a hazard ratio for RFS of 0.77. Thus, in a future interferon adjuvant trial, in order to predict a positive treatment effect on OS, a hazard ratio for RFS of 0.77 or less would need to be achieved.

Internal and External Validation

In order to determine whether the model is robust, we cross-validated it using the leave-one-out method. Figure 3 shows that

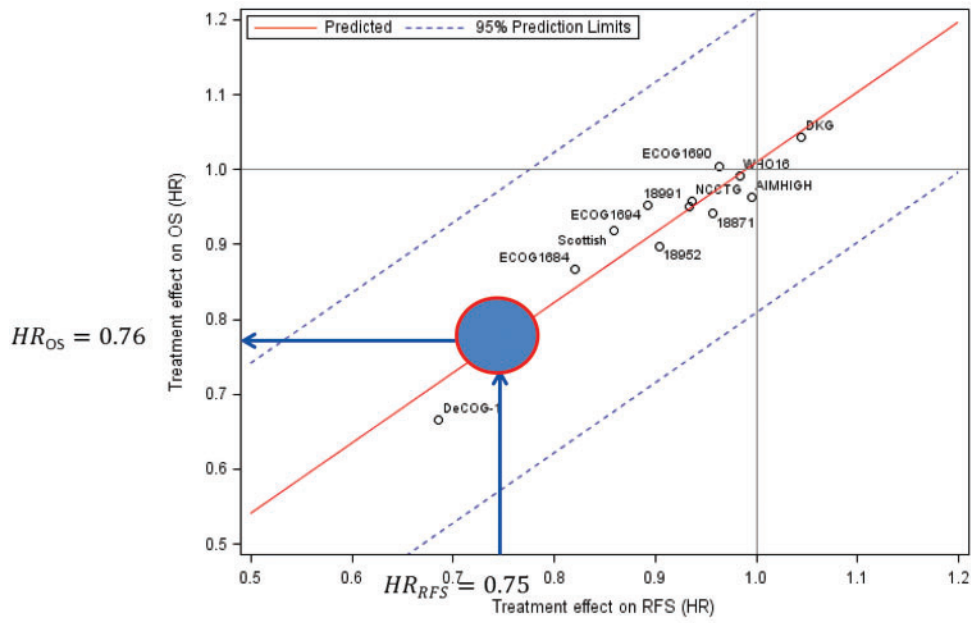


Figure 2. Treatment effects on overall survival (OS) vs treatment effects on relapse-free survival (RFS; 12 trials, n = 6708, Plackett copula) and prediction of hazard ratio for OS for the European Organisation for Research and Treatment of Cancer study 18071 based on the initial estimate of a hazard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

