Progression-Free Survival as a Surrogate for Overall Survival in Advanced/Recurrent Gastric Cancer Trials: A Meta-Analysis

Xavier Paoletti, Koji Oba, Yung-Jue Bang, Harry Bleiberg, Narikazu Boku, Olivier Bouché, Paul Catalano, Nozomu Fuse, Stefan Michiels, Markus Moehler, Satoshi Morita, Yasuo Ohashi, Atsushi Ohtsu, Arnaud Roth, Philippe Rougier, Junichi Sakamoto, Daniel Sargent, Mitsuru Sasako, Kohei Shitara, Peter Thuss-Patience, Eric Van Cutsem, Tomasz Burzykowski, Marc Buyse; on behalf of the GASTRIC group

Manuscript received February 10, 2013; revised July 25, 2013; accepted July 25, 2013.

Correspondence to: Xavier Paoletti, PhD, Bisotatistics Dept / INSERM U900, Institut Curie, 26 rue d’Ulm, 75005 Paris, France (e-mail: xavier.paoletti@curie.net).

The traditional endpoint for assessing efficacy of chemotherapies for advanced/recurrent gastric cancer is overall survival (OS), but OS requires prolonged follow-up. We investigated whether progression-free survival (PFS) is a valid surrogate for OS. Using individual patient data from the GASTRIC meta-analysis, surrogacy of PFS was assessed through the correlation between the endpoints and through the correlation between the treatment effects on the endpoints. External validation of the prediction based on PFS was also evaluated. Individual data from 4069 patients in 20 randomized trials were analyzed. The rank correlation coefficient between PFS and OS was 0.853 (95% confidence interval [CI] = 0.852 to 0.854). The R² between treatment effects on PFS and on OS was 0.61 (95% CI = 0.04 to 1.00). Treatment effects on PFS and on OS were only moderately correlated, and we could not confirm the validity of PFS as a surrogate endpoint for OS in advanced/recurrent gastric cancer.


The prognosis of patients with advanced or recurrent gastric cancer (AGC) remains poor, with a 1-year median overall survival (OS) for commonly used chemotherapy regimens, consisting of fluoropyrimidine, platinum, taxane or anthracyclines agents (1). The most important issue in the development of agents for AGC is their ability to prolong OS with acceptable toxicity. Even though median postprogression survival ranges from 5 to 10 months, a validated shorter-term surrogate endpoint would likely reduce drug development costs, sample sizes, or the duration of trials aimed at establishing the benefit of new drugs. Progression-free survival (PFS) is commonly used in phase II and phase III trials. It has been evaluated as a surrogate endpoint for OS in several types of cancers (2–4). The ability to predict clinical benefits on OS from earlier benefits on PFS could be useful at all stages of clinical development. Here, we investigate the surrogacy of PFS for OS within the framework of the GASTRIC meta-analysis (5).

Trials were eligible if they were randomized, closed to accrual before the end of 2006, and collected individual patient data on PFS. To explore the correlation between the treatment effects at the trial level, we relied on the comparison between the experimental arms of the trials included in the meta-analysis with their corresponding control arms. We defined as experimental the treatment that contained the larger number of drugs (eg, triple combinations vs double combinations). In case of equal number of drugs, we defined as experimental the treatment that included the newer agent. When two experimental arms were tested in the same trial, we combined their data for the purposes of the analyses. All data were centrally checked for inconsistencies (6).

We used a meta-analytic validation approach (3,4,7). OS was defined as the time from randomization to death from any cause or to the last follow-up. PFS was the time to tumor progression or death from any cause or time to the last follow-up assessment. A detailed description of statistical methods used is provided in the Supplementary Material (available online). For external validation, we applied the identified relation to predict the hazard ratio (HR) for OS (HR\textsubscript{OS}) from the hazard ratio for PFS (HR\textsubscript{PFS}) in randomized trials published since 2000 for which we had not obtained the individual patient data. We extracted the summary statistics for both endpoints (8) and compared the predicted value of HR\textsubscript{OS} to the one reported in the articles. To determine whether surrogacy also applied to other classes of agents, we extended the validation to three published trials of targeted agents (9–11).

