Association between treatment effects on disease progression end points and overall survival in clinical studies of patients with metastatic renal cell carcinoma

TE Delea*,1, A Khuu1, DYC Heng2, T Haas3 and D Soulières4

1PAI (Policy Analysis Inc.), 4 Davis Court, Brookline, MA 02445, USA; 2Tom Baker Cancer Centre, University of Calgary, 1331 29th Street NW Calgary, Alberta T2N 4N2, Canada; 3Novartis Pharma, Basel CH-4002, Switzerland; 4Centre Hospitalier de l’Université de Montréal, Montreal, Quebec, Canada

BACKGROUND: The relationship between progression-free survival and time to progression (PFS/TTP) and overall survival (OS) has been demonstrated in a variety of solid tumours but not in metastatic renal cell carcinoma (mRCC). METHOD: A systematic literature search was conducted to identify controlled trials of cytokine or targeted therapies for mRCC reporting information on treatment effects on PFS/TTP and OS for one or more comparison. The associations between treatment effects on PFS/TTP and OS were analysed using linear regression. RESULTS: Thirty-one studies representing 10,943 patients, 75 treatment groups, and 41 comparisons were identified. The correlation coefficient between the negative log of the hazard ratio (HR) for PFS/TTP (– ln HRPFS/TTP) vs the negative log of the HR for OS (– ln HROS) was 0.80 (P < 0.0001). In linear regression, the coefficient on – ln HRPFS/TTP vs – ln HROS was 0.64 (95% confidence interval (CI): 0.47–0.81; R² = 0.63), suggesting each 10% relative risk reduction (RRR) for PFS/TTP was associated with a 6% RRR for OS. A 1-month gain in median PFS/TTP was associated with a 1.17-month gain in median OS (95% CI: 0.59,1.76; R² = 0.28). CONCLUSION: In trials of treatments for mRCC, treatment effects on PFS/TTP are strongly associated with treatment effects on OS. British Journal of Cancer (2012) 107, 1059–1068. doi:10.1038/bjc.2012.367 www.bjcancer.com

Keywords: metastatic; progression; renal cell carcinoma; survival

Overall survival (OS) is the gold standard for the assessment of efficacy in phase III trials of cancer therapies (Sargent and Hayes, 2008). However, use of OS as the primary end point requires that large numbers of patients be followed for an extended period of time to detect statistically significant differences between the treatment groups, thereby increasing study costs and delaying access to potentially beneficial treatments. Also, for ethical or practical reasons, patients randomised to control therapy are often allowed to crossover to study therapy, or receive an off-study investigational or other active treatment upon disease progression, thereby diluting the observed effect of study treatment on OS. These factors make measures of time to disease progression (e.g., progression-free survival (PFS) or time to progression (TTP)) attractive alternatives to OS. Measures of PFS/TTP generally require fewer patients and/or shorter follow-up to detect statistically significant differences between the treatment groups, and are not confounded by use of subsequent therapies upon disease progression. Moreover, PFS/TTP may be important measures per se, as disease progression may be associated with reduced patient health-related quality of life and increased healthcare costs.

The use of PFS/TTP as a valid surrogate end point for OS requires that treatment effects on OS can be reliably predicted from observed treatment effects on PFS/TTP (Fleming and DeMets, 1996; Tang et al, 2007; Burzykowski et al, 2008). Although the association between treatment effects on PFS/TTP and treatment effects on OS has been examined in a variety of solid tumours (Louvet et al, 2001; Johnson et al, 2006; Buyse et al, 2007; Tang et al, 2007; Sherrill et al, 2008), it has not been rigorously examined in patients with metastatic renal cell carcinoma (mRCC) (Knox, 2008). The objective of this study was to evaluate the association between treatment effects on PFS/TTP and treatment effects on OS in randomized controlled trials of patients with mRCC.

MATERIALS AND METHODS

Search strategy

Medline was searched to identify clinical trials of interleukin-2, interferon (IFN)-α, axitinib, lapatinib, pazopanib, sunitinib, sorafenib, bevacizumab, everolimus, or temsirolimus in mRCC. The search was limited to studies published in English from January 1997 to January 2010, which reported data on survival and/or mortality in the abstract. Abstracts of identified studies were reviewed by two independent reviewers (AK and TED) to identify studies for which full-text articles would be retrieved and reviewed. This search was supplemented with hand searches of the American Society of Clinical Oncology (ASCO) and European Cancer Organisation (ECO) Web sites for abstracts, posters, and/or presentations reported between January 2005 and December 2010, as well as reference lists of retrieved articles and prior

*Correspondence: TE Delea; E-mail: tdelea@pai2.com

Revised 31 May 2012; accepted 24 July 2012; published online 30 August 2012
meta-analyses and systematic reviews (Coppin et al, 2005, 2008; Coppin, 2008; Thompson Coon et al, 2010). Studies were included if they reported median PFS/TTP and median OS for two or more treatment groups or hazard ratios (HRs) for PFS/TTP and HRs for OS for one or more treatment comparison.

Data extraction

For each study selected for inclusion, information was extracted on first author, year of publication, prior treatment (treatment-naive, prior cytokine treatment, prior targeted treatment, mixed prior treatment), treatments evaluated, measures of PFS/TTP used (PFS or TTP), overall response rate (ORR) and whether trial patients were allowed to crossover to other study therapy or other active treatment after progression. For each treatment group, sample sizes for ORR, PFS/TTP and OS, median PFS/TTP, median OS, and corresponding 95% confidence intervals (CIs) for median PFS/TTP and median OS were recorded. Also recorded were HRs (and corresponding 95% CIs) for PFS/TTP and OS. Studies representing duplicate reports of the same trial were excluded, with the report least likely to have been impacted by crossover selected for the analysis (e.g., based on rank-preserving structural failure time (RPSFT) models or inverse probability of censoring weighted (IPCW) analyses, with patients censored at crossover, or at study unblinding before crossover).

Measures of treatment effect

Two measures of treatment effects on PFS/TTP and OS were analysed: (1) the absolute differences between the treatment groups in median PFS/TTP (in months) vs the absolute differences between groups in median OS (in months) and (2) the negative of the natural log of the HR for PFS/TTP (-ln HR_{PFS/TTP}) vs the negative of the natural log of the HR for OS (-ln HR_{OS}). For small treatment effects (relative risk reduction (RRR) \leq \pm 30\%), the -ln (HR) is approximately equal to the RRR. The HR is frequently used as the primary measure of treatment effect in controlled clinical trials. However, the median survival for each treatment group is also frequently reported. The advantage of the HR is that it reflects a comparison of hazards for the entirety of the survival distribution, whereas the difference in medians reflects a comparison at a single point on the distribution. On the other hand, if treatment has no effect on post-progression survival, the gain in median PFS will be an unbiased estimate of the gain in median OS regardless of the duration of PPS, whereas the HR for OS will tend to be greater than that for PPS, and the degree of difference will depend on the duration of PPS (Broglio and Berry, 2009). Both of these measures have been used in prior studies of the association between PFS/TTP and OS in other cancers (Louvet et al, 2001; Johnson et al, 2006; Buyse et al, 2007; Tang et al, 2007; Sherrill et al, 2008).

