Meta-analysis for the Association between Overall Survival and Progression-Free Survival in Gastrointestinal Stromal Tumor

Ipek Özer-Stillman1, Lauren Strand1, Jane Chang2, Ateesha F. Mohamed2, and Katherine E. Tranbarger-Freier1

Abstract

Purpose: Gastrointestinal stromal tumor (GIST) is a relatively rare tumor that is treated with targeted therapies in advanced stages. Randomized clinical trials (RCT) often require long follow-up and large sample sizes to evaluate overall survival (OS), the gold-standard measure of treatment efficacy. However, changes in therapy following disease progression may complicate survival assessments. Establishing surrogate endpoints may facilitate the drug approval and availability of new efficacious treatments; however, no published studies have investigated this topic in unresectable and/or metastatic GIST.

Experimental Design: A systematic literature review identified 14 RCTs and five observational studies of sufficient methodologic quality published between January 1995 and December 2013 (29 treatment arms; 2,189 patients). Weighted linear regression was used to evaluate the relation between median OS and median progression-free survival (PFS) for all arms combined and stratified by treatment line, treatment type, and quality score.

Results: Median OS and PFS were positively related with a correlation of 0.91. The association was still moderate (correlation 0.72) after eliminating four influential data points. In stratified analyses, correlation of OS and PFS was greater in later lines of therapy (first line = 0.52; second line = 0.89; third- and later-line = 0.70) and imatinib showed a stronger association (0.91) than other evaluated treatments (−0.26 to 0.69).

Conclusion: This analysis identified a strong relationship between median OS and PFS, especially in later lines of therapy. Findings suggest that PFS could serve as a surrogate marker for OS; however, analyses of patient-level data are needed to establish its validity in GIST. Clin Cancer Res; 21(2); 1–8. ©2014 AACR.
Translational Relevance

This arm-level meta-analysis pertains directly to the evaluation of survival endpoints in oncology clinical trials. We evaluated the relationship between median overall survival and median progression-free survival in gastrointestinal stromal tumor (GIST) trials. Establishing the validity of PFS as a surrogate marker for OS in advanced GIST may allow for shorter follow-up periods in trials, and help to reduce the costs of clinical trials and facilitate the drug approval and availability of new and efficacious targeted therapies. No published studies have investigated this topic in unresectable and/or metastatic GIST. This analysis identified a strong relationship between median OS and PFS, especially in later lines of therapy. Findings suggest that PFS could serve as a surrogate marker for OS; however, analyses of patient-level data are needed to establish its validity in GIST.

Materials and Methods

Literature review

Two systematic literature searches were conducted in PubMed and Embase according to Cochrane guidelines (20); an initial search from January 1, 1995 through July 1, 2010 was conducted in PubMed, and later an updated search through October 28, 2013 was conducted in both databases. Search terms included “GIST,” “advanced or metastatic,” and “unresectable or stage IV.” The searches were limited to clinical trials, prospective observational studies, and retrospective observational studies of targeted therapies for GIST. Additional analyses explore the impact of treatment line, treatment type, and study quality on the strength of the relationship.

Quality assessment

Two researchers evaluated all included studies using the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach, as outlined and adopted by the Cochrane Collaboration (20. Chapter 12: Interpreting results and drawing conclusions). This approach defines the quality of a body of evidence on the basis of study design and the directness, precision, and consistency of the results. Each study was assigned a score of low (1 to <2), moderate (2 to <3), and high (3 to ≤4) for the randomized controlled trials of high-statistical power and consistent reporting. Studies with an average score of at least 1.75 were considered to be of sufficient quality for inclusion in the analysis; others were excluded to limit the impact of studies with critical problems and unsystematic clinical observations on the study results. Sensitivity analysis was conducted using different groups of quality scores.

Data extraction

For each included study of sufficient quality, data were extracted for study design, year of publication, sample size per treatment arm, treatment, treatment line, and Eastern Cooperative Oncology Group (ECOG; ref. 21) performance status. Data on the median OS and median PFS and outcome definitions were collected; reported point values took precedence over data from Kaplan–Meier curves. Confidence intervals were also collected to elucidate the variation in median outcome. In the event that a study was published in multiple articles or abstracts, the most recent data were used. A second researcher fully validated the extracted data.