Individual data were obtained on 4069 patients from 20 eligible randomized trials (12–30). The characteristics of the trials have been described elsewhere (5). Thirteen trials defined the progression using radiological criteria, whereas seven used both clinical and radiological assessments. Overall and at the trial level, the treatment effect on PFS (HR = 0.79; 95% confidence interval [CI] = 0.74 to 0.85) tended to be larger than on OS (HR = 0.85; 95% CI = 0.79 to 0.92) as shown on the forest plot of Supplementary Figures 1 and 2 (available online).

The individual-level association, as measured by the rank correlation coefficient, was 0.853 (95% CI = 0.852 to 0.854), indicating substantial correlation between PFS and OS for a given patient. The association at the trial level between log HR\textsubscript{OS} and log HR\textsubscript{PFS} was only moderate, with a coefficient of determination, R², adjusted for the estimation errors (31), of 0.61 (95% CI = 0.04 to 1.00). The large confidence interval reflects the uncertainty around this estimate. The linear regression model that relates the treatment effect on PFS and on OS adjusted for estimations errors was

\[
\log(\text{HR}_{\text{OS}}) = 0.042 + 0.779 \times \log(\text{HR}_{\text{PFS}})
\]
where the standard errors of the intercept and the slope were 0.79 and 0.295, respectively. This is shown as a straight line in Figure 1. The 95% prediction limits indicate the range of effect on OS that can be expected for a given effect on PFS. The moderate predictive accuracy at the trial level is reflected by the large interval width and a surrogate threshold effect of 0.56; hence, one should observe an HR_{PFS} less than 0.56 to predict, with 95% probability, an HR_{OS} less than 1.

Validation on independent literature data (9–11, 32–39) is shown in Table 1 and Supplementary Figure 3 (available online). The larger the number of progressions, the more precise the prediction; however, precision is limited by the variability of the regression line. The observed HR_{OS} fell within the prediction interval in all trials, even in trials using humanized monoclonal antibodies [Trastuzumab (10), bevacizumab (9), matuzumab (11)]. However, in the trial that concluded a statistically significant benefit of trastuzumab on OS (10), the effect on PFS was smaller than the surrogate threshold effect and therefore could not have been used to predict a statistically significant effect on OS.

This is the first study based on individual patient data to evaluate whether PFS is a reasonable surrogate endpoint to use for randomized trials in AGC. Our results show a high correlation of PFS and OS in individual patients but only a modest correlation (R^2 = 0.61) between treatment effects on PFS and OS. It is lower than that found in trials of 5-fluorouracil–based therapies for advanced colorectal cancer (4). The correlation was also lower than in the adjutant setting (40).

Possible limitations that may explain the moderate correlation observed in our analysis include the numerous processes involved in the progression of stomach cancer (eg, local or distant metastasis, peritoneum involvement), the use of clinical and radiological assessments for progression, and the impact of our definition of investigational treatment related to the heterogeneity in chemotherapies considered here; variability in the investigated treatments and in the effects of the treatments is a condition to generalize any results to future trials. Last, patients included in more recent trials received second-line treatments, including crossover (30), which may have diluted the effect of first-line treatment on OS (2). Because not all trials reported the same information at baseline, we could not assess the surrogacy in clinically relevant subset analyses.

All in all, we would not conclude that PFS is an adequate surrogate for OS in AGC. No precise prediction of the effect of a treatment on OS can be reliably drawn from the effect estimated on PFS.

![Figure 1](https://academic.oup.com/jnci/article-abstract/105/21/1667/2517578)

**Figure 1.** Trial-level association between treatment effects. Log scale was used for the x and y axes; the horizontal line (circles) corresponds to the hazard ratio (HR) on overall survival of 1, which indicates the absence of effect on the overall survival. At the crossing point, the vertical line corresponds to the minimum amount of effect on PFS that will predict a hazard ratio on OS below 1 with 95% probability. This indicates the surrogate threshold effect.

