Including the Ductal Carcinoma-In-Situ (DCIS) Score in the Development of a Multivariable Prediction Model for Recurrence After Excision of DCIS

Lawrence Paszat,[1,2,3] Rinku Sutradhar,[1,2] Limei Zhou,[2] Sharon Nofech-Mozes,[1,3] Eileen Rakovitch[1,2,3]

Abstract

Clinical care for women who have undergone breast-conserving surgery for ductal carcinoma-in-situ (DCIS) is controversial. Accurate prediction of the risks of recurrence would help inform decision making. Predictive models including the DCIS score multigene expression assay were developed using a complete case series of 1102 women; results indicate that such a model allows computation of more accurate risk of recurrence.

Introduction: Individual prediction of local recurrence (LR) risk after breast-conserving surgery (BCS) for ductal carcinoma-in-situ (DCIS) is needed to identify women at low risk, for whom radiotherapy may be omitted. **Patients and Methods:** Three predictive models of LR—clinicopathologic factors (CPF) alone; CPF + estrogen receptor (ER) + human epidermal growth factor receptor 2 (HER2); and CPF + DCIS score (DS)—were developed among 1102 cases of DCIS in patients with complete covariate and outcome data. Categorizations of discrete variables and transformations of continuous variables were examined in Cox models; 2-way interactions and interactions with time were assessed. Internal validation was performed by bootstrapping. Individual predicted 10-year LR risks were computed from covariate values, estimated regression parameters, and estimated baseline survival function. Accuracy was assessed by c statistics and calibration plots. **Results:** The strongest prediction model incorporated CPF + DS. The c statistics for CPF + DS, CPF + ER + HER2, or CPF-alone models were 0.7025, 0.6879, and 0.6825, respectively. The CPF + DS model was better calibrated at predicting low ($\leq 10\%$) individual 10-year LR risks after BCS alone than those incorporating CPF + ER + HER2 or CPF alone, evidenced by c statistics and plots of observed by predicted risks. Among women aged ≥ 50 with no adverse CPF, the CPF + DS model identified the greatest proportion of women (62.3%) with predicted individual 10-year LR $\leq 10\%$ without radiotherapy compared to the CPF + ER + HER2 (50.9%) or CPF alone (46.5%) models. **Conclusion:** Individual prediction of LR incorporating DS is more accurate and identifies a higher proportion of women with low predicted risk of LR after BCS alone, for whom radiotherapy may be omitted.

[1]University of Toronto, Toronto, Ontario, Canada
[2]Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada
[3]Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

Address for correspondence: Lawrence Paszat, MS, MD, University of Toronto, T2-156 2075 Bayview Ave, Toronto, ON M4N3M5, Canada
E-mail contact: lawrence.paszat@sunnybrook.ca

Introduction

Ductal carcinoma-in-situ (DCIS) of the female breast is a non-obligate precursor lesion for infiltrating ductal carcinoma. The prevalence of DCIS has greatly increased in the era of mammography.[1,2] Currently, treatment guidelines recommend that women with DCIS undergo treatment to lower the risk of local recurrence (LR). Most women will receive breast-conserving surgery (BCS) followed by whole-breast radiotherapy (RT), leading to assertions that it DCIS overdiagnosed[3] and overtreated.[4-6] However,

randomized clinical trials[7-11] and meta-analyses[12,13] have failed to identify subsets of women, defined by clinicopathologic factors (CPF) such as age at diagnosis, lesion size, nuclear grade, comedonecrosis, multifocality, and margin width, whose risk of LR not reduced by RT. Improved, accurate prediction of the risk of LR for the individual woman is needed so that those with the lowest predicted risks may forego RT, thereby mitigating some of the consequences of overdiagnosis and reducing overtreatment.

The Oncotype DX DCIS score (DS) is a multigene expression assay based on 7 cancer-related genes (*Ki-67, STK15, Survivin, cyclin B1, MYBL2, PR, GSTM1*) and 5 reference genes.[14] DS is associated with the 10-year risk of LR and invasive recurrence (IR) after BCS for DCIS.[14,15] In multivariate analyses, DS, age, and tumor size have been independently associated with LR and IR.[15,16] In order to derive equations for the computation of the risks of LR and IR after BCS with or without RT, we developed 3 full prediction models: CPF + DS, CPF + estrogen receptor (ER) + human epidermal growth factor receptor 2 (HER2), 3 CPF alone.

We hypothesized that the equations including DS would be more accurate and would identify more women at very low risk of LR and IR compared to equations derived from prediction models excluding DS.

Patients and Methods

Study Population, Covariates, and Outcomes

The study population comprised a complete case series of 1102 women diagnosed with DCIS derived from the Ontario population-wide cohort of pure DCIS confirmed by expert pathologic review[17] and treated by BCS for whom complete information was available for all of the following CPF: age at diagnosis, lesion size (mm), presence or absence of observed multifocality, nuclear grade (high vs. low or moderate), comedonecrosis (present or absent in > 30% or > 50% of involved ducts), margin status (negative or positive), margin width (mm) if margin negative, and adjuvant RT (received or not). DS, ER, and HER2 expression were also available as continuous and categorical variables. The outcomes of interest were LR and IR within 10 years after the date of diagnosis of DCIS.[15-17] All methods described below were implemented separately, determining the best models incorporating CPF + DS, CPF + ER + HER2, and CPF alone for each outcome (LR and IR). This work was approved by the research ethics board of Sunnybrook Health Sciences Centre.

Model Building

We created adjusted Cox models of time to LR and time to IR. Various transformations of continuous variables were examined, and we investigated the effects of substituting different forms of each covariate into the regression models. Model selection procedures (forward, backward, and stepwise) were explored but discarded after inspection of c statistics from reduced models.

Best models were identified according to maximal discriminative ability, as determined by the c statistic. Concordance for survival data was computed based on information gathered from pairs of patients within the cohort. Specifically, the concordance index (c statistic) was calculated as the proportion of pairs in which the patient who experiences LR or IR has a higher predicted probability than the patient who does not experience the event. The index is measured on a scale of 0.5 to 1, where 0.5 implies that the discriminative ability is no better than chance and 1.0 implies perfect discrimination. The log-likelihood estimate and Akaike information criterion were also examined.