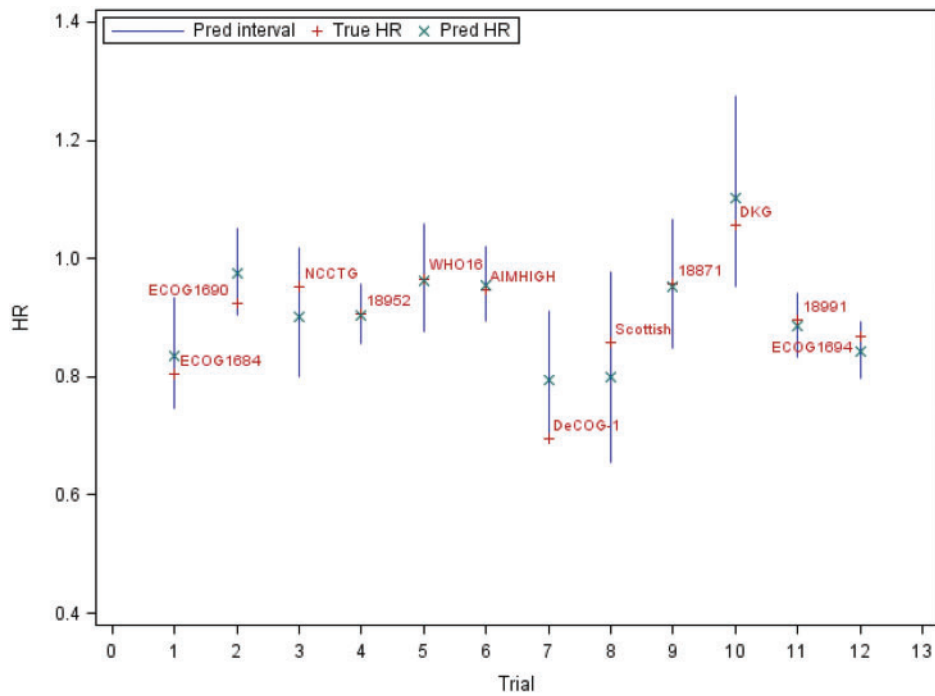


Figure 3. Internal validation through leave-one-out analysis: observed hazard ratio (HR) for OS for left-out trial (True HR) vs predicted hazard ratio for OS (Pred HR) and 95% prediction interval (Pred interval) for predicted hazard ratio for OS. To assess model accuracy, a leave-one-out cross-validation 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_{OS} = hazard ratio for IFN vs no IFN comparison regarding overall survival; HR_{RFS} = hazard ratio for IFN vs no IFN comparison regarding relapse-free survival; IFN = interferon; OS = overall survival.

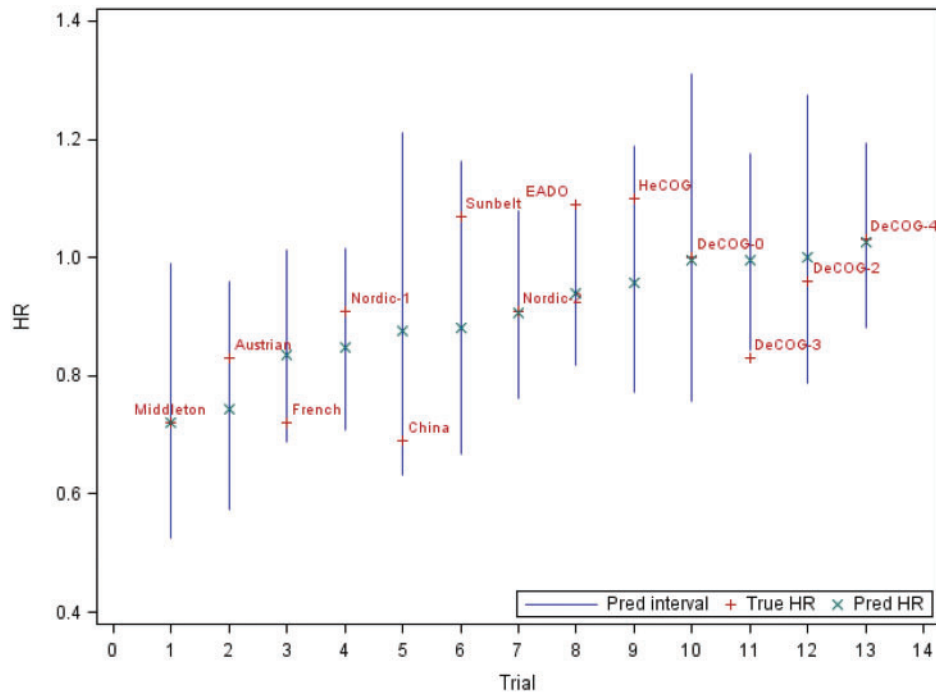


Figure 4. External validation: observed hazard ratio (HR) for OS for external trial (True HR) vs predicted hazard ratio for OS (Pred HR) and 95% prediction interval (Pred Interval) for predicted hazard ratio for OS. Predicted hazard ratio for OS was predicted by the model included in this paper, given the observed hazard ratio for relapse-free survival as indicated in each publication. HR_{OS} = hazard ratio for IFN vs no IFN comparison regarding overall 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 HR_{OS} Based on HR_{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 IIb/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

(0.91), indicating that a large proportion of the variability in the treatment effect on OS is explained by the variability of the treatment effects on RFS. The estimated STE was 0.77, suggesting that for future similar trials, a true treatment effect hazard ratio on RFS of 0.77 or less should predict a treatment effect on OS. In principle, a trial should be designed for detecting a relatively low hazard ratio for RFS (eg, 0.70) with a very high statistical power (eg, >90%) and a higher hazard ratio for OS (eg, 0.75) with an acceptable power (eg, 80%).

