Patient-Reported Outcomes in a Phase III Study of Everolimus Versus Placebo in Patients with Metastatic Carcinoma of the Kidney That Has Progressed on Vascular Endothelial Growth Factor Receptor Tyrosine Kinase Inhibitor Therapy

JENNIFER L. BEAUMONT,^a ZEESHAN BUTT,^a JEANFRANCOIS BALADI,^b ROBERT J. MOTZER,^c TOMAS HAAS,^d NORBERT HOLLAENDER,^d ANDREA KAY,^b DAVID CELLA^a

^aDepartment of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; ^bNovartis Oncology, Florham Park, New Jersey, USA; ^cMemorial Sloan-Kettering Cancer Center, New York, New York, USA; ^dNovartis Oncology, Basel, Switzerland

Key Words. Everolimus • Renal cell carcinoma • Quality of life • Metastatic

Disclosures: Jennifer L. Beaumont: None; Zeeshan Butt: None; Jeanfrancois Baladi: *Employment/leadership position:* Novartis Pharmaceuticals; *Ownership interest:* Novartis; Robert J. Motzer: *Research funding/contracted research:* Novartis (money to MSKCC); Tomas Haas: *Employment/leadership position:* Novartis Pharmaceuticals; *Ownership interest:* Novartis; Norbert Hollaender: *Employment/leadership position:* Novartis Pharma AG, CH-Basel; Andrea Kay: *Employment/leadership position:* Novartis Pharmaceuticals; *Ownership interest:* Novartis; David Cella: *Consultant/advisory role:* Novartis, Bayer, Pfizer, GlaxoSmithKline; *Research funding/contracted research:* Novartis, Bayer, Pfizer, GlaxoSmithKline.

The content of this article has been reviewed by independent peer reviewers to ensure that it is balanced, objective, and free from commercial bias. No financial relationships relevant to the content of this article have been disclosed by the independent peer reviewers.

ABSTRACT

Purpose. A phase III, randomized, double-blind, placebo-controlled trial was conducted in patients with metastatic renal cell carcinoma. The focus of this paper is to evaluate the patient-reported outcomes.

Methods. Patients were randomly assigned (2:1) to receive oral everolimus 10 mg once daily or placebo. The Functional Assessment of Cancer Therapy Kidney Symptom Index—Disease-Related Symptoms (FKSI-DRS) and European Organization for the Research and Treatment of Cancer (EORTC) QLQ-C30 were administered before randomization and on day 1 of each cycle. The FKSI-DRS and the EORTC QLQ-C30 Physical Functioning and Global Quality of Life scores were the primary endpoints examined. Longitudinal models

were used to compare treatment arms. Sensitivity analyses were conducted to explore the impact of missing data assumptions.

Results. Longitudinal trends for FKSI-DRS scores did not differ by treatment arm. Taking nonignorable missing data into account, there were significant differences between treatment arms in the trend over time for physical functioning and global quality of life, with the everolimus arm exhibiting greater decreases. All three of these measures of health-related quality of life were significantly related to progression-free survival.

Conclusions. There was no evidence of a difference between everolimus and placebo in longitudinal patterns of disease-related symptoms, and little difference

Correspondence: Jennifer L. Beaumont, MS, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 710 N Lake Shore Drive, Room 713, Chicago, Illinois 60611, USA. Telephone: (312) 503-9805; Fax: 312-503-9800; e-mail: j-beaumont@northwestern.edu Received August 31, 2010; accepted for publication February 18, 2011; first published online in *The Oncologist Express* on April 1, 2011. ©AlphaMed Press 1083-7159/2011/\$30.00/0 doi: 10.1634/theoncologist.2010-0299

between the arms in physical functioning or global quality of life trends. This supports the conclusion that delay in tumor progression demonstrated by everolimus is as-

sociated with minimal impact on symptoms, physical functioning, or quality of life, as reported by patients. *The Oncologist* 2011;16:632–640