For studies that reported both PFS and TTP, we recorded PFS. For those that reported TTP only, we combined TTP results with those for PFS from other studies. Although TTP and PFS are different measures, in the setting of mRCC, whether survival is short and death due to reasons other than mRCC is rare, the HRs and differences in median survival are likely similar for TTP and PFS. TTP and PFS have been combined in prior studies of the association between disease progression end points and OS in other tumours (Tang et al, 2007; Sherrill et al, 2008). In an evaluation of studies of metastatic breast cancer patients in which TTP and PFS were not combined but were analysed separately, the associations between TTP and PFS on the one hand and OS on the other were similar (Burzykowski et al, 2008).

For studies that did not report HRs for PFS/TTP or OS, HRs were estimated using data from Kaplan–Meier curves or numbers of events and log-rank statistics (Tierney et al, 2007). For treatment arms for which median OS was not reached but for which Kaplan–Meier survival curves were reported, median survival was estimated by fitting Weibull survival functions to reported Kaplan–Meier curves (Carroll, 2003). For studies that included more than the two treatment groups, treatment effects on PFS/TTP and OS were calculated for \( k - 1 \) of \( k \) potential comparisons (e.g., for a study with treatments A, B, and C, we calculated two comparisons: A vs B and A vs C). In cases with an obvious control arm, this arm was selected as the reference arm for all comparisons.

Statistical analyses

The possibility of publication bias was assessed by examining asymmetry of a funnel plot of estimates of -ln HR_{OS} vs its s.e. and using Egger’s test (Egger et al, 1997). Pearson correlation coefficients between treatment effects on PFS/TTP and treatment effects on OS were calculated. In calculating Pearson correlation coefficients, each treatment comparison was weighted by the sum of the number of patients in the two treatment groups compared (non-parametric Spearman correlation coefficient were also calculated and were virtually the same as Pearson correlations and were not reported). The associations between treatment effects on PFS/TTP and treatment effects on OS also were examined using ordinary least squares regression with each treatment comparison weighted by the sum of the number of patients in the two treatment groups. Ninety-five percent prediction limits were calculated from weighted regressions using the mean number of patients per comparison as a weight.

Analyses were conducted separately by prior treatment for mRCC (none vs any), PFS/TTP endpoint reported (PFS vs TTP), whether crossover to active therapy after disease progression was allowed, and year of publication. Analyses also were conducted using all potential comparisons from trials with more than two treatment arms (e.g., for a study with treatments A, B, and C, we calculated three comparisons: A vs B, A vs C, and B vs C), setting the intercept terms in regression models to zero, and using all comparisons and setting intercept terms to zero. An analysis also was conducted to assess the association between ORR and OS for studies that reported ORR. In this analysis, comparisons involving arms with zero or missing response data were excluded. The treatment effect on ORR was measured in terms of the natural log of the relative risk of the response (ln RR_{ORR}) and the treatment effect on OS was measured in terms of the -ln HR_{OS}. An analysis also was conducted of the association between the \(-ln HR_{PFS/TTP}\) and \(-ln HR_{OS}\) in which each comparison was weighted by inverse of the variance of the \(-ln HR_{OS}\) rather than the number of subjects.

RESULTS

Search results

The search identified 235 potential studies. From these, as well as hand searches of reference lists of retrieved studies, ASCO and ECCO web sites, and prior systematic reviews, a total of 31 studies were identified, representing 10 943 patients, 75 treatment groups, and 41 potential treatment comparisons that reported sufficient information for either the analysis of correlation between differences in median PFS/TTP and differences in median OS or between -ln HR_{PFS/TTP} and -ln HR_{OS} (Table 1) (Kruit et al, 1997; Negrier et al, 1998, 2000, 2007, 2008; Medical Research Council Renal Centre Collaborators, 1999; Pyrhonen et al, 1999; Motzer et al, 2000, 2007, 2008, 2009, 2010; Atzpodien et al, 2001, 2002, 2004, 2006; Dutcher et al, 2003; Yang et al, 2003; Atkins et al, 2004; Aass et al, 2005; Donskov et al, 2005; McDermott et al, 2005; Tannir et al, 2006; Bukowski et al, 2007; Escudier et al, 2007a, b; Hudes et al, 2007; Amato et al, 2008; Figlin et al, 2008; Sternberg
| Author (Year) | Treatment | Cross-over | PFS (months) | OS (months) | N (w) | HR (95% CI) | Median months | PFS | OS |
|--------------|------------|------------|-------------|------------|-------|-------------|---------------|-----|----|
| Dutcher (2001) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2001) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2005) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2005) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2008) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2008) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2009) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2009) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2010) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2010) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2011) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2011) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2012) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2012) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2013) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2013) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2014) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2014) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2015) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2015) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2016) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2016) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Dutcher (2017) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
| Escudier (2017) | IFNA + VBL | 63 | 314.5 | 0.94 | 1.02 | 1.02 | 1.04 | 1.05 | 1.00 |
### Table 1 (Continued)