The transformational impact of immune checkpoint inhibitors on the outcomes for patients with advanced melanoma has led to these being evaluated in the adjuvant setting. Most studies have RFS as either 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 predicted 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 ongoing 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 adjuvant treatment on the RFS as compared with observation/placebo 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 placebo 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 nivolumab 3 mg/kg i.v. q. two weeks to ipilimumab 10 mg/kg, completed accrual in September 2015; its primary end point is RFS. The SWOG S1404 study (NCT02506153), comparing pembrolizumab (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 predict 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 relevant patient-level data. A theoretical limitation of this approach is that it establishes correlation rather than causation. A practical limitation is that a range of treatment effects must be observed 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 treatments. Overall survival evaluation, performed later, should support 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) secretariat 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 Suciuc; 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: Natale Cascinelli (deceased), Rosaria Bufalino.

References

- Burzykowski T, Molenberghs G, Buyse M. *The Evaluation of Surrogate Endpoints*. New York: Springer; 2005.
- Leo A Di, Bleiberg H, Buyse M. Overall survival is not a realistic end point for clinical trials of new drugs in advanced solid tumors: A critical assessment based on recently reported phase III trials in colorectal and breast cancer. *J Clin Oncol*. 2003;21(10):2045–2047.
- Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst*. 2009;101(23):1642–1649.
- Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: A meta-analysis. *Lancet*. 2000;356(9227):373–378.
- Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol*. 2008;26(12):1987–1992.
- Ives NJ, Stowe RL, Lorigan P, Wheatley K. Chemotherapy compared with bi-chemotherapy for the treatment of metastatic melanoma: A meta-analysis of 18 trials involving 2,621 patients. *J Clin Oncol*. 2007;25(34):5426–5434.
- Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol*. 2007;25(33):5218–5224.
- Korn EL, Liu PY, Lee SJ, et al. Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. *J Clin Oncol*. 2008;26(4):527–534.
- Flaherty KT, Hennis M, Lee SJ, et al. Surrogate endpoints for overall survival in metastatic melanoma: A meta-analysis of randomised controlled trials. *Lancet Oncol*. 2014;15(3):297–304.
- Halabi S, Vogelzang NJ, Ou S-S, Owzar K, Archer L, Small EJ. Progression-free survival as a predictor of overall survival in men with castrate-resistant prostate cancer. *J Clin Oncol*. 2009;27(17):2766–2771.
- Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Control Clin Trials*. 2002;23(6):607–625.
- Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med*. 1997;16(17):1965–1982.
- Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000;1(1):49–67.
- Eggermont AMM, Spatz A, Robert C. Cutaneous melanoma. *Lancet*. 2014;383(9919):816–827.
- Balch CM, Gershenwald JE, Soong S-J, et al. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol*. 2009;27(36):6199–6206.
- Balch CM, Gershenwald JE, Soong S-J, et al. Multivariate analysis of prognostic factors among 2,313 patients with stage III melanoma: Comparison of nodal micrometastases versus macrometastases. *J Clin Oncol*. 2010;28(14):2452–2459.
- Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group Trial EST 1684. *J Clin Oncol*. 1996;14(1):7–17.
