This is a PREPRINT version of: Groen EJ, Hudecek J, Mulder L, et al. Prognostic value of histopathological DCIS features in a large-scale international interrater reliability study. Breast Cancer Res Treat. 2020 Jul 30. doi: 10.1007/s10549-020-05816-x

Prognostic value of histopathological DCIS features in a large-scale international interrater reliability study

Emma J. Groen MD[1], Jan Hudecek MSc[2], Lennart Mulder BASc[3], Maartje van Seijen MD[3], Mathilde M. Almekinders MD[3], Stoyan Alexov MD[4], Anikó Kovács MD[5], Ales Ryska MD[6], Zsuzsanna Varga MD[7], Francisco-Javier Andreu Navarro MD[8], Simonetta Bianchi MD[9], Willem Vreuls MD[10], Eva Balslev MD[11], Max V. Boot MD[12], Janina Kulka MD[13], Ewa Chmielik MD[14], Ellis Barbé MD[12], Mathilda J. de Rooij MD[15], Winand Vos MD[16], Andrea Farkas MD[17], Natalja E. Leeuwis-Fedorovich MD[18], Peter Regitnig MD[19], Pieter J Westenend MD[20], Loes F.S. Kooreman MD[21], Cecily Quinn MD[22], Giuseppe Floris MD[23], Gábor Cserni MD[24], Paul J. van Diest MD[25], Esther H. Lips PhD[3], Michael Schaapveld PhD[26]*, Jelle Wesseling MD[1,3]* also on behalf of the Grand Challenge PRECISION consortium   ***Joint last authors**

[1] Department of Pathology, Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, the Netherlands

[2] Department of Research IT, Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, the Netherlands

[3] Department of Molecular Pathology, Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, the Netherlands

[4] Department of Pathology, Oncology Hospital, Sofia, Bulgaria

[5] Department of Clinical Pathology, Sahlgrenska University Hospital, Gothenburg, Sweden

[6] The Fingerland Department of Pathology, Charles University Medical Faculty, University Hospital, Hradec Kralove, Czech Republic

[7] Institute of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland

[8] Atryshealth Co, S.L., Barcelona, Spain

[9] Division of Pathological Anatomy, Department of Health Sciences, University of Florence, Florence, Italy

[10] Department of Pathology, Canisius Wilhelmina Hospital, Nijmegen, the Netherlands

[11] Department of Pathology, Herlev University Hospital, Herlev, Denmark

[12] Department of Pathology, Amsterdam University Medical Center, location VUmc, Amsterdam, the Netherlands

[13] 2nd Department of Pathology, Semmelweis University, Budapest, Hungary

[14] Tumor Pathology Department, Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice Branch, Gliwice, Poland

[15] Symbiant Pathology Expert Centre, location ZMC, Zaandam, the Netherlands

[16] Department of Pathology, Zuyderland Medical Center, location Sittard-Geleen, Sittard-Geleen, the Netherlands

[17] Department of Pathology, Gävle Hospital, Gävle, Sweden

[18] Department of Pathology, Deventer Hospital, Deventer, the Netherlands

[19] Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria

[20] Laboratory for Pathology Dordrecht, Dordrecht, the Netherlands

[21] Department of Pathology and GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, the Netherlands.

[22] Department of Pathology and Laboratory Medicine, St. Vincent's University Hospital, Dublin, Ireland

[23] Department of Imaging and Pathology, Laboratory of Translational Cell & Tissue Research, KU Leuven - University of Leuven, Leuven, Belgium; Department of Pathology, University Hospitals Leuven, Leuven, Belgium

[24] Department of Pathology, Bács-Kiskun County Teaching Hospital, Kecskemét, Hungary; Department of Pathology, University of Szeged, Szeged, Hungary

[25] Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands

[26] Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, the Netherlands

<u>Corresponding author</u>

Jelle Wesseling

Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Email: <u>j.wesseling@nki.nl</u>; Telephone: +31 20 5122778

Grand Challenge PRECISION consortium

Email: <u>precision@nki.nl</u>

Abstract

Purpose: For optimal management of ductal carcinoma in situ (DCIS), reproducible histopathological assessment is essential to distinguish low-risk from high-risk DCIS. Therefore, we analyzed interrater reliability of histopathological DCIS features and assessed their associations with subsequent ipsilateral invasive breast cancer (iIBC) risk.

Methods: Using a case-cohort design, reliability was assessed in a population-based, nation-wide cohort of 2,767 women with screen-detected DCIS diagnosed between 1993-2004, treated by breast conserving surgery with/without radiotherapy (BCS+/-RT) using Krippendorff's alpha (KA) and Gwet's AC2 (GAC2). Thirty-eight raters scored histopathological DCIS features including grade (2-tiered and 3-tiered), growth pattern, mitotic activity, periductal fibrosis and lymphocytic infiltrate in 342 women. Using majority opinion-based scores for each feature, their association with subsequent iIBC-risk was assessed using Cox regression.

Results: Interrater reliability of grade using various classifications was fair to moderate, and only substantial for grade 1 versus 2+3 when using GAC2 (0.78). Reliability for growth pattern (KA 0.44, GAC2 0.78), calcifications (KA 0.49, GAC2 0.70) and necrosis (KA 0.47, GAC2 0.70) was moderate using KA and substantial using GAC2; for (type of) periductal fibrosis and lymphocytic infiltrate fair to moderate estimates were found and for mitotic activity reliability was substantial using GAC2 (0.70). Only in patients treated with BCS-RT, high mitotic activity was associated with a higher iIBC-risk in univariable analysis (Hazard Ratio (HR) 2.53, 95% Confidence Interval (95%CI) 1.05-6.11); grade 3 versus 1+2 (HR 2.64, 95%CI 1.35-5.14) and a cribriform/solid versus flat epithelial atypia/clinging/(micro)papillary growth pattern (HR 3.70, 95%CI 1.34-10.23) were independently associated with a higher iIBC-risk.

Conclusions: Using majority opinion-based scores, DCIS grade, growth pattern and mitotic activity are associated with iIBC-risk in patients treated with BCS-RT, but interrater variability is substantial. Semi-quantitative grading, incorporating and separately evaluating nuclear pleomorphism, growth pattern and mitotic activity, may improve the reliability and prognostic value of these features.

4

Keywords

ductal carcinoma in situ, invasive breast cancer, interrater reliability, risk stratification

List of abbreviations

DCIS = ductal carcinoma in situ

IBC = invasive breast cancer

iIBC = ipsilateral invasive breast cancer

95%CI = 95% Confidence Interval

BCS+/-RT = breast conserving surgery with or without radiotherapy

NCR = Netherlands Cancer Registry

PALGA = the nationwide network and registry of histology and cytopathology in the Netherlands.

KA = Krippendorff's alpha

GAC2 = Gwet's AC2

PHA = proportional hazard assumptions

HR = Hazard Ratio

FEA = flat epithelial atypia

HER2 = human epidermal growth factor receptor 2

ER = estrogen receptor

5

Background

Ductal carcinoma in situ (DCIS) of the breast is a non-obligate precursor of invasive breast cancer (IBC). Since the introduction of organized population-based breast screening, the incidence of DCIS has increased manyfold [1–3]. Although DCIS is almost always treated to avoid progression to IBC, this has not led to a reduced IBC incidence. Breast screening programs are therefore criticized by some for being associated with overdiagnosis and overtreatment of DCIS [4–6]. It has been reported that a large proportion of untreated DCIS will not progress to IBC [7,8]. Ryser et al. reported a 10-year net risk of ipsilateral IBC (iIBC) of 12.2% (95% Confidence Interval (95%CI) 8.6-17.1%) for women with DCIS grade 1/2 and 17.6% (95%CI 12.1-25.2%) for grade 3 [8]. Although based on selected patients, these results underline that at least some DCIS lesions have a low risk of progression and may thus be overtreated. However, reliably distinguishing high from low risk DCIS to guide treatment is still challenging.

Many studies have tried to find histopathological markers that could predict progression of DCIS [9,10]. So far, no single marker ended up being used in clinical practice due to lack of conclusive evidence of predictive ability, in part due to suboptimal biased study designs in particular due to insufficient handling of confounders and poorly described study groups [10]. Especially grade has been extensively studied as a biomarker for the invasive potential of DCIS. The use of many different grading systems with partly unclear criteria and often only poor to modest interrater reliability makes it difficult to evaluate the role of grade in risk stratification [11–21].

