



Published in final edited form as:

Mod Pathol. 2020 March ; 33(3): 354–366. doi:10.1038/s41379-019-0367-9.

Inter-observer variability in upfront dichotomous histopathological assessment of ductal carcinoma *in situ* of the breast: the DCISion study.

Hélène Dano¹, Serdar Altinay^{*2}, Laurent Arnould^{*3}, Noella Bletard^{*4}, Cecile Colpaert^{*5}, Franceska Dedeurwaerdere^{*6}, Benjamin Dessauvage^{*7}, Valérie Duwel^{*8}, Giuseppe Floris^{*9,10}, Stephen Fox^{*11}, Clara Gerosa^{*12}, Shabnam Jaffer^{*13}, Eline Kurpershoek^{*14}, Magali Lacroix-Triki^{*15}, Andoni Laka^{*16}, Kathleen Lambein^{*17}, Gaëtan Marie MacGrogan^{*18}, Caterina Marchió^{*19,20}, Dolores Martin Martinez^{*21}, Sharon Nofech-Mozes^{*22}, Dieter Peeters^{*23,24}, Alberto Ravarino^{*12}, Emily Reisenbichler^{*25}, Erika Resetkova^{*26}, Souzan Sanati^{*27}, Anne-Marie Schelfhout^{*28}, Vera Schelfhout^{*23}, Abeer M. Shaaban^{*29}, Renata Sinke^{*14}, Claudia Maria Stanciu-Pop^{*30}, Claudia Stobbe^{*14}, Carolien H.M. van Deurzen^{*31}, Koen Van de Vijver^{*32}, Anne-Sophie Van Rompuy^{*9}, Stephanie Verschuere^{*6}, Anne Vincent-Salomon^{*33}, Hannah Wen^{*34}, Caroline Bouzin³⁵, Christine Galant^{1,36}, Mieke R. Van Bockstal^{1,36}

¹Department of pathology, Cliniques universitaires Saint-Luc Bruxelles, Avenue Hippocrate 10, Woluwé-Saint-Lambert 1200, Belgium. ²Department of Pathology, University of Health Sciences, Bakirköy Dr. Sadi Konuk Health Application and Research Center, 34147 Istanbul, Turkey. ³Département de Biologie et de Pathologie des Tumeurs, Centre George-François Leclerc, 1 Rue Pr. Marion, 21000 Dijon, France. ⁴Department of Pathology, CHU de Liège - site Sart Tilman, Avenue de l'Hôpital 1, 4000 Liège, Belgium. ⁵Department of Pathology, GZA/ZNA Hospitals, 2610 Wilrijk, Belgium. ⁶Department of Pathology, AZ Delta, Westlaan 123, 8800 Roeselare, Belgium. ⁷Division of Pathology and Laboratory Medicine, Medical School, The University of Western Australia, Crawley, WA 6009, Australia. ⁸Department of pathology, AZ Klina Brasschaat, Augustijnslei 100, 2930 Brasschaat, Belgium. ⁹Department of pathology, University Hospitals Leuven, KU Leuven – University of Leuven, Herestraat 49, 3000 Leuven, Belgium. ¹⁰Department of Imaging and Pathology, Laboratory of Translational Cell & Tissue Research, KU Leuven – University of Leuven, Leuven, Belgium. ¹¹Department of Pathology, Peter MacCallum Cancer Center and the University of Melbourne, Melbourne, Vic 3000, Australia. ¹²Department of Pathology, University of Cagliari, AOU San Giovanni di Dio, Via Ospedale 54, 09124 Cagliari, Italy. ¹³Department of Pathology, Mount Sinai Hospital and Icahn School of Medicine, New York, New York, NY10029 USA. ¹⁴Pathan BV, Kleiweg 500, 3045 PM Rotterdam, The Netherlands. ¹⁵Department of Pathology, Gustave-Roussy Cancer Campus, 114 Rue Edouard-Vaillant, 94805

ADDRESS FOR CORRESPONDENCE: Dr. Mieke Van Bockstal, Department of Pathology – Tour Franklin -1, Cliniques Universitaires Saint-Luc Bruxelles, Avenue Hippocrate 10, Woluwé-Saint-Lambert 1200, Belgium, Tel.: 0032 2764 6707, Fax: 0032 2764 6934, mieke.vanbockstal@uclouvain.be.

*These authors equally contributed to this work.

Supplementary information is available at *Modern Pathology*'s website.

Disclosure/Conflict of Interest

The authors declare they have no conflicts of interest concerning this work.

Villejuif, France. ¹⁶Department of Pathology, Clinique Notre-Dame de Grâce (CNDG), Chaussée de Nivelles 212, 6041 Gosselies, Belgium. ¹⁷Department of pathology, St Lucas Hospital, Groenebriel 1, 9000 Ghent, Belgium. ¹⁸Surgical Pathology Unit, Department of Pathobiology, Institut Bergonié, F-33076 Bordeaux, France. ¹⁹Department of Medical Sciences, University of Turin, 10126 Torino, Italy. ²⁰Pathology Unit, FPO-IRCCS, Candiolo Cancer Institute, Candiolo, Italy. ²¹Curepath, Rue de Borfilet 12A, 6040 Jumet, Belgium. ²²Department of Laboratory Medicine and Molecular Diagnostics, Sunnybrook Health Sciences Center, University of Toronto, Toronto, Ontario, ON M4N 3M5, Canada. ²³Department of Pathology, AZ St Maarten, Liersesteenweg 435, 2800 Mechelen, Belgium. ²⁴Histopathology, Imaging and Quantification Unit, HistoGeneX, Sint-Bavostraat 78, 2610 Antwerp, Belgium. ²⁵Department of Pathology, Yale School of Medicine, Yale New Haven Hospital, 310 Cedar Street, New Haven, CT06510, United States. ²⁶The University of Texas MD Anderson Cancer Center, Houston TX77030, Texas, USA. ²⁷Department of Pathology and Immunology, Washington University School of Medicine, 660 S Euclid Boulevard, St Louis, Missouri, MO 63110, United States. ²⁸Department of Pathology, Onze-Lieve-Vrouwziekenhuis Aalst, Moorselbaan 164, 9300 Aalst, Belgium. ²⁹Department of Cellular Pathology, Queen Elizabeth Hospital Birmingham, University of Birmingham, Birmingham B15 2GW, United Kingdom. ³⁰Department of Pathology, CHU UCL Namur, Site Godinne, Avenue Docteur G. Thérassé 1, 5530 Yvoir, Belgium. ³¹Department of Pathology, Erasmus Medical Center Rotterdam, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. ³²Department of Pathology, Ghent University Hospital, C. Heymanslaan 10, 9000 Ghent, Belgium. ³³Pôle de Médecine Diagnostique & Théranostique, INSERM U934, Institut Curie, 26 Rue d'Ulm, 75248 Paris Cedex 05, France. ³⁴Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065. ³⁵2IP IREC Imaging Platform, Institute of Clinical and Experimental Research (IREC), Université catholique de Louvain, Avenue Hippocrate 55, 1200 Brussels, Belgium. ³⁶Institute of Clinical and Experimental Research (IREC), Université catholique de Louvain, Avenue Hippocrate 55, 1200 Brussels, Belgium.

