














## Reevaluating Disease-Free Survival as an Endpoint vs Overall Survival in Stage III Adjuvant Colon Cancer Trials

Jun Yin, PhD <sup>1,†</sup> Mohamed E. Salem, MD,<sup>2,†</sup> Jesse G. Dixon, MS <sup>1</sup> Zhaohui Jin, MD <sup>3</sup>  
 Romain Cohen, MD, PhD <sup>4</sup> Aimery DeGramont, MD,<sup>5</sup> Eric Van Cutsem, MD,<sup>6</sup> Julien Taieb, MD, PhD <sup>7</sup>  
 Steven R. Alberts, MD <sup>3</sup> Norman Wolmark, MD,<sup>8</sup> Hans-Joachim Schmoll, MD,<sup>9</sup> Leonard B. Saltz, MD,<sup>10</sup>  
 Thomas J. George, MD <sup>11</sup> Richard R. M. Goldberg, MD <sup>12</sup> Rachel Kerr, MD <sup>13</sup> Sara Lonardi, MD <sup>14</sup>  
 Takayuki Yoshino, MD,<sup>15</sup> Greg Yothers, PhD <sup>16</sup> Axel Grothey, MD,<sup>17</sup> Thierry Andre, MD <sup>4</sup> Qian Shi, PhD <sup>1,\*</sup>

<sup>1</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; <sup>2</sup>Levine Cancer Institute, Charlotte, NC, USA; <sup>3</sup>Department of Oncology, Mayo Clinic, Rochester, MN, USA; <sup>4</sup>Department of Medical Oncology, Saint-Antoine Hospital, Sorbonne University, Paris, France; <sup>5</sup>Department of Medical Oncology, Franco-British Institute, Levallois-Perret, France; <sup>6</sup>Digestive Oncology, University Hospitals Gasthuisberg Leuven and University of Leuven, Leuven, Belgium; <sup>7</sup>Sorbonne Paris Cité, Paris Descartes University Georges Pompidou European Hospital, Paris, France; <sup>8</sup>Department of Clinical Trials, Alleghany Health Network, Pittsburgh, PA, USA; <sup>9</sup>University Clinic Halle (Saale), Halle Martin-Luther-University, Halle, Germany; <sup>10</sup>Memorial Sloan Kettering Cancer Center, New York, NY, USA; <sup>11</sup>University of Florida Health Cancer Center, Gainesville, FL, USA; <sup>12</sup>West Virginia University Cancer Institute, the Mary Babb Randolph Cancer Center, Morgantown, WV, USA; <sup>13</sup>Adjuvant Colorectal Cancer Group, University of Oxford, Oxford, UK; <sup>14</sup>Department of Oncology, Veneto Institute of Oncology IRCCS, Padova PD, Italy; <sup>15</sup>Department of Gastrointestinal Oncology, National Cancer Center, Tokyo, Japan; <sup>16</sup>Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA and <sup>17</sup>West Cancer Center, Germantown, TN, USA;

<sup>†</sup>Authors contributed equally to this work.

\*Correspondence to: Qian Shi, PhD, Department of Quantitative Health Sciences, Mayo Clinic, 200 First St, SW, Rochester, MN 55905, USA (e-mail: shi.qian2@mayo.edu).

### Abstract

**Background:** Disease-free survival (DFS) with a 3-year median follow-up (3-year DFS) was validated as a surrogate for overall survival (OS) with a 5-year median follow-up (5-year OS) in adjuvant chemotherapy colon cancer (CC) trials. Recent data show further improvements in OS and survival after recurrence in patients who received adjuvant FOLFOX. Hence, reevaluation of the association between DFS and OS and determination of the optimal follow-up duration of OS to aid its utility in future adjuvant trials are needed. **Methods:** Individual patient data from 9 randomized studies conducted between 1998 and 2009 were included; 3 trials tested biologics. Trial-level surrogacy examining the correlation of treatment effect estimates of 3-year DFS with 5 to 6.5-year OS was evaluated using both linear regression ( $R^2_{WLS}$ ) and Copula bivariate ( $R^2_{Copula}$ ) models and reported with 95% confidence intervals (CIs). For  $R^2$ , a value closer to 1 indicates a stronger correlation. **Results:** Data from a total of 18 396 patients were analyzed (median age = 59 years; 54.0% male), with 54.1% having low-risk tumors (T1-3 and N1), 31.6% KRAS mutated, 12.3% BRAF mutated, and 12.4% microsatellite instability high or deficient mismatch repair tumors. Trial-level correlation between 3-year DFS and 5-year OS remained strong ( $R^2_{WLS} = 0.82$ , 95% CI = 0.67 to 0.98;  $R^2_{Copula} = 0.92$ , 95% CI = 0.83 to 1.00) and increased as the median follow-up of OS extended. Analyses limited to trials that tested biologics showed consistent results. **Conclusions:** Three-year DFS remains a validated surrogate endpoint for 5-year OS in adjuvant CC trials. The correlation was likely strengthened with 6 years of follow-up for OS.

Colon cancer (CC) is the third-most frequently diagnosed type of cancer and the fourth leading cause of cancer-related death worldwide (1). Three to 6 months of adjuvant chemotherapy with fluoropyrimidines and oxaliplatin improves survival and is the standard of care for stage III and some stage II CC

patients. Although overall survival (OS) is the gold-standard endpoint in oncology trials to evaluate the efficacy of novel therapies, it requires extended trial duration and is frequently affected by nonprotocol treatments. Because the objective of adjuvant therapy is to avoid disease recurrence, disease-free

survival (DFS) has been developed as a surrogate endpoint for OS. DFS reduces trial duration and cost with the potential to accelerate the time from therapeutic innovation to patient care.