Data analysis

Analyses were defined prospectively in a statistical analysis plan. To evaluate the relationship between median OS and median PFS, linear regression analyses were performed. Each treatment arm was weighted in the least-squares regression models by its respective sample size. In addition, the corresponding coefficient of determination ($R^2$) and Pearson correlation coefficient were recorded for each model fit. Difference in Beta Scaled (DFBETAS; ref. 22) was used to identify influential data points within the key regression analyses and additional models were then run to show the effect of excluding all highly influential points. Furthermore, to explore the impact of other study variables, additional analyses were performed stratifying treatment arms by (i) treatment line (first-line, second- and line 2.5, third- and later-line); (ii) treatment type (imatinib, sunitinib, sorafenib, combination, and other TKIs); and (iii) assigned study quality score (1.75, 2–2.75, 3–3.75). All reported $P$ values reflect two-sided tests, with a value of $P$ < 0.05 indicative of statistical significance. Analyses were performed using R version 3.0.2. Studies that were designated as line 2.5 (2.5-) were those that allowed patients with second-line and later treatments but were largely considered to be second-line studies.

Results

Following the systematic literature review (Fig. 1), a total of 51 studies were evaluated for quality. Nineteen studies were considered to be of sufficient quality for inclusion in analyses (5–10, 23–35). Most studies were phase II or III clinical trials, though three retrospective and one prospective observational study were included (25, 29, 32, 36). These studies included 29 treatment arms and 2,189 patients (Table 1). Across study treatment arms, there were between 15 and 349 patients with unresectable and/or metastatic GIST; the mean and the median for reported median OS were 18.7 and 11.8 months, respectively. The mean of the reported median PFS was 7.9 (median median PFS was 4.1) for the 29 treatment arms. Evaluated targeted therapies included imatinib, sunitinib, regorafenib, and nilotinib. Treatment line and ECOG status of evaluated patients varied across the studies.
Overall analysis
Median OS was strongly associated with median PFS ($R^2 = 0.84, r = 0.91$; Fig. 2A and Table 2). When analyzing influential data points, any DFBETA larger than $2/\sqrt{n}$ or smaller than $-2/\sqrt{n}$ was flagged as influential (where $n$ is the number of points in the model; ref. 22). Ryu and colleagues (35), McAuliffe and colleagues (31), and both low- and high-dose treatment arms from Blanke and colleagues (5) stood out as having a high degree of influence on the fitted model. Figure 2B shows the modest change in slope of the regression with removal of the influential data points. Although the strength of the relationship was also reduced in this sensitivity analysis, the result is still reflective of a strong linear association between median OS and median PFS ($R^2 = 0.52, r = 0.72$; Table 2).

Analysis by treatment line
Four treatment arms evaluated a first-line population (5, 31, 35), whereas eight arms evaluated a second- or 2.5-line [treatment arms designated as line 2.5 (2½) were those that allowed patients receiving second-line and later treatments; refs. 6, 9, 23, 24, 28–30, 33], and 17 arms evaluated a third- and later-line population (7–10, 25–27, 32, 34). In a stratified analysis, the linear association was strongest for second-line therapy ($R^2 = 0.66, r = 0.80$) followed by third- and later-line therapy ($R^2 = 0.39, r = 0.70$; Fig. 2C and Table 2). The association was less strong in first-line therapy ($R^2 = 0.08, r = 0.52$), which may be related to the variable treatment options available to patients with GIST following progression on first-line treatment.

Ryu and colleagues (35) was identified as an influential point in the first-line therapy model and Raut and colleagues (33) was identified for second line. In the third-line regression model, George and colleagues (7), Trent and colleagues (10), and the sorafenib fourth-line arm of Montemurro and colleagues (32) each contributed highly influential data points. The sensitivity analysis with removal of these data points showed a weaker relationship in second line but still reflected moderate correlation ($R^2 = 0.51, r = 0.66$; Fig. 2D and Table 2).

This was also true of the third- and later-line analysis ($R^2 = 0.44, r = 0.63$). For first-line therapy, removal of the influential data points by Ryu and colleagues (35) changed the slope from positive to negative ($R^2 = 0.98, r = -1$), suggesting that no significant association between median OS and median PFS is present in first-line therapy.