| Author (Year) | Prior treatment | PFS/ TTP end point | Crossover | Treatment | N | ORR (%) | Median months (95% CI) | N for comparison* | RR ORRd | PFS | OS | Median months (95% CI) | HR (95% CI) |
|---------------|-----------------|--------------------|-----------|-----------|---|--------|------------------------|-----------------|----------|--------|-----|------------------------|--------------|
| Negrier       | None            | IFS                | IFNα + IL-2 | 70 | 1.4 | 2.0 | (3.0–3.9) | 13.0 | IFNα + IL + OTH        | 61 | 8.2 | 3.0 | 13.0 | 131 | 5.74 | 1.0 | 0.0 | 0.81 (0.56–1.17) |
| Negrier       | None            | IFS                | Non+IFNα | 248 | 32 | 32   | 3.3 (12.1) | 15.1 | IFNα + IL + OTH        | 244 | 7.0 | 3.5 | (3.1–3.3) | 145 | 1.54 | 246 | 2.16 | 0.3 | 0.0 | (0.69–1.57) |
| Negrier       | None            | IFS                | Non+IL   | 245 | 33 | 33   | (12.38) | 14.9 | IL + OTH              | 247 | 6.9 | 3.5 | (3.1–3.3) | 144 | 1.84 | 246 | 2.11 | 0.3 | 0.0 | 1.04 |
| Negrier       | None            | IFS                | IFNα + IL-2 | 80 | 17.9 | 7.2 | (6.0–36) | 37.7 | IFNα + IL-2 + OTH     | 75 | 21.3 | 6.2 | (5.1–5.3) | 131 | 0.81 | 155 | 1.1 | 1.0 | 0.8 | (0.87–1.33) |
| Pyrhonen      | Mixed           | IFS                | VBL       | 81 | 2.4 | 2.1 | 8.7 | 16.0 | IFNα + VBL            | 79 | 16.5 | 3.0 | 15.3 | 160 | 6.88 | 0.9 | 0.6 | 1.04 |
| Ravalard      | Mixed           | TTP                | HT        | 207 | 0.5 | 3.5  | 9.9 | 209 | IFNα + VBL            | 209 | 1.4 | 3.5 | 10.8 | 416 | 2.80 | 0.0 | 0.0 | 0.71 |
| Rini          | None            | IFNα               | IFNα + IL-2 | 363 | 13.1 | 5.2 | (3.1–5.6) | 17.4 | BEV + IFNα            | 369 | 25.5 | 8.5 | (7.5, 9.7) | 163 | 2.23 | 722 | 1.95 | 3.3 | 0.9 | (0.66–1.13) |
| Stenberg      | Mixed           | IFS                | BEV       | 145 | 3.4 | 42   | 18.7 | 290 | IFNα + VBL            | 290 | 30.3 | 9.2 | (9.1–21.0) | 435 | 8.80 | 5.0 | 2.4 | 0.46 |
| Yann          | Mixed           | TTP                | BEV-10 mg | 59 | 6.8 | 37   | (20-55) | 25.5 | IFNα + VBL            | 59 | 6.8 | 3.4 | (2-1.55) | 147 | 2.66 | 118 | 1.00 | 0.0 | 0.0 | (0.79–1.66) |
| Yang          | Mixed           | TTP                | BEV-3mg   | 40 | 0.0 | 2.5  | 13.2 | 40 | IFNα + VBL            | 39 | 10.3 | 4.8 | 15.5 | 59 | 2.3 | 2.3 | 0.39 | 0.8 |

Abbreviations: BEV = bevacizumab; CAP = capetabine; CI = confidence interval; CRA = 13-cis-retinoic acid; ERL = erlotinib; EVE = everolimus; 5FU = 5-fluorouracil; GEF = gefitinib; HD = hormone deprivation; HR = hazard ratio; HT = hormone therapy; IFNα = interferon-α; IFNG = Interferon-γ; IL = interleukin-2; IL-IV = inhaled interleukin-2; ILIV = intravenous interleukin-2; MPA = medroxyprogesterone acetate; MRCRCC = Medical Research Council Renal Cancer Collaborators; MR = not reached; ORR = overall response rate; OS = overall survival; PAZ = pazopanib; PLC = placebo; PFS = progression-free survival; RD = recommended dose; RR = relative risk; TAM = tamoxifen; TTP = time to disease progression; TEM = temsirolimus; TTP = time to progression; SOR = sorafenib; SUT = sunitinib; VBL = vinblastine. For each trial with number of treatment arms (k), k – 1 comparisons shown. *N for comparison adjusted for number of times each arm contributes to a comparison. aCrossover permitted from placebo to sorafenib after single planned analysis of PFS showed a statistically significant benefit of sorafenib over placebo. bCrossover not allowed but analyses potentially confounded by off-label use of bevacizumab. cIFNα patients allowed to crossover to sunitinib following documented disease progression. dHR (95% CI) not reported; estimated from Kaplan–Meier survival curves. eSixteen patients from tamoxifen arm crossed over to immunochemotherapy regimen after progressive disease. fTwo patients in the matinib arm crossed over to the gefitinib arm after experiencing disease progression. gHR (95% CI) not reported; estimated from Kaplan–Meier survival curves. hSixteen patients from tamoxifen arm crossed over to immunochemotherapy regimen after progressive disease. iMedian not reported; values reported were obtained by fitting Weibull survival functions to published survival curve data. Values base on Weibull functions were used in sensitivity analysis only. jCrossover permitted from placebo to sorafenib after single planned analysis of PFS showed a statistically significant benefit of sorafenib over placebo. kCrossover not allowed but analyses potentially confounded by off-label use of bevacizumab. lIFNα patients allowed to crossover to sunitinib following documented disease progression. m50% patients in both the groups received other therapies post progression; Ns and results for OS based on patients who did not receive any post-study treatment. nPlacebo patients allowed to crossover to everolimus after documented progression. Median OS was based on analysis using rank-preserving structural failure time model (Korhonen (2009)); HR for OS was based on analysis using inverse probability of censoring weighted (IPCW) analysis (Weiderkehr (2009)). oPlacebo patients allowed to crossover to pazopanib after documented progression. HR for OS was based on analysis using rank-preserving structural failure time model (Stenberg (2010)); 95% CI for HR based on value. Median OS based on unadjusted analysis.
et al, 2009, 2010a, b; Gore et al, 2010; Korhonen and Malangone, 2010; Rini et al, 2010; Korhonen et al, 2011; Wiedekerke et al, 2011). The great majority of the studies that were excluded for lack of information on both PFS and TTP and OS.

**Study characteristics**

Fifteen studies (48%) were published before 2006; 17 (55%) were in treatment-naive patients; seven (23%) allowed crossover to active treatment after disease progression. Ten studies (32%) included one or more targeted treatments. For the phase III trial of sunitinib vs IFN, several analyses of OS were conducted, which might be differentially affected by crossover from IFN to sunitinib. In our base case, we used the results from the analysis in which patients who received any post-study treatment were excluded (HR 0.647, 95% CI: 0.483–0.870, median OS 28.1 months vs 14.1 months for sunitinib (n = 193) vs IFN (n = 162)) (Figlin et al, 2008; Motzer et al, 2009). The HR from this analysis was virtually identical to that reported in the interim analysis of the ITT population before patients were allowed to crossover (HR 0.65, 95% CI: 0.449–0.942, median OS not reached for sunitinib (n = 375) or IFN (n = 375)) (Motzer et al, 2007). We used the values from the former because median OS was not reached for the latter. For the phase III trial of everolimus, placebo patients were allowed to crossover to everolimus after documented progression (McDermott et al, 2005; Korhonen and Malangone, 2010; Motzer et al, 2010; Korhonen et al, 2011; Wiedekerke et al, 2011). For this study we used median OS based on analysis using the RPSFT model to control for crossover (Korhonen and Malangone, 2010; Korhonen et al, 2011); the HR for OS was based on analysis using IPCW analysis (Wiedekerke et al, 2011). For phase III trial of pazopanib, the HR for OS was based on the analysis using RPSFT to control for crossover (Sternberg et al, 2010b).