DFS with a median of 3 years of follow-up (3-year DFS) was validated as a surrogate endpoint of 5-year OS in trials with 5-fluorouracil (5-FU)-based regimens (2–4). The current standard of care that was established around 2004, the FOLFOX regimen (folinic acid, 5-FU, and oxaliplatin), was approved based on the treatment effect established from the 3-year DFS endpoint (5,6). Nevertheless, the 3-year DFS surrogacy may be hampered by the major therapeutic improvements that have been realized for metastatic diseases, including improved supportive care, enhanced metastasectomies, and novel targeted therapies based on the cancers' molecular profiles (7). Notably, immunotherapy is currently under investigation for localized CC with microsatellite instability (MSI), a biomarker predictive of the efficacy of immune checkpoint inhibitors with a controversial prognostic impact in metastatic settings. BRAF-targeted treatment or intensified adjuvant chemotherapy regimens for BRAF<sup>V600E+</sup> CC patients are proposed and ongoing.

Recent research from the Adjuvant Colon Cancer Endpoints (ACCENT) group showed improved survival after recurrence in recent years compared with patients treated with an oxaliplatin-based regimen over a previous 10-year period (1998–2009), potentially due to changes in options for salvage treatment at relapse (7). Based on these observations, the aim of this study was to reevaluate 3-year DFS as a surrogate endpoint for 5-year OS based on studies testing chemotherapy with or without biologics in stage III CCs, with a particular focus on RAS/RAF mutations, and MSI. The second objective of this study was to identify the optimal follow-up duration of OS for evaluating the benefits of adjuvant therapy in the current treatment era.

## Methods

### Trial Selection and Comparison Definition

We selected randomized, multicenter trials that enrolled stage III CC patients and had a median follow-up of over 5 years in the ACCENT database. The studies with a median follow-up shorter than 5 years were not suitable due to insufficient follow-up for OS endpoints considered in this analysis. In addition, single-agent 5-FU or capecitabine trials were excluded because they are no longer part of the standard of care. In total, 9 studies from the ACCENT database met the selection criteria as of July 2019. The meta-analytic unit for surrogacy estimation was predefined as the comparison between 2 arms (experimental vs control) nested within trials. When more than 1 experimental regimen was evaluated within a trial, the control arm patients were duplicated to form multiple 2-arm comparisons. In addition, 2 studies (PETACC8 and N0147) (8,9) evaluated anti-epidermal growth factor receptor agents with KRAS status. KRAS wild-type and mutant tumors were considered as separate cohorts for testing the treatment effects of anti-epidermal growth factor receptor agents. As such, 14 comparison units were predefined (Table 1). Most of the comparisons were not included in prior ACCENT surrogacy analyses (2–4).

Individual trials were approved through countries' mechanisms at the time trials were conducted. All patients provided written, informed consent at enrollment in the respective trials. The ACCENT database collaboration research protocol was approved by the Mayo Clinic Institution Review Board. Individual

patient data of all trials were collected, and the analyses were conducted at an independent statistical center at Mayo Clinic (Rochester, MN, USA).

### Endpoints

The 3-year DFS was evaluated as a potential surrogate candidate for OS. Although 5 years of OS follow-up has been considered in most of the recent adjuvant trials in CC, Salem et al. (7) reported prolonged OS in modern-era trials. With the consideration of data availability, various median follow-up durations (5, 6, and 6.5 years) for OS were evaluated in order to identify optimal OS follow-up times. DFS was defined as the time from random assignment to disease recurrence or death from any cause, whichever occurred first. The primary clinical endpoint was OS, defined as the time from random assignment to death from any cause. Because different patients were enrolled at different calendar dates, efforts were made to reproduce the actual clinical trial procedure and evaluate various OS follow-up times. For each specific OS follow-up time, the patients were censored within each trial at the point in time after full accrual, for which the median follow-up was the specified OS follow-up time estimated using the reverse Kaplan-Meier method.

### Statistical Analysis

The effect of treatment and 95% confidence intervals (CI) for DFS and OS were quantified using hazard ratios (HRs) estimated by the Cox proportional hazard or Copula bivariate survival models (15).

The standard 2-level surrogacy evaluation method, using individual patient data from a large collection of randomized trials, was applied to evaluate the surrogacy. The primary surrogacy evaluation method was trial-level surrogacy, which measured how precisely treatment effect on the true endpoint may be predicted based on observed treatment effects on the surrogate endpoint. At the trial level, 2 commonly used surrogacy measures were considered:  $R^2_{\text{Copula}}$  and  $R^2_{\text{WLS}}$  (which is based on the weighted least square [WLS] regression method using the number of patients included per comparison as weights), where  $R^2_{\text{Copula}}$  considers patient-level correlation between the 2 endpoints and  $R^2_{\text{WLS}}$  does not. Values of these 2  $R^2$  measures approaching 1 indicate a strong correlation between DFS and OS at the trial level. The predefined rule for declaring trial-level surrogacy required  $R^2_{\text{WLS}}$  or  $R^2_{\text{Copula}}$  of 0.80 or greater with a lower 95% confidence interval bound of at least 0.6 and neither estimate less than 0.7. Supplemental trial-level surrogacy measures included the surrogate threshold effect, the minimum treatment effect on the DFS endpoint required to confidently predict a statistically significant treatment effect on OS in a future trial (16). For patient-level surrogacy, the correlation between DFS and OS endpoints was quantified by the rank correlation coefficient ( $\rho_{\text{Copula}}$ ) via a bivariate copula model. Patient-level correlation was considered a supportive but not sufficient condition for surrogacy validation. The rank correlation coefficient closer to 1.0 indicated a stronger correlation at the patient level. Analyses were conducted with SAS (version 9.4; SAS Institute, Cary, NC, USA) and R (version 2).