Analysis by treatment type
Eight treatment arms evaluated single-agent imatinib (5, 25, 27, 29, 31, 35), three evaluated sunitinib (6, 24, 33), five evaluated sorafenib (8, 25, 28, 32), five evaluated imatinib combinations (9, 25, 30), and seven evaluated other TKIs (7, 10, 23, 25, 26, 34), including regorafenib and nilotinib. (These analyses did not include arm from Reichardt and colleagues, as it was not able to be categorized as a single therapy. Patients received BSC, BSC + imatinib, or BSC + sunitinib.) In a stratified analysis, imatinib was associated with the strongest linear association between median OS and median PFS ($R^2 = 0.72, r = 0.91$), followed by other TKIs ($R^2 = 0.39, r = 0.69$) and sorafenib ($R^2 = 0.44, r = 0.66$; Fig. 3A and Table 2). Sorafenib ($R^2 = 0.03, r = 0.29$) and imatinib combination therapy ($R^2 = 0.16, r = -0.26$) demonstrated little or no association.

Analysis by quality score
The impact of quality grading on relation between median OS and median PFS was evaluated by analyzing data in three groups: studies with score 3 to 3.75, score 2 to 2.75, and score 1.75. Three treatment arms were graded with a score of 1.75 (6, 33, 35). The majority of the treatment arms (arms n = 21) received scores of 2 to 2.75 (7–10, 23–26, 28–32). Five treatment arms were given scores of 3 to 3.75 (5, 27, 34).

In a stratified analysis, the higher-quality studies demonstrated the strongest linear relationship between median OS and median PFS ($R^2 = 0.96, r = 0.99$), though the analyses of moderate-quality ($R^2 = 0.73, r = 0.85$) and lower-quality ($R^2 = 0.96, r = 0.98$) studies also showed strong association (Fig. 3B and Table 2).
Discussion

Targeted therapies have markedly improved survival in patients with unresectable and metastatic GIST. Although clinical trials in GIST have been characterized by long periods of follow-up for survival endpoints, the pivotal S0033 trial of first-line imatinib required a median follow-up of 4.5 years (5). Receipt of additional therapies after treatment progression further complicates the evaluation of OS in clinical trials. PFS offers an advantage over OS because it requires patients to be followed only until their disease progresses and therefore, measures only the effect of the study drug and is not diluted by subsequent treatments patients receive, as OS may be (17).

This analysis demonstrates potential for PFS as a valid surrogate for OS based on a strong relationship between median OS and median PFS in 29 study arms of targeted agents used to treat advanced or metastatic GIST.

The strong correlation between median OS and median PFS persisted even with the removal of four highly influential treatment arms (5, 31, 35). Two of these influential arms were from the S0033 trial, which evaluated the largest number of patients with advanced or metastatic GIST. The other two removed had relatively high estimates of median OS and PFS compared with others; each had a median PFS longer than 2 years. Moreover, all four of these arms represent sequential treatments patients receive, as OS may be (17).
the only first-line GIST populations evaluated in this analysis and therefore their corresponding estimates of OS are most impacted by the use of any later therapies applied after disease progression.

In a stratified analysis, the associations between median OS and median PFS differed by increasing line of therapy. Second-line (eight arms) and third- and later-line (17 arms) therapies all demonstrated a moderate to strong relationship, whereas first-line therapy (four arms) showed only moderate correlation. The first-line sensitivity analysis with removal of an influential data point reflected the absence of a relationship (35). This finding and the nonsignificant slope indicate that there is insufficient evidence for the association between median OS and PFS within first-line therapy alone. The moderate correlation in the main analysis may be related to the variability in treatment options following progression. The power to detect a survival benefit using derivations from PFS declines substantially with increased survival postprogression (37). Thus, since survival after GIST progression in the first-line setting may be long, reduced correlation would be expected in this model.