Abstract

Histopathological assessment of ductal carcinoma *in situ*, a non-obligate precursor of invasive breast cancer, is characterized by considerable inter-observer variability. Previously, *post hoc* dichotomization of multi-categorical variables was used to determine the 'ideal' cut-offs for dichotomous assessment. The present international multi-center study evaluated inter-observer variability among 39 pathologists who performed upfront dichotomous evaluation of 149 consecutive ductal carcinomas *in situ*.

All pathologists independently assessed nuclear atypia, necrosis, solid ductal carcinoma *in situ* architecture, calcifications, stromal architecture and lobular cancerization in one digital slide per lesion. Stromal inflammation was assessed semi-quantitatively. Tumor-infiltrating lymphocytes were quantified as percentages and dichotomously assessed with a cut-off at 50%. Krippendorff's alpha (KA), Cohen's kappa and intraclass correlation coefficient were calculated for the appropriate variables.

Lobular cancerization (KA = 0,396), nuclear atypia (KA = 0,422) and stromal architecture (KA = 0,450) showed the highest inter-observer variability. Stromal inflammation (KA = 0,564),

dichotomously assessed tumor-infiltrating lymphocytes (KA = 0,520) and comedonecrosis (KA = 0,539) showed slightly lower inter-observer disagreement. Solid ductal carcinoma *in situ* architecture (KA = 0,602) and calcifications (KA = 0,676) presented with the lowest inter-observer variability. Semi-quantitative assessment of stromal inflammation resulted in a slightly higher inter-observer concordance than upfront dichotomous tumor-infiltrating lymphocytes assessment (KA = 0,564 *versus* KA = 0,520). High stromal inflammation corresponded best with dichotomously assessed tumor-infiltrating lymphocytes when the cut-off was set at 10% (kappa = 0,881). Nevertheless, a *post hoc* tumor-infiltrating lymphocytes cut-off set at 20% resulted in the highest inter-observer agreement (KA = 0,669).

Despite upfront dichotomous evaluation, the inter-observer variability remains considerable and is at most acceptable, although it varies among the different histopathological features. Future studies should investigate its impact on ductal carcinoma *in situ* prognostication. Forthcoming machine learning algorithms may be useful to tackle this substantial diagnostic challenge.

Keywords

Ductal carcinoma *in situ*; breast; inter-observer variability; interrater agreement; tumor-infiltrating lymphocytes

Ductal carcinoma *in situ* (DCIS), a non-obligate precursor of invasive breast cancer, is a health problem worldwide.¹ The introduction of breast screening programmes in Western countries resulted in a significant rise of its diagnosis with a current incidence of approximately 20% of all breast cancers.²⁻⁴ Prior to those screening programmes, symptomatic ductal carcinoma *in situ* accounted for around 1% of all breast cancer diagnoses.³ Little is known about its natural history, and hence its appropriate management.⁵

Many screen detected ductal carcinoma *in situ* patients are likely over-treated, as a consequence of identifying low risk ductal carcinoma *in situ* cases that would not otherwise be diagnosed outside the screening setting. To address this issue, four randomized clinical trials (LORIS, LORD, COMET and LARRIKIN) are currently ongoing to investigate the potential non-inferiority of active surveillance for patients with a biopsy diagnosis of low risk ductal carcinoma *in situ* by comparing watchful waiting with the current surgical standard of care.⁶ The single-arm Japanese LORETTA trial will investigate the value of endocrine therapy without surgery for estrogen receptor-positive, HER2-negative low risk ductal carcinoma *in situ*.⁷ The definitions of 'low risk' ductal carcinoma *in situ* vary slightly among these trials, but all take into account the degree of nuclear atypia (i.e. ductal carcinoma *in situ* grade).⁵⁻⁷ This inclusion criterion signifies a major challenge, since previous studies have shown that histopathological assessment of nuclear grade is characterized by substantial inter-observer variability, regardless of the grading system used.⁸⁻¹²

This variability is not restricted to study settings, as a nationwide evaluation of ductal carcinoma *in situ* grading in The Netherlands revealed significant variation among different laboratories.¹³ Similar variations at the population level were reported for grading of invasive breast cancers, and the presence of borderline features for grade resulting in

discordant grades was associated with decreased disease-free survival.^{14–16} Although the inclusion criteria of the aforementioned non-inferiority trials encompass strict theoretical definitions of ‘low risk’ ductal carcinoma *in situ*,⁶ the actual implementation of these definitions will probably result in variable eligibility rates among different laboratories, should their inclusion criteria be generalized to routine practice. The LORIS trial aims to reduce this variation by performing central review of all cases diagnosed as ductal carcinoma *in situ*.¹⁷ While this approach applies uniform histological criteria for confirming/refusing patient eligibility, it cannot be generalized to routine practice, where biopsies are often evaluated by only one or two pathologists.