To evaluate the performance of the predicting treatment effect on OS, based on the measured DFS, leave-one-out cross-validation was applied. Each time, 1 of the 14 comparisons was used as the testing dataset, and the other comparisons were used as the training dataset. Predictive models were built based on the training dataset to predict the hazard ratio of OS based on the hazard ratio of DFS. Predicted values of the hazard ratio

Table 1. Trials included<sup>a</sup>

Study	Citation	Accrual year	Median OS follow-up, y	Comparison	Sample size, No.
MOSAIC	André et al., 2004 (5)	1998-2001	9.6	C: LV5FU2 E: LV5FU2 + oxaliplatin	672 669
C07	Kuebler et al., 2007 (6)	2000-2002	8.0	C: 5FU/LV E: FLOX	860 853
C89803	Saltz et al., 2007 (10)	1999-2001	7.7	C: 5FU/LV E: 5FU/LV + irinotecan	606 620
PETACC3	Van Cutsem et al., 2009 (11)	1999-2002	5.7	C: 5FU/FA E: 5FU/FA + irinotecan	1139 1120
XELOXA	Schmoll et al., 2015 (12)	2003-2004	7.0	C: 5FU/LV E: capecitabine + oxaliplatin	909 907
C08	Allegra et al., 2011 (13)	2004-2006	6.4	C: mFOLFOX6 E: mFOLFOX6 ++ bevacizumab	994 985
PETACC8	Taieb et al., 2014 (8)	2005-2009	7.5	Comparison 1: KRAS WT C: FOLFOX4 E: FOLFOX4 + cetuximab	644 652
				Comparison 2: KRAS MT C: FOLFOX4 E: FOLFOX4 + cetuximab	322 316
				Comparison 3: KRAS Unknown C: FOLFOX4 E: FOLFOX4 + cetuximab	312 312
N0147	Alberts et al., 2012 (9)	2004-2009	6.6	Comparison 1: KRAS WT C: mFOLFOX6 E: mFOLFOX6 + cetuximab	951 944
				Comparison 2: KRAS MT C: mFOLFOX6 E: mFOLFOX6 + cetuximab	391 344
				Comparison 3: KRAS WT C: FOLFIRI E: FOLFIRI + cetuximab	71 27
AVANT	de Gramont et al., 2012 (14)	2004-2007	6.2	Comparison 1: C: FOLFOX4 E: FOLFOX4 + bevacizumab	901 948
				Comparison 2: C: FOLFOX4 E: XELOX + bevacizumab	901 927

<sup>a</sup>C = control arm; E = experimental arm; FA = folinic acid; FLOX = 5-fluorouracil, leucovorin, oxaliplatin; FOLFIRI = folinic acid, 5-fluorouracil, irinotecan; FOLFOX = folinic acid, 5-fluorouracil, oxaliplatin; KRAS MT = KRAS mutation; KRAS WT = KRAS wild type; LV5FU2 = leucovorin and 5-fluorouracil; XELOX = capecitabine plus oxaliplatin.

of OS were generated by the model in the testing dataset to compare them with the actual values. To evaluate the effect of outliers, each time, 1 of the 14 comparisons was removed, and the remaining comparisons were used to estimate the surrogacy relationship.

Confirmatory surrogacy analyses were conducted using trials involving biologic agents (bevacizumab or cetuximab) and trials without irinotecan. Exploratory analyses were performed within subpopulations defined by age, risk group classified by T and N stage (low risk defined as T1-3 and N1; high risk defined as T4 or N2), primary tumor location, and mutation status (BRAF, KRAS, and MSI).

## Results

### Trial and Patient Characteristics

A total of 18 396 patients with a median age of 59 years were included (32.6% older than 65 years); 54.0% were male, 54.1% had low-risk stage III tumor (ie, T1-3 and N1), 45.9% had high-risk stage III CC (ie, T4 or N2), 44.7% had only proximal

(cecum, ascending colon, hepatic flexure, and transverse colon), 54.7% had only distal (splenic flexure, descending colon and sigmoid colon, rectosigmoid segment, rectum), and 0.6% had both proximal and distal primary tumors, and 31.6%, 12.3%, and 12.4% harbored KRAS mutations, BRAF mutations, and high levels of MSI/deficient mismatch repair phenotypes, respectively. Patients were included according to the intention-to-treat principle whenever possible. Overall, patient and disease characteristics were well balanced between the experimental and control arms (Table 2), with the largest differences being that patients on control arms were less likely to have 12 and more examined lymph nodes (59.2% vs 61.1%), more likely to be in the low-risk category (45.1% vs 46.7%; also T4 stage, 13.2% vs 14.7%), and more likely to have distal only disease (55.4% vs 53.9%) compared with patients on the experimental arms.

The median follow-up, estimated using the reverse Kaplan-Meier method, was 80.1 months among patients alive at the time of data cutoff, with 5, 6, and 6.5 years follow-up estimates of 90.4%, 68.0%, and 55.9% of patients, respectively. A total of 6641 patients (36.1%) received biologic agents in combination with chemotherapy.