Analyses stratified by study quality score reinforced the result from the analysis of all treatment arms, suggesting that the relationship between median OS and median PFS is not impacted by differences in study quality as assessed by researchers. In the analysis stratified by therapy, imatinib was the most-represented treatment (eight arms: 4 first-line, 1 second-line, 1 third-line, 2

**Figure 2.**
Weighted linear regression analyses for the relationship between median OS and median PFS in 29 treatment arms evaluating patients \( (n = 2,189) \) with unresectable and/or metastatic GIST in observational studies \( (n = 4) \) and clinical trials \( (n = 15) \) of targeted therapies. Influential data points were identified using DFBETAS. A, regression for all arms. B, regression for all arms with four influential points removed. C, regression stratified by treatment line. D, regression stratified by treatment line with five influential points removed.
and expected to show the strongest association between median OS and median PFS. This assumption was confirmed, and the model in “other TKIs” (seven arms) demonstrated moderate correlation. It should be noted that “other TKIs” consisted of arms from studies of second- or later-line therapies. As these showed strong association in the analysis stratified by treatment line, this may confound the interpretation of the analysis by therapy type.

Table 2. Weighted linear regression analyses for the relationship between median OS and median PFS in 29 treatment arms evaluating patients (n = 2,189) with unresectable and/or metastatic GIST in observational studies (n = 4) and clinical trials (n = 15) of targeted therapies.

<table>
<thead>
<tr>
<th>Group</th>
<th>Correlation (95% CI)</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Slope (SE), P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.91 (0.82–0.96)</td>
<td>0.84</td>
<td>0.83</td>
<td>2.08 (0.18), P &lt; 0.0001</td>
</tr>
<tr>
<td>All arms, with removal of four influential points</td>
<td>0.72 (0.46–0.87)</td>
<td>0.52</td>
<td>0.50</td>
<td>1.33 (0.27), P &lt; 0.0001</td>
</tr>
<tr>
<td>Stratified by treatment line</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First line</td>
<td>0.52 (–0.88–0.99)</td>
<td>0.082</td>
<td>–0.38</td>
<td>0.25 (0.59), P = 0.71</td>
</tr>
<tr>
<td>First line, with removal of influential point</td>
<td>–1.0 (not available)*</td>
<td>0.98</td>
<td>0.97</td>
<td>–1.57 (0.20), P = 0.062</td>
</tr>
<tr>
<td>Second and 2.5-line</td>
<td>0.80 (0.21–0.96)</td>
<td>0.66</td>
<td>0.60</td>
<td>0.98 (0.29), P = 0.015</td>
</tr>
<tr>
<td>Second and 2.5-line, with removal of influential point</td>
<td>0.66 (–0.19–0.94)</td>
<td>0.51</td>
<td>0.41</td>
<td>1.32 (0.58), P = 0.072</td>
</tr>
<tr>
<td>Third and later lines</td>
<td>0.70 (0.33–0.88)</td>
<td>0.39</td>
<td>0.35</td>
<td>1.27 (0.41), P = 0.0074</td>
</tr>
<tr>
<td>Third and later lines, with removal of three influential points</td>
<td>0.62 (0.14–0.87)</td>
<td>0.44</td>
<td>0.40</td>
<td>1.77 (0.57), P = 0.0093</td>
</tr>
<tr>
<td>Stratified by treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imatinib</td>
<td>0.91 (0.56–0.98)</td>
<td>0.72</td>
<td>0.67</td>
<td>1.75 (0.45), P = 0.008</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>0.65 (not available)*</td>
<td>0.44</td>
<td>–0.13</td>
<td>0.62 (0.71), P = 0.54</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>0.29 (–0.80–0.93)</td>
<td>0.03</td>
<td>–0.29</td>
<td>0.58 (1.90), P = 0.78</td>
</tr>
<tr>
<td>Combination therapy</td>
<td>–0.26 (–0.93–0.81)</td>
<td>0.16</td>
<td>–0.13</td>
<td>–1.62 (2.18), P = 0.51</td>
</tr>
<tr>
<td>Other TKIs</td>
<td>–0.69 (–0.13–0.95)</td>
<td>0.39</td>
<td>0.27</td>
<td>1.28 (0.71), P = 0.13</td>
</tr>
<tr>
<td>Stratified by quality score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>0.98 (not available)*</td>
<td>0.96</td>
<td>0.91</td>
<td>1.52 (0.26), P = 0.15</td>
</tr>
<tr>
<td>2–2.75</td>
<td>0.85 (0.66–0.94)</td>
<td>0.73</td>
<td>0.72</td>
<td>1.22 (0.17), P &lt; 0.0001</td>
</tr>
<tr>
<td>3–3.75</td>
<td>0.99 (0.83–1.00)</td>
<td>0.96</td>
<td>0.95</td>
<td>2.66 (0.53), P = 0.0032</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.
*The confidence interval was not generated by the linear regression model due to inclusion of only three data points.

