Statistical analysis plan

In this study, baseline characteristics will be summarized using proportions for categorical variables (cancer treatment, revealed cardiotoxic complications, etc.), and median (interquartile range) be presented for continuous variables (biomarker values, etc.). The Pearson χ^2 criterion will be applied to identify intergroup differences for categorical variables. Quantitative variables will be compared using the nonparametric U Mann-Whitney test in terms of rhythm disturbances (ECG / HM), dynamics in the level of the biomarkers (0, 3, 6, 9, 12 months), the degree of myocardial dysfunction identified by EF and GLS (0, 3, 6, 9, 12), etc., for two unrelated groups (with and without the development of cardiotoxicity). A comparison of three or more independent groups will be carried out using parametric analysis of variance (one-way ANOVA) or a nonparametric Kruskal-Wallis H-test, as well as analysis of variance with Greenhouse-Geisser corrections for repeated measurements (for related samples).

To characterize the changes in biomarker levels according to treatment groups, mean estimated changes from baseline will be plotted over time. Mean changes will be determined using repeated-measures linear regression estimated via generalized estimating equations. Each model will be adjusted for the baseline values of the biomarker under consideration and the time since the treatment started. Contemporaneous associations between changes in biomarkers from baseline and changes in LVEF will also be determined using repeated-measures linear regression.

To assess the influence of independent factors on the binary variable of response (cardiotoxicity), multiple logistic regression analysis (LRA) will be used by the sequential exclusion of variables (age, biomarkers, arrhythmias (or other identified cardiotoxic conditions), LVEF according to Simpson and GLS data, etc.). The presence of a statistically significant relationship with the predicted event in the one-dimensional analysis will be the criterion for inclusion in the multivariate analysis. The results will be presented as unadjusted (uOR), adjusted odds ratios (aOR), and 95% CI. The incidence of cardiotoxic complications of chemotherapy will also be presented with 95% CI calculated according to the Wilson method.

Indices of sensitivity (Sn), specificity (Sp), PPV, and PVN, will be calculated for each diagnostic test under consideration (biomarkers, TTE, rhythm disturbances, etc.), as well as for predictive models in general.

The risk of interrupting a course of chemotherapy due to cardiotoxicity in the two analyzed groups will be assessed through Kaplan-Meier survival curves. The relative risk of interrupting a full course of chemotherapy due to cardiotoxic events, adjusted (aOR), given the influence of potential confounders, will be assessed using multiple analyses of proportional Cox risks. The results will be presented as aOR with a 95% CI.

Associations between baseline biomarker values and time to CTRCD will also be assessed using Cox proportional hazards models. Besides, partly conditional survival models will be applied to determine associations between repeated assessments of changes in biomarkers from baseline and time to CTRCD. All models will be adjusted for cancer therapy regimen, baseline LVEF and group at risk (and cardiac protection if administered), baseline biomarker values, age, comorbidities index, and body mass index. Associations between changes from baseline in biomarker levels over time and subsequent changes in LVEF and longitudinal strain will be assessed similarly.

Differences in the associations between changes in biomarkers and LVEF/GLS across the different treatment groups will be evaluated by including biomarker-treatment interaction terms, as shown in the recent papers.

Due to plenty of variables subjected to analysis, and the anticipated presence of multiple outcomes, the included tests may not occur enough to solve a task to build a clinically relevant decision algorithm, and a decision-tree method will be applied.

Two-sided a levels <0.05 are assumed to be statistically significant. For statistical processing, software packages SPSS (IBM, Armonk, USA, v.25), Statistica (StatSoft-Russia, v.10), and R 3.3.2 (R Foundation for Statistical Computing) will be applied.