Table 2. Patient and disease characteristics<sup>a</sup>

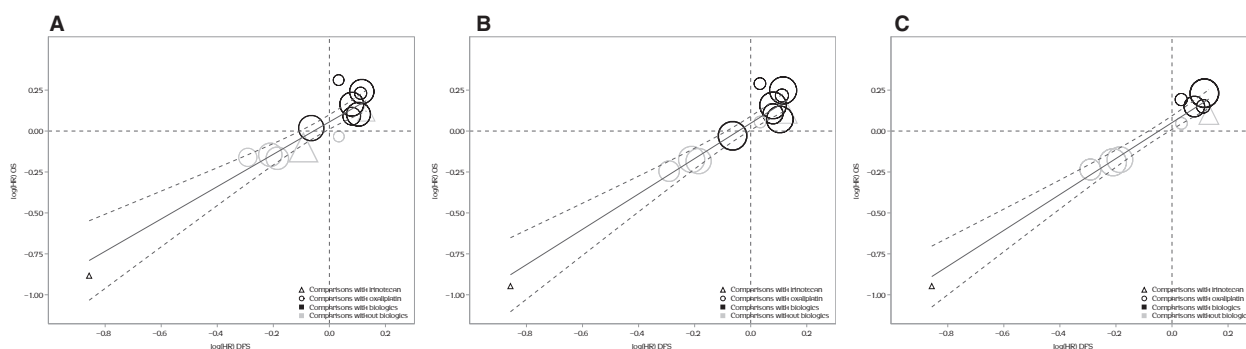
Characteristic	Control arm (n = 8722)	Experimental arm (n = 9624)	Total (n = 18 396)
Age, No. (%), y			
<50	1775 (20.2)	1969 (20.5)	3744 (20.4)
50-64	4114 (46.9)	4548 (47.3)	8662 (47.1)
≥65	2883 (32.9)	3107 (32.3)	5990 (32.6)
BMI, No. (%), kg/m <sup>2</sup>			
<18.5	180 (2.1)	227 (2.4)	407 (2.2)
18.5-25	3394 (38.8)	3820 (39.9)	7214 (39.4)
>25	5172 (59.1)	5538 (57.8)	10710 (58.4)
Missing	26	39	65
Sex, No. (%)			
Female	4026 (45.9)	4438 (46.1)	8464 (46.0)
Male	4746 (54.1)	5186 (53.9)	9932 (54.0)
Tumor grade, No. (%)			
Low grade	6529 (78.9)	7288 (79.8)	13817 (79.4)
High grade	1745 (21.1)	1850 (20.2)	3595 (20.6)
Missing	498	486	984
Performance score, No. (%)			
0	6645 (76.4)	7327 (76.8)	13972 (76.6)
1	1952 (22.4)	2084 (21.9)	4036 (22.1)
2	100 (1.1)	124 (1.3)	224 (1.2)
Missing	75	89	164
T-Stage, No. (%)			
T1 or 2	1049 (12.4)	1094 (11.8)	2143 (12.1)
T3	6281 (74.4)	6833 (73.5)	13114 (73.9)
T4	1111 (13.2)	1368 (14.7)	2479 (14.0)
Missing	331	329	660
N-Stage, No. (%)			
N1	5526 (63.0)	5964 (62.0)	11490 (62.5)
N2	3246 (37.0)	3660 (38.0)	6906 (37.5)
Tumor location, No. (%)			
Distal only	3749 (55.4)	3618 (53.9)	7367 (54.7)
Proximal only	2975 (44.0)	3051 (45.5)	6026 (44.7)
Distal and proximal	38 (0.6)	42 (0.6)	80 (0.6)
Missing	2010	2913	4923
Risk group, No. (%)			
High risk (T4 or N2)	3867 (45.1)	4397 (46.7)	8264 (45.9)
Low risk (T1-3 and N1)	4700 (54.9)	5022 (53.3)	9722 (54.1)
Missing	205	205	410
Examined nodes, No. (%)			
0-7	1175 (15.7)	1252 (15.0)	2427 (15.4)
8-12	1877 (25.1)	1985 (23.9)	3862 (24.4)
>12	4428 (59.2)	5082 (61.1)	9510 (60.2)
Missing	1292	1305	2597
Age			
No.	8772	9624	18 396
Mean (SD)	58.5 (10.89)	58.3 (10.94)	58.4 (10.92)
Median (range)	59 (19.0, 85.0)	59 (17.0, 86.0)	59 (17.0, 86.0)
MSI/MMR status, No. (%)			
MSS/MSI-low/pMMR	3951 (87.5)	3922 (87.6)	7873 (87.6)
MSI-high/dMMR	563 (12.5)	556 (12.4)	1119 (12.4)
Missing	4258	5146	9404
KRAS status, No. (%)			
WT	2715 (67.9)	2707 (68.9)	5422 (68.4)
MT	1282 (32.1)	1223 (31.1)	2505 (31.6)
Missing	4775	5694	10469
BRAF status, No. (%)			
WT	4068 (87.9)	3988 (87.5)	8056 (87.7)
MT	562 (12.1)	570 (12.5)	1132 (12.3)
Missing	4142	5066	9208

<sup>a</sup>BMI = body mass index; MMR = mismatch repair; MSI = microsatellite instability; MSS = microsatellite stable; MT = mutation; N = node; pMMR = proficient mismatch repair; T = tumor; WT = wild type.

**Table 3.** Trial- and individual patient-level surrogacy

True endpoint	No. of comparison units (No. of patients <sup>a</sup> )	$R^2_{WLS}$ (95% CI)	$R^2_{Copula}$ (95% CI)	$\rho_{Copula}$ (95% CI)	STE (HR)
<b>All trials included</b>					
5-y OS	14 (19 279)	0.82 (0.67 to 0.98)	0.92 (0.83 to 1.00)	0.90 (0.90 to 0.91)	0.79
6-y OS	13 (17 020)	0.88 (0.70 to 1.00)	0.97 (0.93 to 1.00)	0.91 (0.90 to 0.91)	0.80
6.5-y OS	10 (11 382)	0.94 (0.81 to 1.00)	0.99 (0.97 to 1.00)	0.91 (0.90 to 0.91)	0.80
<b>Excluding 1 small comparison (n &lt; 100)</b>					
5-y OS	13 (19 181)	0.79 (0.58 to 1.00)	0.74 (0.49 to 0.98)	0.90 (0.90 to 0.91)	—
6-y OS	12 (16 922)	0.86 (0.60 to 1.00)	0.89 (0.78 to 1.00)	0.91 (0.90 to 0.91)	—
6.5-y OS	9 (11 284)	0.93 (0.67 to 1.00)	0.97 (0.94 to 1.00)	0.91 (0.90 to 0.91)	—

<sup>a</sup>AVANT (14) trial included 2 experimental arms. The control arm patients (n = 901) were duplicated for surrogacy analyses. HR = hazard ratio; OS = overall survival;  $\rho_{Copula}$  = rank correlation coefficient; STE = surrogate threshold effect; WLS = weighted least squares.