We have previously shown that pathologists are better at distinguishing grade 2 from grade 3 ductal carcinoma *in situ*, than distinguishing grade 1 from grade 2 ductal carcinoma *in situ*.¹⁸ Similar differences between different cut-offs were noted upon *post hoc* dichotomization of the multi-categorical assessment of comedonecrosis, stromal inflammation and stromal architecture.¹⁸ We hypothesized that *ad hoc* dichotomization of multi-categorical histopathological features according to the ‘ideal’ cut-off might result in acceptable degrees of inter-observer variability. The goal of the current study was therefore to perform upfront dichotomous histopathological assessment of ductal carcinoma *in situ* and to explore the degree of inter-observer variability. Additionally, we aimed to assess pathologists’ concordance in quantifying tumor-infiltrating lymphocytes as a percentage and determine the cut-off for dichotomization characterized by the highest inter-observer agreement.

MATERIALS AND METHODS

Patient samples

A consecutive series of ductal carcinoma *in situ* was selected, based on organ (breast) and lesion codes in the electronic histopathological reports (LIS DaVinci, MIPS, Ghent, Belgium). All patients underwent breast-conserving surgery or mastectomy for ductal carcinoma *in situ* between January 1, 2014 and August 31, 2018 at the Cliniques universitaires Saint-Luc (Brussels, Belgium). Needle biopsies and vacuum-assisted core biopsies were excluded. Ductal carcinoma *in situ* with associated micro-invasive foci (1 mm) or frankly invasive carcinoma (invasive component >1 mm) were excluded. Hematoxylin and eosin stained slides were retrieved from the archives of the Department of Pathology (Cliniques universitaires Saint-Luc) and reviewed by one pathologist (H.D.), who selected one representative slide for each lesion.

Slides containing the biopsy site were avoided, as biopsy reactions hamper the assessment of myxoid stromal periductal changes and stromal inflammation. Resection specimens with limited amounts of residual ductal carcinoma *in situ* (i.e. one duct with ductal carcinoma *in situ*) were excluded from this study. Selected slides were scanned by an automated slide scanner with Z-stack feature (NanoZoomer 2.0-RS, Hamamatsu Photonics K.K., Hamamatsu City, Japan). Digital images were available on a pass-word protected online platform (DIH, Leica Biosystems, Dublin, Ireland). This study was approved by the Ethics Committee of the Cliniques universitaires Saint-Luc (2018/21NOV/443).

Participants

All participating pathologists (pseudonyms P1-P39) had to meet the following criteria: 1) being a board-certified pathologist with a special interest in breast disease, or equivalent; 2) actively working as a pathologist, either in an academic or non-academic pathology laboratory, or both; and 3) assess at least fifty primary oncologic breast cancer resection specimens per year, according to the EUSOMA-criteria for dedicated breast pathologists.¹⁹

DCISion set-up

All participants were invited to complete the DCISion questionnaire (Supplementary Fig. 1), which was partially based on the study of van Doonijeweert et al.¹³ Eleven questions were used to assess the experience (number of years in practice), work environment (academic and/or non-academic laboratory), daily work method (conventional light microscopy and/or digital pathology), weekly amount of time dedicated to breast pathology, the system used for ductal carcinoma *in situ* grading, and the habit of routinely reporting specific histopathological features. No training set was used. Instead, all pathologists were provided with the study protocol, which contained written definitions for each histopathological feature (Supplementary Fig. 2), the DCISion poster with exemplary photographs based on the previous study (Supplementary Fig. 3), and relevant literature concerning the applied histopathological definitions.^{20, 21} All participants received a log-in and password which allowed access to the digital slides during four months. Scores were entered in an Excel template. Completion of the informed consent form was a prerequisite for subsequent data processing.

Definitions for dichotomous histopathological assessment

Eight histopathological features were assessed dichotomously, based on previously determined cut-offs, and were illustrated in the DCISion poster.¹⁸ Nuclear grade was assessed as non-high *versus* high grade, after adaptation of the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) protocol for examination of ductal carcinoma *in situ* specimens.²² While the architectural types include solid, cribriform, papillary and micropapillary growth,²³ in the current study, ductal carcinoma *in situ* architecture was assessed as predominantly solid (50% solid growth) or predominantly non-solid (<50% solid growth), as previously described.^{18, 24} Both ductal carcinoma *in situ* architecture and nuclear grade were assessed regardless of the presence or absence of necrosis. Necrosis was classified into two categories as previously described: no or single cell necrosis *versus* any amount of comedonecrosis.¹⁸ Comedonecrosis was defined by areas of confluent dirty necrosis, i.e. confluent eosinophilic material, often containing ghost cells and karyorrhectic debris, and easily detected at low magnification. Intraductal calcifications within the ductal carcinoma *in situ* were scored as present or absent, regardless of the size and the number of ducts with calcifications.

The architecture of the periductal stroma was recorded as either sclerotic or myxoid. Sclerotic stroma resembles the regular fibrous mammary stroma and consists of regularly arranged collagen fibers. Myxoid stroma was defined as loosely arranged collagen fibers, often interspersed with an amorphous, slightly basophilic substance, as illustrated in the DCISion poster and in previous reports.^{18, 25, 26} Stromal architecture was divided into 2

categories (< 33% or ≥ 33% of ducts surrounded by myxoid stroma), and preferentially assessed at low magnification. If the ductal carcinoma *in situ* was located within adipose tissue without any surrounding fibrous tissue, the case was considered as predominantly sclerotic. Lobular cancerization was defined as the presence of ductal carcinoma *in situ* tumor cells within breast lobules, with preservation of the normal lobular architecture. Lobular cancerization was assessed as either absent or present in the digital slide, regardless of its extent.