- Kirkwood JM, Manola J, Ibrahim J, Sondak V, Ernstoff MS, Rao U. A pooled analysis of Eastern Cooperative Oncology Group and Intergroup trials of adjuvant high-dose interferon for melanoma. *Clin Cancer Res*. 2004;10(5):1670–1677.
- Grob JJ, Dreno B, de la Salmoniere P, et al. Randomised trial of interferon α -2a as adjuvant therapy in resected primary melanoma thicker than 1.5 mm without clinically detectable node metastases. *Lancet*. 1998;351(9120):1905–1910.
- Eggermont AMM, Suciú S, Santinami M, et al. Adjuvant therapy with pegylated interferon alfa-2b versus observation alone in resected stage III melanoma: Final results of EORTC 18991, a randomised phase III trial. *Lancet*. 2008;372(9633):117–126.
- Eggermont AMM, Suciú S, Testori A, et al. Long-term results of the randomized phase III trial EORTC 18991 of adjuvant therapy with pegylated interferon alfa-2b versus observation in resected stage III melanoma. *J Clin Oncol*. 2012;30(31):3810–3818.
- Wheatley K, Ives N, Eggermont A, et al. Interferon- α as adjuvant therapy for melanoma: An individual patient data meta-analysis of randomised trials. In: *ASCO Annual Meeting*; 2007; Chicago (IL), USA: Proceedings Part I. Vol. 25, No. 18S.
- 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.
- Robert C, Schachter J, Long GV, et al. Pembrolizumab versus ipilimumab in advanced melanoma. *N Engl J Med*. 2015;372(26):2521–2532.
- Eggermont AMM, Chiarion-Sileni V, Grob J-J, et al. Adjuvant ipilimumab versus placebo after complete resection of high-risk stage III melanoma (EORTC 18071): A randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2015;16(5):522–530.
- Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*. 2006;5(3):173–186.
- Eggermont AMM, Chiarion-Sileni V, Grob J-J, et al. Prolonged survival in stage III melanoma with ipilimumab adjuvant therapy. *N Engl J Med*. 2016;375(19):1845–1855.
- Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J R Stat Soc Ser C (Applied Stat)*. 2001;50(4):405–422.
- van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Stat Med*. 2002;21(4):589–624.
- Michiels S, Le Maître A, Buyse M, et al. Surrogate endpoints for overall survival in locally advanced head and neck cancer: Meta-analyses of individual patient data. *Lancet Oncol*. 2009;10(4):341–350.
- Ng R, Pond GR, Tang PA, MacIntosh PW, Siu LL, Chen EX. Correlation of changes between 2-year disease-free survival and 5-year overall survival in adjuvant breast cancer trials from 1966 to 2006. *Ann Oncol*. 2008;19(3):481–486.
- Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Oncotargets Ther*. 2012;5:287–296.
- Interferon as therapy for multiple myeloma: An individual patient data overview of 24 randomized trials and 4012 patients. *Br J Haematol*. 2001;113(4):1020–1034.
- Kirkwood JM, Ibrahim JG, Sondak VK, et al. High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. *J Clin Oncol*. 2000;18(12):2444–2458.
- Creagan ET, Dalton RJ, Ahmann DL, et al. Randomized, surgical adjuvant clinical trial of recombinant interferon alfa-2a in selected patients with malignant melanoma. *J Clin Oncol*. 1995;13(11):2776–2783.
- Eggermont AMM, Suciú S, MacKie R, et al. Post-surgery adjuvant therapy with intermediate doses of interferon alfa 2b versus observation in patients with stage IIB/III melanoma (EORTC 18952): Randomised controlled trial. *Lancet*. 2005;366(9492):1189–1196.
- Cascinelli N, Belli F, MacKie RM, Santinami M, Bufalino R, Morabito A. Effect of long-term adjuvant therapy with interferon alpha-2a in patients with regional node metastases from cutaneous melanoma: A randomised trial. *Lancet*. 2001;358(9285):866–869.
