<Study protocol with SAP>

Locoregional recurrence after neoadjuvant versus adjuvant chemotherapy based on tumor subtypes in patients with early-stage breast cancer: A multi-institutional retrospective cohort study

(Brief Title: Locoregional Recurrence After Neoadjuvant Versus Adjuvant Chemotherapy)

Version 1.0

Document Date: 2022-04-22

1. Research Hypothesis and Objectives

: The hypothesis of this study is to investigate whether there is a difference in ipsilateral breast recurrence between patients who received preoperative chemotherapy and those who received postoperative chemotherapy after breast-conserving surgery and radiation therapy for breast cancer. The objective is to conduct a multi-institutional retrospective analysis to examine this difference.

2. Criteria for Selection and Exclusion of Study Subjects

*Selection Criteria:

- Patients who underwent breast-conserving surgery for invasive breast cancer at three hospital in Korea between 2004 and 2018.

* Exclusion Criteria:

- Patients who underwent mastectomy.

- Patients who did not receive pre- or postoperative chemotherapy.

- Patients who did not undergo radiation therapy.

- Patients with bilateral breast cancer.

- Male breast cancer patients.

- Patients with secondary cancer.

- Metachronous or synchronous cancers.

3. Number of Study Subjects

- 10328 patients

4. Specific Research Methods

- Retrospective analysis of clinical information and pathological findings of patients who underwent breast-conserving surgery at the hospital from January 2004 to December 2018.

- Confirmation of ipsilateral recurrence in all patients, noting the date of recurrence if applicable, or the date of last outpatient visit or death if no recurrence occurred.

- Assessment of significant differences in IBTR-free survival rates between patients who received neoadjuvant chemotherapy (NAC) and those who received adjuvant chemotherapy (CTx).

- Subgroup analysis based on patient age, molecular subtype (Luminal A, Luminal B, HER2-enriched, basal-like), and clinical-pathological characteristics.

- Survival analysis conducted using the log-rank test and Cox proportional hazards model, with calculations made within a 95% confidence interval.

5. Data Collection Items

- Patient information: Age at the time of surgery.

- Surgical information: Date of surgery, type of breast-conserving surgery.

- Pathological information: T stage, N stage, pathological complete response (pCR), hormone receptor

- status, HER2 receptor status, Ki-67%, histologic grade.

- Treatment information: NAC and chemotherapy regimen, hormone therapy, Herceptin administration.

- Prognostic information: Occurrence and date of local recurrence, occurrence date of other metastases

- or relapse, date of last outpatient follow-up

6. Informed consent

The requirement for informed consent was waived owing to the retrospective nature of the study.

7. IRB No.

- Seoul National University Hospital, 2204-074-1316

- Samsung Medical Center, 2023-09-076-001

- Boramae medical center, 10-2023-73

8. Statistical analysis plan

- Categorical variables are compared using Pearson's chi-square test.

- Continuous variables are compared using Mann-Whitney U and Kruskal-Wallis tests.

- Survival analysis would be conducted using the log-rank test with calculations made within a 95% confidence interval.

- Cox proportional hazards regression analysis are performed to adjust for related clinicopathological variables and estimate hazard ratios. Variables with a two-sided p-value of <005 in the univariate analysis would be included in the multivariate analysis, and those that showed a variance inflation factor of >4·0 in the multicollinearity test would be excluded. The model's goodness of fit was

evaluated using the Hosmer-Lemeshow test and a two-sided p-value of >0·05 indicated an appropriate model.

- The propensity scores are estimated using a logistic regression model, and 1:1 matching would be conducted using the greedy nearest-neighbor matching method without replacement using a caliper width of 0·10 standard deviation of the logit of the propensity scores. Variables that are differently distributed between the two groups were used for matching analysis. The standardized mean differences are calculated to evaluate the balance of variables between the two groups, and a value of <0·1 indicates an adequate balance.

- Missing data are treated using a complete case analysis approach.

- A two-sided p-value of <0·05 was considered significant.

- Analyses would be conducted using SPSS (version 26.0; SPSS Inc., IBM, Armonk, NY, USA) or R Statistical software (version 3.6.3; R Foundation for Statistical Computing, Vienna, Austria). Kaplan-Meier curves would be drawn using GraphPad Prism (version 10.0; GraphPad Software, San Diego, USA).

- The patterns of annual recurrence incidence would be drawn using the kernel smoothing method with the *ksmooth* function in R.