INTRODUCTION

Renal cell carcinoma (RCC), the most common form of kidney cancer, is associated with poor prognosis and reduced health-related quality of life [1, 2]. Unfortunately, 25%–30% of patients with RCC present with metastatic disease (mRCC) at the time of diagnosis [1, 3]. Furthermore, advanced RCC is one of the most drug-resistant cancers, leaving many with a symptomatic course [4]. Disease symptoms and treatment side effects can negatively impact the physical, social, and emotional well-being of patients and interfere with the ability to engage in normal activities of daily living. Indeed, the influence of symptoms associated with mRCC and treatment-related adverse events on health-related quality of life may often be underestimated [2].

For those patients with incurable mRCC, relief of disease-related symptoms is considered a chief objective. Improvement in patient symptoms, function, and overall quality of life is arguably best assessed using patient-reported outcomes (PROs). Such outcomes are associated with patient survival and tumor response [5] and have been linked to meaningful, clinical changes [6–9]. Assessment of health-related quality of life can provide substantial information regarding a treatment's net effect on patients' disease burden experience. As such, it is important to measure PROs in the evaluation of treatments for mRCC [10].

Everolimus (RAD001; Novartis Oncology, Florham Park, NJ) is an orally administered inhibitor of the mammalian target of rapamycin, a therapeutic target for mRCC. A phase III, randomized, double-blind, placebo-controlled trial of everolimus was conducted in patients with mRCC whose disease had progressed on vascular endothelial growth factor-targeted therapy. Primary results of this trial, with progression-free survival (PFS) as the primary endpoint, were published by Motzer et al. [11] and have been recently updated [12]. Everolimus was associated with a 67% risk reduction for PFS, based on central radiology review, compared with placebo.

The present paper focuses on longitudinal trends in disease-related symptoms, physical functioning, and global quality of life in this phase III, randomized, double-blind, placebo-controlled trial of everolimus in patients with mRCC. Exploratory analyses examine the association between the PROs and PFS. Time to definitive deterioration was analyzed and reported previously [11, 12].

METHODS

Eligible participants were adults (aged 18 years and older) with mRCC that showed a clear-cell component and had progressed on or were within 6 months of stopping treatment with sunitinib and/or sorafenib. Other eligibility criteria have been reported elsewhere [11]. The protocol was approved by the institutional review boards of the participating institutions, and the study was done in accordance with international standards of good clinical practice. All patients provided written informed consent.

Patients were stratified according to Memorial Sloan-Kettering Cancer Center (MSKCC) prognostic score and previous anticancer therapy. They received either continuous treatment with oral everolimus (10 mg) once daily or placebo, based on random assignment in a 2:1 ratio, both in conjunction with the best supportive care. Each cycle was considered as 28 days of treatment. Treatment in both groups was continued until disease progression, unacceptable toxicity, death, or discontinuation for any other reason. After documented progression on the basis of investigator assessment, patients who were initially randomized to placebo were allowed to crossover to receive open-label everolimus. This element of the study design was incorporated to address both ethical and recruitment considerations. The trial was stopped at the second interim analysis when efficacy of everolimus was demonstrated by pre-specified stopping rules.

Health-related quality of life was assessed with the Functional Assessment of Cancer Therapy Kidney Symptom Index—Disease-Related Symptoms (FKSI-DRS) [13] and European Organization for the Research and Treatment of Cancer (EORTC) QLQ-C30 [14] questionnaires. All questionnaires were administered before randomization, on day 1 of each cycle, and on discontinuation from the study. Only data collected during the double-blind part of the study were used.

