Surrogate Endpoints in Second-Line Trials of Targeted Agents in Metastatic Colorectal Cancer: A Literature-Based Systematic Review and Meta-Analysis

Chiara Cremolini, MD¹
Carlotta Antoniotti, MD¹
Filippo Pietrantonio, MD²
Rosa Berenato, MD²
Marco Tampellini, MD³
Chiara Baratelli, MD³
Lisa Salvatore, MD¹
Federica Marmorino, MD¹
Beatrice Borelli, MD¹
Federico Nichetti, MD²
Paolo Bironzo, MD³
Cristina Sonetto, MD⁴
Maria Di Bartolomeo, MD³
Filippo de Braud, MD²
Fotios Loupakis, MD, PhD⁴
Alfredo Falcone, MD⁴
Massimo Di Maio, MD⁴

¹Unit of Medical Oncology, Azienda Ospedaliero-Universitaria Pisana, Department of Translational Research and New Technologies in Medicine, University of Pisa, Pisa;
²Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan;
³Department of Oncology, University of Turin, A.O.U. San Luigi Gonzaga, Turin;
⁴Department of Oncology, University of Turin, A.O. Ordine Mauriziano, Turin, Italy

Purpose
The purpose of this study was to evaluate progression-free survival (PFS) and objective response rate (ORR) as surrogate endpoints of overall survival (OS) in modern clinical trials investigating the efficacy of targeted agents in the second-line treatment of metastatic colorectal cancer (mCRC).

Materials and Methods
A systematic search of literature pertaining to randomized phase II and III trials evaluating targeted agents as second-line treatments for mCRC was performed. The strength of the correlation between both PFS and ORR and OS was assessed based on the Pearson’s correlation coefficient (R) and the coefficient of determination (R²).

Results
Twenty trials, including a total of 7,571 patients, met the search criteria. The median duration of post-progression survival (PPS) was 7.6 months. The median differences between experimental and control arms were 0.65 months (range, −2.4 to 3.4) for the median PFS and 0.7 months (range, −5.8 to 3.9) for the median OS. PFS and ORR showed moderate (R=0.734, R²=0.539, p < 0.001) and poor correlation (R=0.169, R²=0.029, p=0.476) with OS, respectively. No differences between anti-angiogenic agents and other drugs were evident.

Conclusion
Targeted agents investigated in the second-line treatment of mCRC provided minimal PFS gains translating into modest OS improvements. Considering both the moderate correlation between PFS and OS and the short duration of PPS, the OS should remain the preferred primary endpoint for randomized clinical trials in the second-line treatment of mCRC.

Key words: Colorectal neoplasms, Biomarkers, Molecular targeted therapy
Introduction

The choice of the primary endpoint is essential to the design of clinical trials. While overall survival (OS) actually reflects the ultimate goal of cancer treatments, and is therefore regarded as a preferred choice in the metastatic setting, the surrogacy of other endpoints was investigated in different malignancies. The identification of valuable surrogate endpoints, which are potentially reachable in a shorter time and with a lower number of patients, would allow notable decreases in trial duration, thus expediting drug development and making new options more rapidly available for cancer patients.

With regard to metastatic colorectal cancer (mCRC), the reliability of response parameters and progression-free survival (PFS) during first-line treatments as surrogate endpoints of OS has previously been evaluated. While surrogacy for OS has not been formally proven for the objective response rate (ORR) [1,2], nor for the new parameter of early tumor shrinkage [3], PFS was shown to achieve strong surrogacy for OS in trials conducted before the introduction of targeted agents [2,4]. In a recently published literature-based analysis of surrogate endpoints in second-line treatment for mCRC, PFS was considered a reliable surrogate for OS [5]. However, about half of the clinical trials included in that systematic review compared chemotherapy only regimens, without targeted agents. In recent years, the adoption of new drugs with different mechanisms of action and the availability of multiple effective treatments after progression has enabled extension of post-progression survival (PPS), and is challenging the role of PFS as a surrogate of OS. Even though a previous analysis suggested that in modern trials OS could be better associated with PPS than with PFS [6], significant surrogacy for PFS was confirmed, justifying its adoption as a primary endpoint in first-line studies in mCRC [7-9]. However, in a systematic review and meta-analysis of 101 randomized controlled trials conducted in advanced colorectal cancer, none of the surrogate endpoints considered (ORR, PFS, time to progression) achieved the level of evidence required to qualify correlation levels as high or excellent by means of common surrogate evaluation tools [10].