**Figure 1.** Trial-level treatment effect correlation between disease-free survival (DFS) and overall survival (OS). A) Three-year DFS vs 5-year OS. B) Three-year DFS vs 6-year OS. C) Three-year DFS vs 6.5-year OS. The circle or triangle size is proportional to the number of patients in each comparison. The solid line indicates fitted weighted least squares (WLS) regression line; gray dashed lines indicate 95% prediction limits. Triangles indicate trial comparisons with irinotecan, circles indicate comparisons with oxaliplatin, and black and gray colors indicate comparisons with and without biologics, respectively. HR = hazard ratio.

### Trial- and Individual Patient-Level Surrogacy

As summarized in Table 3 and Figure 1, A, trial-level surrogacy for the 3-year DFS vs 5-year OS remained strong in the overall population ( $R^2_{WLS} = 0.82$ , 95% CI = 0.67 to 0.98;  $R^2_{Copula} = 0.92$ , 95% CI = 0.83 to 1.00) and met the predefined criteria for surrogacy. This indicates a strong prediction of the treatment effect (measured by HR) on the 5-year OS based on the observed treatment effect on the 3-year DFS. In addition, the patient-level correlation between the 3-year DFS and 5-year OS was strong, with a rank correlation coefficient of 0.90 (95% CI = 0.90 to 0.91; Table 3). The surrogate threshold effect had a hazard ratio of 0.79, which indicates that an observed hazard ratio of 0.79 for the 3-year DFS would predict a statistically significant treatment effect on the 5-year OS in a future trial. Therefore, the trial-level surrogacy for 3-year DFS vs 5-year OS remains validated.

Furthermore, surrogacy improves as the median follow-up increases. For the 6-year OS, the  $R^2_{WLS}$  estimate was 0.88 (95% CI = 0.70 to 1.00), and the  $R^2_{Copula}$  estimate was 0.97 (95% CI = 0.93 to 1.00) (Figure 1, B). For the 6.5-year OS, the  $R^2_{WLS}$  estimate was 0.94 (95% CI = 0.81 to 1.00), and the  $R^2_{Copula}$  estimate was 0.99 (95% CI = 0.97 to 1.00) (Figure 1, C). Table 3 shows the correlation between 3-year DFS and OS with different follow-up lengths.

Leave-one-out cross-validation demonstrated consistency between observed and predicted OS treatment effects for each comparison unit based on DFS, with different follow-up lengths (Supplementary Figure 1, available online). In most of the settings, the observed hazard ratio of OS lies within the 95%

confidence interval of the predicted values. In addition, no comparisons were identified as outliers when evaluating the reestimated  $R^2$  when 1 comparison at a time was excluded (Supplementary Figure 2, available online). The surrogacy estimates, when excluding 1 comparison at a time, remain strong and meet the predefined criteria in all settings.

### Subgroup Analyses

Table 4 includes the trial- and individual patient-level surrogacy estimates in exploratory analyses in the subgroups. Patient-level surrogacy remained strong for all subgroup analyses. Limiting to 6 comparisons of biologics, including bevacizumab or cetuximab, the trial-level surrogacy remained adequate and increased as the length of the follow-up of OS increased. Similar findings were observed when irinotecan was excluded. The number of comparisons and sample size were reduced in exploratory subgroup analyses, which could have led to wider confidence intervals of surrogacy estimates. However, the trial-level surrogacy point estimates remained adequate ( $\geq 0.7$  for at least 1 of the 2  $R^2$  measures) for most subgroups, such as subsets defined by age, tumor location, and KRAS and BRAF mutation status. The trial-level surrogacy estimates in subsets defined by risk groups were slightly below 0.7, but they increased again as the length of OS follow-up increased.

We further conducted a sensitivity analysis by excluding small comparisons having less than 100 total patients (Table 3). Only 1 comparison from N0147 was removed due to small

Table 4. Trial- and individual patient-level surrogacy for subgroups for 5, 6, and 6.5 years of median follow-up, respectively<sup>a</sup>

