Meta-Analysis of the Validity of Progression-Free Survival as a Surrogate Endpoint for Overall Survival in Metastatic Colorectal Cancer Trials

Costel Chirila,[1] Dawn M. Odom,[1] Giovanna Devercelli,[2] Shahnaz Khan,[1] Bintu N. Sherif,[1] James A. Kaye,[1] Istvan Molnar,[2] Beth H. Sherrill[1]

[1]RTI Health Solutions, Research Triangle Park, NC, United States; [2]Bayer HealthCare Pharmaceuticals Inc, Montville, NJ, United States

INTRODUCTION

- Overall survival (OS) is viewed as the gold standard clinical endpoint in trials of new cancer therapies. However, there are limitations in using OS as the primary endpoint. Not only are large sample sizes and long-term follow-up required, but the use of subsequent therapies after disease progression (either by planned crossover or outside of the trial design) may obscure the antineoplastic effect of cancer therapies under investigation.
- Progression-free survival (PFS) can be assessed earlier than OS and is not confounded by subsequent therapies, so researchers continue to assess its viability as a surrogate for OS in multiple cancer indications.
- Validating a surrogate endpoint in oncology is challenging, and the methodology is still evolving.[1]
- Numerous published meta-analyses have examined the validity of PFS as a surrogate for OS in clinical trials of treatment for patients with metastatic colorectal cancer (mCRC), primarily in the setting of first-line treatment.[2-4]
- The objectives of this study were as follows:
  - Extend previous meta-analytic work by examining the relationship between PFS and OS in a larger number of mCRC studies, including first-line and later lines of therapy.
  - Explore a range of thresholds for sensitivity and specificity of using PFS as a surrogate for OS by employing receiver-operator characteristic (ROC) curves.

METHODS

Systematic Literature Review

- Sources: PubMed, Embase, and Cochrane databases (no date limit) and American Society of Clinical Oncology (ASCO) conference abstracts from 2008 and 2009.
- Search terms: "colorectal," "metastasis," and "survival," which were combined through Boolean operators "AND," "NOT," and "OR" with standard search terms for randomized controlled trials.
- Inclusion criteria:
  - Phase 2 or 3 clinical trials in patients with mCRC.
  - At least two treatments arms and at least 20 patients in each arm.
  - Hazard ratios (HR) for OS and PFS (or TTP) or median times to events for each treatment.
  - First-, second-, or third-line therapy for patients with mCRC.
- Exclusion criteria:
  - Efficacy of an agent of interest was analyzed as part of a sequential drug regimen.
  - Study design was randomized discontinuation.
  - Study presented the results of an interim analysis when a later analysis was available.
  - Agent of interest was used for adjuvant therapy (nonmetastatic disease).
  - Study participants did not have metastatic disease.
  - Study was an animal or in vitro study.

METHODS (CONT.)

Data Extraction

- Data were extracted by one reviewer and checked by another reviewer; any disagreement between reviewers was discussed and resolved.
- Extracted data included publication information, patient factors/characteristics, treatment information, other trial characteristics, and efficacy information.

Outcome Measures

- Treatment effects were defined for PFS and OS as the ratio of median times (m1/m2) where the m1 is the median time to event of group 1 and m2 is the median time to event of group 2.
  - Used the ratio of median times as an estimate of HR[5] because only about a third of the studies reported HR and associated variance.
  - Analyzed PFS and TTP endpoints concurrently (referred to as PFS_TTP), and then conducted a sensitivity analyses by each endpoint separately.
  - Included only two treatment arms from studies with multiple arms and determined in advance which arm of each study would be "control" (group 1) versus "experimental" (group 2) based on clinical judgment.