In the last few years, several targeted agents have been tested in second- and further-lines of treatment and shown to produce significant, although only incremental, gains in OS. Today, as previously shown for first-line treatments, the effectiveness of new drugs in third and later lines might dilute the impact of second-line regimens on OS. Moreover, the frequent adoption of cross-over designs, especially in clinical settings with no other effective options, deeply influences OS findings, making the choice of earlier endpoints extremely appealing.

The present literature-based analysis was conducted to evaluate the correlation of both PFS and ORR with OS in modern clinical trials investigating the efficacy of targeted agents in the second-line treatment of mCRC. Since the relevance of surrogate endpoints may differ according to the mechanisms of action of investigated drugs, this analysis also separately evaluated the correlation of PFS and ORR with OS for anti-angiogenic agents relative to drugs with other mechanisms of action.

Materials and Methods

1. Literature search

A literature search was performed in October of 2015 to identify all randomized phase II and phase III trials evaluating molecular-targeted agents as second-line treatments for advanced colorectal cancer. The literature search was performed using PubMed, and the following keywords: “(colorectal cancer) AND (pretreated OR “previously treated” OR “second line”) AND random”. Following a comment by a reviewer, a second search “(colorectal cancer) AND (pretreated OR “previously treated” OR “second line”) AND randomized controlled trial [Publication type]” was performed to verify that all records included in the latter search had already been included in the former search. References of the selected articles were also checked to identify further eligible trials. Moreover, the proceedings of the American Society of Clinical Oncology (ASCO) annual meeting and European Society of Medical Oncology meeting were searched from 2012 onwards for relevant abstracts. When more than one report describing the results of the same trial was available, the most recent information (corresponding to a longer follow-up and a higher number of events) was utilized. Trials randomizing patients to receive or not receive an anti-epidermal growth factor receptor monoclonal antibody were included only if results in the RAS (or at least KRAS) wild-type subgroup were available.

2. Data abstraction

For each eligible trial, the following data were collected, if available:
- Study phase (II or III).
- Details of study treatment: control arm; experimental arm (or arms if more than one experimental treatment). Control arms were identified based on the null hypothesis of the statistical design underlying each single trial as reported in full manuscripts or presented abstracts.
- Details regarding cross-over (administration of experimental treatment to patients assigned to the control arm after disease progression).
- Study primary endpoint.
- Patients’ enrolment: number of enrolled patients, number of patients assigned to control arm, number of patients assigned to experimental arm.
- ORR: proportion of objective responses in the control arm, proportion of objective responses in the experimental arm; relative risk of response (calculated as the ratio between the response rate in the experimental arm and in the control arm).
- PFS: median PFS in the control arm, median PFS in the experimental arm, hazard ratio (HR) with 95% confidence interval, p-value.
- OS: median OS in the control arm, median OS in the experimental arm, HR with 95% confidence interval, p-value.
- PPS: absolute PPS was calculated as the difference between median OS and median PFS; relative PPS was calculated as the ratio between median PPS and median OS. For instance, in a treatment arm with a median PFS of 4 months and a median OS of 10 months, absolute PPS was 6 months (10−4) and relative PPS was 60% (6/10).

For trials with more than two treatment arms, multiple records were completed, one for each comparison. Two investigators independently abstracted the data from the publications, and subsequently compared their results. All data were checked for internal consistency, and disagreements were resolved by discussions among the investigators.

3. Statistical analysis

To analyze the correlation between PFS and OS, two different regression analyses were performed: (1) correlation between the HR for PFS and HR for OS and (2) correlation between the difference in median PFS and the difference in median OS between arms. Similarly, to analyze the correlation between ORR and OS, two different regression analyses were performed: (1) correlation between the relative risk of response and HR for OS and (2) correlation between the difference in ORR between arms and the difference in median OS between arms.