In addition, various studies have assessed reproducibility of histopathological evaluation of DCIS lesions. Unfortunately, these studies were frequently based on highly selected case sets, assessed by expert breast pathologists often after having received instructions or tutorials beforehand and using reference diagnoses without follow-up data [17,18,22–28]. The interpretation of results and evaluation of potential bias is further complicated by inadequate reporting [29].

This study assesses the interrater reliability of various histopathological features in DCIS in a setting which as closely as possible reflects daily practice. We subsequently evaluate whether these

6

features, based on a more robust majority opinion of 38 raters, are associated with risk of development

of subsequent iIBC.

Methods

Patient selection

We assembled a population-based, nation-wide cohort of screen-detected primary and pure DCIS, treated with breast conserving surgery with or without adjuvant radiotherapy (BCS+/-RT) between January 1st 1993 and December 31st 2004, by linkage of data from the Netherlands Cancer Registry (NCR) with data from the Dutch breast cancer screening program [30]. From 1989, the Dutch biennial screening program was gradually introduced, inviting women aged 50-69 years and from 1998 aged 50-75 years. Screen-detected DCIS was defined as DCIS detected within 30 months after a first or subsequent positive screening examination. The cohort was supplemented with data from the nationwide network and registry of histology and cytopathology in the Netherlands (PALGA) [31]. Information on age and date at diagnosis, treatment, and if applicable subsequent iIBC and vital status was provided by the NCR (follow-up data available until January 1, 2011). Patients diagnosed with a prior malignancy, other than non-melanoma skin cancer, were excluded. The review boards of the NCR, PALGA and the Dutch breast cancer screening organization approved this study.

Interrater reliability analysis

We first assessed the interrater reliability of histopathological DCIS features in this cohort using a case-cohort design [32]. From the cohort of 2,767 women, we randomly sampled 357 women (subcohort; 13%) and additionally selected all 177 patients who subsequently developed an iIBC but were not included in the random sample for a total of 534 patients. Fig. 1 shows the selection of patients with exclusions at pathology report review (n = 27) and slide review (n = 76). Slide review was based on freshly cut slides stained with hematoxylin and eosin and in case of uncertainty about the in-situ nature of the lesion also with cytokeratin 14 by EJG (clone LL002; 1/3200 dilution, 32 minutes at 37^0C + amplification, Neomarkers / Thermo Scientific).

For 353 patients the diagnosis of pure DCIS could be confirmed and from each lesion a single slide was selected with the highest quantity of DCIS. These slides were digitized using an Aperio AT2

8

scanner (Leica Biosystems) at 20x magnification and uploaded on an online viewing platform (https://www.slidescore.com/). For each DCIS lesion a scoring form (Supplementary methods) was built-in with the items: DCIS diagnosis, grade (2- and 3-tiered), growth pattern and mitotic activity of DCIS, calcifications, necrosis, periductal fibrosis and lymphocytic infiltrate.

European raters with varying expertise were invited to participate in the study. Each rater was assigned a study set of 146 cases to score independently, blinded to subject information. Raters were not given instructions regarding the (interpretation of) histopathological features and were requested to score as they would in daily practice to provide an unbiased baseline measure of reliability. Further details on rater selection, participation and the scoring process are described in Supplementary methods.

Statistical analysis

In total 11 patients were excluded from reliability analysis because >50% of raters considered their lesion as no DCIS /not assessable (n = 5) or >25% commented on suboptimal slide quality (n = 6). If DCIS was not confirmed, any scores for following histopathological features were ignored. Scores for type of fibrosis were only considered when periductal fibrosis was present according to the majority opinion. Raters were excluded for the analysis of single histopathological features when they scored an item as 'not assessable' in >50% of their study set.

Krippendorff's alpha (KA), Gwet's AC2 (GAC2) and percentage agreement were calculated to assess interrater reliability ('not assessable' scores were excluded) [33,34]. KA and GAC2 are applicable to studies involving nominal/ordinal data and multiple raters scoring different subsets. A weighted analysis using linear weights was used for ordinal variables with >2 categories. Interpretation was performed according to Landis and Koch [35]. Recategorization of grade, periductal fibrosis, and lymphocytic infiltrate was undertaken during analysis to evaluate reliability using different cut-offs.

9

For the analysis of subsequent iIBC risk an additional 10 patients were excluded, because >25% of the raters considered IBC to be present adjacent to DCIS (n=8) or because the patient underwent a mastectomy before developing iIBC (n=2). For a detailed comparison of clinical characteristics between in- versus excluded patients see Supplementary Table 1.

Associations of histopathological features, treatment, age at diagnosis and period of diagnosis (1993-1998, reflecting the screening implementation phase, versus 1999-2004, reflecting full nationwide coverage) with risk of iIBC was assessed using Cox models. Analyses were performed irrespective of treatment as well as separately for BCS alone and BCS+RT. Interactions with treatment were also considered. Proportional hazard assumptions (PHA) were tested using residual-based and graphical methods. In case the PHA was violated, a time factor was added, and the associations were estimated for different time-periods (i.e. for the first 5 years and after 5 years). For the histopathological features the majority opinion, i.e. the most frequently assigned category, was used in the analysis ('not assessable' scores were excluded). In case of equal frequencies, the presence of a histopathological feature was chosen over absence, the highest grade, the most complex growth pattern (i.e. cribriform/solid), many over sparse mitoses, prominent over subtle presence for periductal fibrosis and lymphocytic infiltrate and the least common type of fibrosis (i.e. myxoid). Clinicopathological factors were entered in multivariable models including treatment, based on a P value ≤0.15 in univariable analyses. Barlow's inverse probability weights were used to adjust the partial likelihood function for case-cohort analysis with robust variance estimation [32]. Fit of non-nested models was compared using Akaike's and Bayesian information criteria. Two-sided P values ≤0.05 were considered statistically significant. All statistical analyses were performed using Stata/SE (version 13.1, Statacorp).

Results

Interrater reliability

The mean number of scores per slide was 14 (range 12-15) (Supplementary Table 2). The raters consisted of a mixed group (Supplementary Table 3), about half of them working in the Netherlands and half in other European countries within a wide range of laboratories regarding size and degree of specialization. Forty-seven percent of raters were members of the European Working Group of Breast Screening Pathologists. The diagnosis of DCIS was confirmed in 98.6% of the patients based on the majority opinion.

The interrater reliability for the 3-tiered grading system (grade 1, 2 or 3), the most commonly used histopathological feature, was only fair (KA 0.34; 95%CI 0.30-0.39) to moderate (GAC2 0.52; 95%CI 0.50-0.55; Table 1). Using a 2-tiered grading system (either low versus high grade or grade 1+2 versus grade 3) did not improve reliability. When the 3-tiered grading was recategorized into a category for grade 1 and a category for grade 2+3 combined, the reliability was substantial using GAC2 (0.78; 95%CI 0.74-0.82).

Comparable moderate (KA) to substantial (GAC2) reliability was found for growth pattern, necrosis and calcifications, which are all features assessed in daily practice within the context of DCIS. A striking discrepancy in reliability was found for the assessment of mitotic activity with only fair reliability when considering KA (0.24) but substantial reliability based on GAC2 (0.70). In a 3-tiered system (absent, subtle or prominent presence) lymphocytic infiltrate showed moderate reliability, which was slightly better than the interrater reliability for periductal fibrosis. Recategorization, comparing periductal fibrosis presence with absence led to a moderate reliability (GAC2 0.53).

Risk of subsequent iIBC after DCIS

Subcohort patients were diagnosed with DCIS at a median age of 58.4 (interquartile range 53.4-64.0) and treated by BCS alone in 40.5% (87 patients) and by BCS+RT in 59.5% (128 patients). After a median follow-up of 11.2 years (interquartile range 8.6-14.1), 20 patients developed an iIBC in

11

the subcohort. DCIS was assigned grade 1 in 10.7%, grade 2 in 53.5% and grade 3 in 35.8%, based on the majority opinion. Median time to iIBC was 5.3 years (interquartile range 3.3-7.6 years). Table 2 shows clinicopathological characteristics of the subcohort and of all patients who developed an iIBC.

In univariable analysis, patients treated with BCS alone had a much higher risk of iIBC than patients treated with BCS+RT with a Hazard Ratio (HR) of 4.80 (95%CI 2.49-9.24) in the first 5 years and a HR of 2.47 after 5 years (95%CI 1.42-4.30; Supplementary Table 4). In patients treated with BCS alone, grade 3 (versus grade 1+2 combined), a cribriform/solid growth pattern (versus flat epithelial atypia (FEA), clinging and (micro)papillary growth pattern) and mitotically active DCIS (versus DCIS with low mitotic activity) was also associated with a higher iIBC risk, whereas in patients treated with BCS+RT these associations were not found. In univariable analysis, a significant interaction with treatment was found for grade 3 versus 1+2 (P=0.028) and for growth pattern (P=0.023).