Before this analysis, it was assumed that there would be few high-quality studies of patients with unresectable and/or metastatic GIST, as it is a relatively rare cancer. This expectation was confirmed by the results of the systematic review. Included studies were characterized by limitations in design, the most notable of which was small sample sizes. To improve the strength of this arm-level analysis, only higher-quality studies with at least 15 patients per arm were included. The introduction of the GRADE...
quality assessment feature of the review and the strict cutoff for inclusion attempted to improve the level of evidence. A relatively low sample cutoff was chosen as several key GIST trials evaluated small sample sizes and it was necessary to include these data. Nonetheless, this analysis still considered four observational studies, which differs from analyses in more common cancers that only include data from clinical trials (14, 15, 38, 39).

One limitation of this analysis is the potential for a high degree of heterogeneity among the analyzed studies. Studies utilized differing inclusion and exclusion criteria, progression criteria (e.g., RECIST and SWOG), and time intervals between radiologic and clinical assessments. These differences may have contributed to a wider range of survival outcomes, which could bias the findings toward a stronger relationship between endpoints. However, the results of this study were similar in magnitude to estimates demonstrated in other cancers (39, 40). Future analyses should confirm the strength of the association in GIST using patient-level data from clinical trials, which has been widely regarded as necessary for the validation of endpoint surrogacy (41).

The inherent relationship between OS and PFS should also be considered. By definition, PFS is contained within OS, thus median OS is always expected to be longer than median PFS. Overlapping definitions of PFS and OS may account for part of the relationship between the values, especially if the length of survival after progression is short (37). Part of this problem may be solved by performing a study-level analysis with relative estimates of survival efficacy (i.e., HR). However, few randomized trials have evaluated a GIST population, which precludes an analysis of this type.

Relying on data from observational studies and clinical trials in patients with GIST, this analysis demonstrated a strong correlation between median OS and PFS. The association was more apparent in later treatment lines compared with first-line therapy. These findings provide some insight into the use of PFS as a surrogate marker for OS, which may be strengthened by future trial results; however, further patient-level analyses are needed to fully establish its validity in GIST.

This study highlights the importance and the need for further research to establish validity of PFS as a surrogate marker, which may ultimately lead to earlier treatment decisions, help to reduce the considerable costs of clinical trials, and to facilitate approval, reimbursement, and availability of new efficacious treatments for GIST.

**Disclosure of Potential Conflicts of Interest**

J. Chang and A. Mohamed are employees of Bayer HealthCare Pharmaceuticals. I. Ozer-Stillman is an employee of and a consultant/advisory board member for Evidera. L. Strand is an employee of Evidera. K. Tranbarger-Freier was an employee of Evidera. No other conflicts of interest were disclosed by the other authors.

**Authors’ Contributions**

Conception and design: I. Ozer-Stillman, J. Chang
Development of methodology: I. Ozer-Stillman, J. Chang, K.E. Tranbarger-Freier
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): I. Strand
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Strand, A.F. Mohamed, K.E. Tranbarger-Freier
Writing, review, and/or revision of the manuscript: I. Ozer-Stillman, L. Strand, J. Chang, A.F. Mohamed, K.E. Tranbarger-Freier
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Ozer-Stillman, L. Strand, J. Chang, A.F. Mohamed
Study supervision: I. Ozer-Stillman, J. Chang

**Acknowledgments**

The authors thank Dr. Kyle Fahrbach of Evidera for statistical advising and review of the manuscript and Mina Jeong, Emily Shore, and Christopher Ngai for review of the manuscript.

**Grant Support**

This work was supported by Bayer HealthCare Pharmaceuticals.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 16, 2014; revised October 29, 2014; accepted November 3, 2014; published OnlineFirst December 4, 2014.

**References**


Clinical Cancer Research

Meta-analysis for the Association between Overall Survival and Progression-Free Survival in Gastrointestinal Stromal Tumor

İpek Özer-Stillman, Lauren Strand, Jane Chang, et al.

Clin Cancer Res  Published OnlineFirst December 4, 2014.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-1779

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.