All analyses were weighted by the sample size of each comparison. In the case of trials with two experimental arms and a single control arm [11-14], two separate comparisons were analyzed (each experimental arm versus the control arm). However, to avoid double-counting of the patients enrolled in the control arm and the risk of clustered data, each comparison was given a lower weight that was obtained by equally dividing the total number of patients of the control arm between the two comparisons.

In each analysis, the strength of the correlation was evaluated by calculating the Pearson’s correlation coefficient (R) and the coefficient of determination (R²). Pearson’s R is a simple measure of the linear correlation between two variables, giving a value between 1 and −1, where 1 is a total positive correlation, 0 is the absence of correlation, and −1 is a total negative correlation. The coefficient of determination is such that 0 ≤ R² ≤ 1. Although there are no specific cut-offs to define a moderate or strong correlation, a higher R² score indicates a stronger association.

Correlations were graphically described by bubble plots, where each bubble represents a comparison between one experimental arm and one control arm, with bubble size proportional to the sample size of each comparison. As all analyses were weighted by the sample size of each trial/comparison, weighted least-squares regression lines were calculated and reported in each graph.

Exploratory subgroup analyses were performed according to the type of experimental drug tested (anti-angiogenic drugs vs. other drugs).

Statistical analyses were conducted using SPlus (S-PLUS 6.0 Professional, release 1, Insightful Corporation, Seattle, WA) and SPSS ver. 22.0 (IBM Corp., Armonk, NY). Graphs were realized using SigmaPlot (Systat Software, San Jose, CA). For all analyses, a p-value of < 0.05 was considered statistically significant.

Results

1. Trial characteristics

Overall, 20 trials were identified (Fig. 1), nine phase III trials and 11 randomized phase II trials (Table 1) [11-30]. A total of 7,571 patients were enrolled in these trials, and the median number of enrolled patients was 197 (range, 75 to 1,226). The primary endpoint was PFS in 12 trials (60%) [11-15,18,19, 21-24,30], OS in six trials (30%) [16,20,26-29] and ORR in one trial (5%) [17]. In one trial (5%), PFS and OS were co-primary endpoints [25]. Four trials had three treatment arms, with two comparisons between each of the two experimental arms and the single control arm [11-14]. In one trial, there were four arms (two experimental arms and two control arms) with two separate comparisons [15]. Overall, 25 comparisons were recorded (Table 1).

Information regarding cross-over was not available for most reports (19 out of 25 comparisons). In the six reports with details about subsequent administration of experimental drugs (or drugs with the same mechanism of action) in patients assigned to control arms, cross-over was quite neg-
Fig. 1. Outline of the search flow diagram. Selection process for randomized controlled trials included in the analysis. ASCO, American Society of Clinical Oncology; ESMO, European Society of Medical Oncology; ECCO, European Cancer Organization; GI, gastrointestinal; EGFR, epidermal growth factor receptor.

ligible (median proportion, 3.5%; range, 0% to 32%).

2. Outcomes

Based on all comparisons with available information, the median value of the OS in the 25 experimental arms was 13.1 months (range, 9.6 to 21.4), and the median value of the OS in the control arms was 13.9 (range, 8.8 to 19.8). The median difference between experimental and control arms was equal to 0.7 months (range, 5.8 to 3.9). In the 21 comparisons with available information, the median HR for OS was 0.90 (range, 0.69 to 1.57).

Based on all comparisons and available information, the median value of the PFS in the 24 experimental arms was 6.4 (range, 2.1 to 8.5), and the median value of the median PFS in the control arms was 5.4 (range, 2.4 to 9.0). The difference in median values between the experimental and control arms was equal to 0.65 months (range, −2.4 to 3.4). In the 23 comparisons with available information, the median HR for PFS was 0.85 (range, 0.61 to 1.45).