In multivariable analysis a model which, besides treatment, included grade 3 versus grade 1+2 and growth pattern (cribriform and solid versus FEA, clinging and (micro)papillary) best predicted the risk of developing iIBC in patients treated with BCS alone, while grade and growth pattern were not associated with iIBC risk in patients treated with BCS+RT (Table 3). The risk of developing iIBC did not differ between patients with DCIS grade 1/2 and FEA, clinging or (micro)papillary growth pattern who were treated with BCS alone or BCS+RT. Fig. 2 shows cumulative risk of iIBC based on categories derived from this model.

12

Discussion

To the best of our knowledge, this is the first study combining a comprehensive interrater reliability study in DCIS, reflecting daily practice as closely as possible, with an analysis of iIBC risk based on the majority opinion of a large group of raters. This approach minimizes the muddling effect of interrater variability and subjectivity on the evaluation of the prognostic value of histopathological features.

In univariable analysis, patients treated with radiotherapy after BCS had a strongly reduced risk of iIBC compared to those treated by BCS alone, as was already shown previously [30,36,37]. Also grade 3 (versus grade 1+2 combined), a high mitotic activity and a cribriform/solid growth pattern (versus FEA, clinging or (micro)papillary growth pattern) were associated with increased iIBC risk in patients treated with BCS alone. In multivariable analysis however, only grade 3 (versus grade 1+2) and a cribriform/solid growth pattern were independently associated with an increased iIBC risk. Mitotic activity did not add any predictive value to grade 3 versus 1+2 and growth pattern in a multivariable model, though this is likely due to collinearity with grade. Another important finding in our study is that no histopathological features were associated with iIBC risk in the patients treated with BCS+RT. Although women in our study were not randomized for treatment arm, this finding may suggest that radiotherapy neutralizes the effect of these classical histopathological features. This is also in line with the fact that within the large randomized controlled trials of RT in DCIS no subgroup could be identified without RT benefit [36].

So far, grade is the sole histopathological feature in DCIS that is used in clinical practice and also has an impact on eligibility in the context of clinical trials investigating the safety of active surveillance in low risk DCIS [38–40]. In general, only women over the age of 45 or 50 with screen-detected calcifications associated with DCIS grade 1 or grade 2 are eligible in these trials. A three-tiered grading system is used for this selection purpose. Our study supports the rationale to distinguish between grade 1+2 versus grade 3 as DCIS grade 3 is independently associated with an increased risk of iIBC in patients treated with BCS alone. Unfortunately, the interrater reliability of assessing grade

13

using either a 3-tiered grading system (grade 1, 2 or 3) or a 2-tiered system differentiating grade 1+2 combined versus grade 3 was only fair when considering KA and at best moderate based on the GAC2.

The interrater reliability for growth pattern was moderate (KA) to substantial (GAC2). The predictive ability of grade and growth pattern has been intensively studied previously, with conflicting results [10]. Factors such as substantial interrater variability, grading system used, bias in designs and relying on histopathological assessments of a single pathologist's opinion may have resulted in these different findings [10]. Interrater reliability based on GAC2 was higher overall, when histopathological features showed strongly skewed distribution and when agreement was already very high (i.e. grade 1 versus 2+3, growth pattern and mitotic activity). Under these circumstances a GAC2 test may result in more accurate reliability coefficients, as was previously shown in comparison with Cohen's kappa, which overestimates the concordance attributed due to chance alone in these situations leading to lower reliability coefficients [41].

In view of the prognostic value and interrater reliability observed in our study, it is questionable whether it is safe to base clinical treatment decisions solely on the assessment of classical histopathological features. Here, we propose four strategies that may improve risk stratification in DCIS.

Within the context of DCIS the three features with reasonable prognostic value (grade 1+2 versus 3, growth pattern and mitotic activity) are currently used in many grading systems, but without clear definitions and rules about how to value each feature. We therefore firstly would suggest to objectify histological grading by using a numerical semi-quantitative scoring system which separately evaluates each of these features, analogous to the modified Bloom and Richardson grading system for IBC [42,43]. Dichotomous scoring systems may further improve reliability and prognostic value and should be further explored evaluating different cut-offs [44,45].

Secondly, performing additional immunohistochemistry to assign specific DCIS profiles may add prognostic value, possibly only in subsets of patients (i.e. grade 2). Previously, associations were reported of human epidermal growth factor receptor 2 (HER2)-positive, estrogen receptor (ER)-

14

negative DCIS and DCIS with high cyclooxygenase 2, p16 and Ki-67 levels with increased iIBC risk [9,10,46,47]. These markers would be good candidates for further exploration. Automated scoring within this context may result in more standardized and objective assessment [48–51]. Previously, a 3-tiered grading system in DCIS, combining nuclear grade according to the Van Nuys criteria with automated Ki-67 count, was reported to show excellent correlation with immunohistochemical markers of reported biological relevance such as ER and HER2 [9,46,47,50].

Thirdly, alternative approaches using pathology information such as artificial intelligence-based methods should also be considered in search for clinically relevant biomarkers in DCIS [52]. Recently, others have developed a whole slide image-based machine learning model, which accurately predicted the risk of an invasive or in situ recurrence and significantly outperformed traditional clinicopathological variables [53].

Lastly, besides pathology, other criteria could also be incorporated in clinical decision schemes, e.g. as in current active surveillance trials requiring DCIS to be screen-detected based on calcifications only without clinical symptoms and diagnosed on representative vacuum-assisted biopsies [38–40].

Our study had several limitations. From our study population each rater scored a different subset of patients. Therefore, we were not able to analyze the association of histopathological DCIS features with iIBC risk per rater or grading system used and to study the effect of interrater variability on risk stratification. However, the resulting immense workload would probably have caused major rater-dropout. Also tissue slides were digitally assessed using research technology producing images of somewhat lower resolution. This may have led to difficulty of assessing histopathological features requiring great detail, such as mitotic activity. Our reliability study was nonetheless performed under conditions as close as possible to clinical practice, as a large set of non-selected DCIS cases from a population-based cohort were reviewed by a large group of raters with varying levels of expertise without provision of instructions or tutorials beforehand. And lastly, data on margin status and DCIS lesion size, factors potentially associated with the risk of iIBC, was not available [10,46,47,54].

15

Conclusions

We evaluated the prognostic value of histopathological DCIS features to inform risk stratification using a unique, combined approach. Our study showed substantial interrater variability in the classification of histopathological DCIS features, while using rater majority opinions, minimizing the muddling effect of interrater variability, DCIS grade, growth pattern and mitotic activity were associated with the risk of subsequent ipsilateral invasive breast cancer after DCIS in patients treated with BCS without radiotherapy. A semi-quantitative grading system incorporating and separately evaluating nuclear pleomorphism, growth pattern and mitotic activity, analogue to IBC grading, may improve the reliability and prognostic value of these histopathological features.

Declarations

Funding

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethics approval

This retrospective study involving human participants was in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the review boards of the NKI-AVL, the Netherlands Cancer Registry, the nationwide network and registry of histology and cytopathology in the Netherlands and the Dutch breast cancer screening organization.

Consent to participate

The study used only unidentifiable patient information, and no informed consent was required.

Consent to publish

Not applicable.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request. Requests should be made to Prof. J. Wesseling: j.wesseling@nki.nl.

Authors' contributions

EJG, EHL, MS and JW were responsible for the study design. EJG coordinated the study. LM provided technical support. EJG revised all slides. JH provided an online platform to enable pathology scoring.

17

References

1.   Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in Situ of the breast: A systematic review of incidence, treatment, and outcomes. J Natl Cancer Inst. 2010;

2.   Netherlands Comprehensive Cancer Organisation (IKNL) [Internet]. Available from: http://www.cijfersoverkanker.nl

3.   Cancer Research UK [Internet]. Available from: https://www.cancerresearchuk.org/

4.   Ripping TM, Verbeek ALM, Fracheboud J, De Koning HJ, Van Ravesteyn NT, Broeders MJM. Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in the Netherlands. Int J Cancer. 2015;

5.   Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast cancer screening, incidence, and mortality across US counties. JAMA Intern Med. 2015;

6.   van Luijt PA, Heijnsdijk EAM, Fracheboud J, Overbeek LIH, Broeders MJM, Wesseling J, et al. The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Res. 2016;

7.   Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: A review. Breast Cancer Research and Treatment. 2006.