Thirty studies representing 40 treatment comparisons reported median PFS/TPP and median OS for one or more comparisons. Median OS was estimated based on fitting of Weibull survival functions to Kaplan–Meier curves for one treatment arm (bevacizumab plus placebo arm in the study by Bukowski et al (2007). This arm was represented in one comparison. Across all studies, median PFS/TPP and OS averaged 4.9 and 16.6 months, respectively. The median difference between the treatment groups in PFS/TPP averaged 1.4 months (s.d. 2.1 months, range 1.4–7.1 months); the median difference between the treatment groups in OS averaged 2.0 months (s.d. 5.7 months, range 18.0–14.0 months). Twenty-eight studies representing 36 treatment comparisons reported sufficient information for the analysis of −ln HR_{PFS/TPP} vs −ln HR_{OS}. The −ln HR_{PFS/TPP} averaged 0.31 (s.d. 0.36, range 0.42–1.17); the −ln HR_{OS} averaged 0.15 (s.d. 0.27, range 0.45–0.84). HRs were estimated from Kaplan–Meier curves or log-rank statistics and event counts in 40 treatment arms represented in 23 comparisons.

The funnel plot of estimates of −ln HR_{OS} vs corresponding s.e.’s provided no strong evidence of publication bias (Figure 1). The estimated intercept on a regression of the inverse s.e.’s provided no strong evidence of publication bias (Figure 1). The weighted Pearson correlation coefficient for the difference in median PFS/TPP and the difference in median OS was 0.54 (P = 0.0002). In linear regression analysis, a 1-month difference in median PFS/TPP was associated with a 1.17-month difference in median OS (95% CI: 0.59, 1.76; adjusted R^2 = 0.28) (Figure 2).

The weighted Pearson correlation coefficient for −ln HR_{PFS/TPP} vs −ln HR_{OS} was 0.68 (P < 0.0001). The coefficient on −ln HR_{PFS/TPP} vs −ln HR_{OS} was 0.64 (95% CI: 0.47, 0.81; adjusted

**Association between treatment effects on PFS/TPP and treatment effects on OS**

The weighted Pearson correlation coefficient for the difference in median PFS/TPP and the difference in median OS was 0.54 (P = 0.0002). In linear regression analysis, a 1-month difference in median PFS/TPP was associated with a 1.17-month difference in median OS (95% CI: 0.59, 1.76; adjusted R^2 = 0.28) (Figure 2).

The weighted Pearson correlation coefficient for −ln HR_{PFS/TPP} and −ln HR_{OS} was 0.80 (P < 0.0001). The coefficient on −ln HR_{PFS/TPP} vs −ln HR_{OS} was 0.64 (95% CI: 0.47, 0.81; adjusted

**Figure 1** Funnel plot of negative log of HR for OS vs corresponding s.e. for each comparison. The funnel plot shows an assessment of publication bias. If there is no publication bias, the coordinates should be scattered symmetrically around the pooled estimate. The vertical line represents the fixed effects pooled estimate of −ln HR_{OS}. The diagonal lines describing the funnel represent the 95% CI for each value of the s.e. The outlier is the coordinate for the pivotal study of pazopanib (−lnHR_{OS} = 0.84, s.e. (−lnHR_{OS}) = 0.62) (Stenburg et al, 2010b). The relatively high degree of imprecision associated with this estimate was due to the RPSFT method used to analyse OS to control for crossover.

**Figure 2** Association between differences in median PFS/TPP and differences in median OS. Abbreviation: R^2 = adjusted R-squared. Area of bubbles is proportional to the number of patients. Solid line is predicted value. Dashed lines are prediction intervals.

R^2 = 0.63) (Figure 3), suggesting that a 10% increase in the RRR for PFS/TPP is associated with an ~6% increase in the RRR for OS.

**Subgroup and sensitivity analyses**

Results in subgroups of studies are presented in Table 2. The correlation between treatment effects on PFS/TPP and treatment effects on OS was greater in studies that did not allow/require crossover, studies that used PFS rather than TTP, and in studies published before 2005 (studies before 2005 were less likely to have allowed crossover). There was no significant association between the treatment effects on PFS/TPP and OS in the subset of trials of vascular endothelial growth factor (VEGF) inhibitors, although there was a trend in the linear regression for −ln HR_{PFS/TPP} vs −ln HR_{OS} (P = 0.0510). Results were similar to those of primary analysis when all potential comparisons from trials with multiple treatment arms were included. The adjusted R^2 for the analysis of differences in median PFS/TPP vs differences in median OS was greater with the exclusion of the study by Bukowski et al (2007), a randomized phase II trial comparing bevacizumab plus erlotinib.
Clinical Studies

several novel treatments for mRCC have been approved for use in a surrogate for OS in pivotal studies (Buyse et al., 2009). However, the proportion of variability in treatment effects on OS that was explained by treatment effects on PFS/TTN was modest. In particular, the adjusted \( R^2 \) was 0.63 for the association between \(-\ln HR_{OS} \) and \(-\ln HR_{PFS/TTN} \). This value is within the range reported in prior analyses of the relationship between treatment effects on PFS/TTN and OS (Sherrill et al., 2008). A high \( R^2 \) is not a necessary criterion for surrogacy, however, as some of the unexplained variation may reflect the sampling error in each trial due to small sample size. Even for a perfect surrogate end point, therefore, \( R^2 \) will be less than one in a set of trials with small samples (Tang et al., 2007). The trials examined in this evaluation were relatively small (median of 96 patients per arm). Moreover, there is no standard value above which an \( R^2 \) (or correlation coefficient) can be claimed to be sufficient. The adjusted \( R^2 \) for the association between differences in median PFS/TTN and differences in median OS was only 0.28. While the difference in median survival times may be a more appropriate measure of treatment effect than HRs if the proportional hazards assumption is violated, median survival times represent only a single point on the survival distribution and are potentially imprecise. It is not surprising therefore that the amount of unexplained variation is greater when treatment effects are measured in terms of differences in median survival. Despite the relatively low \( R^2 \) from this regression, it is useful to note that the results from the regression analysis presented here suggest that, on average, there is an slightly better than 1-month gain in median OS associated with a 1 month gain in median PFS/TTN. This is consistent with the hypothesis that treatment effects on post-progression survival are uncorrelated with treatment effects on PFS/TTN (Bowater et al., 2008).

Not surprisingly, the association between treatment effects was stronger in studies that did not allow crossover to active treatment. Additionally, the association between treatment effects on PFS/TTN and OS were less in trials conducted after 2005, when targeted therapies for treatment of mRCC were more likely to be available as potential off-study second-line treatments. Estimates of the association between treatment effects on PFS/TTN and OS based on the entire sample of trials may therefore be conservative. An increase in response rate was also correlated with OS, although the association was not as strong as that with treatment effects on PFS/TTN measured in terms of \(-\ln (HR)\).