Based on all available information regarding the median OS and median PFS, the median absolute PPS in the experimental arms was 7.6 months (range, 4.4 to 14.6). The relative PPS (expressed as a proportion of OS) ranged from 43.4% to 82.3%, with a median value of 55.7%. In the control arms, the median absolute PPS was 7.6 months (range, 3.6 to 14.3) and
<table>
<thead>
<tr>
<th>Trial</th>
<th>No. of patients</th>
<th>Treatments</th>
<th>Primary endpoint</th>
<th>Anti-angio (Y/N)</th>
<th>HR</th>
<th>Delta PFS (mo)</th>
<th>HR</th>
<th>Delta OS (mo)</th>
<th>RR response</th>
<th>Delta ORR (%)</th>
<th>PPS/OS [exp arm–ctr arm] (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bendell et al. [15]</td>
<td>36</td>
<td>FOLFOX+axitinib</td>
<td>FOLFOX+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.04</td>
<td>1.2</td>
<td>0.69</td>
<td>3</td>
<td>0.97</td>
<td>–0.6</td>
</tr>
<tr>
<td>Bendell et al. [15]</td>
<td>49</td>
<td>FOLFIRI+axitinib</td>
<td>FOLFIRI+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.27</td>
<td>–1.2</td>
<td>1.36</td>
<td>–2.8</td>
<td>1.04</td>
<td>1</td>
</tr>
<tr>
<td>Bennouna et al. [16]</td>
<td>409</td>
<td>Chemo+bevacizumab</td>
<td>Chemo</td>
<td>OS</td>
<td>Y</td>
<td>0.68</td>
<td>1.6</td>
<td>0.81</td>
<td>1.4</td>
<td>1.25</td>
<td>1</td>
</tr>
<tr>
<td>Cao et al. [17]</td>
<td>65</td>
<td>FOLFIRI+bevacizumab</td>
<td>FOLFIRI</td>
<td>ORR</td>
<td>Y</td>
<td>–</td>
<td>3.4</td>
<td>–</td>
<td>3.9</td>
<td>1.67</td>
<td>19.2</td>
</tr>
<tr>
<td>Cardiello et al. [18]</td>
<td>39</td>
<td>FOLFOX+oxaliplatin</td>
<td>FOLFOX</td>
<td>PFS</td>
<td>N</td>
<td>0.80</td>
<td>1.3</td>
<td>0.78</td>
<td>1.6</td>
<td>1.53</td>
<td>8.9</td>
</tr>
<tr>
<td>Cohn et al. [11]</td>
<td>52</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>N</td>
<td>1.01</td>
<td>–0.1</td>
<td>1.27</td>
<td>0.4</td>
<td>4.00</td>
<td>6</td>
</tr>
<tr>
<td>Cohn et al. [11]</td>
<td>51</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>N</td>
<td>0.69</td>
<td>1.9</td>
<td>0.89</td>
<td>0.3</td>
<td>7.00</td>
<td>12</td>
</tr>
<tr>
<td>Cunningham et al. [12]</td>
<td>71</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.28</td>
<td>–2</td>
<td>1.39</td>
<td>–5.3</td>
<td>0.67</td>
<td>–9</td>
</tr>
<tr>
<td>Cunningham et al. [12]</td>
<td>73</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.17</td>
<td>–0.6</td>
<td>1.00</td>
<td>–2.8</td>
<td>0.70</td>
<td>–8.1</td>
</tr>
<tr>
<td>Eng et al. [19]</td>
<td>60</td>
<td>Cetuximab+irinotecan</td>
<td>Cetuximab+irinotecan</td>
<td>PFS</td>
<td>N</td>
<td>0.85</td>
<td>1</td>
<td>0.70</td>
<td>2.9</td>
<td>1.35</td>
<td>11.7</td>
</tr>
<tr>
<td>Gianantonio et al. [20]</td>
<td>286</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI</td>
<td>OS</td>
<td>Y</td>
<td>0.61</td>
<td>2.6</td>
<td>0.75</td>
<td>2.1</td>
<td>2.64</td>
<td>14.1</td>
</tr>
<tr>
<td>Heczko et al. [13]</td>
<td>85</td>
<td>FOLFIRI+oxaliplatin 200 mg</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>N</td>
<td>1.45</td>
<td>–0.4</td>
<td>1.50</td>
<td>–5.8</td>
<td>0.59</td>
<td>–4.1</td>
</tr>
<tr>
<td>Heczko et al. [13]</td>
<td>84</td>
<td>FOLFIRI+oxaliplatin 700 mg</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>N</td>
<td>1.32</td>
<td>–0.3</td>
<td>1.23</td>
<td>–4.9</td>
<td>1.19</td>
<td>1.9</td>
</tr>
<tr>
<td>Hoehl et al. [21]</td>
<td>50</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>N</td>