8.   Ryser MD, Weaver DL, Zhao F, Worni M, Grimm LJ, Gulati R, et al. Cancer Outcomes in DCIS Patients Without Locoregional Treatment. J Natl Cancer Inst. 2019;111(9):952–60.

9.   Lari SA, Kuerer HM. Biological markers in DCIS and risk of breast recurrence: A systematic review. Journal of Cancer. 2011.

10.   Visser LL, Groen EJ, Van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an invasive breast cancer recurrence after DCIS: A Systematic Review and Meta-analyses. Cancer Epidemiol Biomarkers Prev. 2019;28(5):835–45.

11.   Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, Van de Vijver MJ, et al. Ductal carcinoma in situ: A proposal for a new classification. Semin Diagn Pathol. 1994;11(3):167–80.

19

12. Pinder SE, Duggan C, Ellis IO, Cuzick J, Forbes JF, Bishop H, et al. A new pathological system for grading DCIS with improved prediction of local recurrence: Results from the UKCCCR/ANZ DCIS trial. Br J Cancer. 2010;

13. Cserni G, Sejben A. Grading Ductal Carcinoma In Situ (DCIS) of the Breast – What's Wrong with It? Pathol Oncol Res [Internet]. 2019; Available from: http://link.springer.com/article/10.1007/s12253-019-00760-8?utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound

14. Lagios MD. Duct carcinoma in situ. Pathology and treatment. Surg Clin North Am. 1990;70(4):873–83.

15. Silverstein MJ, Poller DN, Waisman JR, Colburn WJ, Barth A, Gierson ED, et al. Prognostic classification of breast ductal carcinoma-in-situ. Lancet. 1995;345(8958):1154–7.

16. Sloane JP, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, et al. Consistency achieved by 23 European pathologists in categorizing ductal carcinoma in situ of the breast using five classifications. European Commission Working Group on Breast Screening Pathology. Hum Pathol. 1998;29(10):1056–62.

17. Wells WA, Carney PA, Eliassen MS, Grove MR, Tosteson ANA. Pathologists' agreement with experts and reproducibility of breast ductal carcinoma-in-situ classification schemes. Am J Surg Pathol. 2000;

18. Bethwaite P, Smith N, Delahunt B, Kenwright D. Reproducibility of new classification schemes for the pathology of ductal carcinoma in situ of the breast. J Clin Pathol. 1998;51:450–4.

19. Lakhani SR, Ellis. I.O., Schnitt SJ, Tan PH, van de Vijver MJ. WHO classification of tumours of the breast. 4th ed. Lyon: International Agency for Research on Cancer; 2012.

20. College of American pathologists [Internet]. Available from: https://documents.cap.org/protocols/cp-breast-dcis-18protocol-4100.pdf

21. Poller DN, Silverstein MJ, Galea M, Locker AP, Elston CW, Blamey RW, et al. Ideas in pathology.

20

Ductal carcinoma in situ of the breast: a proposal for a new simplified histological classification association between cellular proliferation and c-erbB-2 protein expression. Mod Pathol. 1994;7(2):257–62.

22. Elston CW, Sloane JP, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, et al. Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast. Eur J Cancer. 2000;

23. Scott MA, Lagios MD, Axelsson K, Rogers LW, Anderson TJ, Page DL. Ductal carcinoma in situ of the breast: Reproducibility of histological subtype analysis. Hum Pathol. 1997;

24. Schuh F, Biazús JV, Resetkova E, Benfica CZ, Edelweiss MIA. Reproducibility of three classification systems of ductal carcinoma in situ of the breast using a web-based survey. Pathol Res Pract. 2010;

25. Schuh F, Biazús JV, Resetkova E, Benfica CZ, Ventura A de F, Uchoa D, et al. Histopathological grading of breast ductal carcinoma In Situ: Validation of a web-based survey through intra-observer reproducibility analysis. Diagn Pathol. 2015;

26. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA - J Am Med Assoc. 2015;313(11):1122–32.

27. Verkooijen HM, Peterse JL, Schipper MEI, Buskens E, Hendriks JHCL, Pijnappel RM, et al. Interobserver variability between general and expert pathologists during the histopathological assessment of large-core needle and open biopsies of non-palpable breast lesions. Eur J Cancer. 2003;

28. van Dooijeweert C, van Diest PJ, Willems SM, Kuijpers CCHJ, Overbeek LIH, Deckers IAG. Significant inter- and intra-laboratory variation in grading of ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. Breast Cancer Res Treat [Internet]. 2019;174(2):479–88. Available from: http://dx.doi.org/10.1007/s10549-018-05082-y

21

29. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hroóbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud. 2011;

30. Elshof LE, Schaapveld M, Schmidt MK, Rutgers EJ, van Leeuwen FE, Wesseling J. Subsequent risk of ipsilateral and contralateral invasive breast cancer after treatment for ductal carcinoma in situ: incidence and the effect of radiotherapy in a population-based cohort of 10,090 women. Breast Cancer Res Treat. 2016;

31. Casparie M, Tiebosch ATMG, Burger G, Blauwgeers H, Van De Pol A, Van Krieken JHJM, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19–24.

32. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. J Clin Epidemiol. 1999;52(12):1165–72.

33. Hayes AF, Krippendorff K. Answering the Call for a Standard Reliability Measure for Coding Data. Commun Methods Meas. 2007;1(1):77–89.

34. Gwet KL. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. 4th ed. Gaithersburg, MD: Advanced Analytics, LLC; 2014.

35. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529310. Biometrics. 1977;33(1):159–74.

36. Correa C, McGale P, Taylor C, Davidson N, Gelber R, Piccart M, et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. J Natl Cancer Inst - Monogr. 2010;41(41):162–77.

37. Donker M, Litière S, Werutsky G, Julien JP, Fentiman IS, Agresti R, 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(32):4054–9.

22

38.	Elshof LE, Tryfonidis K, Slaets L, Van Leeuwen-Stok AE, Skinner VP, Dif N, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. Eur J Cancer. 2015;

39.	Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JMS, Brookes C, et al. Addressing overtreatment of screen detected DCIS; The LORIS trial. Eur J Cancer. 2015;

40.	Hwang ES, Hyslop T, Lynch T, Frank E, Pinto D, Basila D, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). BMJ Open. 2019;9(3):e026797.

41.	Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(1):29–48.

42.	Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer. 1957;11(3):359–77.

43.	Elston CW, Ellis IO. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403–10.

44.	Van Bockstal M, Baldewijns M, Colpaert C, Dano H, Floris G, Galant C, et al. Dichotomous histopathological assessment of ductal carcinoma in situ of the breast results in substantial interobserver concordance. Histopathology. 2018;73(6):923–32.

45.	Dano H, Altinay S, Arnould L, Bletard N, Colpaert C, Dedeurwaerdere F, et al. Interobserver variability in upfront dichotomous histopathological assessment of ductal carcinoma in situ of the breast: the DCISion study. Mod Pathol. 2019;

46.	Visser LL, Elshof LE, Schaapveld M, Van De Vijver K, Groen EJ, Almekinders MM, et al. Clinicopathological risk factors for an invasive breast cancer recurrence after ductal carcinoma in situ-a nested case-control study. Clin Cancer Res. 2018;24(15):3593–601.

47.	Kerlikowske K, Molinaro AM, Gauthier ML, Berman HK, Waldman F, Bennington J, et al. Biomarker expression and risk of subsequent tumors after initial ductal carcinoma in situ

23

diagnosis. J Natl Cancer Inst. 2010;102(9):627–37.

48.   Mohammed ZMA, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, et al. Comparison of Visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. Br J Cancer. 2012;106(2):383–8.

49.   Van Velthuysen MLF, Groen EJ, Sanders J, Prins FA, Van Der Noort V, Korse CM. Reliability of proliferation assessment by Ki-67 expression in neuroendocrine neoplasms: Eyeballing or image analysis? Neuroendocrinology. 2014;100(4):288–92.

50.   Stasik CJ, Davis M, Kimler BF, Fan F, Damjanov I, Thomas P, et al. Grading ductal carcinoma in situ of the breast using an automated proliferation index. Ann Clin Lab Sci. 2011;41(2):122–30.

