

Statistical analyses:

The primary endpoint was time to distant recurrence. Distant recurrence was defined as metastatic disease, excluding contralateral disease, and loco-regional and ipsilateral recurrences. The endpoint was censored at last follow-up visit or death before distant recurrence such that risk is a pure risk calculation ignoring deaths.

Cox proportional hazard models were used to create the model in ATAC and the CTS5(ATAC) score was tested in BIG 1-98. A shrinkage factor of 0.980 for the non-nodal part of the CTS5(ATAC) score had been calculated during its derivation using a nested Cox model [14] and applied to allow for the small amount of overfitting. We estimated the shrinkage factor with the following equation:

$$\gamma = ((\text{model } \chi^2 - \text{df}) / \text{model } \chi^2)^{1/2}$$

where model χ^2 is the likelihood ratio χ^2 statistics for testing all predictors and df is degrees of freedom.

To define the relation between CTS5 and 5-10 year DR risk, the logarithm of the baseline cumulative hazard function was fitted. Baseline risk at 5 years was calculated using the “stcox/basesurv” command in STATA to implement the Breslow method. Five to 10-year DR risk was then calculated for each participant by adjusting the baseline risk: $\text{risk}(5-10 \text{ years}) = 1 - ([\text{baseline risk}]^{\exp[\text{linear prediction CTS5}]})$. Proportional assumptions were verified using Schoenfeld residuals.

Likelihood ratio chi-square ($LR-\chi^2$) statistics and Kaplan-Meier survival estimates with corresponding 95% confidence intervals (calculated from the standard error of the cumulative hazards based on a normal approximation) were used to determine the prognostic performance of the CTS5(ATAC) in BIG 1-98. The risk of a DR of events for individual patients in BIG1-98 was estimated using CTS5 or CTS0 and the expected risk compared to the observed events in deciles of expected risk. The observed and expected numbers were compared by the χ^2 test. The overall agreement was assessed by calculating the correlation coefficient across the deciles. Concordance between expected and actual outcomes was also calculated by computing Harrell's C-index.

The 5-10 year distant recurrence risk groups were determined in ATAC and defined as: low risk group <5%, intermediate risk group 5-10%, and high risk group >10%. To compare the prognostic performance of CTS5(ATAC) between ATAC and BIG 1-98 trials, CTS5(ATAC) was normalised to have unit variance and the hazard ratios (HRs) and associated 95% confidence interval (CI) were estimated from Cox models. All statistical analyses were two-sided, and $p < 0.05$ was regarded as statistically significant. We also compared the newly developed CTS5(ATAC) to the published CTS (termed CTS0 below to avoid confusion) that had been developed for estimating prognosis from the time of disease presentation [6] to see whether an improved prognostication for late distant recurrence was achieved. All analyses were performed with STATA version 13.1 (College Station, Texas, USA).

The final model was fitted on the combined ATAC and BIG 1-98 datasets to give an overall calibration of the CTS5. Therefore, new coefficients were fitted in the combined dataset but using the same variables as in the training or validation cohorts (i.e. five nodal groups, continuous age, continuous size, and three grade groups).