Supplementary Materials and Methods

Protocol Summary of ACTS-GC Biomarker Study

Background:
The results of the first interim analysis of the ACTS-GC study, a randomized phase III trial, demonstrated that adjuvant treatment with S-1 after D2 dissection for locally advanced gastric cancer (GC) was more effective than surgery alone in East Asian patients\(^3\).

The clinical significance of most biomarkers in GC remains unclear. Exploratory evaluations are required to better understand the clinical significance of biomarkers.

Objectives:
1) To define and classify patients with stage II/III gastric cancer into high-risk and low-risk groups for recurrence and survival.
2) To define and classify patients with stage II/III gastric cancer into good responders and poor responders to postoperative adjuvant therapy with S-1.

Study design:
A series of biomarker studies designed retrospectively after the completion of the first interim analysis of the ACTS-GC.

Study population:
The 1,059 patients enrolled in the ACTS-GC.

Samples:
Archived formalin-fixed, paraffin-embedded (FFPE) specimens obtained by surgical resection.

Planned sample size:
A minimum of 634 patients (60%) of the total number of enrolled patients.

Treatment regimens:
- Arm A: S-1 (80-120 mg/day, 4 weeks of treatment followed by 2 weeks of rest per course; continued for 12 months)
- Arm B: Surgery alone (no further therapy)

**Endpoints:**
- Relapse-free survival (RFS) and overall survival (OS)
- Histology, sites of metastasis/recurrence, and other clinicopathological factors evaluated in the ACTS-GC study.

**Biomarkers and methods for assessment:**
- Expression of 63 genes in total, including EGFR and HER2, on RT-PCR (Table S1)
- Expression of 7 genes including EGFR and HER2 on immunohistochemistry (IHC) (Table S1)
- HER2 amplification on dual-color *in-situ* hybridization (dual-ISH)
- KRAS mutation on DNA direct sequencing

**Statistical methods:**

Categorical data analysis will be conducted using the chi-square test. Either the Wilcoxon test or the Kruskal-Wallis test will be used to assess correlations between groups. Survival curves will be estimated using the Kaplan-Meier product-limit method, and the statistical significance of differences between survival curves will be assessed using the log-rank test. Univariate and multivariate survival analyses will be performed using a Cox proportional-hazards model. Results will be considered statistically significant at P<0.05. All statistical analyses will be carried out with the SAS software package, version 9.1 and JMP software, version 8.01 (SAS Institute Inc., Cary, NC).

**Sample size estimation:**

Our primary analysis will determine whether these biomarkers, considered as a group, could be used to divide: 1) the population of patients into high-risk and low-risk groups for recurrence and survival (objective #1); and 2) the population of patients into good responders and poor responders to postoperative adjuvant therapy with S-1 (objective #2).

Each of the variables will be used in a proportional-hazards model to determine if high-risk and low-risk patients can be identified. For this purpose, the variables will be defined as binary. At the time of planning this study, we estimated that archived FFPE specimens obtained by surgical resection would be available for 60% (634) of the patients enrolled in the ACTS-GC study. The primary endpoint would be 3-year RFS. Based on the data from ACTS-GC, we estimated that the probability of 3-year RFS in the surgery alone group, S-1 group, and the total population would be 60.1%, 72.2%,
and 66.2%, respectively. When estimating the sample size, adjustment for multiple comparisons was not considered because the analyses were exploratory.

For objective #1, we estimated that a hazard ratio for recurrence of 0.69 would be equivalent to a 10% absolute difference in the 3-year RFS between the high- and low-risk groups. Practically, we expected that the 3-year RFS for the total population would be 71.2% and 61.2% in the high- and low-risk groups, respectively. Given that the sample sizes would be equal (n = 317) and that two-sided \( \alpha = .05 \), the power would be 76% to detect a 10% difference in the 3-year RFS.

For objective #2, we estimated that the hazard ratio for recurrence would be 0.55, which is equivalent to a 14% absolute difference in the 3-year RFS between good responders and poor responders in the S-1 group. Practically, we expect that the 3-year RFS in the good responders and poor responders in the S-1 group would be 65.2% and 79.2%, respectively. Given that sample sizes are equal (n = 158) and that two-sided \( \alpha = .05 \), the power would be 79% to detect a 14% difference in the 3-year RFS.

For individual variables, if the proportion of high- and low-risk patients (or good and poor responders) differs considerably, the statistical power would be lower. Because of potential biological implications, we may perform further exploratory analyses based on a statistical analysis protocol designed by investigators involved in this ACTS-GC biomarker study.

a, In the results, archived FFPE specimens were available for 829 (78.3%) of the 1,059 patients thanks to the cooperation of the 65 institutions. This figure was considerably higher than expected.

b, Sample size estimates were performed before the completion of the 5-year follow-up.