51.   Balkenhol MCA, Tellez D, Vreuls W, Clahsen PC, Pinckaers H, Ciompi F, et al. Deep learning assisted mitotic counting for breast cancer. Lab Investig [Internet]. 2019;99(11):1596–606. Available from: http://dx.doi.org/10.1038/s41374-019-0275-0

52.   Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nat Rev Clin Oncol [Internet]. 2019;16(11):703–15. Available from: http://dx.doi.org/10.1038/s41571-019-0252-y

53.   Klimov S, Miligy IM, Gertych A, Jiang Y, Toss MS, Rida P, et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (DCIS) recurrence risk. Breast Cancer Res. 2019;21(1):1–19.

54.   Collins LC, Achacoso N, Haque R, Nekhlyudov L, Fletcher SW, Quesenberry CP, et al. Risk factors for non-invasive and invasive local recurrence in patients with ductal carcinoma in situ. Breast Cancer Res Treat. 2013;139(2):453–60.

24

Supplementary Material

This is a .pdf file containing:

Supplementary methods. This file gives additional information regarding rater selection, participation and the scoring process of the reliability study.

Supplementary Table 1. Clinical characteristics of included and excluded patients for iIBC risk analysis.

Supplementary Table 2. Number of scores per slide and agreement with the majority opinion per histopathological feature. This table gives background information about the majority opinion used in the analysis of subsequent iIBC risk after DCIS, on how many scores and on what agreement is the majority opinion based on, per histopathological feature.

Supplementary Table 3. Characteristics of raters participating in the study.

Supplementary Table 4. Associations of clinicopathological characteristics with subsequent iIBC in univariable analysis.

Figure Legends

Fig. 1 Flow diagram for patient selection and exclusions

Subcohort = randomly selected patient group; outside subcohort = patients who developed subsequent ipsilateral invasive breast cancer not included in the subcohort; iIBC = ipsilateral invasive breast cancer; [a] 2 outside subcohort patients developed invasive breast cancer after a mastectomy was performed during follow-up, for other reasons than iIBC.

Fig. 2 Kaplan-Meier curve illustrating iIBC incidence after diagnosis of DCIS treated by BCS alone

GP = growth pattern; other = flat epithelial atypia, clinging and (micro)papillary growth pattern

The red dashed reference line depicts the maximum reached incidence in patients with DCIS grade 3 with a cribriform/solid growth pattern treated with BCS+RT.

Table 1. Agreement, Gwet's AC2 (GAC2) and Krippendorff's alpha (KA) coefficients per histopathological feature

Histopathological feature	Agreement, %	95%CI, %	GAC2	95%CI	KA	95%CI
Grade (1, 2 or 3)	76.4	75.27-77.52	0.52	0.50-0.55	0.34	0.30-0.39
Grade (1 versus 2+3)	83.5	81.33-85.68	0.78	0.74-0.82	0.35	0.28-0.42
Grade (1+2 versus 3)	69.3	66.94-71.63	0.43	0.38-0.49	0.34	0.29-0.38
Grade (low versus high)	72.8	70.54-75.12	0.52	0.47-0.57	0.38	0.32-0.44
Dominant growth pattern	84.8	82.58-86.97	0.78	0.75-0.82	0.44	0.37-0.51
Calcifications	81.1	78.81-83.40	0.70	0.65-0.75	0.49	0.43-0.54
Necrosis	81.4	79.12-83.64	0.70	0.66-0.75	0.47	0.41-0.53
Mitotic activity	78.5	76.12-80.97	0.70	0.65-0.74	0.24	0.19-0.29
Periductal fibrosis (absent, subtle or prominent presence)	70.9	69.71-72.13	0.37	0.34-0.39	0.25	0.22-0.29
Periductal fibrosis (present versus absent)	71.2	68.82-73.48	0.53	0.48-0.58	0.23	0.18-0.28
Type of periductal fibrosis (if present)	70.5	67.57-73.37	0.50	0.44-0.57	0.26	0.21-0.31
Lymphocytic infiltrate (absent, subtle or prominent presence)	77.1	75.82-78.36	0.50	0.47-0.53	0.42	0.38-0.47
Lymphocytic infiltrate (present versus absent)	73.0	70.51-75.40	0.51	0.45-0.56	0.38	0.33-0.43

27

GAC2 = Gwet's AC2; KA = Krippendorff's alpha; weighted analysis was performed for ordinal features with more than 2 categories using linear weights (grade 1-3, periductal fibrosis and lymphocytic infiltrate); CI = Confidence Interval

Table 2. Clinicopathological characteristics of the study population

Number of DCIS patients (%)

	All patients with iIBC 137*	Subcohort 215**
Treatment		
BCS+RT	42 (30.7)	128 (59.5)
BCS alone	95 (69.3)	87 (40.5)
Age at DCIS diagnosis, years, median (iqr)	57.5 (53.1-63.6)	58.4 (53.4-64.0)
Age at DCIS diagnosis, years (quartiles)		
≥49.5 - ≤53.4	37 (27.0)	54 (25.1)
>53.4 - ≤58.2	36 (26.3)	50 (23.3)
>58.2 - ≤63.8	32 (23.4)	56 (26.1)
>63.8 - ≤75.6	32 (23.4)	55 (25.6)
Period of DCIS diagnosis[a]		
1993 - 1998	76 (55.5)	82 (38.1)
1999 - 2004	61 (44.5)	133 (61.9)
Median follow-up, years (iqr)		11.2 (8.6-14.1)
Time to iIBC, years, median (iqr)	5.3 (3.3-7.6)	
Grade (1,2 or 3)[b]		
Grade 1	10 (7.3)	23 (10.7)
Grade 2	67 (48.9)	115 (53.5)
Grade 3	60 (43.8)	77 (35.8)
Grade (low versus high)		
Low grade	31 (22.6)	60 (27.9)
High grade	106 (77.4)	155 (72.1)
Dominant growth pattern[c]		
FEA[d], clinging, (micro)papillary	14 (10.2)	34 (15.9)
Cribriform, solid	123 (89.8)	180 (84.1)

Table 2 continued.

Number of DCIS patients (%)

	All patients with iIBC 137*	Subcohort 215**
Calcifications		
Present	103 (75.2)	168 (78.1)
Absent	34 (24.8)	47 (21.9)
Necrosis		
Present	109 (79.6)	167 (77.7)
Absent	28 (20.4)	48 (22.3)
Mitoses		
Sparse	114 (83.2)	198 (92.1)
Many	23 (16.8)	17 (7.9)
Periductal fibrosis		
Absent	28 (20.4)	41 (19.1)
Subtle	73 (53.4)	102 (47.4)
Prominent	36 (26.3)	72 (33.5)
Type of periductal fibrosis[e]		
Sclerotic	80 (73.4)	133 (76.4)
Myxoid	29 (26.6)	41 (23.6)
Lymphocytic infiltrate		
Absent	38 (27.7)	77 (35.8)
Subtle	65 (47.5)	89 (41.4)
Prominent	34 (24.8)	49 (22.8)

subcohort = randomly selected patient group; * six out of all patients with iIBC developed breast cancer metastases only; ** sixteen patients from the subcohort developed an iIBC and four developed breast cancer metastases only; iqr = interquartile range; [a] 1993-1998 reflecting part of the screening implementation phase and 1999-2004 reflecting full nationwide coverage;

[b] histopathological DCIS features were based on the majority opinion; [c] in one patient growth pattern was scored as not assessable by all raters and was therefore excluded (n included patients = 331); FEA = flat epithelial atypia; [d] there is controversy about whether to consider FEA as a subtype of DCIS (clinging, monomorphic type) or not, therefore this option was included as possible DCIS growth pattern; [e] for type of fibrosis patients were only included when according to the majority opinion periductal fibrosis was present, either subtle or prominent (n included patients = 268)

Table 3. Associations of histopathological features with subsequent iIBC in multivariable analysis

Histopathological feature	BCS alone			BCS+RT			Treatment interaction
	n	HR (95%CI)	*P*	n	HR (95%CI)	*P*	*P*
Grade (1+2 versus 3)							0.017
1+2	107 (52)	**REF**		104 (28)	**REF**		
3	62 (43)	2.64 (1.35-5.14)	0.005	58 (14)	0.79 (0.38-1.62)	0.52	
Dominant growth pattern							0.022
FEA/clinging/(micro)papillary	23 (7)	**REF**		23 (7)	**REF**		
Cribriform/Solid	146 (88)	3.70 (1.34-10.23)	0.012	139 (35)	0.77 (0.32-1.85)	0.56	

n = total number (number of patients with subsequent iIBC); HR = Hazard Ratio; CI = Confidence Interval; *P* = P value; REF = reference; FEA = flat epithelial atypia

Fig.1.