The proportional hazards assumption was assessed by including interactions with time and each covariate into the models. All 2-way interactions between covariates were also assessed.

Derivation of Equations to Calculate Individual Predicted Risk

The 10-year individual predicted probabilities of LR and of IR were calculated for each woman using specific covariate values, estimate of regression parameters, and estimate of baseline survival function from the best full models of LR and IR.

Model Calibration and Discrimination

Calibration (that is, how close the estimated risk is to the observed risk) was examined by grouping the individual predicted 10-year risks of LR and IR into deciles. We plotted the observed 10-year probabilities of LR and IR (the mean Kaplan-Meier estimate for patients in each decile) against the predicted 10-year of LR and IR from the equations derived from the best-performing predicted models (mean predicted probability for patients in each decile).[18]

Model Validation

Internal validation of the predictive models was performed by bootstrapping. Women were randomly selected with replacement from the source cohort to create 500 cohorts of the same sample size as the source cohort. The predictive equations derived from the source cohort were applied to obtain the 10-year predicted estimates of LR and IR for each woman in each bootstrap sample, and the concordance index was calculated for each bootstrap sample. The average of the 500 concordance indexes provided a bias-corrected estimate of the concordance index for the prediction model.

Distribution of Individual Predicted Risks

The univariate distribution of individual predicted risks of LR and IR was derived from the best full models incorporating CPF + DS, CPF + ER + HER2, and CPF alone, and stratified by the use of RT after BCS or not. We then compared the proportion of cases with a very low risk of LR (10-year risk of LR ≤ 10%) or a very low risk of IR (10-year risk of IR ≤ 5%) where there would be general agreement that RT can be safely omitted by each of the 3 predictive models. The percentage of women with 10-year individual predicted risk of LR ≤ 10% was computed by age (≥ 50 vs. < 50 years at diagnosis), presence or absence of ≥ 1 adverse CPF (lesion size > 10 mm, high nuclear grade, comedonecrosis, positive resection margins, multifocality), and receipt of RT after BCS or not, and compared them among the 3 models of LR: CPF + DS, CPF + ER + HER2, and CPF alone. The percentage of women with 10-year individual predicted risk of IR ≤ 5% was similarly computed and compared. The thresholds LR ≤ 10% predicted risk and IR ≤ 5% were selected to facilitate the description of very conservative estimates of the proportion of women who might conceivably be spared adjuvant RT after BCS. All analyses were conducted by the statistical software R 2.15 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org/) and SAS 9.3 (SAS Institute, Cary, NC, USA).

Results

The characteristics of the study population are listed in Table 1. Women aged ≥ 50 years on the date of diagnosis of DCIS comprised 78.3% of the population. Nuclear grade was low or moderate among 62.4%, comedonecrosis was present in ≤ 30% of ducts among 77.9%, lesion size was ≤ 10 mm among 50.4%, no multifocality was observed among 74.8%, and final resection margin width was ≥ 2 mm among 51.1%. DS category was low among 55.4%, ER was positive among 93.0%, and HER2 over-expression was negative among 72.1%. RT after BCS was received by 54.4% of the population.

Exploratory modeling of the 10-year risks of LR and IR incorporating DS, with various transformations and categorizations of CPF, yielded 120 models, compared to 384 models that included ER and HER2 with CPF, and that included CPF alone without DS, ER, and HER2 (Supplemental Material in the online version). Model selection procedures were explored, but they were abandoned because the c statistics were consistently lower for reduced models compared to full models.

Best Model of LR

The best full prediction model of the 10-year risk of LR after BCS incorporates CPF + DS [c statistic = 0.7035 (likelihood ratio χ^2 = 95.0378, degrees of freedom (df) = 10, $P < .0001$)]. The CPF + ER + HER2 and CPF-alone models had lower c statistics and lower likelihood ratio chi-square values (CPF + ER + HER2: c statistic = 0.6874, likelihood ratio χ^2 = 79.2108 with df = 10 and $P < .0001$; CPF alone: c statistic = 0.6827, LR = 78.2548 with df = 9 and $P < .0001$) (Table 2). The superior c statistic and likelihood ratio chi-square values indicated that the best full model of LR including DS (CPF + DS) predicted LR more accurately than the models without DS. The mean c statistics from the 500 bootstrapping samples for each model are shown in Table 2; they are concordant with the observation that the prediction model that integrates DS with CPF is the most accurate. There were no significant interactions found, and in particular no interaction between DS and RT.

Individual Predicted Risks of LR After BCS

The equation to predict an individual woman's risk of 10-year LR from the CPF + DS model is

$$
\begin{aligned}
0.014 * \text{DCIS score} + 0.486 * I(\text{age} < 50 \text{ years}) + 0.416 \\
* I(\text{comedo necrosis invloving} > 30\% \text{ of involved ducts}) \\
+ (-0.006) * I(\text{margin size 0 to 2 mm}) + (-0.130) \\
* I(\text{margin size} \geq 2\text{mm}) + (-0.679) \\
* I(\text{received adjuvant RT}) + 0.701 * I(\text{multifocality present}) \\
+ (-0.149) * I(\text{nuclear grade high}) + 0.211 * \sqrt{\text{tumor size}}
\end{aligned}
$$

where $I(.)$ represents the indicator function. The equations from the CPF + ER + HER2 model and the CPF-alone model are presented in the Supplemental Material (available in the online version).

Individual predicted 10-year risks of LR from CPF + DS ranged from 2.8% to 85.9% compared to the range 3.4% to 94.9% from CPF + ER + HER2 and the range 3.7% to 84.2% from CPF alone (data not shown). The univariate distribution of individual predicted risks of LR from the 3 models, from the 1st to 99th percentiles, are shown in Table 3, overall as well as stratified by age, the presence or absence of ≥ 1 adverse CPF (lesion size > 10 mm, high nuclear grade, comedonecrosis, observed multifocality, or positive resection margins), and whether BCS was followed by RT. By definition, this presents the individual predicted risk for the individual whose predicted risk falls at each percentile.