<td>0.84</td>
<td>–0.4</td>
<td>1.57</td>
<td>–3.1</td>
<td>2.10</td>
<td>13.4</td>
</tr>
<tr>
<td>Iwamoto et al. [22]</td>
<td>187</td>
<td>FOLFIRI+oxaliplatin 10 mg/kg</td>
<td>FOLFIRI+bevacizumab (5 mg/kg)</td>
<td>PFS</td>
<td>Y</td>
<td>0.95</td>
<td>0.3</td>
<td>1.08</td>
<td>0.7</td>
<td>1.00</td>
<td>0</td>
</tr>
<tr>
<td>Masi et al. [23]</td>
<td>92</td>
<td>FOLFIRI</td>
<td>FOLFIRI</td>
<td>PFS</td>
<td>Y</td>
<td>0.70</td>
<td>1.8</td>
<td>0.77</td>
<td>–1.4</td>
<td>1.24</td>
<td>4</td>
</tr>
<tr>
<td>O’Neill et al. [14]</td>
<td>50</td>
<td>FOLFIRI+oxaliplatin low dose</td>
<td>FOLFIRI+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.45</td>
<td>–2.4</td>
<td>–</td>
<td>-4.5</td>
<td>0.69</td>
<td>–10.7</td>
</tr>
<tr>
<td>O’Neill et al. [14]</td>
<td>49</td>
<td>FOLFIRI</td>
<td>FOLFIRI+bevacizumab</td>
<td>PFS</td>
<td>Y</td>
<td>1.26</td>
<td>–1.3</td>
<td>–</td>
<td>-0.1</td>
<td>0.65</td>
<td>–12.3</td>
</tr>
<tr>
<td>Peeters et al. [24]</td>
<td>95</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>PFS</td>
<td>Y</td>
<td>1.23</td>
<td>–1.7</td>
<td>0.90</td>
<td>3.1</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td>Peeters et al. [25]</td>
<td>208</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI</td>
<td>PFS/OS</td>
<td>N</td>
<td>0.70</td>
<td>1.8</td>
<td>0.81</td>
<td>2.3</td>
<td>4.10</td>
<td>31</td>
</tr>
<tr>
<td>Seymour et al. [26]</td>
<td>230</td>
<td>Irinotecan+oxaliplatin</td>
<td>Irinotecan</td>
<td>OS</td>
<td>N</td>
<td>0.78</td>
<td>-</td>
<td>1.01</td>
<td>–0.5</td>
<td>2.93</td>
<td>22.6</td>
</tr>
<tr>
<td>Tabernero et al. [27]</td>
<td>536</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>OS</td>
<td>Y</td>
<td>0.79</td>
<td>1.2</td>
<td>0.84</td>
<td>1.6</td>
<td>1.07</td>
<td>0.9</td>
</tr>
<tr>
<td>Van Cutsen et al. [28]</td>
<td>426</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>OS</td>
<td>Y</td>
<td>0.83</td>
<td>1.4</td>
<td>1.00</td>
<td>1.2</td>
<td>1.06</td>
<td>1</td>
</tr>
<tr>
<td>Van Cutsen et al. [29]</td>
<td>612</td>
<td>FOLFIRI+oxaliplatin</td>
<td>FOLFIRI+placebo</td>
<td>OS</td>
<td>Y</td>
<td>0.76</td>
<td>2.23</td>
<td>0.82</td>
<td>1.44</td>
<td>1.78</td>
<td>8.7</td>
</tr>
<tr>
<td>Vieitez et al. [30]</td>
<td>38</td>
<td>Raltitrexed+gefitinib</td>
<td>Raltitrexed+placebo</td>
<td>PFS</td>
<td>N</td>
<td>–</td>
<td>–0.3</td>
<td>2.3</td>
<td>2.6</td>
<td>1.49</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Exp arm, experimental arm; Ctr arm, control arm; Anti-angio, anti-angiogenic agent; HR, hazard ratio; PFS, progression-free survival; OS, overall survival; RR, relative risk; ORR, objective response rate; PPS, post-progression survival; FOLFIRI, oxaliplatin, 5-fluouracil, and leucovorin; FOLFIRI, irinotecan, folic acid, and 5-fluorouracil; Chemo, chemotherapy, combination of fluoropyrimidine and oxaliplatin or irinotecan. a)In the case of trials with two experimental arms and a single control arm [11-14], two separate comparisons were analyzed (each experimental arm versus the control arm), b)Co-primary endpoints. c)In KRAS codon 12-13-61 wild-type tumors.
expressed as a proportion of OS, while the relative PPS ranged from 40.9% to 75.0%, with a median value of 60.7% (Table 2). Fig. 2 describes the median PFS and median PPS based on all comparisons included in the analysis, scattered by the type of experimental drug (anti-angiogenics vs. other drugs).