Limitations of this study should be noted. First, this study was based on published results of controlled trials which may be subject to publication bias. To the extent that only studies showing positive effects on both PFS and OS were published, then our estimates may overstate the true association between PFS and OS. However, a funnel plot analysis of the \( R^2 \) in HRs provided no evidence of publication bias (more results would be expected around the mean effect size and Egger’s test was not significant).

Ially, the assessment of association of PFS/TTN and OS should be demonstrated over different stages of the disease (as the causal pathways of the disease process might differ depending on the stage) and across classes of drug (as drugs with different modes of action may have different pathways of intervention) (Fleming and DeMets, 1996). It is possible that the association reported here could only apply to specific recognised prognostic groups, but analyses by prognostic groups were not feasible based on data reported in study publications (Molina and Motzer, 2008; Heng et al., 2009b). The majority of studies included in this analysis involved comparisons of two or more cytokine therapies. The association between treatment effects on PFS/TTN and those on OS studies of sunitinib and sorafenib have demonstrated that the introduction of these treatment has resulted in increased survival (Heng et al., 2009a; Warren et al., 2009), a rigorous examination of the association between PFS/TTN end points and OS has yet to be undertaken.

The analysis presented here suggests that treatment effects on measures of PFS/TTN are strongly associated with treatment effects on OS in patients with mRCC. However, the proportion of variability in treatment effects on OS that was explained by treatment effects on PFS/TTN was modest. In particular, the adjusted \( R^2 \) was 0.63 for the association between \(-\ln HR_{OS} \) and \(-\ln HR_{PFS/TTN} \). This value is within the range reported in prior analyses of the relationship between treatment effects on PFS/TTN and OS (Sherrill et al., 2008). A high \( R^2 \) is not a necessary criterion for surrogacy, however, as some of the unexplained variation may reflect the sampling error in each trial due to small sample size. Even for a perfect surrogate end point, therefore, \( R^2 \) will be less than one in a set of trials with small samples (Tang et al., 2007). The trials examined in this evaluation were relatively small (median of 96 patients per arm). Moreover, there is no standard value above which an \( R^2 \) (or correlation coefficient) can be claimed to be sufficient. The adjusted \( R^2 \) for the association between differences in median PFS/TTN and differences in median OS was only 0.28. While the difference in median survival times may be a more appropriate measure of treatment effect than HRs if the proportional hazards assumption is violated, median survival times represent only a single point on the survival distribution and are potentially imprecise. It is not surprising therefore that the amount of unexplained variation is greater when treatment effects are measured in terms of differences in median survival. Despite the relatively low \( R^2 \) from this regression, it is useful to note that the results from the regression analysis presented here suggest that, on average, there is an slightly better than 1-month gain in median OS associated with a 1 month gain in median PFS/TTN. This is consistent with the hypothesis that treatment effects on post-progression survival are uncorrelated with treatment effects on PFS/TTN (Bowater et al., 2008).

Not surprisingly, the association between treatment effects was stronger in studies that did not allow crossover to active treatment. Additionally, the association between treatment effects on PFS/TTN and OS were less in trials conducted after 2005, when targeted therapies for treatment of mRCC were more likely to be available as potential off-study second-line treatments. Estimates of the association between treatment effects on PFS/TTN and OS based on the entire sample of trials may therefore be conservative. An increase in response rate was also correlated with OS, although the association was not as strong as that with treatment effects on PFS/TTN measured in terms of \(-\ln (HR)\).

Limitations of this study should be noted. First, this study was based on published results of controlled trials which may be subject to publication bias. To the extent that only studies showing positive effects on both PFS and OS were published, then our estimates may overstate the true association between PFS and OS. However, a funnel plot analysis of the \( R^2 \) in HRs provided no evidence of publication bias (more results would be expected around the mean effect size and Egger’s test was not significant).

Ially, the assessment of association of PFS/TTN and OS should be demonstrated over different stages of the disease (as the causal pathways of the disease process might differ depending on the stage) and across classes of drug (as drugs with different modes of action may have different pathways of intervention) (Fleming and DeMets, 1996). It is possible that the association reported here could only apply to specific recognised prognostic groups, but analyses by prognostic groups were not feasible based on data reported in study publications (Molina and Motzer, 2008; Heng et al., 2009b). The majority of studies included in this analysis involved comparisons of two or more cytokine therapies. The association between treatment effects on PFS/TTN and those on OS

**Clinical Studies**

**Figure 3** Association between negative log of HR for PFS/TTN and negative log of HR for OS. Abbreviation: \( R^2 \) = adjusted R-squared. Area of bubbles is proportional to the number of patients. Solid line is predicted value. Dashed lines are 95% prediction intervals.

vs bevacizumab plus placebo that was an extreme outlier, with a positive treatment effect on PFS/TTN and a negative treatment effect on OS (difference in median PFS/TTN 1.4 vs difference in median OS – 18.0 (the latter was estimated based on fitting a Weibull survival function to Kaplan–Meier curves) and HR for PFS/TTN 0.86 vs HR for OS 1.57). The observed negative effect of erlotinib on OS in this study may have been due to the relatively high utilisation of non-study treatment post progression in the placebo group (Bukowski et al., 2007). The associations between treatment effects on PFS/TTN and treatment effects on OS were less strong when we used the results for OS from trials of sunitinib, everolimus, and pazopanib that were not adjusted for crossover from placebo to active therapy. The weighted Pearson correlation coefficient for the natural log of the relative risk of the OR (i.e., \( \ln R_{OR} \)) vs \( \ln HR_{OS} \) was 0.78 (P < 0.0001). In linear regression, the coefficient on \( \ln R_{OR} \) vs \( \ln HR_{OS} \) was 0.30 (95% CI: 0.20, 0.39, adjusted \( R^2 = 0.59 \) (Figure 4). In the analysis of \( \ln HR_{PFS/TTN} \) vs \( \ln HR_{OS} \) in which comparisons were weighted by the inverse variance of \( \ln HR_{OS} \) (35 comparisons), the weighted Pearson correlation coefficient for \( \ln HR_{PFS/TTN} \) vs \( \ln HR_{OS} \) was 0.76 (P < 0.0001). The coefficient on \( \ln HR_{PFS/TTN} \) was 0.53 (95% CI: 0.37, 0.68, adjusted \( R^2 = 0.56 \)). These results are qualitatively similar to those in which the results are weighted by the numbers of subjects.