The FKSI-DRS consists of nine items that experts and patients have indicated are important targets for the treatment of advanced kidney cancer and that clinical experts have indicated are primarily disease-related, as opposed to treatment-related. Symptoms assessed on the FKSI-DRS include pain, fatigue, shortness of breath, fevers, weight loss, coughing, and blood in urine. The total score can range from 0 (worst) to 36 (best). Although Motzer et al [11] considered a decrease by at least 2 score units (with no later

increase above this threshold) as deterioration, subsequent research has estimated the minimally important difference (MID) for individual change on the FKSI-DRS at 3 points [13]. The EORTC QLQ-C30 is a questionnaire that was developed to assess cancer patients' physical, emotional, cognitive, social, and role function, global quality of life, and several specific symptoms. The literature is inconsistent on the appropriate MID for the EORTC QLQ-C30 scales; [15, 16] 10% change from baseline was used for this study. The primary health-related quality of life endpoints selected from these questionnaires were the FKSI-DRS score and the Physical Functioning and Global Quality of Life subscales of the EORTC QLQ-C30. Other subscales of the EORTC QLQ-C30 were analyzed as secondary endpoints, and data are not presented here. Higher scores on all scales presented indicate better health-related quality of life.

Statistical Analyses

Missing assessments present a challenge in analyzing and interpreting PRO data, especially when the data are missing nonrandomly. For example, patients who experience increased toxicity, disease progression, or death are more likely to miss one or more assessments. Similarly, in a longitudinal study data may be missing at later assessments because of death or withdrawal due to disease progression. This leaves the purportedly healthier subjects, with better PRO scores, on study longer and displaying a misleading trajectory of higher scores over time. If the reasons for missing data are ignored and the analyses are based only on the observed data, the estimates may be biased in that they do not accurately reflect the entire population of study patients [17, 18].

Missing data can be classified into three main types. Data are considered missing completely at random (MCAR) if the missingness is completely independent of both the observed and unobserved (i.e., missing) data. This is a very strong assumption and is rarely met in practice. Examples of scenarios resulting in MCAR data include a site inadvertently forgetting to administer a PRO questionnaire or the patient missing their clinic visit for something non-health-related. It is important to note that many classic statistical procedures, such as multivariate analysis of variance, rely on the unrealistic MCAR assumption. A slightly more lenient missing data assumption is the missing at random (MAR) classification. Data are MAR if they depend on observed data only. Incorporating the predictors of missing data, for example, stage at diagnosis or baseline PRO scores, allows for an unbiased analysis using all available data. Mixed effects models for repeated measures data based on maximum likelihood theory provide valid results when data are MAR. In the present example, mixed effects

models describe the rate of change in PRO scores over time for each treatment arm (fixed effect), taking into account the between-patient variability by incorporating each patient's individual starting point and individual rate of change (random effect).

If the probability of missingness depends on the unobserved PRO scores, then missing data are considered non-ignorable, or missing not at random (MNAR) [18, 19]. This type of missing data can occur in association with death, disease progression, or toxicity. Formally distinguishing between MAR and MNAR is not trivial and relies on assumptions that are themselves untestable because the required data are missing [20]. Applications of mixed models have been developed that can assist in the analysis of non-ignorable missing data. One such application is the pattern-mixture model [21]. In this method, patients are stratified by the missing data pattern, and mixed models are created within each stratum. Parameter estimates for each stratum are then combined into a weighted average for the entire study population. This strategy is advantageous because it provides descriptive information about treatment effects within different groups of patients, it is not necessary to define a model for the missing data mechanism, and the models can be fit with standard software. It can be difficult, however, to determine how to create the strata; possibilities include stratification based on the duration of follow-up, such as <3 months, 3–6 months, >6 months, or those with disease progression versus those without. The amount of, and reasons for, missing data need to be evaluated to determine the best stratification strategies.

We examined the patterns of missing PRO data in the present trial and classified patients into groups according to timing of study dropout. We then used χ^2 tests and two-sample *t*-tests, as appropriate, to compare baseline characteristics between dropout groups. Because it is not possible to definitively determine whether missing data are ignorable, we conducted sensitivity analyses (using a mixed effects model and pattern-mixture model, described below) to ensure that treatment group effects were consistent across different analytical methods with different missing data assumptions and to evaluate the range of possible treatment group differences.