Based on all available information, the median ORR in the 25 experimental arms was 19% (range, 5% to 48%), and the median ORR in the control arms was 12% (range, 0% to 35%). The median difference in the ORR between experimental and control arms was equal to 2.6% (range, −12.3% to 31%). The median relative risk of response was 1.24 (range, 0.59 to 7.00).

3. Association between PFS and OS

Information regarding HRs for PFS and OS was available for 21 trials. Overall, there was a moderate correlation (R=0.734, R²=0.539, p < 0.001) (Table 3, Fig. 3A). The slope of the regression line (0.739) suggests that a 0.1 improvement in PFS HR corresponds to a 0.074 improvement in OS HR. The correlation between HRs for PFS and OS was significant for the 13 comparisons investigating anti-angiogenic drugs (R=0.655, R²=0.429, p=0.015) and the eight comparisons investigating other drugs (R=0.857, R²=0.734, p<0.007) (Table 3, Fig. 3B and C). There was no significant interaction between drug categories and the correlation between HRs for PFS and OS (p=0.775) (Table 3).

Similar results were observed when the correlation between PFS and OS was analyzed for both endpoints based on the difference in median values between study arms. This information was available for 24 comparisons (Table 3, S1 Fig. A). Overall, there was a moderate correlation between PFS and OS (R=0.632, R²=0.399, p < 0.001). The slope of the regression line (1.065) suggests that a one month increase in the difference in median PFS corresponds to a 1.06 month increase in the difference in median OS. The correlation between PFS and OS based on the difference in median values between study arms was significant for both the 16 comparisons evaluating anti-angiogenic drugs (R=0.651, R²=0.423, p=0.006) and the eight comparisons evaluating other drugs (R=0.724, R²=0.525, p=0.042) (Table 3, S1 Fig. B and C). The interaction between drug categories and the correlation between PFS and OS was not significant (p=0.110) (Table 3).

4. Association between ORR and OS

Information regarding the relative risk of objective response and HR for OS was available for 20 comparisons. Overall, there was a weak correlation that was not statistically significant (R=0.169, R²=0.029, p=0.476) (Table 4, Fig. 4A). The correlation between relative risks of response and HRs for OS was not significant for the 12 comparisons.
evaluating anti-angiogenic drugs ($R=0.361$, $R^2=0.131$, $p=0.249$) or the eight comparisons evaluating other drugs ($R=0.441$, $R^2=0.195$, $p=0.274$) (Table 4, Fig. 4B and C). There was no significant interaction between drug categories and the association between the relative risk of response and the HR for OS ($p=0.654$) (Table 4).

Information regarding the difference in ORR and the median OS between study arms was available for 25 comparisons. Based on these parameters, a weak correlation was found ($R=0.345$, $R^2=0.119$, $p=0.092$) (Table 4, S2 Fig. A). The correlation between response and OS considering the difference in ORR and in the median OS between study arms was weak to moderate for the 16 comparisons of anti-angiogenic drugs ($R=0.522$, $R^2=0.272$, $p=0.038$) and the nine comparisons investigating other drugs ($R=0.632$, $R^2=0.399$, $p=0.068$) (Table 4, S2 Fig. B and C). The interaction between this correlation and drug categories was not significant ($p=0.904$) (Table 4).