**DISCUSSION**

Advances in understanding the biology and genetics of renal cell carcinoma have led to novel approaches for treatment of mRCC that target the VEGF receptor. With the growing therapeutic arsenal against mRCC, it is now feasible for patients to receive multiple lines of potentially beneficial treatment. Indeed, a recent trial reported on a study population that had received three to five prior lines of therapy (Motzer et al., 2010). With the increasing number of effective treatments available (Soulieres, 2009), the effect of first-line therapies on OS are more likely to be confounded by the effects of subsequent therapies. The question of whether PFS/TTN rather than OS should be employed as a primary outcome measure in pivotal studies of new treatments for mRCC is therefore important. This situation is similar to that with metastatic colorectal cancer, in which there was rapid development of novel treatments, necessitating the consideration of using PFS as a surrogate for OS in pivotal studies (Buyse et al., 2007). Although several novel treatments for mRCC have been approved for use in the United States with TTP or PFS as the primary end point in pivotal studies, and results of population-based historical cohort
were significant in trials evaluating targeted and non-targeted therapies. The association between treatment effects on PFS/TTP and OS was not significant for comparisons involving VEGF inhibitors, although there was a trend towards an association ($P = 0.0510$). The number of such comparisons was small, however, and these comparisons may have been more likely to have been confounded by crossover and receipt of other non-study therapies post progression. It is reasonable to assume that results presented here can be generalised to evaluations of agents such as axitinib, that have similar mechanisms of action to the therapies included in this analysis (Rugo et al, 2005; Rini et al, 2007; Rixe et al, 2007).

For studies that allowed for crossover from control to active therapy, we used the reported measure of treatment effect that was considered to be least likely to be subject to confounding by such crossover. While it would be desirable to use a common measure of treatment effect for all studies, it is well established that crossover from control to active treatment may attenuate observed treatment effects on OS relative to what would have been observed in the absence of such crossover (Finkelstein and Schoenfeld, 2011; Saad and Buyse, 2012). To include results of studies with extensive crossover without controlling for crossover would add no useful information to the analyses. The RPSFT and IPCW methods used in the analyses of everolimus (Korhonen and Malangone, 2010; Table 2

| Subgroup/Sensitivity analysis | Weighted Pearson correlation | Estimate | 95% CI | $R^2$ | Weighted Pearson correlation | Estimate | 95% CI | $R^2$
|-----------------------------|-----------------------------|---------|-------|------|-----------------------------|---------|-------|------|
| All                         | 41                          | 0.54    | 1.17  | 0.59 | 1.76 | 0.28 | 0.80 | 0.64 | 0.47 | 0.81 | 0.63|
| Prior treatment             |                             |         |       |      |     |     |     |     |     |     |     |
| None                        | 20                          | 0.57    | 1.22  | 0.35 | 2.08 | 0.29 | 0.84 | 0.61 | 0.39 | 0.82 | 0.69|
| Any                         | 21                          | 0.49    | 1.04  | 0.14 | 1.94 | 0.20 | 0.78 | 0.62 | 0.37 | 0.88 | 0.58|
| Targeted therapy            |                             |         |       |      |     |     |     |     |     |     |     |
| No                          | 27                          | 0.65    | 1.42  | 0.74 | 2.10 | 0.40 | 0.80 | 0.60 | 0.40 | 0.80 | 0.62|
| Yes                         | 14                          | 0.38    | 0.85  | -0.44| 2.13 | 0.08 | 12   | 0.79 | 0.70 | 0.32 | 1.09 | 0.59|
| Measure of disease progression |                           |         |       |      |     |     |     |     |     |     |     |
| PFS                         | 35                          | 0.55    | 1.21  | 0.56 | 1.86 | 0.28 | 31   | 0.81 | 0.68 | 0.49 | 0.86 | 0.65|
| TTP                         | 6                           | -0.10   | -0.21 | -2.98| 2.56 | -0.24| 5    | 0.64 | 0.17 | -0.20| 0.53 | 0.21|
| Crossover allowed           |                             |         |       |      |     |     |     |     |     |     |     |
| No                          | 33                          | 0.50    | 1.29  | 0.47 | 2.11 | 0.23 | 30   | 0.70 | 0.69 | 0.42 | 0.97 | 0.47|
| Yes                         | 8                           | 0.28    | 0.82  | -1.95| 3.59 | -0.07| 6    | 0.61 | 0.63 | -0.49| 1.76 | 0.22|
| Year of publication         |                             |         |       |      |     |     |     |     |     |     |     |
| <2000                       | 22                          | 0.80    | 1.78  | 1.15 | 2.41 | 0.62 | 21   | 0.69 | 0.55 | 0.27 | 0.83 | 0.45|
| >2000                       | 19                          | 0.59    | 1.22  | 0.37 | 2.08 | 0.31 | 15   | 0.84 | 0.68 | 0.41 | 0.95 | 0.68|
| Number of subjects          |                             |         |       |      |     |     |     |     |     |     |     |
| <200                        | 20                          | 0.33    | 2.15  | -0.95| 5.26 | 0.06 | 17   | 0.70 | 0.51 | 0.23 | 0.80 | 0.46|
| ≥200                        | 21                          | 0.66    | 1.13  | 0.51 | 1.74 | 0.40 | 19   | 0.82 | 0.67 | 0.43 | 0.90 | 0.65|
| HR estimated from Kaplan–Meier curves |               |         |       |      |     |     |     |     |     |     |     |
| No                          | 12                          | 0.82    |       |     |     |     | 12   | 0.82 | 0.61 | 0.31 | 0.91 | 0.64|
| Yes                         | 24                          | 0.72    |       |     |     |     | 24   | 0.72 | 0.69 | 0.40 | 0.98 | 0.50|
| Drug class                  |                             |         |       |      |     |     |     |     |     |     |     |
| Cytokines                   | 27                          | 0.67    | 1.75  | 0.94 | 2.55 | 0.42 | 22   | 0.76 | 0.66 | 0.40 | 0.93 | 0.56|
| VEGF inhibitors             | 9                           | 0.50    | 1.43  | -0.80| 3.65 | 0.14 | 9    | 0.66 | 0.65 | 0.00 | 1.30 | 0.36|
| MTOR inhibitors             | 5                           | 0.87    | 1.65  | -0.06| 3.36 | 0.68 | 5    | 0.93 | 0.70 | 0.21 | 1.19 | 0.83|
| Exclude Bukowski et al (2007) | 40                      | 0.60    | 1.16  | 0.65 | 1.67 | 0.34 | 35   | 0.83 | 0.63 | 0.47 | 0.79 | 0.64|
| No adjustment for crossover in sunitinib, everolimus, and pazopanib trials |         |         |       |      |     |     |     |     |     |     |     |
| No intercept                | 41                          | 0.54    | 1.20  | 0.76 | 1.65 | 0.28 | 36   | 0.80 | 0.58 | 0.46 | 0.69 | 0.69|
| All potential comparisons   | 48                          | 0.55    | 1.25  | 0.68 | 1.82 | 0.24 | 42   | 0.79 | 0.64 | 0.48 | 0.80 | 0.80|
| All comparisons, no intercept | 48                      | 0.55    | 1.33  | 0.90 | 1.76 | 0.27 | 42   | 0.79 | 0.56 | 0.46 | 0.67 | 0.67|
| All comparisons, no intercept, exclude Bukowski et al (2007) |         |         |       |      |     |     |     |     |     |     |     |

Abbreviations: CI = confidence interval; HR = hazard ratio; MTOR = mammalian target of rapamycin; N = number of comparisons; OS = overall survival; PFS = progression-free survival; $R^2$ = adjusted R-squared; TTP = time to progression; VEGF = vascular endothelial growth factor. Adjusted $R^2$ for regressions without intercept may not be comparable to those with intercept and are not reported.