Table 1	Characteristics of Study Population
Characteristic	**Value**
DCIS Score	
Low	611 (55.4%)
Intermediate or high	491 (44.6%)
Mean (SD)	37.49 (23.29)
Estrogen Receptor	
Positive	1025 (93.0%)
Negative	77 (7.0%)
Mean (SD)	8.50 (1.42)
HER2 Overexpression	
Positive	212 (19.2%)
Equivocal	95 (8.6%)
Negative	795 (72.1%)
HER2 gene expression level, mean (SD)	10.51 (1.31)
Age at Diagnosis, y	
<50	239 (21.7%)
≥50	863 (78.3%)
Multifocality	
Not observed	824 (74.8%)
Observed	278 (25.2%)
Nuclear Grade	
Low or moderate	688 (62.4%)
High	414 (37.6%)
Comedonecrosis	
≤30% of ducts	859 (77.9%)
>30% of ducts	243 (22.1%)
Comedonecrosis	
≤50% of ducts	1006 (91.3%)
>50% of ducts	96 (8.7%)
Lesion Size	
≤10 mm	555 (50.4%)
>10 mm	547 (49.6%)
Margin Width	
0 mm (positive)	121 (11.0%)
>0 <2 mm	418 (37.9%)
≥2 mm	563 (51.1%)
Whole-Breast Radiotherapy	
No	503 (45.6%)
Yes	599 (54.4%)

Abbreviations: DCIS = ductal carcinoma-in-situ; HER2 = human epidermal growth factor receptor 2.

Multivariable Prediction Model in DCIS

Table 2	Best Full Predictive Models of Local Recurrence							
Characteristic	**Univariate Estimates and HRs**		**Best Model Including DS (CPF + DS)**		**Best Model Including ER and HER2 (CPF + ER + HER2)**		**Best Model Without DS, ER, or HER2 (CPF Alone)**	
c statistic			0.7025		0.6879		0.6825	
Mean bootstrapping c statistic (min, max) among 500 samples			0.7104 (0.6590, 0.7610)		0.6981 (0.6381, 0.7479)		0.6903 (0.6351, 0.7408)	
Likelihood ratio (P by chi-square test)			88.2998, 9 df (<.0001)		81.1531, 11 df (<.0001)		74.2756, 8 df (<.0001)	
	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)
DS per 1-unit increase	0.01587 (<.0001)	1.02 (1.01, 1.02)	0.01384 (.0001)	1.01 (1.01, 1.02)				
Estrogen Receptor								
Positive	−0.74039 (0.0012)	0.48 (0.30, 0.75)			−0.54462 (.0295)	0.58 (0.36, 0.95)		
Negative	Ref	Ref			Ref	Ref		
HER2 Overexpression								
Positive	−0.3216 (.2)	1.16 (0.67, 2.00)			−0.30856 (.2927)	0.73 (0.41, 1.30)		
Equivocal	Ref	Ref			Ref	Ref		
Negative	0.14433 (.6)	0.72 (0.44, 1.19)			−0.41483 (.1043)	0.66 (0.40, 1.09)		
Age, y								
<50	0.38171 (.02)	1.46 (1.06, 2.03)	0.48603 (.004)	1.63 (1.17, 2.26)	0.48614 (.004)	1.63 (1.17, 2.26)	0.47341 (.0051)	1.61 (1.15, 2.24)
≥50	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Multifocality								
Not observed	0.72107 (<.0001)	2.06 (1.52, 2.79)	Ref	Ref	Ref	Ref	Ref	Ref
Observed	Ref	Ref	0.70104 (<.0001)	2.02 (1.46, 2.78)	0.69033 (<.0001)	1.99 (1.44, 2.75)	0.67152 (<.0001)	1.96 (1.42, 2.70)
Nuclear Grade								
Low or moderate	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
High	0.37047 (.015)	1.45 (1.07, 1.95)	−0.14906 (.4051)	0.86 (0.61, 1.22)	−0.01457 (.9351)	0.99 (0.69, 1.40)	0.08005 (.6353)	1.08 (0.78, 1.51)
Comedonecrosis								
≤30% of ducts	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
>30% of ducts	0.66924 (<.0001)	1.95 (1.43, 2.67)	0.41599 (.02)	1.52 (1.07, 2.15)	0.48197 (.0076)	1.62 (1.14, 2.31)	0.55052 (.0017)	1.73 (1.23, 2.45)
Square root of lesion size per 1-unit increase	0.28249 (<.0001)	1.33 (1.20, 1.47)	0.21098 (.0003)	1.23 (1.10, 1.38)	0.23062 (<.0001)	1.26 (1.12, 1.41)	0.23307 (<.0001)	1.25 (1.13, 1.41)
Margin width								
00 mm (positive)	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
>0 < 2 mm	−0.16343, (.48)	0.85 (0.54, 1.34)	−0.00608 (.9797)	0.99 (0.62, 1.59)	0.01711 (.9427)	1.02 (0.64, 1.62)	0.03565 (.8807)	1.04 (0.65, 1.65)
≥ 2 mm	−0.40953 (.08)	0.66 (0.42, 1.04)	−0.13031 (.5827)	0.88 (0.55, 1.40)	−0.13751 (.5618)	0.87 (0.55, 1.39)	−0.13096 (.5795)	0.88 (0.55, 1.39)

Table 2 Continued

	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)
RT								
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	−0.47056 (.002)	0.63 (0.46, 0.84)	−0.67902 (<.0001)	0.51 (0.37, 0.69)	−0.6869 (<.0001)	0.50 (0.37, 0.68)	−0.65185 (<.0001)	0.52 (0.38, 0.71)

Abbreviations: CI = confidence interval; CPF = clinicopathologic factor; DS = ductal carcinoma-in-situ score; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.

Figure 1 shows the calibration plots of the observed 10-year risks of LR against the individual predicted 10-year risks of LR for each of the 3 prediction models. The plot for the predicted risks derived from CPF + DS (Figure 1A) shows tight calibration for the lower 5 deciles of risk, whereas the plots for the predicted risks derived from CPF + ER + HER2 (Figure 1B) and from CPF alone (Figure 1C) are only well calibrated for the lower 3 deciles of risk. High individual predicted risks from all models are less well calibrated than are low predicted risks.