We first implemented a mixed effects model for repeated measures to evaluate the FKSI-DRS and EORTC QLQ-C30 scores longitudinally. This model describes the rate of change in PRO scores over time for each treatment arm, taking into account the between-patient variability by incorporating each patient's individual starting point and individual rate of change. Because missing data were suspected to be related to the outcomes of interest, a pattern-mixture model, to assist in the analysis of the potentially

nonignorable missing data, was also implemented as a sensitivity analysis [21, 22]. For this method, patients were stratified into groups based on timing of study dropout (generally due to disease progression in this study), and mixed effects models were created within each stratum. Parameter estimates for each stratum were then combined into a weighted average for the entire treatment arm. All of the longitudinal models were performed using only the first 8 months of follow-up to ensure adequate precision of model estimates. Two-sided tests with $\alpha = 0.05$ were used for longitudinal models.

In an exploratory analysis, Cox proportional hazard models were used to evaluate the association between PFS and the longitudinally assessed health-related quality of life scores. The health-related quality of life scores were included in PFS models as time-dependent repeated measure variables taking values according to a step function that retains the value of the health-related quality of life score until the time that the next observed score is reached [23]. The model was stratified by baseline risk strata and treatment arm.

RESULTS

The numbers of patients in each group with PRO data available for analysis are shown in Figure 1. Of the 277 patients who were randomized to everolimus, 242 (87%) had a PRO assessment at baseline whereas 128 of 139 patients (92%) randomized to placebo provided baseline PRO data. Baseline demographic and disease characteristics were much the same in the two groups (Table 1). Most patients were functioning well at baseline with Karnofsky performance status of 90 or 100 in 65% of patients. The MSKCC risk stratification was intermediate or poor in 71%. The numbers of patients on study and providing PRO data steadily declined over time. Approximately 60% of the original patients were still on study at the 3-month assessment, with 32% at the 6-month assessment. Disease progression was the primary reason for dropout. Patients were classified as “dropouts” if their last PRO assessment was prior to 3 months and as “completers” if their last assessment was 3 months or later. Dropout groups are compared in Table 2. Dropout groups differed by treatment assignment with 56% of placebo patients off study prior to month 3 and only 31% of everolimus patients off study at that point ($p < .001$). Higher rates of dropout were also associated with worse baseline risk strata ($p = .021$), younger age ($p = .010$), and worse baseline EORTC QLQ-C30 Physical Functioning score ($p = .004$). Disease progression was the primary reason for dropout. The probability of a missing assessment was strongly dependent on the score at the previous time point. Specifically, a score decrease of ~1 standard deviation (5 points

Figure 1. CONSORT diagram for patient-reported outcome (PRO) portion of study, based on February 28, 2008, data cut-off. Abbreviations: EORTC, European Organization for the Research and Treatment of Cancer; FKSI-DRS, Functional Assessment of Cancer Therapy Kidney Symptom Index—Disease-Related Symptoms; PF, progression free; QoL, quality of life.

Table 1. Demographic and disease characteristics of patients with baseline patient-reported outcomes assessment

	Everolimus (N = 242)	Placebo (N = 128)
Age, years	61 (27–85)	60 (33–79)
Sex		
Male	189 (78%)	95 (74%)
Female	53 (22%)	33 (26%)
Baseline Karnofsky performance status		
100	56 (23%)	34 (27%)
90	94 (39%)	51 (40%)
80	63 (26%)	29 (23%)
70	28 (12%)	14 (11%)
Missing	1 (<1%)	0
MSKCC risk strata		
Favorable	73 (30%)	34 (27%)
Intermediate	132 (55%)	75 (59%)
Poor	37 (15%)	19 (15%)
Past VEGF-R inhibitor		
One	187 (77%)	97 (76%)
Two	55 (23%)	31 (24%)

Numbers in table are median (range) or N (%). Abbreviations: VEGF-R, vascular endothelial growth factor receptor.

for FKSI-DRS, 20 points for EORTC QLQ-C30 scores) was associated with a 32–53% increase in the odds of missing the subsequent assessment. These analyses indicate that the missing data due to dropout were not completely random and analyses that do not correctly account for this may be biased.