The presence and extent of chronic inflammatory infiltrates in the periductal stroma (regardless of its architecture) was recorded in a semi-quantitative manner as previously described and was preferentially assessed at low magnification, distant from the biopsy site (if present).^{18, 24, 25, 27} Low stromal inflammation was defined as periductal stroma that is not infiltrated by lymphocytes, or that was infiltrated by few loosely arranged lymphocytes with apparent intervening stroma.²⁷ High stromal inflammation was defined as periductal stroma containing a chronic inflammatory infiltrate that consists of at least one lymphoid aggregate (i.e. any infiltrate that consists of lymphocytes abutting one-another without intervening collagenous stroma).²⁷ Lymphoid follicle formation could be present but was not a prerequisite. Assessment of the stromal architecture could be hampered by the density of the inflammatory infiltrate.

Assessment of tumor-infiltrating lymphocytes

Tumor-infiltrating lymphocytes were assessed according to the standardized method proposed by the International Immuno-oncology Biomarkers Working Group.^{20, 21, 28} The 'supplementary Figure 1' of Pruneri et al. was used as a visual aid during assessment.²⁹ Tumor-infiltrating lymphocytes percentages signified the percentage of lymphocytes related to the total periductal stromal surface area, which served as a denominator.^{20, 28} The percentage of tumor-infiltrating lymphocytes was assessed without considering the score for 'stromal inflammation'. Participants were asked to provide an average percentage for tumor-infiltrating lymphocytes surrounding all ducts affected by ductal carcinoma *in situ* in that particular slide.²⁸ Hotspots were not taken into account. Participants were also asked to assess tumor-infiltrating lymphocytes dichotomously, using an upfront cut-off (<50% versus ≥ 50%), as previously described by Pruneri et al.²⁹

Statistical analysis

Statistical analyses were performed with IBM SPSS statistics 25.0 (IBM Chicago, IL, USA) as previously reported.¹⁸ Pie charts were constructed in Excel (Excel Windows 10, Microsoft Corporation, Redmond, WA, VS). The arithmetic mean for each histopathological characteristic was calculated per lesion, to evaluate the distribution of each characteristic within the ductal carcinoma *in situ* cohort. The mean corresponds to the most commonly addressed category for a specific characteristic per lesion. Percentages of absolute agreement were determined, signifying the number of lesions with 100% concordance. Krippendorff's alpha (KA) reliability estimates were calculated using the 'Kalpha' macro provided by Hayes and Krippendorff (<http://afhayes.com/spss-sas-and-mplus-macros-and-code.html>). This macro was introduced in SPSS to compute Krippendorff's alpha for all dichotomous

data to investigate overall inter-observer variability per characteristic.^{30, 31} The number of bootstrap samples was set at 10.000.

The intraclass correlation coefficient was calculated for tumor-infiltrating lymphocytes assessed as a percentage and was interpreted according to Koo and Li.³² Intraclass correlation coefficient settings were: two-way mixed, single measures, absolute agreement. Cohen's kappa (*K*) values were calculated for all dichotomous variables for each observer duo (i.e. 741 kappa values for each dichotomous histopathological characteristic). The kappa's distribution was visualized by box-and-whisker plots. Interpretation of the kappa values was performed according to Landis and Koch.³³ Pearson's and Spearman's correlation tests were performed when appropriate, to investigate correlations between the degree of inter-observer variability and any possible confounder mentioned in the DCISion questionnaire. Multiple linear regression analysis was performed to correct for multiple potential confounders. All tests were two-sided. The statistical significance level was set at 0,05.

RESULTS

Characteristics of the DCISion participants

In total, 47 pathologists were invited to participate in this study. Thirty-nine pathologists (83%) from nine different countries representing three continents and thirty different laboratories responded. The DCISion questionnaire was completed by 38 pathologists (97%); one participant had skipped four questions regarding habits of reporting (Table 1). The participants had been practicing as certified pathologists for 13,7 years on average (range 1–30 years, excluding the years of training). Twenty-five pathologists (64%) work in an academic laboratory, twelve pathologists (31%) work in a non-academic laboratory and two pathologists (5%) work in both an academic and non-academic laboratory. The average weekly time dedicated to breast pathology amounts to less than one day for seven participants (18%), one to two days for ten participants (26%), and two to three days for four participants (10%). Eight participants (20%) spend between three and four days on breast pathology, and ten participants spend more than four days on breast pathology (26%). Three participants (8%) use both conventional and digital microscopy in daily practice; thirty-six participants (92%) only use conventional light microscopy on a daily basis.

Case selection

Since a previous report has shown that histopathological diagnosis based on a single slide is characterized by substantial inter-observer variability regarding the distinction between benign breast disease, ductal carcinoma *in situ* and (micro-)invasive breast cancer,³⁴ a consecutive series of 149 slides encoded as pure ductal carcinoma *in situ* was provided to all participants. Upon completion of the digital histopathological assessment, all the participants' comments regarding the eligibility of the ductal carcinoma *in situ* lesions were gathered and reviewed together with the digital images and the original glass slides and available archived immunohistochemical stained slides (performed by HD, CG and MRVB). In total, twenty-one lesions (14%) were removed from the series because of: 1) insufficient amount of ductal carcinoma *in situ* (i.e. one single duct) in one case; 2) insufficient size

(<2mm), favoring a diagnosis of atypical ductal hyperplasia, in six cases; 3) presence of micro-invasive carcinoma in six cases; 4) presence of a 2 mm focus of invasive lobular carcinoma in one case; 5) lesions falling short of being designated ductal carcinoma *in situ*, favoring a diagnosis of usual ductal hyperplasia, flat epithelial atypia or apocrine atypia in seven cases. Eventually, 128 unequivocal pure ductal carcinoma *in situ* (86%) remained for the final analyses. This method aimed to prevent the introduction of a selection bias, which was likely to occur upon removal of specific cases judged by a single pathologist.³⁴

Absolute agreement

The mean histopathological scores were calculated and designated Px (Figure 1A). Absolute agreement reflects the number of cases that were rated identically by all pathologists (100% agreement) or all but one pathologist (97% agreement; Table 2). Absolute agreement was lowest for nuclear atypia and lobular cancerization (6% and 7%, respectively).