We then compared the ability of each model to detect the proportion of cases with a very low individual predicted risk of LR (10-year risk of LR ≤ 10%) or a very low individual predicted risk of IR (10-year risk of IR ≤ 5%). Among women aged ≥ 50 years at diagnosis with no adverse CPF and who did not receive RT after BCS, the CPF + DS model identified 62.3% of cases, with individual predicted 10-year risks of LR ≤ 10% compared to 50.9% identified from the CPF + ER + HER2 model and 46.5% identified from the CPF-alone model. Therefore, the 2 models without DS overestimated the risk of some individual women. Additionally, the 2 models without DS underestimated the individual predicted risk for other women; among women aged ≥ 50 years without any adverse CPF with a predicted risk of LR > 10% by the CPF + DS model, 20.7% had predicted 10-year LR risks ≤ 10% according to the CPF + ER + HER2 model, and 13.2% of had predicted 10-year LR risks ≤ 10% according to the CPF-alone model. Furthermore, we applied the 3 prediction models to the cases in the cohort selected to receive RT and set the RT covariate to 0 to determine the proportion of women with low predicted 10-year risks of LR (≤ 10%) after BCS alone that might have avoided treatment. When the prediction equations were applied to women aged ≥ 50 years with no adverse CPF who had been selected to receive RT, the CPF + DS model identified 44.4% of cases that would have had individual predicted risks of LR ≤ 10% without RT compared to 39.4% from the CPF + ER + HER2 model, and 32.3% from the CPF-alone model.

Very few women aged ≥ 50 years at diagnosis with lesions > 10 mm in diameter or with ≥ 1 adverse CPF (multifocality, comedonecrosis, high nuclear grade, positive resection margins) had individual predicted 10-year risks of LR ≤ 10% without RT (7.8% from the CPF + DS model, 4.7% from the CPF + ER + HER2 model, and 2.4% from the CPF-alone model). Setting the RT covariate to 0 and applying the prediction equations to similar women who had received RT, 5.6% would have had individual predicted 10-year risks of LR ≤ 10% without RT from the CPF + DS model, compared to 2.8% from both the CPF + ER + HER2 model and the CPF-alone model. Among women aged < 50 years at diagnosis, 1.1% had individual predicted risk of LR < 10% without RT from the CPF + DS model, but none from the other models. Setting the RT covariate to 0 and applying the equations to women aged < 50 years who had received RT, 1.4% would have had individual predicted risk of LR ≤ 10% without RT from the CPF + DS model, but none from the other models.

Best Model of IR

The CPF + DS prediction model was also the best full model predictive of the 10-year risk of IR [c statistic = 0.6838 (likelihood ratio χ^2 = 43.3544, degrees of freedom (df) = 9, P < .0001)]

Table 3	Univariate Distributions of Individual Predicted 10-Year Risks (%) of Local Recurrence								
Characteristic	1st Percentile	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	99th Percentile
Overall									
CPF + DS	2.8	4.8	6.1	8.4	12.6	20.2	30.4	38.2	58.1
CPF + ER + HER2	4.3	5.2	6.0	8.8	12.7	19.7	29.8	38.1	56.7
CPF alone	4.7	5.5	6.4	9.1	13.1	20.4	29.7	36.3	53.5
≥50, No Adverse CPF,[a] No RT									
CPF + DS	5.8	6.1	6.3	7.9	9.1	11.4	12.6	14.7	18.5
CPF + ER + HER2	7.4	7.9	8.4	8.9	9.8	11.7	12.6	14.9	16.0
CPF alone	7.7	8.2	8.2	9.2	10.1	11.3	12.3	13.2	13.2
≥50, ≥1 Adverse CPF,[a] No RT									
CPF + DS	6.7	8.9	10.1	12.9	19.3	28.1	38.5	45.8	63.6
CPF + ER + HER2	8.8	10.0	11.4	14.4	19.3	26.7	36.7	45.6	59.4
CPF alone	9.4	10.9	12.2	14.3	19.4	26.0	35.7	43.3	58.7
<50, no RT									
CPF + DS	9.7	10.9	12.1	15.1	20.7	29.5	48.3	59.4	85.9
CPF + ER + HER2	10.7	13.3	14.0	16.2	21.5	30.7	48.1	60.3	94.9
CPF alone	11.0	13.7	14.4	16.8	21.6	31.6	44.1	63.4	84.2
≥50, No Adverse CPF,[a] RT									
CPF + DS	2.8	3.5	3.8	4.4	5.6	6.7	9.0	10.5	11.7
CPF + ER + HER2	3.5	4.1	4.6	4.8	5.3	6.1	7.2	8.5	9.7
CPF alone	3.7	4.1	4.7	4.9	5.6	6.1	6.9	7.1	7.1
≥50, ≥1 Adverse CPF,[a] RT									
CPF + DS	3.4	4.8	6.1	7.8	10.9	15.7	22.6	26.0	35.0
CPF + ER + HER2	4.3	5.5	6.1	7.5	10.8	15.4	22.1	29.8	41.8
CPF alone	5.0	6.1	6.6	8.1	10.7	14.6	20.8	26.8	36.2
<50, RT									
CPF + DS	5.1	6.8	7.3	9.9	15.0	21.7	33.4	41.0	57.3
CPF + ER + HER2	5.6	7.2	8.2	10.3	15.1	21.9	31.3	39.2	56.7
CPF alone	5.9	7.8	8.7	10.8	15.6	21.6	30.6	38.4	51.4

Abbreviations: CPF = clinicopathologic factor; DS = ductal carcinoma-in-situ score; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; RT = radiotherapy.

[a]Adverse CPF indicates lesion size > 10 mm, high nuclear grade, comedonecrosis in > 30% of ducts, observed multifocality, and positive resection margin.

compared to CPF + ER + HER2 (*c* statistic = 0.6703, likelihood ratio χ^2 = 38.1764 with *df* = 11 and *P* < .0001), and to CPF alone (*c* statistic = 0.6575, likelihood ratio χ^2 = 30.1594 with *df* = 8 and *P* < .0001) (Table 4). The superior *c* statistic and likelihood ratio chi-square value indicate that the CPF + DS best full model of IR predicts LR more accurately than the models without DS. The mean *c* statistics from the 500 bootstrapping samples for each model are also shown in Table 4, and they are concordant with the observation that the CPF + DS model provides the most accurate estimates of the 10-year risk of IR. There were no significant interactions found, and in particular no interaction between DS and RT.