Table 2. Comparison of dropout groups			
	“Completers” (N = 224)	“Dropouts” (N = 146)	p-value
Treatment group			<.001
Everolimus	168 (69%)	74 (31%)	
Placebo	56 (44%)	72 (56%)	
Sex			.603
Male	174 (61%)	110 (39%)	
Female	50 (58%)	36 (42%)	
Baseline Karnofsky performance status (n = 1 missing)			.417
100	54 (60%)	36 (40%)	
90	92 (63%)	53 (37%)	
80	57 (62%)	35 (38%)	
70	21 (50%)	21 (50%)	
MSKCC risk strata			.021
Favorable	72 (67%)	35 (33%)	
Intermediate	125 (60%)	82 (40%)	
Poor	27 (48%)	29 (52%)	
Age, years	61.4 (9.9)	58.7 (10.1)	.010
Baseline FKSI-DRS	28.9 (5.4)	28.0 (5.2)	.128
Baseline EORTC QLQ-C30 Physical Functioning	79.9 (19.7)	73.5 (23.0)	.004
Baseline EORTC QLQ-C30 Global Quality of Life	61.0 (23.1)	57.9 (21.2)	.200

Numbers in table are N (row %) or mean (SD). “Completers” indicates last PRO assessment at cycle (month) 3 or later. “Dropouts” indicates last PRO assessment prior to cycle (month) 3.

Longitudinal Models

Figure 2 illustrates the longitudinal trends of completers and dropouts. First examining the completers on the FKSI-DRS, the placebo arm started out with slightly higher scores at baseline and both arms remained relatively stable over time. In the dropout groups, the placebo arm similarly began with slightly higher scores at baseline but scores tended to worsen over time, with FKSI-DRS scores of the everolimus dropout group decreasing more than those of the placebo dropout group. The pattern-mixture model results in Table 3 summarize these findings, averaged across the dropout/completer strata. According to these results, the everolimus arm began 1.4 points (standard error [SE] = 0.62) lower than the placebo arm at baseline ($p = .026$), scores decreased at a rate of 0.7 points per month (SE = 0.15) in the placebo arm ($p < .001$), and the rate of decrease in the everolimus arm did not differ from this ($p = .221$). To provide context for these estimated effects, FKSI-DRS scores are on a scale of 0 (worst) to 36 (best) and the MID is 3 points [13]. The results for the mixed effects model,

which assumes the missing data are MAR, are also presented in Table 3 for comparison. Compared to the pattern mixture model, all estimated effects (baseline difference, change over time, difference between arms in change over time) are closer to the null value of zero and p -values are larger, with only the average change over time achieving $p < .05$.

The EORTC QLQ-C30 Physical Functioning and Global Quality of Life scores displayed similar trends (Figure 2). Specifically, among completers, the placebo arm started out with slightly higher scores at baseline and both arms remained relatively stable over time. In the dropout groups, however, both arms started out at the same level at baseline, and scores tended to worsen over time in the everolimus dropout group but remained stable or even increased slightly in the placebo dropout group. According to the pattern-mixture model results in Table 3 for Physical Functioning, the everolimus arm began a nonsignificant 3.0 points (SE = 2.48) lower than the placebo arm at baseline ($p = .229$), scores decreased at a rate of 1.9 points per month (SE = 0.54) in the placebo arm ($p < .001$), and the rate of decrease in the everolimus arm was even greater ($-1.9 + -2.1 = -4.0$ points per month; $p < .001$). EORTC QLQ-C30 scores are on a scale of 0 (worst) to 100 (best), and the MID is a 10% change from baseline, which is equivalent to ~ 8 points in this study. Compared to the pattern-mixture model, in the mixed effects model assuming MAR, all estimated effects are closer to the null value of zero and p -values are larger, with only the change over time achieving $p < .05$.

