

Supplement 3: Calibration plot of the prediction model

Of 2892 DCIS included in the study, 379 (13%) had missing data for one or more potential risk factors. These missing data were accounted for via multiple imputations (20 times). For one of these imputed datasets the calibration plot was drawn.

The predicted risk for an underestimated invasive breast cancer versus the observed rate is plotted in the figure below. The distance between the grouped observations and the 45 degree line is a measure of the error of the prediction model. The statistics for the plot for this imputed dataset are: c-statistic (ROC) of 0.666, R^2 of 0.110, Intercept of -0.019, slope of 0.985.

The histogram of the x-axis reflects the frequency of a predicted risk. 9.0% of DCIS have a risk of <12.0%, 54.6% of DCIS have a predicted risk of <15.0% and 26.0% have a risk of >25.0%.