Individual Predicted Risks of IR

The equation to predict an individual woman's 10-year risk of IR incorporating DS is

$$0.188 * \sqrt{\text{DCIS score}} + 0.417 * I(\text{age} < 50 \text{ years}) + (-0.527)$$
$$* I(\text{comedo necrosis invloving} > 50\% \text{ of involved ducts})$$
$$+ 0.063 * I(\text{margin size 0 to 2 mm}) + (-0.090)$$
$$* I(\text{margin size} \geq 2\text{mm})$$
$$+ (-0.502) * I(\text{received adjuvant RT}) + 0.868$$
$$* I(\text{multifocality present})$$
$$+ (-0.135) * I(\text{nuclear grade high}) + 0.208 * \ln(\text{tumor size})$$

where *I*(.) represents the indicator function and *ln*(.) represents the natural logarithm function.

The equations from the CPF + ER + HER2 model of IR as well as the CPF-alone model of IR are presented in the Supplemental Material (available in the online version).

Individual predicted 10-year risks of IR from CPF + DS ranged from 1.1% to 44.3% compared to the range 1.8% to 62.4% from CPF + ER + HER2 and the range 2.6% to 38.8% from CPF alone (data not shown). The univariate distribution of individual predicted risks of IR from the 3 models, from the 1st to 99th percentiles, are shown in Table 5, overall as well as stratified by age, presence or absence of $\geq$ 1 adverse CPF (lesion size > 10 mm, high nuclear grade, comedonecrosis, observed multifocality, positive resection margins), and whether BCS was followed by RT.

Among women aged $\geq$ 50 years at diagnosis having lesions $\leq$ 10 mm diameter and no adverse CPF, 28.6% had individual predicted 10-year risks of IR $\leq$ 5% from the CPF + DS model compared to 1.6% from the CPF + ER + HER2 model and 1.6% from the CPF-alone model. When the prediction equations were applied to similar women who had received RT, and setting the RT covariate to 0, 19.6% would have had individual predicted risks of IR $\leq$ 5% without RT from the CPF + DS model compared to 3.7% from both the CPF + ER + HER2 model and the CPF-alone model.

Very few women aged $\geq$ 50 years at diagnosis had individual predicted 10-year risks of IR $\leq$ 5% without RT if the lesion was > 10 mm or any of the following were present: multifocality, comedonecrosis > 30%, high nuclear grade, or positive resection margins (8.1% incorporating DS, 1.4% incorporating ER and HER2, and none by prediction incorporating none of the 3). Setting the RT covariate to 0 and applying the prediction equations to similar women who had received RT, 5.2% would have had individual predicted 10-year risks of IR $\leq$ 5% without RT from the CPF +

DS model compared to 2.6% from the CPF + ER + HER2 model and 0.3% from the CPF-alone model. Among women aged < 50 years at diagnosis, 3.2% had an individual predicted risk of IR < 5% without RT from the CPF + DS model but none from the other models. Setting the RT covariate to 0 and applying the equations to women aged < 50 years who had received RT, 4.1% would have had individual predicted risk of IR $\leq$ 5% without RT from the CPF + DS model, but none from the other models.

Discussion

This work has shown that incorporation of DS with clinico-pathologic features of DCIS (CPF + DS) more accurately predicts individual 10-year risks of LR and IR after BCS for DCIS compared to models based on CPF with ER and HER2 status or based on CPF alone without DS.

Treatment decision making relies on estimating an individual's risk (or probability) of developing an event weighed against the benefits of treatment. Prediction models use statistics to predict the probability of an event according to presence or absence of baseline characteristics and help quantify the risks of an event, which help clinicians quantify the benefits of treatment. More accurate prediction of recurrence risks can help clinicians target interventions to subjects at high risk of recurrence and omit treatment to those at low risk. The performance of a prediction model (ie, its ability to accurately predict outcomes) can be evaluated against observed outcomes using measures such as concordance (*c* statistic) or calibration curves. We incorporated clinical and pathologic features of DCIS (CPF) along with ER status, HER2 status, and DCIS score to determine the optimal model to most accurately predict the 10-year risk of LR and IR after BCS for DCIS. We evaluated over 500 prediction models using a well-described population cohort of women with pure DCIS, with complete systematic pathologic evaluation and long-term follow-up. We compared the performance of each prediction model against observed outcomes experienced in this cohort by calculating *c* statistics and calibration curves. We found that the model based on the DCIS score with CPF provides the most accurate prediction of the 10-year risk of LR and the 10-year risk of IR after treatment by BCS alone.

Models that best classify individuals at very low of recurrence have the greatest clinical usefulness because they serve as a guide for clinicians to omit or recommend treatment accordingly. In this regard, the CPF + DS model performed best at risk prediction within the lower 5 deciles of recurrence risk compared to the other models; that is, incorporating the DCIS score into the risk prediction model (with CPF) more accurately identified a significantly greater proportion of women at very low risk of LR and IR for whom RT would have a very small absolute benefit. Although the thresholds of 10-year LR risk $\leq$ 10% and IR $\leq$ 5% were arbitrarily selected, there is general agreement that women with such low risks of recurrence might be spared RT after BCS. Using these thresholds, we computed the percentage of women with 10-year individual predicted risk of LR $\leq$ 10% among the subgroup of women aged > 50 years at diagnosis with no adverse CPF (lesion size > 10 mm, high nuclear grade, comedonecrosis, positive resection margins, multifocality) for whom overdiagnosis and overtreatment is a significant concern. The CPF + DS model identified 62.3% of this subgroup with a low risk of LR ($\leq$ 10% at 10 years). In