**Discussion**

Different targeted agents recently gained approval for the second-line treatment of mCRC based on relatively small absolute gains in OS. Nevertheless, the impact of these treatments on the overall prognosis of mCRC patients is rather limited [31], and the improvements achieved with novel treatments are below the expectations. Overall, the results from the 20 second-line trials included in the present analysis indicate that the median PFS accounts for 44% and 39% of the median OS in the experimental and control arms, respectively. Although PPS will probably increase in the future, the median absolute duration of PPS in our series was quite short (7.6 months) due to the availability of new effective options in later lines. These findings demonstrate that, at least for the timeframe in which the trials included in this analysis were conducted, the impact of third- and further-line treatments on mCRC patients’ prognosis was rather modest.

We systematically reviewed the inherent literature to focus on clinical trials investigating the efficacy of targeted agents in the second-line treatment of mCRC to assess the correlation of earlier endpoints, PFS and ORR, with OS, and to
Table 3. Correlation between PFS and OS in all comparisons and scattered by the type of experimental drugs.

<table>
<thead>
<tr>
<th>Type of drugs</th>
<th>No. of comparisons</th>
<th>Correlation between hazard ratios</th>
<th>p-value interaction</th>
<th>Slope</th>
<th>R</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>All comparisons</td>
<td>21</td>
<td>0.734</td>
<td>0.001</td>
<td>0.739</td>
<td>0.399</td>
<td>0.422</td>
</tr>
<tr>
<td>Anti-angiogenic drugs</td>
<td>13</td>
<td>0.655</td>
<td>0.001</td>
<td>0.686</td>
<td>0.399</td>
<td>0.423</td>
</tr>
<tr>
<td>Other drugs</td>
<td>8</td>
<td>0.857</td>
<td>0.002</td>
<td>0.757</td>
<td>0.256</td>
<td>0.409</td>
</tr>
</tbody>
</table>

PFS, progression-free survival; OS, overall survival.

Fig. 3. Correlation between overall survival (OS) and progression-free survival (PFS). (A) Correlation between hazard ratios in all comparisons with available information (n=21). (B) Correlation between hazard ratios in all comparisons with available information pertaining to anti-angiogenic drugs (n=13). (C) Correlation between hazard ratios in all comparisons with available information with other drugs (n=8).
Table 4. Correlation between ORR and OS in all comparisons and scattered by the type of experimental drugs

<table>
<thead>
<tr>
<th></th>
<th>Correlation between relative risks and hazard ratios</th>
<th>Correlation between differences in ORR and in median OS values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of comparisons</td>
<td>R</td>
</tr>
<tr>
<td>All comparisons</td>
<td>20</td>
<td>0.169</td>
</tr>
<tr>
<td>Anti-angiogenic drugs</td>
<td>12</td>
<td>0.361</td>
</tr>
<tr>
<td>Other drugs</td>
<td>8</td>
<td>0.441</td>
</tr>
</tbody>
</table>

ORR, objective response rate; OS, overall survival.

Fig. 4. Correlation between objective response rate and overall survival (OS). (A) Correlation between relative risks and hazard ratios in all comparisons (n=20). (B) Correlation between relative risks and hazard ratios in all comparisons with anti-angiogenic drugs (n=12). (C) Correlation between relative risks and hazard ratios in all comparisons with other drugs (n=8).
analyze their surrogacy for OS. While a similar approach was previously pursued by other groups [5], we chose to restrict our analysis to modern trials of targeted agents to put our results in the context of ongoing and future studies in this setting. In fact, previous studies have clearly shown that the reliability of surrogate endpoints must be properly verified within the context in which these endpoints should be subsequently adopted. Namely, out of the 23 trials included in the systematic review by Giessen et al. [5], as many as nine trials compared chemotherapy-only treatment regimens without targeted agents. Furthermore, those authors emphasized that a re-analysis according to the different mechanisms of drug activity should be conducted as soon as a larger set of trials was available. Therefore, we conducted an exploratory subgroup analysis to assess potential differences in surrogacy according to the targeted agents’ mechanisms of action (mainly anti-angiogenic versus directed against other cellular targets), as already suggested in first-line studies [7]. This exploratory subgroup analysis did not produce clear evidence of an interaction between the mechanism of action and surrogacy for the endpoints considered. A clear limitation of this study is that, while the anti-angiogenic group is clearly defined, the “other drugs” group includes agents with heterogeneous mechanisms of action.