For studies that allowed for crossover from control to active therapy, we used the reported measure of treatment effect that was considered to be least likely to be subject to confounding by such crossover. While it would be desirable to use a common measure of treatment effect for all studies, it is well established that crossover from control to active treatment may attenuate observed treatment effects on OS relative to what would have been observed in the absence of such crossover (Finkelstein and Schoenfeld, 2011; Saad and Buyse, 2012). To include results of studies with extensive crossover without controlling for crossover would add no useful information to the analyses. The RPSFT and IPCW methods used in the analyses of everolimus (Korhonen and Malangone, 2010; Rinox et al, 2007).
Figure 4  Association between the log of relative risk of overall response and the negative log of the hazard ratio of OS. $R^2 =$ adjusted R-squared. Area of bubbles is proportional to the number of patients. Solid line is predicted value. Dashed lines are 95% prediction intervals.

Korhonen et al, 2011; Wiederkehr et al, 2011) and pazopanib (Sternberg et al, 2010b) are useful methods for analysing OS in the context of selective crossover (Finkelstein and Schoenfeld, 2011; Morden et al, 2011; Rimawi and Hilsenbeck, 2012).

In unblinded trials, there may be a motivation for clinicians to call a patient's disease progression earlier if the patient is in the control arm than if the same patient had been in the experimental arm (Dodd et al, 2008). To the extent that this inflates the treatment effects on PFS, the association between treatment effects on PFS/TTP and treatment effect on OS might be attenuated (because OS is not impacted by this bias). The use of blinded independent central review (BICR) may reduce any such bias. However, retrospective BICR may necessitate informative censoring on local assessment of progression, which may bias the comparison in favour of control patients (Dodd et al, 2008). This also would attenuate the observed association between treatment effect on PFS/TTP and treatment effect on OS. Treatment assignment was blinded in only six of the studies included in the analyses. Independent review of progression was employed in six studies. As studies that used blinded treatment assignment and/or review of progression tended to be those evaluating novel targeted agents, assessment of the independent effects of blinding of treatment assignment and/or BICR on the association between treatment effects on PFS and treatment effects on OS was infeasible.

In conclusion, results presented in this study suggest that treatment effects on diseases progression end points are strongly associated with treatment effects on OS. Further research is required to establish whether disease progression end points may be used as surrogate end points for OS in clinical trials of novel treatments for mRCC.

ACKNOWLEDGEMENTS

This work was supported by Novartis Pharmaceuticals Corp., Florham Park, NJ, USA. We thank Andrea Kay, Jean Francois Baladi, and Oleg Sofrygin for their assistance in the conduct of this study.

Conflict of interest

TED has received research support and consulting fees from Novartis, GlaxoSmithKline, Amgen, and other companies, which manufacture or market treatments for patients with metastatic renal cell carcinoma. TH is an employee of Novartis and owns stock or stock options in Novartis. DS has participated on Advisory Boards and received Honoraria from Novartis. The remaining authors declare no conflict of interest.

REFERENCES

Aass N, De Mulder PH, Mickisch GH, Mulders P, van Oosterom AT, van PH, Fossa SD, de PL, Sylvester RJ (2005) Randomized phase II/III trial of interferon alfa-2a with and without 13-cis-retinoic acid in patients with progressive metastatic renal cell carcinoma: the European Organisation for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group (EORTC 30951). J Clin Oncol 23(18): 4172–4178

Amato RJ, Jac J, Hernandez-McClain J (2008) Interferon-alpha in combination with either imatinib (Gleevec) or gefitinib (Iressa) in metastatic renal cell carcinoma: a phase II trial. Anticancer Drugs 19(5): 527–533

Atkins MB, Hidalgo M, Stadler WM, Logan TF, Dutcher JP, Hudes GR, Park Y, Liu SH, Marshall B, Boni JP, Dukart G, Sherman ML (2004) Randomized phase II study of multiple dose levels of CCI-779, a novel mammalian target of rapamycin kinase inhibitor, in patients with advanced refractory renal cell carcinoma. J Clin Oncol 22(5): 909–918

Atzpodien J, Hoffmann R, Franzeke M, Stief C, Wandert T, Reitz M (2002) Thirteen-year, long-term efficacy of interferon 2alpha and interleukin-2-based home therapy in patients with advanced renal cell carcinoma. Cancer 95(5): 1045–1050

Atzpodien J, Kirchner H, Illiger HJ, Metzner B, Ukema D, Schott H, Funke PJ, Gramatziki M, Jurgenson S, Wandert T, Patzelt T, Reitz M (2001) IL-2 in combination with IFN- alpha and 5-FU versus tamoxifen in metastatic renal cell carcinoma: long-term results of a controlled randomized clinical trial. Br J Cancer 85(8): 1130–1136

Atzpodien J, Kirchner H, Jonas U, Bergmann L, Schott H, Heynemann H, Fornara P, Loening SA, Roigas J, Muller SC, Bodenstein H, Pomer S, Metzner B, Rebmann U, Oberneder R, Siebels M, Wandert T, Puchberger T, Reitz M (2004) Interleukin-2- and interferon alfa-2a-based immunochemotherapy in advanced renal cell carcinoma: a Prospectively Randomized Trial of the German Cooperative Renal Carcinoma Chemoinmunotherapy Group (DGClCN). J Clin Oncol 22(7): 1188–1194

Atzpodien J, Kirchner H, Rebmann U, Soder M, Gertenbach U, Siebels M, Roigas J, Raschke R, Salm S, Schwind B, Muller SC, Hauser S, Leiber C, Huland E, Heinzer H, Siemer S, Metzner B, Heynemann H, Fornara P, Reitz M (2006) Interleukin-2-interferon-alpha2a/13-retinoic acid-based chemoimmunotherapy in advanced renal cell carcinoma: results of a prospectively randomised trial of the German Cooperative Renal Carcinoma Chemoinmunotherapy Group (DGClCN). Br J Cancer 95(4): 463–469

Bowater RJ, Bridge LJ, Lilford RJ (2008) The relationship between progression-free and post-progression survival in treating four types of metastatic cancer. Cancer Lett 262(1): 48–53

Broglio KR, Berry DA (2009) Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst 101(23): 1642–1649

Information from the trial reports on the frequency of assessments, the criteria used to assess response and/or progression, or the duration of treatment was not extracted. It therefore was not feasible in this analysis to assess how these and other unmeasured factors might affect the association between treatment effects on PFS and treatment effects on OS. Differences in these factors might help explain some variability in observed associations between treatment effects on PFS/TTP and on OS.