Figure 1 Calibration Plots for (A) CPF + DS, (B) CPF + ER + HER2, and (C) CPF Alone
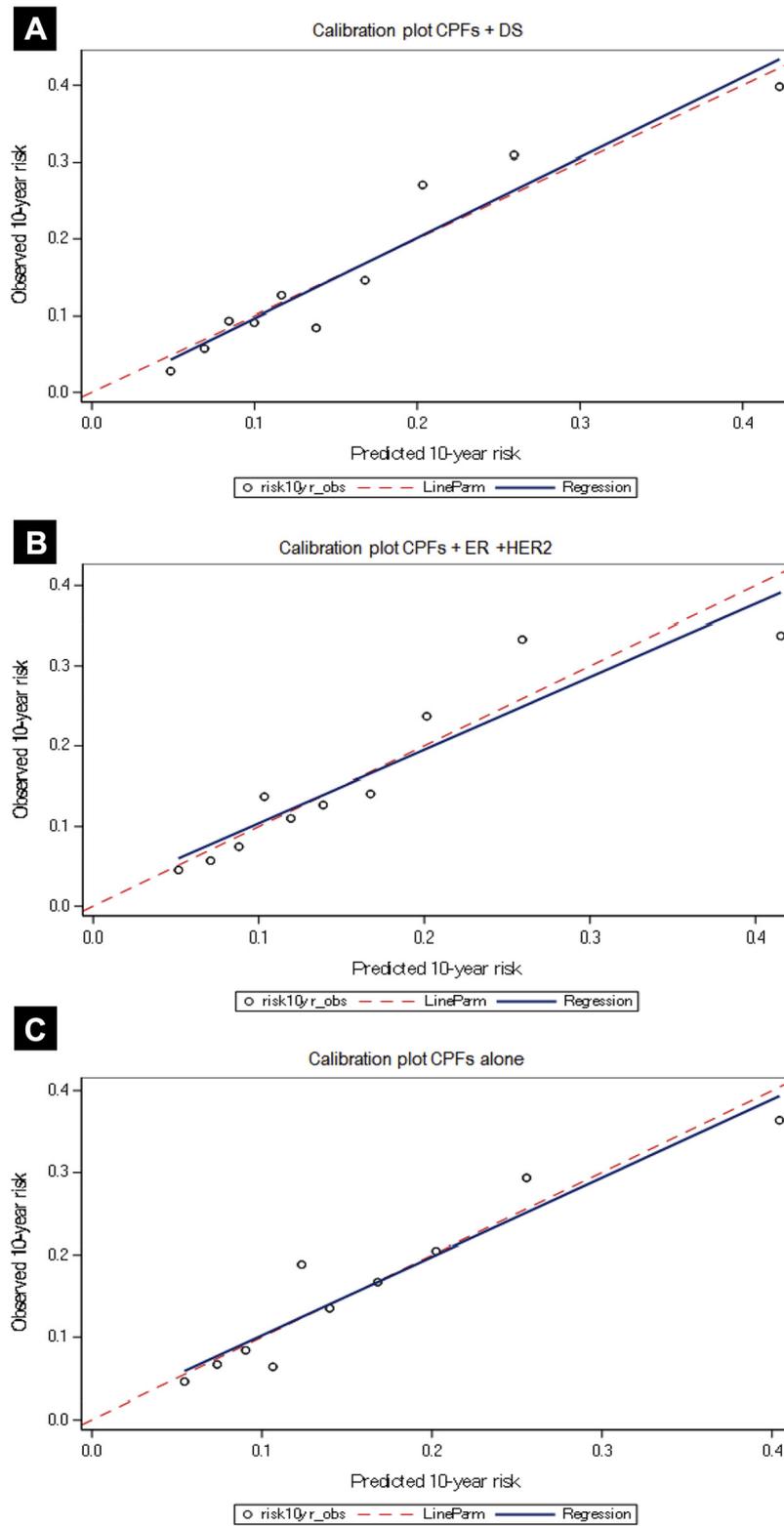
Abbreviations: CPF = clinicopathologic factor; DS = ductal carcinoma-in-situ score; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2.

Table 4 Best Full Predictive Models of Invasive Recurrence

Characteristic	Univariate Estimates and HR		Best Model Including DS (CPF + DS)		Best Model Including ER and HER2 (CPF + ER + HER2)		Best Model Without DS, ER or HER2 (CPF Alone)	
c statistic			0.6838		0.6702		0.6575	
Mean bootstrapping c statistic (min, max) among 500 samples			0.6951 (0.6257, 0.7859)		0.6863 (0.6101, 0.7494)		0.6681 (0.6048, 0.7514)	
Likelihood ratio (P by chi-square test)			43.3544, 9 df (<.0001)		38.1764, 11 df (<.0001)		30.1594, 8 df (.0002)	
	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)
Square root of DS per 1 unit increase	0.15835 (.001)	1.17 (1.06, 1.29)	0.18836 (.0007)	1.21 (1.08, 1.35)				
Estrogen Receptor								
Positive	−0.75525 (.008)	0.47 (0.27, 0.82)			−0.7239 (.0271)	0.48 (0.26, 0.92)		
Negative	Ref	Ref			Ref	Ref		
HER2 Overexpression								
Positive	0.06286 (.89)	1.06 (0.53, 2.16)			−0.24548 (.5205)	0.78 (0.37, 1.65)		
Equivocal	Ref	Ref			Ref	Ref		
Negative	−0.33523 (.3)	0.72 (0.38, 1.35)			−0.44327 (.1751)	0.64 (0.34, 1.22)		
Age, y								
<50	0.32237 (.14)	1.38 (0.90, 2.11)	0.41686 (.0575)	1.52 (0.99, 2.33)	0.41498 (.0584)	1.51 (0.99, 2.33)	0.39919 (.0684)	1.49 (0.97, 2.29)
≥50	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Multifocality								
Not observed	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Observed	0.87773 (<.0001)	2.41 (1.63, 3.55)	0.86802 (.0001)	2.38 (1.58, 3.58)	0.83727 (.0001)	2.31 (1.53, 3.48)	0.8383 (.0001)	2.31 (1.54, 3.48)
Nuclear Grade								
Low or moderate	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
High	0.21709 (.27)	1.24 (0.84, 1.83)	−0.1354 (.5411)	0.87 (0.57, 1.35)	−0.0118 (.9586)	0.99 (0.63, 1.54)	0.06995 (.7473)	1.07 (0.70, 1.64)
Comedonecrosis								
≤30% of ducts	Ref	Ref					Ref	Ref
>30% of ducts	0.21571 (.334)	1.24 (0.80, 1.92)					0.13342 (.5813)	1.14 (0.71, 1.84)
Comedonecrosis								
≤50% of ducts	Ref	Ref	Ref	Ref	Ref	Ref		
>50% of ducts	−0.18638 (.612)	0.83 (0.40, 1.71)	−0.52677 (.1678)	0.59 (0.28, 1.25)	−0.46597 (.2236)	0.63 (0.30, 1.33)		
Log of lesion size per 1 unit increase	0.36629 (.01)	1.44 (1.09, 1.91)	0.20814 (.1664)	1.23 (0.92, 1.65)	0.2307 (.1226)	1.26 (0.94, 1.69)	0.23696 (.1084)	1.27 (0.95, 1.69)
Margin width								
00 mm (positive)	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
>0 < 2 mm	−0.13473 (.65)	0.87 (0.49, 1.57)	0.06301 (.8372)	1.07 (0.58, 1.94)	0.04647 (.8796)	1.05 (0.57, 1.91)	0.05382 (.86)	1.06 (0.58, 1.92)
≥ 2 mm	−0.3864 (.196)	0.68 (0.38, 1.22)	−0.09026 (.7694)	0.91 (0.50, 1.67)	−0.14704 (.6329)	0.86 (0.47, 1.58)	−0.13632 (.6564)	0.87 (0.48, 1.59)