Although our analysis has several limitations, we observed a moderate correlation between PFS and OS, while a poor correlation between ORR and OS was reported, with no relevant differences according to the drugs’ mechanisms of action. It should be noted that, after demonstrating a similar moderate correlation between PFS and OS, other authors concluded that PFS may be considered an appropriate surrogate endpoint in second-line treatments for mCRC [5]. However, when specifically focused on targeted agents, our results can affirm that OS remains the preferred primary endpoint for randomized clinical trials in this setting. However, the following considerations should be taken into account to justify this interpretation. First, only small median absolute gains in PFS were reported in statistically positive trials, making it rather difficult to translate these results into clinically relevant improvements in OS. According to the ASCO perspective, improvements of at least three months in median OS (primary endpoint) or median PFS (secondary endpoint) should be regarded as meaningful for mCRC patients experiencing disease progression with all prior therapies, or not eligible for standard options [32]. However, the slope of the regression line in our analysis suggests that small benefits in PFS, on average, are going to translate into modest OS differences. These achievements can only be regarded as clinically relevant if supported by solid improvements in quality of life, which were rarely assessed in the available literature. While the lack of molecular criteria able to positively select patients more likely to benefit from targeted agents may explain the present findings, the introduction of “precision medicine” principles into clinical research will likely change the present scenario.

Secondly, since the duration of PPS is quite short, the adoption of PFS instead of OS as a primary endpoint would not lead to a dramatic decrease in the duration, sample size, and financial costs of trials, or to a considerable acceleration of a drug’s development. However, the recent availability of new effective drugs in later lines of treatment, i.e., after failure of second-line agents, will probably prolong the duration of PPS. Moreover, only 30%-40% of patients included in second-line clinical trials actually receive treatments after progression. Hopefully, this percentage will increase in response to the introduction of highly effective targeted strategies in earlier lines of treatment. Both of these aspects may further weaken the correlation between the PFS and OS and lead to reconsideration of the surrogacy of second-line PFS in currently ongoing and future trials.

In other settings, cross-over has been shown to play a relevant role in the correlation between PPS and OS. As expected, if a high proportion of patients assigned to the control arm receive the experimental drug after disease progression (or a drug with the same mechanism of action), the difference between treatment arms might be significantly decreased [33]. In the present analysis, information regarding the possibility of cross-over according to study protocol and the proportion of patients actually receiving cross-over was not available in most trials; however, as detailed in the Results, this proportion was quite low in all trials for which this information was available.

A limitation of the present meta-analysis is that it is not based on individual patient data, but rather on data extracted from the publications (or, in some cases, from meeting presentations); therefore, we could only estimate trial-level, but not individual patient-level surrogacy. However, even if analysis of the individual patient-level association can lead to an estimation of how much the endpoints are likely to be causally linked to each other, the trial-level analysis remains useful to show the proportion of the OS effect captured by surrogate endpoints [34]. Although intrinsically limited, this information could facilitate the interpretation of trial results and design of future trials in this specific setting.

In conclusion, caution is needed when assessing the surrogacy of potentially useful endpoints and supporting their adoption in phase III clinical trials. Notably, only five out of 36 drugs approved by the U.S. Food and Drug Administration on the basis of surrogate endpoints were able to provide an OS benefit in subsequent trials [35]. Based on our data, OS should be the primary endpoint for registrative phase III trials in the second line treatment of mCRC. Given its moderate surrogacy for OS, PFS may be adopted in earlier steps of drug development.
Electronic Supplementary Material

Supplementary materials are available at Cancer Research and Treatment website (http://www.e-crt.org).

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

References


19. Eng C, Hart LL, Seversete A, Gladkov O, Mueller L, Kopp MV, et al. A randomized, placebo-controlled, phase I/II study of tivantinib (ARQ 197) in combination with cetuximab and irinotecan in patients (pts) with KRAS wild-type (WT) metastatic colorectal cancer (CRC) who had received previous