DCIS treated by BCS+/-RT in 1993-2004

n = 2767

Subcohort n = 357	Outside subcohort n = 177
n = 534	

Excluded after pathology report review
n = 11

No (pure) DCIS	n = 9
Other	n = 2

Excluded after pathology report review
n = 16

No (pure) DCIS	n = 7
Uncertainty on iIBC occurrence	n =

Patients eligible for study (n = 507)

Material received	
n = 286 (83%)	n = 143 (89%)

Excluded after internal slide review
n = 58

DCIS not confirmed	n = 22
No (pure) DCIS	n =

Excluded after internal slide review
n = 18

DCIS not confirmed	n = 6
No (pure) DCIS	n =

Patients included in reliability study (n =	
n = 228	n = 125

Excluded after external slide review
n = 8

Excluded after external slide review
n = 3

Final analysis of reliability (n = 342)	
n = 220	n = 122

Excluded after external slide review
n = 5

Excluded after external slide review[a]
n = 5

32

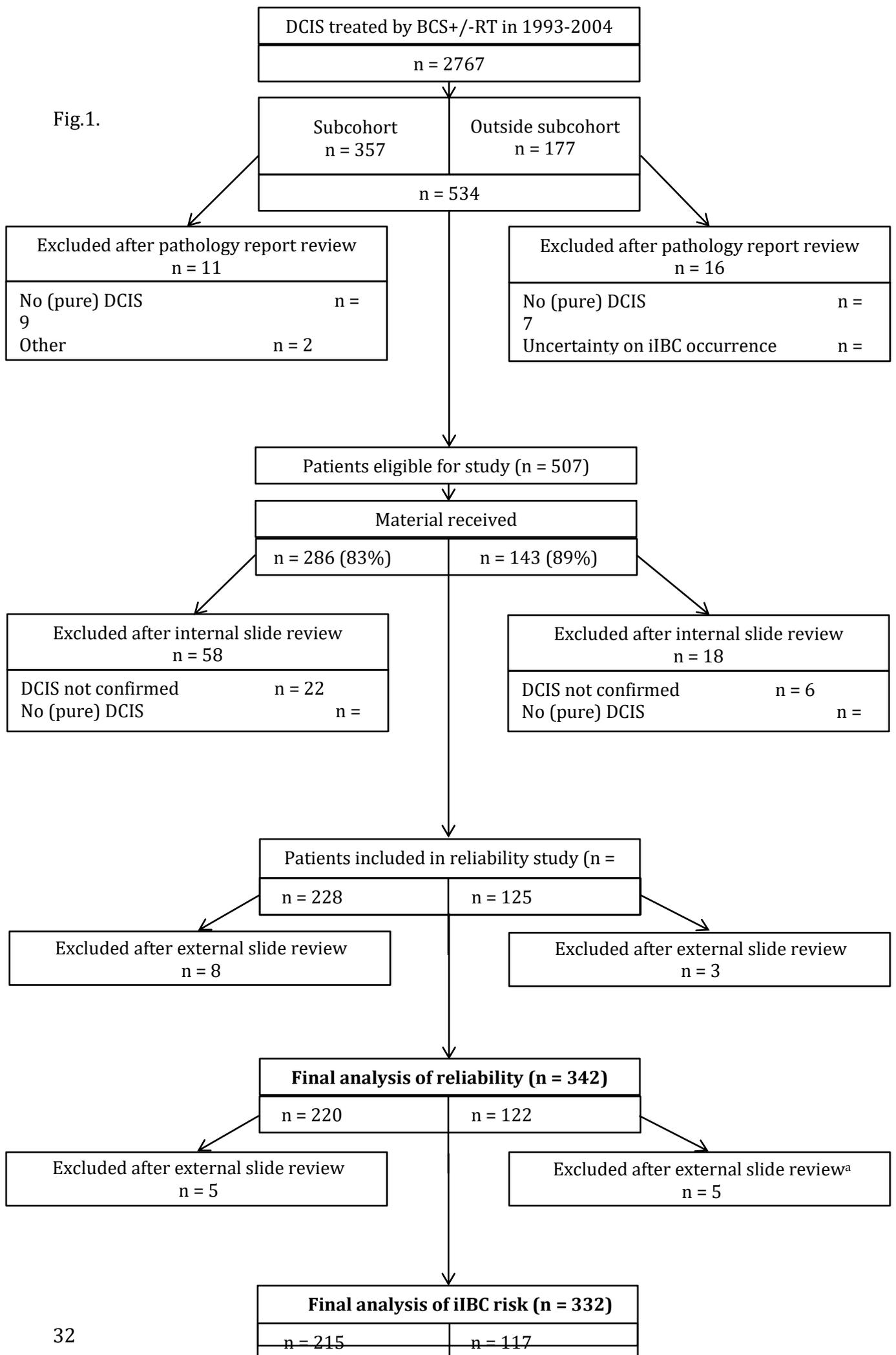| **Final analysis of iIBC risk (n = 332)**	
n = 215	n = 117

Fig. 2.

Supplementary methods

Rater selection and participation

To ensure a mixed group of raters in terms of expertise and experience, a dual selection approach was undertaken. Members of the European Working Group for Breast Screening Pathology, a working group set up in 1993 in order to make the practice of breast pathology more uniform and considered breast pathology experts, were invited to participate by email. Twenty-two members agreed to participate and 17 completed the study.

All participants of the '7th Dutch Breast Pathology Course' (November 2018, Amsterdam; 31 pathologists and 3 residents) with different levels of expertise were also invited to participate in the study. Nineteen pathologists and 2 residents completed the study, for which the first received CME accreditation as compensation.

After study-closure all raters who completed the study received personal feedback by providing an overview comparing their scores with those from the group.

Study sets

To reduce the workload while ensuring enough ratings per case for subsequent analysis, each rater was assigned a personal study set, including in total 146 cases. The study sets were composed in two steps. Firstly, 100 cases were randomly selected from the total cohort of 353 cases and assigned to the study sets of all raters. Secondly, for each rater individually 46 cases out of the remaining 253 cases not yet assigned, were randomly selected and added to their study set.

Fifty out of the 100 cases, which were assigned to all raters, were placed in the beginning of the study set and the other fifty were randomly distributed amongst the remaining cases. Raters were aware of a presumed DCIS diagnosis in this study and were not restricted in scoring time (starting date 15/10/2018 - closing date 08/02/2019).

DCIS scoring form

1. DCIS present? (if not, please give the diagnosis under comments)
   - Yes
   - No
   - Not assessable

1. Dominant growth pattern?
   - Not assessable
   - FEA
   - Clinging
   - (Micro)papillary
   - Cribriform
   - Solid

2. DCIS grade? (1/2/3)
   - Not assessable
   - Well differentiated
   - Moderately differentiated
   - Poorly differentiated

3. DCIS grade? (low/high)
   - Not assessable
   - Low grade
   - High grade

4. Necrosis present?
   - Not assessable
   - Absent
   - Present

5. Calcification present?
   - Not assessable
   - Absent
   - Present

6. Frequency of mitoses?
   - Not assessable
   - Sparse
   - Many

7. Periductal fibrosis present?
   - Not assessable
   - Absent
   - Subtle
   - Prominent

8. Only if fibrosis is present: what is the (dominant) type of stroma?
   - Not assessable
   - Sclerotic
   - Myxoid

9. Lymphocytic infiltrate present?
   - Not assessable
   - Absent
   - Subtle
   - Prominent

Comments (other diagnosis or otherwise)

1 = well differentiated/2 = moderately differentiated/3 = poorly differentiated; FEA = flat epithelial atypia; There is controversy about whether to consider FEA as a subtype of DCIS

(clinging, monomorphic type) or not, therefore this option was included as possible DCIS growth pattern.