Table 4 Continued

	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)	Parameter Estimate (P)	HR (95% CI)
RT								
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	−0.31906 (.1)	0.73 (0.49, 1.07)	−0.50208 (.012)	0.61 (0.41, 0.90)	−0.48163 (.0164)	0.62 (0.42, 0.92)	−0.43762 (.0283)	0.65 (0.44, 0.95)

Abbreviations: CI = confidence interval; CPF = clinicopathologic factor; DS = ductal carcinoma-in-situ score; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.

comparison, the model based on CPF + ER + HER2 only identified 50.9%, and the model based on CPF alone identified only 46.5% of cases with this low risk of LR. This suggests that omitting DS from the risk prediction model would result in overtreatment of 11.3% or 15.8% of women, respectively. If somewhat higher threshold levels of LR and IR are used to recommend treatment, the proportion of women avoiding RT would further increase.

We applied the 3 prediction models to the cases in the cohort who received RT and set the RT covariate to 0 to determine the proportion of women with low predicted 10-year risks of LR ($\leq$ 10%) after BCS alone that might have avoided treatment. The CPF + DS model identified a greater proportion of women (44.4%) who had been selected to receive RT, who, on the basis of their baseline clinicopathologic features and their DCIS score, had a predicted 10-year risk of LR $\leq$ 10% without RT. In comparison, the proportion of women with a predicted low risk of LR was 39.4% from the CPF + ER + HER2 model and 32.3% from the CPF-alone model. The CPF + DS model identified more cases that potentially could have been spared RT compared to models that did not include DS. If the threshold to treat with RT is a 10-year risk of invasive LR > 5%, then 28.6% would have been spared RT using DS + CPF model compared to only 1.6% from the CPF + ER + HER2 model and 1.6% from the CPF-alone model. The absence of significant interaction between DS and RT, or among ER and HER2 and RT, indicates that neither DS nor ER + HER2 status was able to distinguish women who would experience recurrence despite RT from those who would not.

In addition, we found that the CPF + ER + HER2 and CPF-alone models underestimated individual predicted risks in a significant proportion of women. For example, among women aged $\geq$ 50 years without any adverse CPF with a predicted risk of LR > 10% by the DS + CPF model, 20.7% and 13.2% had predicted 10-year LR risks $\leq$ 10%, suggesting that the omission of DS from the prediction model could also lead to insufficient treatment if the risk of recurrence is underestimated.

We have shown that none of the 3 models is particularly well calibrated for the 5 higher deciles of individual predicted risk of LR or IR compared to their calibration for the 5 lower deciles of risk. This is expected because the distribution of individual predicted risks from each of the 3 predictive models is much more dispersed for the upper 5 deciles of individual predicted risks compared to the lower 5 deciles, as shown in the calibration plots and in Tables 3 and 5. Additional data of women with higher risks of LR and IR would be needed to refine higher risk estimates in order to determine where more aggressive treatment, such as higher doses of RT, antiestrogen therapy, or mastectomy, would be recommended.

The generalizability of these risk prediction equations to women who have undergone BCS is supported by the representation of the full range of CPF in this study population (Table 1). However, they are not generalizable to any women who have had biopsy without excision of DCIS, and because adjuvant endocrine therapy was infrequently prescribed in this population, the equations do not account for any effect of such therapy on LR and IR. These risk prediction equations have been internally validated but lack external validation; we encourage clinical investigators who have substantial case series of women who have had BCS for DCIS, and for whom all CPF and DS are known, to test them in these series.

Table 5

| Table 5 | Univariate Distributions of Individual Predicted Risks (%) of Invasive Recurrence by Subcohort |