- Hancock BW, Wheatley K, Harris S, et al. Adjuvant interferon in high-risk melanoma: The AIM HIGH Study—United Kingdom Coordinating Committee on Cancer Research randomized study of adjuvant low-dose extended-duration interferon Alfa-2a in high-risk resected malignant melanoma. *J Clin Oncol*. 2004;22(1):53–61.
- Garbe C, Radny P, Linse R, et al. Adjuvant low-dose interferon [alpha]2a with or without dacarbazine compared with surgery alone: A prospective-randomized phase III DeCOG trial in melanoma patients with regional lymph node metastasis. *Ann Oncol*. 2008;19(6):1195–1201.
- Cameron DA, Combleet MC, Mackie RM, et al. Adjuvant interferon alpha 2b in high risk melanoma – the Scottish study. *Br J Cancer*. 2001;84(9):1146–1149.
- Kleeberg UR, Suciú S, Bröcker EB, et al. Final results of the EORTC 18871/DKG 80-1 randomised phase III trial. *Eur J Cancer*. 2004;40(3):390–402.
- Kirkwood JM, Ibrahim JG, Sosman JA, et al. High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: Results of intergroup trial E1694/S9512/C509801. *J Clin Oncol*. 2001;19(9):2370–2380.
- Kirkwood JM, Ibrahim J, Lawson DH, et al. High-dose interferon alfa-2b does not diminish antibody response to GM2 vaccination in patients with resected melanoma: Results of the Multicenter Eastern Cooperative Oncology Group Phase II Trial E2696. *J Clin Oncol*. 2001;19(5):1430–1436.
- Payne MJ, Argyropoulou K, Lorigan P, et al. Phase II pilot study of intravenous high-dose interferon with or without maintenance treatment in melanoma at high risk of recurrence. *J Clin Oncol*. 2013;32(3):185–190.
- Pehamberger H, Soyer HP, Steiner A, et al. Adjuvant interferon alfa-2a treatment in resected primary stage II cutaneous melanoma. Austrian Malignant Melanoma Cooperative Group. *J Clin Oncol*. 1998;16(4):1425.
- Hansson J, Aamdal S, Bastholt L, et al. Two different durations of adjuvant therapy with intermediate-dose interferon alfa-2b in patients with high-risk melanoma (Nordic IFN trial): A randomised phase 3 trial. *Lancet Oncol*. 2011;12(2):144–152.
- Mao L, Si L, Chi Z, et al. A randomised phase II trial of 1 month versus 1 year of adjuvant high-dose interferon α -2b in high-risk acral melanoma patients. *Eur J Cancer*. 2011;47(10):1498–1503.
- McMasters KM Reintgen DS, et al. Final results of the Sunbelt Melanoma Trial. 2008;(26: suppl 9003).
- Grob JJ, Jouary T, Dréno B, et al. Adjuvant therapy with pegylated interferon alfa-2b (36 months) versus low-dose interferon alfa-2b (18 months) in melanoma patients without macrometastatic nodes: An open-label, randomised, phase 3 European Association for Dermato-Oncology (EADO) study. *Eur J Cancer*. 2013;49(1):166–174.

50. Pectasides D, Dafni U, Bafaloukos D, et al. Randomized phase III study of 1 month versus 1 year of adjuvant high-dose interferon alfa-2b in patients with resected high-risk melanoma. *J Clin Oncol*. 2009;27(6):939–944.
51. Hauschild A, Weichenthal M, Balda B-R, et al. Prospective randomized trial of interferon alfa-2b and interleukin-2 as adjuvant treatment for resected intermediate- and high-risk primary melanoma without clinically detectable node metastasis. *J Clin Oncol*. 2003;21(15):2883–2888.
52. Hauschild A, Weichenthal M, Rass K, et al. Prospective randomized multicenter adjuvant dermatologic cooperative oncology group trial of low-dose interferon alfa-2b with or without a modified high-dose interferon alfa-2b induction phase in patients with lymph node-negative melanoma. *J Clin Oncol*. 2009;27(21):3496–3502.
53. Hauschild A, Weichenthal M, Rass K, et al. Efficacy of low-dose interferon (alpha)2a 18 versus 60 months of treatment in patients with primary melanoma of ≥ 1.5 mm tumor thickness: Results of a randomized phase III DeCOG trial. *J Clin Oncol*. 2010;28(5):841–846.