As the searches upon which this study was based were conducted in 2010, results of randomized controlled trials of systemic therapies for mRCC may have been published since the original literature search for this study was conducted. One such trial is the Renal EFFECT trial, a randomized controlled trial of intermittent vs continuous sunitinib (Motzer et al, 2012). It may be worthwhile in future research to update these analyses using results of this and other recently published studies, and to explore in multivariate analysis the independent effects of study design and other factors on the associations between treatment effects on PFS/TTP and treatment effects on OS.

© 2012 Cancer Research UK
Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, Oudard S, Negrier S, Szczylik C, Kim ST, Chen I, Bycott PW, Baum CM, Figlin RA (2007) Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 356(2): 115–124

Motzer RJ, Murphy BA, Bacik J, Schwartz LH, Nanus DM, Mariani T, Loehrer P, Wilding G, Fairclough DL, Cella D, Mazumdar M (2000) Phase III trial of interferon alfa-2a with or without 13-cis-retinoic acid for patients with advanced renal cell carcinoma. *J Clin Oncol* 18(16): 2972–2980

Negrier S, Caty A, Lesimple T, Douillard JY, Escudier B, Rossi FJ, Viens P, Gomez F (2000) Treatment of patients with metastatic renal carcinoma with a combination of subcutaneous interleukin-2 and interferon alfa with or without fluorouracil. Groupe Francais d’Immunotherapie, Federation Nationale des Centres de Lutte Contre le Cancer. *J Clin Oncol* 18(24): 4009–4015

Negrier S, Escudier B, Lasset C, Douillard JY, Savary J, Chevreau C, Ravaud A, Mercatello A, Pena J, Mousseau M, Philip T, Tursz T, for the Groupe Francais d’Immunotherapie (1998) Recombinant human interleukin-2, recombinant human interferon alfa-2a, or both in metastatic renal-cell carcinoma. *N Engl J Med* 338(18): 1272–1278

Negrier S, Perol D, Ravaud A, Bay JO, Oudard S, Chabaud S, Fargeot P, Delva R, Deplanque G, Gravis G, Escudier B (2008) Randomized study of intravenous versus subcutaneous interleukin-2, and IFNAlphalpha in patients with good prognosis metastatic renal cancer. *Clin Cancer Res* 14(18): 5907–5912

Negrier S, Perol D, Ravaud A, Chevreau C, Bay JO, Delva R, Sevin E, Caty A, Escudier B, for the French Immunotherapy Intergroup (2007) Medroxyprogesterone, interferon alfa-2a, interleukin-2, or combination of both cytokines in patients with metastatic renal carcinoma of intermediate prognosis: results of a randomized controlled trial. *Cancer* 110(11): 2466–2477

Pyhonen S, Salminen E, Ruutu M, Lehtonen T, Nurmi M, Tammela T, Juusela H, Rintala E, Hietanen P, Kellokumpu-Lehtinen PL (1999) Prospective randomized trial of interferon alfa-2a plus vinblastine versus vinblastine alone in patients with advanced renal cell cancer. *J Clin Oncol* 17(9): 2859–2867

Rimawi M, Hilsenbeck SG (2012) Making sense of clinical trial data: is survival the only thing that matters? *J Clin Oncol* 26(12): 1922–1923

Sherill B, Amonkar M, Wu Y, Hirst C, Stein S, Walker M, Cuzick J (2008) Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. *Br J Cancer* 99(10): 1572–1578

Soulieres D (2009) Review of guidelines on the treatment of metastatic renal cell carcinoma. *Curr Oncol* 16(Suppl 1): S67–S70

Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, Barrios CH, Salzman P, Gladkov OA, Kavina A, Zarja JJ, Chen M, McCann L, Pandite L, Roy Chowdhury DF, Hawkins RE (2010a) Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol* 28(6): 1061–1068

Sternberg CN, Hawkins RE, Szczylik C, Davis ID, Wagstaff J, McCann L, Chen, M, Rubin SD (2010b) Randomized, double-blind phase III study of pazopanib in patients with advanced/metastatic renal cell carcinoma (MRCC): final overall survival (OS) results. *Annals Oncol* 21(Suppl 8): viii–viil2

Sternberg CN, Szczylik C, Lee E, Salmon PV, Mardiak J, Davis ID, Pandite L, Chen M, McCann L, Hawkins R (2009) A randomized, double-blind phase III study of pazopanib in treatment-naive and cytokine-pre-treated patients with advanced renal cell carcinoma (RCC). Presentation available at http://www.asco.org/ascov2/Meetings/Abstracts?view=abst_detail_view&confID=65&abstractID=30883 (Accessed 16 April, 2012). *J Clin Oncol* 27: (15s (suppl; abstract 5021)

Tang PA, Bentzen SM, Chen EX, Siu LL (2007) Surrogate end points for overall survival in studies of metastatic breast cancer. *J Clin Oncol* 25: 4562–4568

Tannir NM, Cohen L, Wang X, Thall P, Mathew PF, Jonasch E, Siefer-Kadite A, Pagliaro LG, Ng CS, Logothetis C (2006) Improved tolerability and quality of life with maintained efficacy using twice-daily low-dose interferon-alpha-2b: results of a randomized phase II trial of low-dose versus intermediate-dose interferon-alpha-2b in patients with metastatic renal cell carcinoma. *Cancer* 107(9): 2254–2261

Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, Stein K (2010) Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. *Health Technol Assess* 14(2): 1–184, iii–iv

Tinnerw JK, Stewart LA, Gherzi D, Burdett S, Sydes MR (2007) Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 8: 16

Warren M, Venner PM, North S, Cheng T, Venner C, Ghosh S, Venner AA, Finch D (2009) A population-based study examining the effect of tyrosine kinase inhibitors on survival in metastatic renal cell carcinoma in Alberta and the role of nephrectomy prior to treatment. *Can Urol Assoc J* 3(4): 281–289

Wiederkehr D, Howe CJ, Casciano R, Motzer R, Zheng J, Baladi FJ (2011) Overall survival among metastatic renal cell carcinoma patients corrected for crossover using inverse probability of censoring weights: analyses from the RECORD phase 3 trial. Presented at the Joint 15th Congress of the European CanCer Organisation (ECCO) and 34th Congress of the European Society for Medical Oncology (ESMO), 20–24 September, 2009, Berlin, Germany

Yang JC, Haworth L, Sherry RM, Hwu P, Schwartzentruber DJ, Topalian SL, Steinberg SM, Chen HX, Rosenberg SA (2005) A randomized trial of bevacizumab, an anti-vascular endothelial growth factor antibody, for metastatic renal cancer. *N Engl J Med* 349(5): 427–434