Characteristic	1st Percentile	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	99th Percentile
Overall									
CPF + DS	2.1	2.9	3.6	5.3	7.6	11.8	17.6	22.1	31.6
CPF + ER + HER2	2.9	3.8	4.3	5.6	7.8	11.6	17.1	20.9	31.1
CPF alone	3.6	4.1	4.7	5.9	7.9	11.7	17.4	20.2	25.6
≥50, No Adverse Factors, No RT									
CPF + DS	2.3	2.5	3.0	4.6	6.2	8.4	9.9	10.8	12.4
CPF + ER + HER2	4.6	5.1	5.4	5.7	6.6	7.8	9.4	10.5	12.2
CPF alone	4.7	5.2	5.2	5.9	6.3	7.0	7.8	8.2	8.2
≥50, ≥1 Adverse Factor, No RT									
CPF + DS	2.6	4.5	5.3	7.0	10.1	15.9	23.6	27.3	31.6
CPF + ER + HER2	4.8	5.6	6.2	7.3	9.3	16.0	20.3	22.4	32.6
CPF alone	5.9	6.8	7.3	8.0	9.5	16.8	20.2	22.2	24.9
<50 No RT									
CPF + DS	3.5	5.7	6.8	8.3	11.4	16.1	21.5	37.3	44.3
CPF + ER + HER2	5.9	8.1	8.3	9.1	11.3	14.4	24.2	35.5	62.4
CPF alone	6.0	8.2	8.6	9.8	11.9	14.2	25.4	31.9	38.8
≥50, No Adverse Factors, RT									
CPF + DS	1.3	1.7	2.5	3.3	4.5	5.6	7.2	8.2	8.9
CPF + ER + HER2	2.5	3.2	3.5	3.8	4.2	4.8	5.5	6.6	7.7
CPF alone	2.6	3.1	3.6	3.8	4.3	4.6	5.3	5.3	5.3
≥50, ≥1 Adverse Factor, RT									
CPF + DS	2.2	3.0	3.7	4.9	6.9	10.4	15.2	16.8	20.8
CPF + ER + HER2	2.4	3.5	4.1	4.8	6.4	11.0	14.0	16.4	26.1
CPF alone	3.8	4.4	4.7	5.3	6.3	10.9	14.1	15.7	17.7
<50 RT									
CPF + DS	2.7	3.6	4.9	6.6	9.3	13.2	20.2	26.7	32.9
CPF + ER + HER2	3.7	5.0	5.5	6.8	8.6	14.1	17.7	23.0	38.8
CPF alone	3.9	5.6	6.1	7.3	8.8	14.6	20.2	22.3	25.9

Abbreviations: CPF = clinicopathologic factor; df = degree of freedom; DS = ductal carcinoma-in-situ score; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; RT = radiotherapy.

Conclusion

Individual predicted 10-year risks of recurrence of DCIS after BCS LR incorporating DS + CPF are more accurate and identify a higher proportion of women with low predicted risk of LR and invasive LR after BCS alone for whom RT may be omitted, compared to risk prediction incorporating ER and HER2 or CPF alone.

Clinical Practice Points

- Although survival after BCS for DCIS is excellent, LR may be substantial and may vary with CPF.
- Individual risk prediction incorporating DS along with CPF facilitates the computation of more accurate individual risks of LR than predictive models incorporating CPF alone or with ER and HER2, and identifies a greater proportion of women whose risk of LR is ≤ 10% at 10 years after BCS without RT.
- Accurate identification of a larger number of women with predicted risk of LR ≤ 10% may allow a larger number of women to avoid RT after BCS for DCIS.

Acknowledgments

Disclosure

E.R. reports having received research funds from Genomic Health Inc. The other authors have stated that they have no conflict of interest.

Supplemental Data

Supplemental material comprising equations and tables accompanying this article can be found in the online version https://doi.org/10.1016/j.clbc.2018.07.018.

References

1. Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson C. Incidence of and treatment for ductal carcinoma in situ of the breast. *JAMA* 1996; 275:913-8.
2. Ernster VL, Barclay J, Kerlikowske K, Wilkie H, Ballard-Barbash R. Mortality among women with ductal carcinoma in situ of the breast in the population-based Surveillance, Epidemiology and End Results Program. *Arch Intern Med* 2000; 160: 953-8.
3. Gøtzsche PC. Re: Trends in the treatment of ductal carcinoma in situ of the breast. *J Natl Cancer Inst* 2004; 96:1257.
4. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. *Breast Cancer Res* 2006; 8:204.
5. Kumar AS, Bhatia V, Henderson IC. Overdiagnosis and overtreatment of breast cancer: rates of ductal carcinoma in situ: a US perspective. *Breast Cancer Res* 2005; 7:271-5.
6. Hosseini A, Khoury AL, Esserman LJ. Precision surgery and avoiding over-treatment. *Eur J Surg Oncol* 2017; 43:938-43.
7. Wapnir IL, Dignam JJ, Fisher B, et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst* 2011; 103:478-88.
8. Cuzick J, Sestak I, Pinder SE, et al. Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: long-term results from the UK/ANZ DCIS trial. *Lancet Oncol* 2011; 12:21-9.
9. Donker M, Litière S, Werutsky G, et al. Breast-conserving treatment with or without radiotherapy in ductal carcinoma in situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial. *J Clin Oncol* 2013; 31:4054-9.
10. Wärnberg F, Garmo H, Emdin S, et al. Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS Trial. *J Clin Oncol* 2014; 32:3613-8.
11. McCormick B, Winter K, Hudis C, et al. RTOG 9804: a prospective randomized trial for good-risk ductal carcinoma in situ comparing radiotherapy with observation. *J Clin Oncol* 2015; 33:709-15.
12. Goodwin A, Parker S, Ghersi D, Wilcken N. Post-operative radiotherapy for ductal carcinoma in situ of the breast. *Cochrane Database Syst Rev* 2013; 11: CD000563.
13. EBCTCG (Early Breast Cancer Trialists' Collaborative Group), McGale P, Taylor C, Correa C, et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet* 2014; 383: 2127-35.
14. Solin LJ, Gray R, Baehner FL, et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *J Natl Cancer Inst* 2013; 105:701-10.
15. Rakovitch E, Nofech-Mozes S, Hanna W, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Res Treat* 2015; 152:389-98.
16. Rakovitch E, Nofech-Mozes S, Hanna W, et al. Multigene expression assay and benefit of radiotherapy after breast conservation in ductal carcinoma in situ. *J Natl Cancer Inst* 2017 Apr 1; 109(4). https://doi.org/10.1093/jnci/djw256.
17. Rakovitch E, Nofech-Mozes S, Narod SA, et al. Can we select individuals with low risk ductal carcinoma in situ (DCIS)? A population-based outcomes analysis. *Breast Cancer Res Treat* 2013; 138:581-90.
18. Yi M, Meric-Bernstam F, Kuerer HM, et al. Evaluation of a breast cancer nomogram for predicting risk of ipsilateral breast tumor recurrences in patients with ductal carcinoma in situ after local excision. *J Clin Oncol* 2012; 30: 600-7.