<table>
<thead>
<tr>
<th>Trial label</th>
<th>Trial</th>
<th>Observed HR_{PFS} (95% CI)</th>
<th>Observed HR_{OS} (95% CI)</th>
<th>Predicted HR_{OS} (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Jeung et al. (36)</td>
<td>0.63 (0.28 to 1.05)</td>
<td>0.56 (0.25 to 0.88)</td>
<td>0.73 (0.46 to 1.04)</td>
</tr>
<tr>
<td>B</td>
<td>AIO (33)</td>
<td>0.67 (0.43 to 1.04)</td>
<td>0.82 (0.47 to 1.45)</td>
<td>0.76 (0.53 to 1.07)</td>
</tr>
<tr>
<td>C</td>
<td>ToGA (10)</td>
<td>0.71 (0.59 to 0.85)</td>
<td>0.74 (0.60 to 0.91)</td>
<td>0.80 (0.58 to 1.09)</td>
</tr>
<tr>
<td>D</td>
<td>AVAGAST (9)</td>
<td>0.80 (0.68 to 0.93)</td>
<td>0.87 (0.73 to 1.03)</td>
<td>0.88 (0.76 to 1.14)</td>
</tr>
<tr>
<td>E</td>
<td>Kang et al. (35)</td>
<td>0.80 (0.63 to 1.03)</td>
<td>0.85 (0.64 to 1.13)</td>
<td>0.88 (0.76 to 1.14)</td>
</tr>
<tr>
<td>F</td>
<td>Park et al. (38)</td>
<td>0.86 (0.54 to 1.37)</td>
<td>0.96 (0.60 to 1.52)</td>
<td>0.93 (0.71 to 1.18)</td>
</tr>
<tr>
<td>G</td>
<td>REAL (a)† (34)</td>
<td>0.92 (0.80 to 1.04)</td>
<td>0.92 (0.80 to 1.10)</td>
<td>0.98 (0.77 to 1.22)</td>
</tr>
<tr>
<td>H</td>
<td>REAL (b) (34)</td>
<td>0.92 (0.81 to 1.05)</td>
<td>0.86 (0.80 to 0.99)</td>
<td>0.98 (0.77 to 1.22)</td>
</tr>
<tr>
<td>I</td>
<td>Ross et al. (39)</td>
<td>0.95 (0.80 to 1.09)</td>
<td>0.91 (0.76 to 1.04)</td>
<td>1.00 (0.79 to 1.29)</td>
</tr>
<tr>
<td>J</td>
<td>FLAGS (32)</td>
<td>0.99 (0.86 to 1.14)</td>
<td>0.92 (0.80 to 1.05)</td>
<td>1.03 (0.81 to 1.31)</td>
</tr>
<tr>
<td>K</td>
<td>Rao et al. (11)</td>
<td>1.13 (0.63 to 2.01)</td>
<td>1.02 (0.61 to 1.70)</td>
<td>1.14 (0.89 to 1.46)</td>
</tr>
<tr>
<td>L</td>
<td>Moehler et al. (37)</td>
<td>1.14 (0.59 to 2.21)</td>
<td>0.77 (0.51 to 1.17)</td>
<td>1.15 (0.90 to 1.48)</td>
</tr>
</tbody>
</table>

* HR = hazard ratio; PFS = progression-free survival; CI = confidence interval; OS = overall survival.
† This trial was designed as a factorial 2 × 2 plan to test two comparisons: a platinum comparison (a) and a fluoropyrimidine comparison (b).


**Funding**

This work was partially supported by the French Institut National du Cancer (grant PHRC GASTRIC); the Clinical Research Support Unit; and the Epidemiological and Clinical Research Information Network. A meeting was supported financially by unrestricted grants from GlaxoSmithKline. The GASTRIC Investigators: Secretariat: Marc Buyse, Stefan Michiels, Kenichi Nakamura, Koji Oba, Xavier Paolletti, Philippe Rougier, and Seiichiro Yamamoto. Steering Committee: Yung-Jue Bang (Seoul National University College of Medicine, Seoul, Korea); Harry Bleiberg (Jules Bordet Hospital, Brussels, Belgium); Tomasz Burzykowski (Hasselt University, Diepenbeek, Belgium); Marc Buyse (International Drug Development Institute, Louvain-la-Neuve, Belgium); Catherine Delbaldo (Hôpital Louis Mourier, Colombes, France); Stefan Michiels (Institut Gustave Roussy, Université Paris XI, Villejuif, France); Satoshi Morita (Yokohama City University, Kanagawa, Japan); Koji Oba (Hokkaido University Hospital, Hokkaido, Japan); Yasao Ohashi (University of Tokyo, Tokyo, Japan); Xavier Paolletti (Institut Curie, Paris, France); Jean-Pierre Pignon (Institut Gustave Roussy, Villejuif, France); Philippe Rougier (University Hospital European Georges Pompidou, Paris, France); Junichi Sakamoto (Tokai Central Hospital, Sohara, Japan); Daniel Sargent (Mayo Clinic, Rochester, MN); Mitsuru Sasako (Hyogo College of Medicine, Hyogo, Japan); and Eric Van Cutsem (Digestive Oncology Unit, University Hospital Gasthuisberf, Leuven, Belgium). Collaborators: J. Ajang, N. Boku, O. Bouche, J. Buckner, C. Coombes, S. Cullinan, M. Dank, N. Fuse, B. Gimelius, R. Hawkins, W. Koizumi, M. Moehler, Y. Nio, A. Ohtsu, A. Roth, K. Shitara, P. Thuss-Patience, A. Tsuburaya, E. Van Cutsem, U. Vanhoever, J. Wils, and Y. Yamamura.