According to the pattern-mixture model results in Table 3 for Global Quality of Life, the everolimus arm began a nonsignificant 3.1 points (SE = 2.48) lower than the placebo arm at baseline ($p = .210$), scores decreased at a nonsignificant rate of 0.5 points per month (SE = 0.66) in the placebo arm ($p = .412$), and the rate of decrease in the everolimus arm was significantly greater ($-0.5 + -1.6 = -2.1$ points per month; $p = .030$). The MID is a 10% change from baseline, which is equivalent to ~ 6 points in this study for the Global Quality of Life scale. Compared to the pattern-mixture model, in the mixed effects model assuming MAR, all estimated effects are closer to the null value of zero and p -values are larger, with none of them achieving statistical significance.

Predictors of PFS

When included as time-dependent variables in exploratory analysis, all three measures of health-related quality of life were significant predictors of PFS in models stratified by baseline risk strata and treatment arm. The hazard ratio for the time-dependent FKSI-DRS scores was 0.82 [95% con-

Figure 2. (A): Functional Assessment of Cancer Therapy Kidney Symptom Index—Disease-Related Symptoms (FKSI-DRS) scores stratified by treatment arm and dropout group. (B): European Organization for the Research and Treatment of Cancer (EORTC) QLQ-C30 Physical Functioning scores stratified by treatment arm and dropout group. (C): EORTC QLQ-C30 Global Quality of Life scores stratified by treatment arm and dropout group.

Table 3. Mixed effect and pattern-mixture model results

Effect	Mixed effects model (N = 370)		Pattern-mixture model (N = 370)	
	Estimate (SE)	p-value	Estimate (SE)	p-value
FKSI-DRS				
Time (months)	−0.3 (0.11)	.009	−0.7 (0.15)	<.001
Arm: everolimus versus placebo	−0.8 (0.57)	.142	−1.4 (0.62)	.026
Time*arm interaction	−0.0 (0.13)	.998	−0.2 (0.16)	.221
EORTC QLQ-C30 Physical Functioning				
Time (months)	−1.0 (0.44)	.024	−1.9 (0.54)	<.001
Arm: everolimus versus placebo	−0.7 (2.31)	.764	−3.0 (2.48)	.229
Time*arm interaction	−0.8 (0.50)	.121	−2.1 (0.59)	<.001
EORTC QLQ-C30 Global Quality of Life				
Time (months)	−0.1 (0.49)	.906	−0.5 (0.66)	.412
Arm: everolimus versus placebo	−1.5 (2.27)	.512	−3.1 (2.48)	.210
Time*arm interaction	−0.7 (0.55)	.228	−1.6 (0.74)	.030

Abbreviations: SE, standard error.

fidence interval (CI), 0.74–0.92; $p = .001$], indicating that, for every standard deviation increase on the FKSI-DRS, the instantaneous hazard of death or progression was reduced by 18%. Results were very similar for the EORTC QLQ-C30 Physical Functioning (hazard ratio = 0.84; 95% CI, 0.75–0.94; $p = .001$) and Global Quality of Life scores (hazard ratio = 0.85; 95% CI, 0.76–0.96; $p = .006$).

DISCUSSION

We analyzed patient-reported outcome data from a phase III, randomized, double-blind, placebo-controlled trial of everolimus in patients with metastatic renal cell carcinoma to determine longitudinal trends in disease-related symptoms, physical functioning, and global quality of life. We also examined the association of PFS with these outcomes.

Because the missing data in this study were quite possibly related to patients' health state and therefore nonignorable or MNAR, pattern-mixture models were fit to the PRO endpoints. Pattern-mixture models attempt to adjust for the potential impact of dropout by first stratifying the sample according to dropout pattern, then fitting separate mixed effects models within each stratum. This provides separate estimates of treatment effect for "dropouts" and "completers." In the final step, a weighted average of these separate treatment effects is calculated, resulting in an overall estimate of the effect of treatment, averaged across dropout strata. In these models, scores for overall quality of life and physical functioning deteriorated more quickly in patients assigned to everolimus than in patients assigned to placebo. However, there was no difference between everolimus and

placebo arms in the rate of change in disease-related symptoms.