Dichotomously assessed tumor-infiltrating lymphocytes showed the highest absolute agreement (59%). Of note, the high concordance was due to the rarity of ductal carcinoma *in situ* cases with 50% tumor-infiltrating lymphocytes (i.e. 12% of the cohort on average), resulting in high agreement on the cases with <50% tumor-infiltrating lymphocytes and no agreement at all on the cases with 50% tumor-infiltrating lymphocytes.

Overall inter-observer variability

Krippendorff's alpha was lowest for lobular cancerization (Krippendorff's alpha = 0,396), and slightly higher for stromal architecture (Krippendorff's alpha = 0,450) and nuclear atypia (Krippendorff's alpha = 0,422). The Krippendorff's alphas were fairly high for stromal inflammation (Krippendorff's alpha = 0,564), dichotomously assessed tumor-infiltrating lymphocytes (Krippendorff's alpha = 0,520), and comedonecrosis (Krippendorff's alpha = 0,539). The highest Krippendorff's alphas were observed for solid ductal carcinoma *in situ* architecture (Krippendorff's alpha = 0,602) and presence of intraductal calcifications (Krippendorff's alpha = 0,676).

Comparison of pathologists duos

As Krippendorff's alpha does not permit detailed evaluation of inter-observer variability among different pathologists, Cohen's kappa values were determined for each pathologist duo (Supplementary Tables 1–8). These kappa values confirmed that evaluation of calcifications and lobular cancerization was characterized by the highest and lowest concordance, respectively (Table 3). The narrowest interquartile range was observed for evaluation of calcifications, whereas stromal inflammation showed the highest dispersion of kappa values (Figure 1B).

Comparison of different cut-offs for tumor-infiltrating lymphocytes

The intraclass correlation coefficient was calculated for tumor-infiltrating lymphocytes assessed as a percentage, showing good overall agreement with an average of 0,821 (range 0,566 – 0,933; Supplementary Table 10). Since intraclass correlation coefficient and Krippendorff's alpha are different statistical measures, their values cannot be mutually compared. Therefore, all participants were asked to rate stromal tumor-infiltrating

lymphocytes *ad hoc* as a dichotomous variable by using a cut-off at 50%.²⁹ The tumor-infiltrating lymphocytes percentages were also dichotomized *post hoc* according to different cut-offs with 10% increments (10%, 20%, 30% and 40%). The number of cases with average low and high tumor-infiltrating lymphocytes according to each cut-off was compared with the number of ductal carcinoma *in situ* presenting with low and high 'stromal inflammation' (Figure 2), resulting in the following kappa values: 0,881 (10% cut-off), 0,837 (20% cut-off), 0,637 (30% cut-off), 0,437 (40% cut-off) and 0,245 (50% cut-off).

Similarly to the other dichotomously assessed histopathological features, Krippendorff's alphas were calculated for tumor-infiltrating lymphocytes according to the different cut-offs: the Krippendorff's alpha amounted 0,512 for the 10% cut-off, 0,669 for the 20% cut-off, 0,641 for the 30% cut-off, 0,604 for the 40% cut-off and 0,520 for the 50% cut-off. The kappa values were calculated for each cut-off per pathologist duo, which confirmed that the highest level of concordance was observed for the 20% cut-off (Figure 3).

Concordance for a combined risk score

Previously, a combined risk score for ductal carcinoma *in situ*, based on dichotomously assessed nuclear atypia, stromal architecture and stromal inflammation, was shown to be associated with recurrence risk after breast-conserving surgery. We calculated the combined risk score for each participant, based on the available individual histopathological characteristics: ductal carcinoma *in situ* with a combination of high nuclear grade, predominantly myxoid stromal architecture and high stromal inflammation were considered to have a high combined risk score. When any of these features was lacking in a ductal carcinoma *in situ* lesion, it was considered to have a low combined risk score. The overall agreement for this *post hoc* determined dichotomous variable was rather low (Krippendorff's alpha = 0,482). The kappa values were calculated for each cut-off per pathologist duo (Supplementary Table 9). Descriptive values for the kappa value of the combined risk score assessment are shown in Table 3.

Confounders

The influence of potential confounders on the (dis)agreement among pathologists was investigated, including: experience, time dedicated to breast pathology, work environment, classification system used for ductal carcinoma *in situ* grading (for nuclear atypia only), and habit of reporting the characteristic of interest. No significant associations were observed ($p > 0,05$), except for two. Inter-observer variability for lobular cancerization was significantly lower for pathologists claiming to always mention this feature, and highest for pathologists stating never to report this feature ($p = 0,014$). This observation was independent of the laboratory environment, experience and time dedicated to breast pathology. The degree of concordance for stromal inflammation was significantly higher for pathologists with more time dedicated to breast pathology ($p = 0,031$), independent of the laboratory environment and experience. This association was not observed for tumor-infiltrating lymphocytes assessment. The influence of a 'center effect' (i.e. pathologists working together in one center might show higher concordance) was not investigated as the number of colleagues was too low to allow sufficiently powered statistical analysis.