Subgroup and true endpoint	No. of comparison units (No. of patients)	R <sup>2</sup> <sub>WLS</sub> (95% CI)	R <sup>2</sup> <sub>Copula</sub> (95% CI)	ρ <sub>Copula</sub> (95% CI)
Age <65 y				
5-y OS	13 (12 943)	0.73 (0.52 to 0.94)	0.66 (0.37 to 0.96)	0.90 (0.90 to 0.91)
6-y OS	12 (11 382)	0.76 (0.51 to 1.00)	0.73 (0.46 to 0.99)	0.91 (0.90 to 0.91)
6.5-y OS	9 (7444)	0.83 (0.61 to 1.00)	0.81 (0.58 to 1.00)	0.91 (0.90 to 0.91)
Age ≥65 y				
5-y OS	13 (6238)	0.71 (0.47 to 0.94)	0.56 (0.21 to 0.92)	0.91 (0.91 to 0.92)
6-y OS	12 (5540)	0.78 (0.60 to 0.95)	0.69 (0.41 to 0.98)	0.92 (0.91 to 0.92)
6.5-y OS	9 (3840)	0.75 (0.49 to 1.00)	0.69 (0.36 to 1.00)	0.91 (0.90 to 0.92)
KRAS WT				
5-y OS	7 (5406)	0.86 (0.53 to 1.00)	0.95 (0.88 to 1.00)	0.91 (0.90 to 0.92)
6-y OS	6 (4856)	0.84 (0.39 to 1.00)	0.95 (0.87 to 1.00)	0.91 (0.90 to 0.92)
6.5-y OS	5 (4600)	0.85 (0.42 to 1.00)	0.96 (0.90 to 1.00)	0.91 (0.90 to 0.92)
KRAS MT				
5-y OS	6 (2491)	0.90 (0.71 to 1.00)	0.93 (0.81 to 1.00)	0.90 (0.89 to 0.92)
6-y OS	5 (2133)	0.87 (0.46 to 1.00)	0.90 (0.74 to 1.00)	0.91 (0.89 to 0.92)
6.5-y OS	4 (2009)	0.90 (0.34 to 1.00)	0.97 (0.91 to 1.00)	0.91 (0.89 to 0.92)
BRAF WT				
5-y OS	9 (7979)	0.70 (0.27 to 1.00)	0.71 (0.39 to 1.00)	0.89 (0.88 to 0.90)
6-y OS	8 (7142)	0.78 (0.38 to 1.00)	0.78 (0.51 to 1.00)	0.89 (0.88 to 0.90)
6.5-y OS	7 (5985)	0.83 (0.27 to 1.00)	0.80 (0.53 to 1.00)	0.89 (0.88 to 0.90)
BRAF MT				
5-y OS	7 (1095)	0.92 (0.78 to 1.00)	0.94 (0.86 to 1.00)	0.96 (0.95 to 0.97)
6-y OS	6 (1018)	0.94 (0.80 to 1.00)	0.96 (0.89 to 1.00)	0.96 (0.95 to 0.97)
6.5-y OS	5 (856)	0.92 (0.78 to 1.00)	0.96 (0.88 to 1.00)	0.96 (0.95 to 0.97)
High risk				
5-y OS	14 (8662)	0.60 (0.30 to 0.90)	0.56 (0.22 to 0.91)	0.89 (0.89 to 0.90)
6-y OS	13 (7661)	0.59 (0.28 to 0.90)	0.60 (0.27 to 0.94)	0.90 (0.89 to 0.90)
6.5-y OS	10 (5070)	0.68 (0.34 to 1.00)	0.65 (0.29 to 1.00)	0.89 (0.89 to 0.90)
Low risk				
5-y OS	12 (10 153)	0.64 (0.35 to 0.93)	0.60 (0.25 to 0.95)	0.90 (0.90 to 0.91)
6-y OS	11 (8896)	0.69 (0.43 to 0.94)	0.65 (0.32 to 0.98)	0.91 (0.90 to 0.92)
6.5-y OS	8 (5863)	0.73 (0.47 to 0.99)	0.64 (0.24 to 1.00)	0.91 (0.90 to 0.92)
dMMR				
5-y OS	6 (933)	0.86 (0.67 to 1.00)	0.96 (0.91 to 1.00)	0.93 (0.92 to 0.95)
6-y OS	5 (828)	0.87 (0.49 to 1.00)	0.94 (0.83 to 1.00)	0.93 (0.91 to 0.95)
6.5-y OS	4 (675)	0.90 (0.71 to 1.00)	0.97 (0.92 to 1.00)	0.93 (0.91 to 0.95)
pMMR				
5-y OS	9 (7756)	0.64 (0.16 to 1.00)	0.78 (0.52 to 1.00)	0.90 (0.89 to 0.91)
6-y OS	8 (6995)	0.66 (0.19 to 1.00)	0.80 (0.54 to 1.00)	0.90 (0.89 to 0.91)
6.5-y OS	7 (6052)	0.80 (0.29 to 1.00)	0.79 (0.52 to 1.00)	0.90 (0.89 to 0.91)
Distal				
5-y OS	11 (7277)	0.72 (0.51 to 0.93)	0.65 (0.32 to 0.98)	0.88 (0.87 to 0.89)
6-y OS	10 (5885)	0.75 (0.45 to 1.00)	0.62 (0.25 to 0.99)	0.88 (0.87 to 0.89)
6.5-y OS	8 (4896)	0.85 (0.60 to 1.00)	0.80 (0.56 to 1.00)	0.88 (0.87 to 0.89)
Proximal				
5-y OS	10 (5968)	0.67 (0.37 to 0.97)	0.54 (0.11 to 0.96)	0.92 (0.92 to 0.93)
6-y OS	9 (5116)	0.86 (0.74 to 0.98)	0.77 (0.51 to 1.00)	0.92 (0.92 to 0.93)
6.5-y OS	8 (4158)	0.86 (0.72 to 0.99)	0.77 (0.48 to 1.00)	0.93 (0.92 to 0.93)
Excluding IRI				
5-y OS	11 (15 696)	0.81 (0.57 to 1.00)	0.71 (0.41 to 1.00)	0.91 (0.90 to 0.91)
6-y OS	11 (15 696)	0.87 (0.65 to 1.00)	0.89 (0.78 to 1.00)	0.90 (0.90 to 0.91)
6.5-y OS	8 (10 058)	0.96 (0.77 to 1.00)	0.97 (0.94 to 1.00)	0.90 (0.90 to 0.91)
Biologics				
5-y OS	6 (6623)	0.82 (0.37 to 1.00)	0.97 (0.93 to 1.00)	0.90 (0.89 to 0.91)
6-y OS	6 (6623)	0.87 (0.41 to 1.00)	0.98 (0.95 to 1.00)	0.90 (0.89 to 0.91)
6.5-y OS	5 (4662)	0.94 (0.34 to 1.00)	1.00 (0.99 to 1.00)	0.90 (0.89 to 0.91)

<sup>a</sup>CI = confidence interval; dMMR = deficient mismatch repair; IRI = irinotecan; MT = mutation; OS = overall survival; pMMR = proficient mismatch repair; WLS = weighted least squares; WT = wild-type; ρ<sub>Copula</sub> = rank correlation coefficient.

sample size. Sensitivity results showed that the  $R^2_{WLS}$  was robust due to weighting by sample size, whereas  $R^2_{COPULA}$  was substantially reduced (from 0.92 to 0.74); however, trial-level surrogacy remained adequate ( $\geq 0.7$  for at least 1 of the 2  $R^2$  measures) and increased as the length of OS follow-up increased.