**Writing committee:** Xavier Paolletti, Koji Oba, Tomasz Burzykowski, Yung-Jue Bang, Harry Bleiberg, Narikazu Boku, Olivier Bouché, Paul Catalanou, Nozomu Fuse, Stefan Michiels, Markus Moehler, Satoshi Morita, Yasuo Ohashi, Atsushi Ohtsu, Arnaud Roth, Philippe Rougier, Junichi Sakamoto, Daniel Sargent, Mitsuru Sasako, Kohei Shitara, Peter Thuss-Patience, Eric van Cutsem, and Marc Buyse.

This project was initiated under the auspice of the French Institut National du Cancer, who served as a sponsor. The French Institut National du Cancer did not participate in the design of the study. It participated in the conduct of the study at an earlier stage by centralizing all the databases and by providing administrative and data management support. The sponsor had no role in the preparation, review, or approval of the manuscript.

The GASTRIC Group thanks all patients who took part in the trials and contributed to this research. The meta-analysis would not have been possible without their participation or without active participation of the collaborating institutions that provided their trial data (East Central Oncology Group; European Organization of Research and Treatment of Cancer; Fédération Francophone de Cancérologie Digestive; Italian Trials in Medical Oncology; Japanese Cooperative Oncology Group; North Central Cancer Treatment Group; South West Oncology Group; the V325 Study Group; the Swiss Group for Clinical Cancer Research; the Kyoto Research Group for Chemotherapy of Gastric Cancer, the Medizinische Klinik mit Schwerpunkt Haematologie und Onkologie, Charité, Universitätsmedizin Berlin; the Center for Caring Sciences, University of Uppsala Sweden. We thank Nicolas Thannavong for the data management.

**Affiliations of authors:** Biostatistics Department, INSEED U900 Institut Curie, Paris, France (XP); Translational Research and Clinical Trial Center, Hokkaido University Hospital, Hokkaido, Japan (KO); Seoul National University College of Medicine, Oncology Division, Seoul, Korea (Y-JB); Jules Bordet Hospital, Brussels, Belgium (HB); St. Marianna University School of Medicine, Kawasaki, Japan (NB); Hôpital Robert Debré, Reims, Department of Clinical Oncology, France (OB); Dana-Farber Cancer Institute and Harvard School of Public Health, Department of Biostatistics, Boston, MA (PC); National Cancer Center Hospital East, Department of Gastrointestinal Oncology, Kashiwa, Japan (NF AO); Institut Gustave Roussy, Université Paris XI, Biostatistics and Epidemiology Department, Villejuif, France (SMi); Johannes Gutenberg University, Medical Department, Mainz, Germany (MM); Yokohama City University, Department of Biostatistics and Epidemiology, Kanagawa, Japan (SMO); University of Tokyo, Tokyo, Japan (YO); University Hospital, Department of Surgery, Geneva, Switzerland (AR); University Hospital European Georges Pompidou, gastro-enterology Department, Paris, France (PR); Tokai Central Hospital, Sohara, Japan (IJS); Mayo Clinic, Division of Biomedical Statistics and Informatics, Rochester, MN (DS); National Cancer Center Hospital East, Kashiwa, Japan (MS); Aichi Cancer Center Hospital, Department of Gastrointestinal Oncology, Aichi, Japan (KS); Charité-Universitätsmedizin Berlin, Department of Haematology, Oncology, and Tumorimmunology, Berlin, Germany (PFP); University Hospital Gasthuisberg, Digestive Oncology Unit, Leuven, Belgium (EVC); Hasselt University, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Diepenbeek, Belgium (TB MB); International Drug Development Institute, Louvain-la-Neuve, Belgium (MB).