Statistical Methods

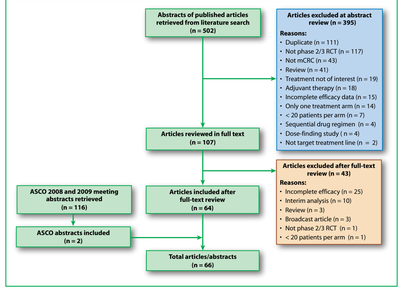
- Assessed correlation between median time to PFS_TTP and OS using weighted Pearson correlation by single treatment arm and correlation of treatment effects by study.
- Used meta-regression to explore and quantify the relationship between treatment effects on PFS_TTP and OS.
  - Ratio of median OS from each study was outcome in a weighted least squares meta-regression model, which was weighted by the study sample size.
  - Examined model diagnostics and performed leave-one-out cross-validation.
  - Used statistically significant factors to create subgroup analyses.
- Implemented ROC analysis, typically used to evaluate classification properties of diagnostic measures, to evaluate the association of study effects.
  - Defined the threshold for clinical benefit as an OS effect size no greater than 0.8 (a 1.25-fold relative improvement in OS).[6]
  - Considered the magnitude of effect on progression that would be required for a clinically meaningful survival benefit.
  - Constructed an ROC curve by varying a cutoff value for PFS_TTP and also for PFS only as a sensitivity analysis; the ROC curve is a graphical display of the trade-off between sensitivity and specificity at each cutoff value across a range of values.

RESULTS

Literature Search (Figure 1)

- Identified a total of 502 published articles and 116 ASCO abstracts.
- Extracted data from 66 articles/abstracts that met inclusion/exclusion criteria.
- Analyzed 62 articles/abstracts that presented median values for PFS and/or TTP, and OS.

Figure 1. Selection of Articles



RCT = randomized controlled trial.

Characteristics of Included Studies

The 62 included studies comprised a total of 23,527 patients and had the following characteristics:

- Publication year ranged from 1991 to 2009.
- Most of the studies (n = 46) were phase 3 studies, and the rest were phase 2.
- For 61 studies, the site of primary tumor was "colon or rectum"; only one study had colon only as the site of primary tumor.
- 56 studies reported results from an intention-to-treat analysis.
- Drug therapies could be broadly classified as fluorouracil alone, fluorouracil plus other cytotoxic chemotherapy, biological and targeted therapies, and other cytotoxic chemotherapy regimens.
- Where line of therapy was not explicitly stated in the article, it was determined by clinical judgment based in part on patient eligibility criteria:
  - First line or mostly first: 48 studies.
  - Second line, mostly second, or second and later: 13 studies.
  - Third line: 1 study.

Statistical Analysis

Correlation Analysis

- Found high positive correlation between:
  - Median PFS_TTP and median OS within treatment arms: Pearson coefficient 0.87 (95% confidence interval [CI], 0.82-0.91).
  - Treatment effects for OS and PFS_TTP by study: Pearson coefficient 0.69 (95% CI, 0.53-0.80).

Meta-Regression Analysis

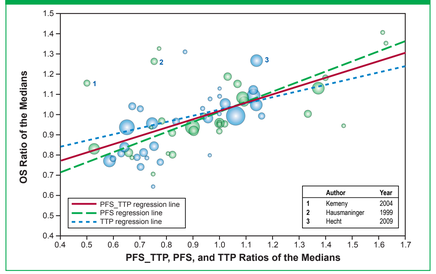
- Meta-regression results for the different fitted models are presented in Table 1.

Table 1. Model Parameter Estimates by Regression Model

Subgroup	Number of Studies	Intercept (95% CI)	Slope (95% CI)	R^2
Primary model (PFS_TTP)	62	0.60 (0.49-0.71)	0.41 (0.30-0.52)	0.48
Studies of first line (PFS_TTP)	48	0.52 (0.39-0.66)	0.49 (0.36-0.62)	0.54
Studies of second line (PFS_TTP)	13	0.71 (0.54-0.88)	0.25 (0.04-0.46)	0.37
Studies with PFS endpoint only	35	0.52 (0.39-0.66)	0.49 (0.35-0.64)	0.59
Studies with TTP endpoint only	27	0.71 (0.53-0.90)	0.31 (0.12-0.49)	0.32

- Primary model (PFS_TTP):
  - One-unit increase in PFS_TTP treatment effect predicted a 0.41 unit increase in OS treatment effect, and about half (R^2 = 0.48) of the variation in OS effect size was explained by the PFS_TTP effect size.
  - Figure 2 is a scatter plot with regression lines showing a linear relationship between the effect size for PFS_TTP, PFS only, and TTP only on the x-axis and effect size for OS on the y-axis. Each study is represented by a sphere with size proportional to study sample size.
  - Of the three outliers, one was the only third-line study in the analysis, which allowed for therapy crossover,[7] but the other two[8,9] did not appear to have any unusual characteristics from a clinical point of view.
- The only statistically significant factor was line of therapy (P = 0.03) with a higher R^2 for first-line (R^2 = 0.54) compared with second-line studies (R^2 = 0.37).
- The higher R^2 value for the model of studies reporting PFS seems to indicate a stronger association between PFS and OS than between TTP and OS.