DISCUSSION

Grading systems for ductal carcinoma *in situ* classify these lesions in three categories, analogous to the Nottingham grading system for invasive breast cancer.^{35–38} Previous studies showed that ductal carcinoma *in situ* grading is characterized by substantial inter-observer variability, regardless of the grading system used.^{8–12, 39} The current *ad hoc* dichotomization as non-high *versus* high nuclear grade resulted in moderate agreement with an average kappa of 0,430, which is lower than the kappa values previously reported for *post hoc* dichotomization of ductal carcinoma *in situ* grade (0,55 by Rakha et al. and 0,53 by Van Bockstal et al.).^{18, 40} Nuclear atypia represents a biological spectrum, ranging from monotonous, slightly atypical nuclei to extreme pleomorphism. The upper ends of this spectrum are easily assessable, but a large grey zone exists in between. Heterogeneous morphology throughout a single ductal carcinoma *in situ* lesion can further hamper the adequacy and reproducibility of morphological assessment. Molecular and genomic studies provided evidence for a two-tier (low grade *versus* high grade) pathway in breast cancer development, wherein morphologic grade 2 ductal carcinoma *in situ* either cluster together with morphologic grade 1 or grade 3.^{41–43} Two-tier classification systems for morphological evaluation of dysplasia are currently common practice in other organ systems, such as gastrointestinal epithelial neoplasia, cervical and vulvar squamous intraepithelial lesions.^{44, 45} These classifications improved interrater concordance.^{44, 45} Logically, the discordance among observers is generally larger when the number of available categories increases. We therefore hypothesized that two-tier assessment might improve the current inter-observer variability in ductal carcinoma *in situ* grading.

We previously examined the inter-observer variability among 13 pathologists for multi-categorical histopathological features and reported the ‘ideal’ cut-offs for two-tier assessment after *post hoc* dichotomization.¹⁸ The prognostic value of this two-tier assessment was subsequently investigated in a cohort of 211 ductal carcinoma *in situ* patients, wherein high grade nuclear atypia, high stromal inflammation and myxoid stromal architecture were associated with increased overall recurrence risk after breast-conserving surgery.²⁷

The DCISion study applied these binary cut-offs upfront, which resulted in overall moderate agreement for all individual histopathological features, as well as for the *post hoc* combined risk score assessment (i.e. average kappa values between 0,41–0,60). Despite dichotomization, the inter-observer variability remains considerable for all features except intraductal calcifications. Of note, the current study setting lacked some essential features of the ‘real-life setting’ of histopathological diagnosis, e.g. immunohistochemistry, deeper levels and multiple tissue blocks. In addition, only three pathologists reported that they used digital pathology on a regular basis. These factors might have negatively influenced the degree of inter-observer variability. Nevertheless, the observed discordance might explain the variable prognostic power of different histopathological features in previous reports. For example, some large retrospective study cohorts and randomized trials could not confirm the association between high nuclear grade and increased recurrence risk.^{46–49} We were unable to investigate the impact of inter-observer variability on the recurrence risk stratification in this cohort, as most patients were recently diagnosed and adequate follow-up data were not

available. However, it is likely that inter-observer variability has an effect on the association with recurrence risk, i.e. some pathologists' scores might be associated with recurrence after breast-conserving surgery, and others might not. We aim to investigate this in the future, when the median follow-up time of this cohort reaches at least 5 years. Such a large-scale study would enable the definition of an allowed margin of error that is associated with a limited degree of discordance among pathologists, without significantly affecting recurrence risk stratification.

Robust prognostic histopathological markers require reproducibility of their assessment. Although features assessed as a continuous measure (such as stromal tumor-infiltrating lymphocytes) show overall acceptable concordance rates, the clinical decision making is often dichotomous (e.g. to treat or not to treat) and therefore generally requires a particular cut-off. For example, positive or negative hormone receptor status and HER2 status in invasive breast cancer will determine eligibility for (neo)adjuvant hormonal treatment and HER2-targeted therapy. Ductal carcinoma *in situ* patients who undergo breast-conserving surgery are often treated with adjuvant radiotherapy, and in some countries also with selective estrogen receptor-modulators or aromatase inhibitors. The choice for adjuvant treatment is mainly led by the recurrence risk, as up to 30% of non-irradiated ductal carcinoma *in situ* patients will develop a loco-regional recurrence after .⁵⁰ To date, margin size and histopathological features remain the cornerstone for recurrence risk stratification, although molecular tests such as Oncotype DX® DCIS (Genomic Health, Redwood City, CA, USA) are emerging.⁵¹ It is therefore important that pathologists speak the same language, to ensure that patients are treated in a consistent manner throughout different hospitals.

The DCISion study illustrates that the current histopathological evaluation, though acceptable, should improve. It is challenging to identify the precise causes of the observed disagreement. As mentioned above, discordance might have partly been induced by the use of digital histological images, since the majority of the participants commonly uses glass slides in daily practice. In general, diagnostic concordance between glass slides and digital slides is reported as high, but these studies usually investigate intra-observer equivalency.^{52, 53} However, inter-observer variability using glass slides will probably result in inter-observer variability using digital slides. Thus, the use of digital slides cannot entirely explain the observed discordance in the DCISion study. Intra-tumor heterogeneity, differences in selected regions of interest and different interpretations of the provided definitions may also account for increased discordance rates. The current study was limited to the evaluation of morphological features in a single H&E slide per lesion. Multiple H&E slides, additional levels and/or immunohistochemical markers might increase the degree of concordance in ductal carcinoma *in situ* grading. Future studies should explore whether there is a role for hormone receptor status and HER2 status in ductal carcinoma *in situ* risk stratification, as well as for immunohistochemical characterization of the stromal immune response.

Reproducibility and robustness of assessment are of particular importance for inclusion in the ongoing randomized trials that investigate non-inferiority of active surveillance. In these trials, risk assessment and eligibility are based on histopathological morphological features, although the LARRIKIN, COMET and LORETTA trials also take into account hormone

receptor and HER2 status.^{6, 7} It will take several years before these trials' findings are translated to routine clinical practice, but it is likely that watchful waiting will become a valid 'treatment' option. Deep learning algorithms or so-called 'artificial intelligence' might offer a solution to cope with the inter-observer variability in histopathological assessment, but we should be careful not to introduce inter-observer variability into these deep learning networks. If all the available deep learning algorithms are initially taught by only one or two pathologists, we might end up with similar inter-observer variability in artificial intelligence. It would be of interest to use the results of a multi-headed panel evaluation to enable the creation of a robust deep learning algorithm. The current DCISion study cohort might therefore be of interest, as it has been assessed by 39 different observers, representing thirty different laboratories. We aim to train an algorithm based on the ductal carcinoma *in situ* cases with 100% absolute agreement for a particular feature, and to explore its prognostic power in an independent ductal carcinoma *in situ* patient cohort with available follow-up data.