We also fit mixed effects models, which assume a stricter missing data assumption (MAR) and found that all effects were attenuated. All estimates of baseline differences, rates of change, and differences between arms in rates of change, while larger in magnitude for the pattern-mixture models, were in the negative direction for all models. Even with differing missing data assumptions, this indicates some consistency of results between the models. Because the primary reason for dropout was disease progression, it is reasonable to suppose that the MNAR assumption is more appropriate than the MAR assumption for this study, although it is not possible to test this directly. On the basis of this, the pattern-mixture models are likely *less* biased than the ordinary mixed effects models. These pattern-mixture model results may still be biased, however, because the required assumption of ignorable dropout *within* strata may still be violated. Models utilizing three dropout strata were also fit, with similar findings. Additional strata would not be feasible due to the rapidly diminishing sample size in the placebo arm. In all models, the baseline differences, rates of change, and differences between arms in rates of change, even though statistically significant in some instances, were small relative to the minimally important differences for these scales and, therefore, of little practical significance.

The exploratory analyses were undertaken to explore the association between PROs and disease progression. When included as time-dependent covariates, all three mea-

sures of health-related quality of life were significantly related to PFS. These results demonstrate that, even after controlling for baseline risk strata, PRO scores are associated with PFS, are meaningful secondary endpoints that could be expected to demonstrate a similar treatment effect, and may be useful predictors of PFS.

CONCLUSION

In summary, there was no evidence of a difference between everolimus and placebo in longitudinal patterns of disease-related symptoms, and only small, although statistically significant, differences in longitudinal patterns of physical functioning and global quality of life. The pattern-mixture models demonstrated a greater decline in physical functioning and global quality of life scores in the everolimus group relative to placebo. This difference was primarily driven by differences between arms in the early dropout strata, that is, the patients whose disease was progressing more rapidly. As Figure 2 demonstrates, there was very little change in average PRO scores over the course of the trial for those who remained on study for at least 3 months. Exploratory analyses indicated that there was an association between PRO scores and PFS. Many patients were asymptomatic or experiencing very low levels of symptoms at study entry. Karnofsky performance status was 90 or 100 in 65% at baseline; FKSI-DRS scores were only 6–8 points below the best possible score for that scale; EORTC QLQ-C30 scores were around 60–80 points on a scale of 0–100. It is likely that, due to frequent tumor assessments for the primary endpoint, patients were taken off study before disease-related symptoms manifested or quality of life worsened. Unfortunately, PRO data were not collected after progression. In future trials, the continued collection of PRO data after disease progression could aid in the meaningful interpretation of PRO results in this stage of the disease continuum.

As is true in several other advanced solid tumor trials, the evidence in the present trial supports the conclusion that delay in tumor progression is associated with at most min-

imal impact on symptoms, functioning, and quality of life reported by patients with mRCC. Patients' preference for available therapies can vary based on a number of considerations, including but not limited to expected impact on health-related quality of life. For patients receiving first-line treatment with a targeted agent, health-related quality of life may be less important than the prospect for clinical benefit. In contrast, the second- and third-line settings bring health-related quality of life considerations forward more prominently for some patients, their families, and their providers. Continued research to confirm and expand upon these observations will further our understanding of the positive and negative outcomes associated with new treatments for mRCC.

ACKNOWLEDGMENTS

Research support was provided by Novartis Pharmaceuticals and in part by Grant KL2-RR-0254740 from the National Center for Research Resources, National Institutes of Health.

These findings have been previously presented in the Proceedings of the 2009 American Society for Clinical Oncology annual meeting and as a poster presentation at the Joint 15th Congress of the European Cancer Organisation (ECCO) and 34th Congress of the European Society for Medical Oncology (ESMO), September 2009, Berlin, Germany. The primary clinical outcomes of this clinical trial have been reported by Motzer et al. in *Lancet* 2008; 372(9637):449–456. Updated results are also reported by Motzer et al. in *Cancer* 2010;116:4256–4265.