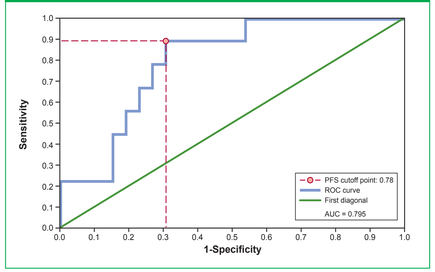
Figure 2. Relationship Between Effect Sizes for PFS_TTP, PFS, TTP, and OS



ROC Analysis

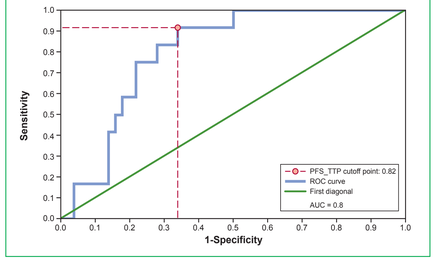
- Figure 3 presents an ROC curve, where sensitivity is the proportion of trials with OS clinical benefit that achieved PFS clinical benefit (true positives), and specificity is the proportion of trials without OS clinical benefit that did not achieve PFS clinical benefit (true negatives).
- Using an effect size of OS no greater than 0.8 as the threshold for success, we chose an optimal cutoff for a successful trial of PFS effect size equal to 0.78 (Figure 3), which corresponds to a sensitivity of 0.89 and a specificity of 0.69 of the surrogate measure.
- PFS results are substantially more likely to predict OS accurately than would be predicted by chance alone (P = 0.0097).
- The ROC analysis for PFS_TTP is very similar (Figure 4).

Figure 3. ROC Curve for PFS



AUC = area under the curve.

Figure 4. ROC Curve for PFS_TTP



CONCLUSIONS

- These results confirm and extend results reported by other meta-analyses of the relationship between PFS or TTP and OS in clinical trials of patients with mCRC.
- We found a strong relationship between the two endpoints and a clear, consistent, linear relationship between the treatment effect sizes of PFS or TTP and those of OS.
- Of the various characteristics tested, only the line of therapy and surrogate endpoint choice (PFS or TTP) showed potentially different regression lines by subgroups.
  - PFS_TTP seemed to be better correlated with OS in first-line therapy (R^2 = 0.54) than in second-line therapy (R^2 = 0.37).
  - PFS seemed to be better correlated with OS than TTP.
- Using the effect size for PFS as a surrogate endpoint for OS results in a 79.5% chance of correctly identifying the clinically effective trial from a random pair of trials (i.e., identifying the trial in which the experimental treatment is more effective than the control treatment).
- We present a novel application of the ROC curve, typically used for evaluating classification properties of diagnostic measures, as a useful visual tool for evaluating surrogate endpoints in oncology.

REFERENCES

1. Saad ED, et al. Ann Oncol. 2010 Jan;21(1):7-12.
2. Louvet C, et al. Cancer. 2001 Jun 1;91(11):2033-8.
3. Tang PA, et al. J Clin Oncol. 2007 Oct 10;25(29):4562-8.
4. Buyse M, et al. J Clin Oncol. 2007 Nov 20;25(33):5218-24.
5. Buyse M, et al. Meta-Analysis Group in Cancer. Lancet. 2000 Jul 29; 356(9227):373-8.
6. Fleming TR, et al. J Clin Oncol. 2009 Jun 10;27(17):2874-80.
7. Kemeny N, et al. J Clin Oncol. 2004 Dec 1;22(23):4753-61.
8. Hausmaninger H, et al. Eur J Cancer. 1999 Mar;35(3):380-5.
9. Hecht JR, et al. J Clin Oncol. 2009 Feb 10;27(5):672-80.

CONTACT INFORMATION

Costel Chirila, PhD
Research Statistician

RTI Health Solutions
200 Park Offices Drive
Research Triangle Park, NC 27709

Phone: 919-541-8083
Fax: 919-541-7222
E-mail: cchirila@rti.org

Presented at: 35th European Society for Medical Oncology Congress
October 8-12, 2010
Milan, Italy