## Discussion

Here, we obtained and analyzed data from 9 major international trials conducted from 1998 to 2009, which have relatively mature OS follow-up (all with a median follow-up of  $>5$  years), and the majority were not included in the previous ACCENT surrogacy analysis (2–4). These analyses demonstrate that, for trials with stage III CC patients using oxaliplatin (or irinotecan)-based therapy with and without biologics, the association between DFS assessed after 3 years of median follow-up and OS was confirmed: 3-year DFS remains a valid surrogate endpoint for OS in stage III adjuvant CC trials. Compared with previous validation results based on testing 5-FU regimens published in 2005 (2) and 2007 (3), the correlation between 3-year DFS and 5-year OS in the current treatment era was weaker. Longer follow-up of OS was assessed in the current analyses, whereas only 5-year OS was assessed in the previous analyses. Additional analyses showed that the trial-level correlation increases numerically with longer follow-up of OS (6 and 6.5 years). This is consistent with previous findings from the MOSAIC study (17), which showed that the OS benefit of oxaliplatin-based adjuvant treatment increased over time and that at least 6 years of follow-up for OS is required to detect its benefit. This further highlights the need to consider 3-year DFS as the primary endpoint in future stage III CC trials.

In addition to the surrogacy analyses based on all comers, we conducted subgroup analyses to test the robustness of this surrogacy regarding subpopulations defined by biomarkers, age, tumor location, and risk groups. Overall, subpopulation analyses showed that subset patients depended on the mutation status of KRAS, BRAF, and MSI, and proximal or distal sites showed consistently strong levels of surrogacy at both trial and patient levels. However, the sample size of the BRAF-mutated and the MSI-high groups were limited. More importantly, most of the patients included in this analysis were less likely to receive targeted therapy for known mutations after recurrences given the time era when these trials were conducted. These newer targeted agents treating recurrence disease could potentially alter the correlation between DFS and OS. Thus, cautions should be considered when applying the 3-year DFS surrogacy results to trials with enriched enrollments according to these 2 populations, for example, the ongoing ATOMIC trial (NCT02912559) testing the treatment effect of standard chemotherapy plus Atezolizumab vs chemotherapy alone within stage III CCs with deficient DNA mismatch repair tumors. It is worth noting that the trial-level surrogacy is relatively weak in high-risk stage III patients. These results are of major interest because the number of randomized phase III trials dedicated to specific subgroups of stage III CC patients is growing. Additional examples include the ongoing IROCAS trial (NCT02967289) and ADAGE trial (NCT02355379). These studies should therefore be interpreted along with the results we provide here.

Clinical trials in adjuvant therapy for CC now include the possibility of both chemotherapeutic and biologic agents. We further conducted sensitivity analyses to evaluate the surrogacy of biologic agents that have different underlying mechanisms of action than oxaliplatin-based chemotherapy alone. Strong trial-

and individual patient-level surrogacy were observed and met the predefined criteria, although a wider confidence interval was observed for 1 of the  $R^2$  measures. This was probably due to the small number of comparison units available ( $n = 6$ ) for the meta-analysis. Analyses of surrogate endpoints with a limited number of trials are known to suffer from large variability in estimation, which is likely manifested in our analysis where the model-based  $R^2$  ( $R^2_{COPULA}$ ) and simple  $R^2$  ( $R^2_{WLS}$ ) measures differed greatly (18). The limited number of trials or comparison units testing biologic agents is a major limitation of this study.

We also demonstrate the value of assembling meta-databases with individual patient data, which not only evaluates and identifies surrogate endpoints for efficient clinical trials but also provides a rich repository to explore the biology and treatment of CC. In this regard, the ACCENT collaboration continuously provides rich data for high-impact research, which can potentially advance trial design and treatment development for this disease.

In summary, in this analysis of 9 trials of 18 396 stage III CC patients, we have validated that, for trials testing adjuvant chemotherapies with or without biologics, that results based on 3-year DFS are valid and appropriate primary endpoints. As standard care has shifted from 5-FU-based to oxaliplatin-based cytotoxic chemotherapy regimens, median survival following recurrence continues to improve (7). Longer follow-up (at least 6 years) in stage III adjuvant trials may be necessary to optimally evaluate the benefits of new adjuvant therapy and to allow a complete assessment of OS. The strong association observed, based on previous and current surrogacy analyses as well as the longer follow-up needed for OS in recent trials further support 3-year DFS as the primary endpoint in future stage III adjuvant CC trials.

## Funding

This analysis was supported by a grant from NIH: U10 CA180882.

## Notes

**Role of the funder:** The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

**Disclosures:** QS reports consulting/advisory role from Yiviva Inc, Boehringer Ingelheim Pharmaceuticals, Inc, Regeneron Pharmaceuticals, Inc (to myself), Honorarium/speaker role from Chugai Pharmaceutical Co, Ltd, stock from Johnson & Johnson, Amgen, and Merck & Co (to myself), research funds from Celgene/BMS, Roche/Genentech, Janssen, Novartis (to institution). TA reports consulting/advisory role and or received honoraria from Amgen, Bristol-Myers Squibb, Chugai, Clovis, Gristone Oncology, HaliDx, MSD Oncology, Pierre Fabre, Roche/Ventana, Sanofi, Servier, and GSK and has received travel, accommodations, and expenses from Roche/Genentech, MSD Oncology, and Bristol-Myers Squibb. RC declares honoraria from MSD Oncology and Servier. AG reports grants, personal fees and non-financial support from Bayer, grants, personal fees and non-financial support from Genentech/Roche, grants and personal fees from Array/Pfizer, grants and personal fees from Boston Biomedicals, grants from OBI Pharmaceuticals, grants from Merck, during the conduct of the study. TY reports grants

from Novartis Pharma K.K., grants from MSK K.K., grants from Sumitomo Dainippon Pharma Co, Ltd, grants from Chugai Pharmaceutical Co, Ltd, grants from Sanofi K.K., grants from Daiichi Sankyo Company, Limited, grants from PAREXEL International Inc, grants from Ono Pharmaceutical Co, Ltd, grants from Glaxo SmithKline K.K., outside the submitted work. JT reports consulting/advisory role and or received honoraria from Amgen, Haliodx, MSD Oncology, Astra-Zeneca, Pierre Fabre, Roche, Sanofi, Lilly, Servier and Merck KGAA and has received travel, accommodations, and expenses from Roche/Genentech, Celgene, Pierre Fabre, Servier and Merck KGAA. All remaining authors have declared no conflicts of interest.