DCIS Interobserver Study – rater background questionnaire

1. Your email address [a]

2. In which country are you working?

3. In which hospital/pathology lab are you working?

4. Where did you receive your pathology training? (hospital/place/country)

5. How many years are you working as a pathologist?

> 0-5 years
>
> 6-10 years
>
> 11-15 years
>
> 16-20 years
>
> >20 years

6. How many years are you looking at breast cases?

> 0-5 years
>
> 6-10 years
>
> 11-15 years
>
> 16-20 years
>
> >20 years

7. Do your colleagues consider you an expert in breast pathology?

> Yes
>
> No

8. How many pathologists are working in your lab?

9. How many pathologists are looking at breast cases in your lab?

10. How many breast cases are seen annually in your lab (estimate, biopsies + surgical specimens)

11. Do you look at revision or consult cases?

> Yes
>
> No

12. Which DCIS grading system do you use in daily practice?

Holland et al (1994; 3-tiered; based on nuclear grade and cell polarization)

Pinder et al (2010; 4-tiered; very high = high nuclear grade + >50% solid growth & comedo-necrosis)

Van Nuys (1995; 3-tiered; high grade, non-high grade with necrosis, non-high grade without necrosis)

Poller et al. (1994; 2-tiered; pure comedo, non comedo)

Lagios (1990; 3-tiered; based on nuclear features & frequency of mitoses)

College of American Pathologists Guidelines

WHO

Intuition

Other:

13. In case of a heterogeneous DCIS, how did you grade in this study?

I gave the highest grade

I gave the predominant grade

Other:

14. Comments regarding your interpretation of specific items in the study

15. How would you rate the slide viewing platform 'Slide Score'?

16. Comments/feedback for Slide Score

A questionnaire was sent to all 38 raters who finished their complete study set with questions regarding their working environment, experience and their method of DCIS grading. Thirty-five pathologists and 2 residents completed the questionnaire.

[a] Required

Supplementary Table 1. Clinical characteristics of included and excluded patients for iIBC risk analysis

	Subcohort patients			Patients outside subcohort with subsequent iIBC		
	Included patients n (%) 215 (60.2)	Excluded patients n (%) 142 (39.8)	P^a	Included patients n (%) 117 (66.1)	Excluded patients n (%) 60 (33.9)	P^a
Patient group						
Subcohort, no iIBC	195 (90.7)	131 (92.3)				
Subcohort, iIBC	20 (9.3)	11 (7.8)	0.61			
Treatment						
BCS+RT	128 (59.5)	77 (54.2)		34 (29.1)	24 (40.0)	
BCS alone	87 (40.5)	65 (45.8)	0.32	83 (70.9)	36 (60.0)	0.14
Age at DCIS diagnosis, years, median (iqr)	58.4 (53.4-64.0)	58.3 (53.3-64.2)	0.68	57.5 (53.2-63.6)	59.0 (54.5-62.0)	0.63
Age at DCIS diagnosis, years (quartiles)						
≥49.5 - ≤53.4	54 (25.1)	38 (26.8)		30 (25.6)	13 (21.7)	
>53.4 - ≤58.3	53 (24.7)	33 (23.2)		32 (27.4)	15 (25.0)	
>58.3 - ≤63.7	53 (24.7)	31 (21.8)		27 (23.1)	23 (38.3)	
>63.7 - ≤75.6	55 (25.6)	40 (28.2)	0.88	28 (23.9)	9 (15.0)	0.16

Period of DCIS diagnosis[b]						
1993 – 1998	82 (38.1)	58 (40.9)		63 (53.9)	43 (71.7)	
1999 – 2004	133 (61.9)	84 (59.2)	0.61	54 (46.2)	17 (28.3)	**0.022**

Subcohort = randomly selected patient group; n = number; *P* = P value; [a] For categorical variables the P value was calculated by a chi-square test, for age at diagnosis by a Wilcoxon rank-sum test; iIBC = ipsilateral invasive breast cancer; iqr = interquartile range; [b] 1993-1998 reflecting part of the screening implementation phase and 1999-2004 reflecting full nationwide coverage

Supplementary Table 2. Number of scores per slide and agreement with the majority opinion per histopathological feature

| Histopathological feature | n of scores per slide | | Agreement with the majority opinion score (%) | |
	Mean	Median (iqr)	Mean	Median (iqr)
Grade (1,2 or 3)	14	7 (6-32)	70.1	69.4 (57.1-83.3)
Grade (1 versus 2+3)	14	7 (6-32)	89.8	97.3 (83.3-100)
Grade (1+2 versus 3)	14	7 (6-32)	79.4	83.3 (66.7-100)
Grade (low versus high)	14	7 (6-30)	83.0	85.3 (71.4-100)
Dominant growth pattern[a]	15	7 (6-32)	90.4	100 (83.3-100)
Calcifications	15	7 (6-32)	88.2	97.1 (80.0-100)
Necrosis	15	7 (6-33)	88.2	95.4 (80.0-100)
Mitotic activity	13	7 (6-29)	86.4	93.8 (75.0-100)
Periductal fibrosis (absent, subtle or prominent presence)	15	7 (6-32)	65.1	62.5 (54.1-75.0)
Periductal fibrosis (present versus absent)	15	7 (6-32)	81.6	83.3 (71.4-100)
Type of periductal fibrosis[b]	12	6 (5-24)	81.7	83.3 (66.7-100)
Lymphocytic infiltrate (absent, subtle or prominent presence)	15	7 (6-31)	71.1	67.6 (57.1-83.3)
Lymphocytic infiltrate (present versus absent)	15	7 (6-31)	82.4	83.8 (66.7-100)

n = number; iqr = interquartile range; [a] in one patient growth pattern was scored as not assessable by all raters and was therefore excluded (n included patients = 341); [b] for type of fibrosis patients were only included when according to the majority opinion periductal fibrosis was present, either subtle or prominent (n included patients = 276)

Supplementary Table 3. Characteristics of raters participating in the study[a,b]

Experience, years	n (%)
0-5	5 (15.2)
6-10	2 (6.1)
11-15	3 (9.1)
16-20	5 (15.2)
> 20	18 (54.6)
Country of work	
the Netherlands	17 (48.6)
Europe, other	18 (51.4)
EWGBSP-member	
Yes	17 (47.2)
No	19 (52.8)
Considered expert in breast pathology by colleagues	
Yes	30 (88.2)
No	4 (11.8)
Experience with breast revision/consult cases	
Yes	26 (74.3)
No	9 (25.7)
DCIS grading system used	
WHO[1]	9 (25.0)
Holland[2]	10 (27.8)
Van Nuys[3]	4 (11.1)
WHO & Van Nuys	4 (11.1)

WHO & Holland	2 (5.6)
WHO & Holland & Lagios[4]	1 (2.8)
WHO & CAP[5]	1 (2.8)
Lagios	1 (2.8)
Pinder[6]	1 (2.8)
Other	3 (8.3)
Grading in case of heterogeneous DCIS	
Highest grade	33 (94.3)
Predominant grade	2 (5.7)

Supplementary Table 3 continued.

Characteristics of the raters' laboratories	
n of pathologists, median (iqr)	13 (8-15)
n of breast pathologists, median (iqr)	4 (3-5)
Laboratory specialization[c], median (iqr)	2.6 (1.8-4.6)
n of breast cases seen annually, median (iqr)	1200 (600-2000)

[a] the questionnaire was not filled in (completely) by all raters, percentages are based on the responders; [b] Residents are included only in questions regarding their grading of DCIS

n = number; iqr = interquartile range; EWGBSP = members of the European Working Group for Breast Screening Pathology; [c] Laboratory specialization = number of pathologists in rater's laboratory/number of breast pathologists in rater's laboratory

Supplementary Table 4. Associations of clinicopathological characteristics with subsequent iIBC in univariable analysis

Clinicopathological characteristic	All patients			BCS alone			BCS+RT			Interaction
	n	HR (95% CI)	*P*	n	HR (95% CI)	*P*	n	HR (95% CI)	*P*	*P*
Grade (1,2 or 3)[a]										
1	31 (10)	**REF**		21 (8)	**REF**		10 (2)	**REF**		
2	172 (67)	1.28 (0.58-2.83)	0.54	84 (43)	1.61 (0.63-4.08)	0.32	88 (24)	1.39 (0.27-7.15)	0.69	0.94
3	129 (60)	1.69 (0.75-3.80)	0.20	65 (44)	3.19 (1.21-8.37)	**0.019**	64 (16)	1.10 (0.20-5.89)	0.91	0.33
Grade (1 versus 2+3)										
1	30 (10)	**REF**		21 (8)	**REF**		9 (2)	**REF**		
2+3	302 (127)	1.35 (0.62-2.91)	0.45	149 (87)	2.15 (0.88-5.22)	0.092	153 (40)	0.98 (0.19-5.14)	0.99	0.50
Grade (1+2 versus 3)										
1+2	211 (80)	**REF**		107 (52)	**REF**		104 (28)	**REF**		
3	121 (57)	1.41 (0.90-2.20)	0.13	63 (43)	2.34 (1.24-4.42)	**0.009**	58 (14)	0.74 (0.35-1.56)	0.42	**0.028**
Grade (low versus high)										
Low	87 (31)	**REF**		54 (27)	**REF**		33 (4)	**REF**		
High	245 (106)	1.33 (0.81-2.20)	0.26	116 (68)	1.47 (0.79-2.76)	0.23	129 (38)	2.68 (0.88-8.21)	0.084	0.34
Dominant growth pattern[b]										
FEA/clinging/(micro)papillary	46 (14)	**REF**		23 (7)			23 (7)	**REF**		

44

	All patients			BCS alone			BCS+RT			Interaction
Cribriform/solid	285 (123)	1.76 (0.92-3.36)	0.087	146 (88)	3.44 (1.33-8.91)	**0.011**	139 (35)	0.70 (0.29-1.72)	0.44	**0.023**

Supplementary Table 4 continued.