AUTHOR CONTRIBUTIONS

Conception/Design: Robert J. Motzer

Data analysis and interpretation: Jennifer L. Beaumont, Zeeshan Butt, Jeanfrancois Baladi, Tomas Haas, Norbert Hollaender, Andrea Kay, David Cella

Manuscript writing: Jennifer L. Beaumont, Zeeshan Butt, Jeanfrancois Baladi, Tomas Haas, Norbert Hollaender, Andrea Kay, David Cella

Final approval of manuscript: Jennifer L. Beaumont, Zeeshan Butt, Jeanfrancois Baladi, Robert J. Motzer, Tomas Haas, Norbert Hollaender, Andrea Kay, David Cella

REFERENCES

- Gupta K, Miller JD, Li JZ et al. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer Treat Rev* 2008;34:193–205.
- Cella D. Quality of life in patients with metastatic renal cell carcinoma: the importance of patient-reported outcomes. *Cancer Treat Rev* 2009;35:733–737.
- Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. *N Engl J Med* 1996;335:865–875.
- Coppin C, Le L, Porzsolt F et al. Targeted therapy for advanced renal cell carcinoma. *Cochrane Database Syst Rev* 2008;(2):CD006017.
- Victorson D, Soni M, Cella D. Metaanalysis of the correlation between radiographic tumor response and patient-reported outcomes. *Cancer* 2006; 106:494–504.
- Cella D, Nichol MB, Eton D et al. Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy–Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. *Value Health* 2009;12:124–129.
- Eton DT, Fairclough DL, Cella D et al. Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. *J Clin Oncol* 2003;21:1536–1543.
- Steel JL, Eton DT, Cella D et al. Clinically meaningful changes in health-

- related quality of life in patients diagnosed with hepatobiliary carcinoma. *Ann Oncol* 2006;17:304–312.
- 9 Yost KJ, Cella D, Chawla A et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. *J Clin Epidemiol* 2005;58:1241–1251.
 - 10 Wagner LI, Wenzel L, Shaw E et al. Patient-reported outcomes in phase II cancer clinical trials: lessons learned and future directions. *J Clin Oncol* 2007;25:5058–5062.
 - 11 Motzer RJ, Escudier B, Oudard S et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008;372(9637):449–456.
 - 12 Motzer RJ, Escudier B, Oudard S et al. Phase 3 trial of everolimus for metastatic renal cell carcinoma: final results and analysis of prognostic factors. *Cancer* 2010;116:4256–4265.
 - 13 Cella D, Yount S, Brucker PS et al. Development and validation of a scale to measure disease-related symptoms of kidney cancer. *Value Health* 2007;10:285–293.
 - 14 Aaronson NK, Ahmedzai S, Bergman B et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–376.
 - 15 Osoba D, Rodrigues G, Myles J et al. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998;16:139–144.
 - 16 Dubois D, Dhawan R, van de Velde H et al. Descriptive and prognostic value of patient-reported outcomes: the bortezomib experience in relapsed and refractory multiple myeloma. *J Clin Oncol* 2006;24:976–982.
 - 17 Little RJA, Rubin DB. *Statistical Analysis with Missing Data*, 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2002.
 - 18 Troxel AB, Fairclough DL, Curran D et al. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med* 1998;17:653–666.
 - 19 Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, 1987.
 - 20 Curran D, Bacchi M, Schmitz SF et al. Identifying the types of missingness in quality of life data from clinical trials. *Stat Med* 1998;17:739–756.
 - 21 Little RJA. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *J Am Stat Assoc* 1995;90:1112–1121.
 - 22 Brown H, Prescott R. *Applied Mixed Models in Medicine*. West Sussex: John Wiley & Sons Ltd, 1999.
 - 23 Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* 1999;20:145–157.

Manfred Hauben, MD, MPH

Sulfur Mountain, British Columbia