**Author contributions:** Conceptualization: JY, MES, ZJ, RC, TA, QS; data curation: JY, JGD, QS; formal analysis: JY, JGD, QS; investigation: JY, MES, ZJ, RC, TA, QS; resources: ADG, EVC, JT, SRA, NW, H-JS, LBS, TJG, RRMG, RK, SL, TY, GY, AG, TA, QS; supervision: RRMG, RK, SL, TY, GY, AG, TA, QS; writing—original draft: JY, MES, ZJ, RC, TA, QS; writing—review and editing: all co-authors.

**Acknowledgements:** This work is dedicated to the memory of Daniel J. Sargent. Dan was one of the world's foremost experts in biostatistics and oncology who brought together disparate investigators and established data sharing across academia and industry internationally. His groundbreaking initiatives of integrating large collections of databases enabled research to answer questions otherwise beyond statistical possibility, to design important new clinical studies, to make regulatory observations, and to set new standards. He pushed these innovations farther to prospectively plan internationally combined analyses that answered questions previously believed to be impossible. The world of oncology statistics and analysis will not be the same without him, but his legacy continues. We would like to thank Editage for English language editing.

**Disclaimer:** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Prior presentations:** Presented in part at the Annual Meeting of the American Society of Clinical Oncology, Chicago, IL, May 31–June 4, 2019 (Oral presentation, abstract 3502).

## Data Availability

The data sharing of individual patient data from each participating trial will be subject to the policy and procedures of the institutions and groups who conducted the original study .

## References

1. Araghi M, Soerjomataram I, Jenkins M, et al. Global trends in colorectal cancer mortality: projections to the year 2035. *Int J Cancer*. 2019;144(12):2992–3000.
2. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol*. 2005;23(34):8664–8670.
3. Sargent DJ, Patiyil S, Yothers G, et al. ACCENT Group. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT group. *J Clin Oncol*. 2007;25(29):4569–4574.
4. Sargent D, Shi Q, Yothers G, et al. Adjuvant Colon Cancer End-points (ACCENT) Group. Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C. *Eur J Cancer*. 2011;47(7):990–996.
5. André T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med*. 2004;350(23):2343–2351.
6. Kuebler JP, Wieand HS, O'Connell MJ, et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. *J Clin Oncol*. 2007;25(16):2198–2204.
7. Salem ME, Yin J, Goldberg RM, et al. Evaluation of the change of outcomes over a 10-year period in patients with stage III colon cancer: pooled analysis of 6501 patients treated with fluorouracil, leucovorin, and oxaliplatin in the ACCENT database. *Ann Oncol*. 2020;31(4):480–486.
8. Taieb J, Tabernero J, Mini E, et al. Oxaliplatin, fluorouracil, and leucovorin with or without cetuximab in patients with resected stage III colon cancer (PETACC-8): an open-label, randomised phase 3 trial. *Lancet*. 2014;15(8):862–873.
9. Alberts SR, Sargent DJ, Nair S, et al. Effect of oxaliplatin, fluorouracil, and leucovorin with or without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial. *JAMA*. 2012;307(13):1383–1393.
10. Saltz LB, Niedzwiecki D, Hollis D, et al. Irinotecan fluorouracil plus leucovorin is not superior to fluorouracil plus leucovorin alone as adjuvant treatment for stage III colon cancer: results of CALGB 89803. *J Clin Oncol*. 2007;25(23):3456–3461.
11. Van Cutsem E, Labianca R, Bodoky G, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol*. 2009;27(19):3117–3125.
12. Schmoll HJ, Tabernero J, Maroun J, et al. Capecitabine plus oxaliplatin compared with fluorouracil/folinic acid as adjuvant therapy for stage III colon cancer: final results of the NO16968 randomized controlled phase III trial. *J Clin Oncol*. 2015;33(32):3733–3740.
13. Allegra CJ, Yothers G, O'Connell MJ, et al. Phase III trial assessing bevacizumab in stages II and III carcinoma of the colon: results of NSAMP protocol C-08. *J Clin Oncol*. 2011;29(1):11–16.
14. de Gramont A, Van Cutsem E, Schmoll HJ, et al. Bevacizumab plus oxaliplatin-based chemotherapy as adjuvant treatment for colon cancer (AVANT): a phase 3 randomised controlled trial. *Lancet Oncol*. 2012;13(12):1225–1233.
15. Burzykowski T, Molenberghs G, Buyse M, et al. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J R Stat Soc*. 2001;50(4):405–422. doi:10.1111/1467-9876.00244
16. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*. 2006;5(3):173–186.
17. André T, de Gramont A, Vernerey D, et al. Adjuvant fluorouracil, leucovorin, and oxaliplatin in stage II to III colon cancer: updated 10-year survival and outcomes according to BRAF mutation and mismatch repair status of the MOSAIC study. *J Clin Oncol*. 2015;33(35):4176–4187.
18. Renfro LA, Shi Q, Bot B, et al. An assessment of meta-analytic measures for evaluating time-to-event surrogate endpoints in clinical trials. *JSM*. 2009; (abstr 304742).