One limitation of the current study is the lack of a gold standard, since nobody knows who is 'right' or 'wrong' regarding the ductal carcinoma *in situ* cases with total lack of absolute agreement for a specific characteristic, such as a ductal carcinoma *in situ* lesion that was regarded as non-high grade by nineteen participants and as high grade by twenty other participants (Supplementary Fig. 4). A deep learning algorithm trained by using cases with 100% absolute agreement for a particular characteristic, might act as an artificial reference. This study also highlights that poor concordance (i.e. low kappa values) between two observers does not give any information on who is 'right' and who is 'wrong'. It only indicates that this particular feature has been perceived differently by two observers, and this poor concordance might be due to unclear definitions of the investigated characteristics. For instance, Harrison et al. recently showed that the threshold for comedonecrosis is highly variable, even among experienced breast pathologists.⁵⁴ A discussion and developing national / international guidance on how to define or re-define particular histopathological features, including their applied cut-offs, should be highly ranked when setting a research agenda for ductal carcinoma *in situ*.

This is of particular interest for stromal inflammation or tumor-infiltrating lymphocytes. The International Immuno-oncology Biomarkers Working Group proposed to quantify tumor-infiltrating lymphocytes in as much detail as possible, i.e. as a continuous variable by using percentages.²⁸ This method allows in-depth analysis of potentially clinically relevant cut-offs. In the DCISion study, we asked all participants to quantify tumor-infiltrating lymphocytes both dichotomously and as a percentage by using previously published photographs as a visual aid (such as supplementary Figure S1 of Pruneri et al.).^{20, 21, 29} By using 10% increments, we identified the 20% tumor-infiltrating lymphocytes cut-off as the cut-off associated with the highest inter-observer agreement. However, the 10% tumor-infiltrating lymphocytes cut-off corresponded best with semi-quantitative assessment of stromal inflammation, implying that most ductal carcinoma *in situ* with 10% stromal tumor-infiltrating lymphocytes present at least one dense aggregate of lymphocytes (designated 'high stromal inflammation'). Presence or absence of dense lymphoid aggregates (with or without tertiary follicle formation) might serve as a visual aid for pathologists to classify a particular ductal carcinoma *in situ* lesion as having low or high

stromal tumor-infiltrating lymphocytes. Since inter-observer agreement seems to be the highest for this 20% tumor-infiltrating lymphocytes cut-off, future work should also focus on investigating its prognostic value for ductal carcinoma *in situ* recurrence risk stratification after breast-conserving surgery. Overall agreement on stromal inflammation and stromal tumor-infiltrating lymphocytes was moderate in the DCISion study, but we may improve the current degree of concordance by providing clear definitions and adequate visual aids to enable training of pathologists. The website of the International Immuno-Oncology Biomarker Working Group (www.tilsinbreastcancer.org) provides such a useful tumor-infiltrating lymphocytes training tool for pathologists. The added value of immunohistochemical characterization of the immune infiltrate should also be explored, as well as the role of the location of stromal tumor-infiltrating lymphocytes. Toss et al. have recently demonstrated that dense touching tumor-infiltrating lymphocytes (i.e. tumor-infiltrating lymphocytes touching the basement membrane or away from it with maximum one lymphocyte of thickness) are associated with decreased recurrence-free survival.⁵⁵ This promising observation warrants validation in an independent patient cohort, as it indicates that spatial arrangement of tumor-infiltrating lymphocytes might be more important than the quantity of tumor-infiltrating lymphocytes. Moreover, the assessment of touching tumor-infiltrating lymphocytes seems reproducible, because it showed higher inter-observer concordance than assessment of tumor-infiltrating lymphocytes as a percentage.⁵⁵

CONCLUSIONS

Despite upfront dichotomous evaluation, the inter-observer variability for histopathological assessment of ductal carcinoma *in situ* remains considerable and is at most acceptable, although it varies between the evaluated features. This large-scale international multicenter study allowed us to compare two different methods to assess the inflammatory response in the periductal stroma. This comparison suggests that a semi-quantitative method (absence or presence of lymphoid aggregates) corresponds best with tumor-infiltrating lymphocytes assessed as a percentage when a cut-off at 10% is used for the latter. Nevertheless, a *post hoc* applied cut-off at 20% for stromal tumor-infiltrating lymphocytes results in the highest inter-observer concordance. Future research should validate the upfront use of this 20% cut-off, and investigate its relation with post-operative outcomes. Furthermore, the impact of the current degree of inter-observer variability on ductal carcinoma *in situ* prognostication should be explored. Forthcoming machine learning algorithms and additional immunohistochemistry might be useful to tackle these substantial diagnostic challenges.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

M. Van Bockstal is supported by the non-for-profit organization 'Foundation Against Cancer' (Clinical postdoctoral mandate 2019-089, Brussels, Belgium).

ABBREVIATIONS

ASCO/CAP	American Society of Clinical Oncology/College of American Pathologists
COMET	comparison of operative to monitoring and endocrine therapy trial
DCIS	ductal carcinoma <i>in situ</i>
Fig	figure
LARRIKIN	Low And Inter-mediate RIsK ductal carcinoma <i>IN situ</i> study
LORD	low-risk ductal carcinoma <i>in situ</i> trial
LORIS	low-risk ductal carcinoma <i>in situ</i> trial