Clinicopathological characteristic	All patients			BCS alone			BCS+RT			Interaction
	n	HR (95% CI)	*P*	n	HR (95% CI)	*P*	n	HR (95% CI)	*P*	*P*
Calcifications										
Present	256 (103)	REF		131 (71)	REF		125 (32)	REF		
Absent	76 (34)	1.23 (0.75-2.04)	0.41	39 (24)	1.31 (0.65-2.65)	0.45	37 (10)	1.13 (0.51-2.53)	0.77	0.76
Necrosis										
Present	260 (109)	REF		126 (72)	REF		134 (37)	REF		
Absent	72 (28)	0.87 (0.52-1.46)	0.59	44 (23)	0.80 (0.41-1.56)	0.51	28 (5)	0.60 (0.22-1.65)	0.32	0.59
Mitotic activity										
Sparse	294 (114)	REF		141 (74)	REF		153 (40)	REF		
Many	38 (23)	2.42 (1.20-4.91)	**0.014**	29 (21)	2.53 (1.05-6.11)	**0.038**	9 (2)	0.79 (0.15-4.15)	0.78	0.21
Periductal fibrosis[a]										

Clinicopathological characteristic	All patients			BCS alone			BCS+RT			Interaction
	n	HR (95% CI)	P	n	HR (95% CI)	P	n	HR (95% CI)	P	P
Absent	64 (28)	REF		42 (24)	REF		22 (4)	REF		
Subtle	165 (73)	1.02 (0.58-1.78)	0.95	84 (48)	1.01 (0.50-2.05)	0.98	81 (25)	1.98 (0.63-6.20)	0.24	0.33
Prominent	103 (36)	0.70 (0.38-1.31)	0.27	44 (23)	0.84 (0.36-1.91)	0.67	59 (13)	1.29 (0.39-4.30)	0.68	0.56
Periductal fibrosis present/ absent										
Present (subtle/prominent)	275 (113)	REF		134 (75)	REF		141 (38)	REF		
Absent	57 (24)	1.06 (0.61-1.84)	0.84	36 (20)	0.97 (0.48-1.96)	0.94	21 (4)	0.67 (0.22-2.02)	0.48	0.56

Supplementary Table 4 continued.

Clinicopathological characteristic	All patients			BCS alone			BCS+RT			Interaction
	n	HR (95% CI)	P	n	HR (95% CI)	P	n	HR (95% CI)	P	P
Type of periductal fibrosis[c]										
Sclerotic	202 (80)	REF		101 (54)	REF		101 (26)	REF		
Myxoid	66 (29)	1.29 (0.74-2.24)	0.37	27 (17)	2.23 (0.86-5.76)	0.099	39 (12)	1.18 (0.53-2.62)	0.68	0.34
Lymphocytic infiltrate[a]										
Absent	108 (38)	REF		58 (30)	REF		50 (8)	REF		
Subtle	144 (65)	1.48 (0.90-2.44)	0.12	77 (42)	1.11 (0.58-2.14)	0.75	67 (23)	2.74 (1.12-6.69)	**0.027**	0.11
Prominent	80 (34)	1.35 (0.75-2.41)	0.32	35 (23)	1.91 (0.80-4.54)	0.14	45 (11)	1.56 (0.58-4.21)	0.38	0.79

Clinicopathological characteristic	All patients n	HR (95% CI)	P	BCS alone n	HR (95% CI)	P	BCS+RT n	HR (95% CI)	P	Interaction P
Lymphocytic infiltrate present/absent										
Present (subtle/prominent)	227 (100)	REF		113 (66)	REF		114 (34)	REF		
Absent	105 (37)	0.71 (0.45-1.13)	0.15	57 (29)	0.73 (0.39-1.35)	0.31	48 (8)	0.50 (0.22-1.16)	0.11	0.46
Age at diagnosis, years (quartiles)										
≥49.5 - ≤53.4	84 (37)	REF		38 (20)	REF		46 (17)	REF		
>53.4 - ≤58.2	82 (36)	0.97 (0.53-1.76)	0.92	43 (24)	1.12 (0.47-2.64)	0.80	39 (12)	0.73 (0.30-1.79)	0.49	0.51
>58.2 - ≤63.8	83 (32)	0.81 (0.44-1.48)	0.49	43 (26)	1.24 (0.53-2.90)	0.61	40 (6)	0.33 (0.11-0.92)	**0.035**	**0.048**
>63.8 - ≤75.6	83 (32)	0.84 (0.46-1.53)	0.57	46 (25)	1.02 (0.45-2.34)	0.96	37 (7)	0.46 (0.17-1.26)	0.13	0.20

Supplementary Table 4 continued.

Clinicopathological characteristic	All patients n	HR (95% CI)	P	BCS alone n	HR (95% CI)	P	BCS+RT n	HR (95% CI)	P	Interaction P
Age at diagnosis (cont.)		0.98 (0.95-1.02)	0.38		1.00 (0.96-1.05)	0.90		0.93 (0.86-1.00)	0.053	0.079
Period of DCIS diagnosis										
1993 – 1998	145 (76)	REF		104 (63)	REF		41 (13)	REF		
1999 – 2004	187 (61)	0.61 (0.39-0.96)	**0.032**	66 (32)	0.75 (0.41-1.37)	0.35	121 (29)	1.44 (0.58-3.57)	0.44	0.66
Treatment										

BCS+RT / 0-5 years	162 (14)	**REF**	
BCS+RT / >5 years	142 (28)	0.51 (0.24-1.12)	0.093
BCS alone / 0-5 years	170 (43)	4.80 (2.49-9.24)	**0.000**
BCS alone / >5 years	118 (52)	2.47 (1.42-4.30)	**0.001**
		$P_{heterogeneity}$	**0.000**

n = total number (number of patients with subsequent iIBC); HR = Hazard Ratio; CI = Confidence Interval; *P* = P value; Interaction = interaction with treatment; REF = reference; cont. = Continuous; [a] Recategorizations of grade, periductal fibrosis, and lymphocytic infiltrate may have led to small differences in the majority opinion (for example when considering the histopathological feature grade 1-3 with a distribution of grade 1 -30%, grade 2 -30% and grade 3 -40% with grade 3 as majority opinion will lead to a categorical shift when recategorizing grade 1-3 into grade 1+2 versus 3 with an adjusted distribution of grade 1 or 2 - 60% and grade 3 -40% with grade 1+2 as majority opinion); [b] in one patient growth pattern was scored as not assessable by all raters and was therefore excluded (n included patients = 331); [c] for type of fibrosis patients were only included when according to the majority opinion periductal fibrosis was present, either subtle or prominent (n included patients = 268)

References

1. Lakhani SR, Ellis. I.O., Schnitt SJ, Tan PH, van de Vijver MJ. WHO classification of tumours of the breast. 4th ed. Lyon: International Agency for Research on Cancer; 2012.

2. Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, Van de Vijver MJ, et al. Ductal carcinoma in situ: A proposal for a new classification. Semin Diagn Pathol. 1994;11(3):167–80.

3. Silverstein MJ, Poller DN, Waisman JR, Colburn WJ, Barth A, Gierson ED, et al. Prognostic classification of breast ductal carcinoma-in-situ. Lancet. 1995;345(8958):1154–7.

4. Lagios MD. Duct carcinoma in situ. Pathology and treatment. Surg Clin North Am. 1990;70(4):873–83.

5. College of American pathologists [Internet]. Available from: https://documents.cap.org/protocols/cp-breast-dcis-18protocol-4100.pdf

6. Pinder SE, Duggan C, Ellis IO, Cuzick J, Forbes JF, Bishop H, et al. A new pathological system for grading DCIS with improved prediction of local recurrence: Results from the UKCCCR/ANZ DCIS trial. Br J Cancer. 2010;