REFERENCES

- Hanna WM, Parra-Herran C, Lu FI, Slodkowska E, Rakovitch E, Nofech-Mozes S. Ductal carcinoma in situ of the breast: An update for the pathologist in the era of individualized risk assessment and tailored therapies. *Mod Pathol* 2019; 32; 916–928. [PubMed: 30778169]
- Guo F, Kuo YF, Shih YCT, Giordano SH, Berenson AB. Trends in breast cancer mortality by stage at diagnosis among young women in the united states. *Cancer* 2018;124;3500–3509. [PubMed: 30189117]
- Jacklyn G, Morrell S, McGeechan K, Houssami N, Irwig L, Pathmanathan N et al. Carcinoma in situ of the breast in new south wales, australia: Current status and trends over the last 40 year. *Breast* 2018;37;170–178. [PubMed: 28882419]
- van Maaren MC, Lagendijk M, Tilanus-Linthorst MMA, de Munck L, Pijnappel RM, Schmidt MK et al. Breast cancer-related deaths according to grade in ductal carcinoma in situ: A dutch population-based study on patients diagnosed between 1999 and 2012. *Eur J Cancer* 2018;101;134–142. [PubMed: 30059817]
- Van Bockstal MR, Agahozo MC, Koppert LB, van Deurzen CHM. A retrospective alternative for active surveillance trials for ductal carcinoma in situ of the breast. *Int J Cancer* 2019; doi: 10.1002/ijc.32362.
- Toss M, Miligy I, Thompson AM, Khout H, Green AR, Ellis IO et al. Current trials to reduce surgical intervention in ductal carcinoma in situ of the breast: Critical review. *Breast* 2017;35;151–156. [PubMed: 28743035]
- Kanbayashi C, Iwata H. Current approach and future perspective for ductal carcinoma in situ of the breast. *Jp J Clin Oncol* 2017;47;671–677.
- Schnitt SJ, Connolly JL, Tavassoli FA, Fechner RE, Kempson RL, Gelman R et al. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. *Am J Surg Pathol* 1992;16;1133–1143. [PubMed: 1463092]
- 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;450–454. [PubMed: 9771444]
- Douglas-Jones AG, Morgan JM, Appleton MAC, Attanoos RL, Caslin A, Champ CS et al. Consistency in the observation of features used to classify duct carcinoma in situ (DCIS) of the breast. *J Clin Pathol* 2000;53;596–602. [PubMed: 11002762]
- 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;1056–1062. [PubMed: 9781641]

12. Sneige N, Lagios MD, Schwarting R, Colburn W, Atkinson E, Weber D et al. Interobserver reproducibility of the lagios nuclear grading system for ductal carcinoma in situ. *Hum Pathol* 1999;30;257–262. [PubMed: 10088542]
13. van Dooijeweert C, van Diest PJ, Willems SM, Kuijpers C, 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* 2019;174;479–488. [PubMed: 30539380]
14. van Dooijeweert C, van Diest PJ, Willems SM, Kuijpers CCHJ, van der Wall E, Overbeek LIH et al. Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the netherlands. *Int J Cancer* 2019; doi: 10.1002/ijc.32330.
15. Rakha EA, Aleskandarany MA, Toss MS, Mongan NP, ElSayed ME, Green AR et al. Impact of breast cancer grade discordance on prediction of outcome. *Histopathology* 2018;73;904–915. [PubMed: 29999192]
16. Dalton LW, Gerds TA. The advantage of discordance: An example using the highly subjective nuclear grading of breast cancer. *Am J Surg Pathol* 2017;41;1105–1111. [PubMed: 28614207]
17. Rea D, Francis A, Wallis M, Thomas J, Bartlett J, Bowden S et al. Confusion over differences in registration and randomization criteria for the loris (low-risk DCIS) trial. *Ann Surg Oncol* 2017;24;566–567. [PubMed: 29185098]
18. 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;923–932. [PubMed: 30168167]
19. Wilson AR, Marotti L, Bianchi S, Biganzoli L, Claassen S, Decker T et al. The requirements of a specialist breast centre. *Eur J Cancer* 2013;49;3579–3587. [PubMed: 23968730]
20. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B et al. Assessing tumor-infiltrating lymphocytes in solid tumors: A practical review for pathologists and proposal for a standardized method from the international immunooncology biomarkers working group: Part 1: Assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. *Adv Anat Pathol* 2017;24;235–251. [PubMed: 28777142]
21. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: Recommendations by an international TILs working group 2014. *Ann Oncol* 2015;26;259–271. [PubMed: 25214542]
22. Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL et al. Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. *Arch Pathol Lab Med* 2009;133;15–25. [PubMed: 19123730]
23. Siziopikou KP. Ductal carcinoma in situ of the breast: Current concepts and future directions. *Arch Pathol Lab Med* 2013;137;462–466. [PubMed: 23544935]
24. 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;103;94–100. [PubMed: 20517310]
25. Van Bockstal M, Lambein K, Gevaert O, De Wever O, Praet M, Cocquyt V et al. Stromal architecture and periductal decorin are potential prognostic markers for ipsilateral locoregional recurrence in ductal carcinoma in situ of the breast. *Histopathology* 2013; 63; 520–533. [PubMed: 23889174]
26. Van Bockstal M, Libbrecht L, Floris G, Lambein K. The Baader-Meinhof phenomenon in ductal carcinoma in situ of the breast. *Histopathology* 2016;69;522–523. [PubMed: 27028714]
27. Van Bockstal M, Lambein K, Smeets A, Slembrouck L, Neven P, Nevelsteen I et al. Stromal characteristics are adequate prognosticators for recurrence risk in ductal carcinoma in situ of the breast. *Eur J Surg Oncol* 2019;45;550–559. [PubMed: 30454971]
28. Dieci MV, Radosevic-Robin N, Fineberg S, van den Eynden G, Ternes N, Penault-Llorca F et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: A report of the international immuno-oncology biomarker working group on breast cancer. *Semin Cancer Biol* 2018;28;16–25. [PubMed: 29024776]

29. Pruneri G, Lazzeroni M, Bagnardi V, Tiburzio GB, Rotmensz N, DeCensi A et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (TILs) in ductal carcinoma in situ of the breast. *Ann Oncol* 2017;28;321–328. [PubMed: 28426105]
30. Krippendorff K Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research* 2004;30;411–433.
31. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 2007;1;77–89.
32. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15;155–163. [PubMed: 27330520]
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33;159–174. [PubMed: 843571]
34. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* 2015;313;1122–1132. [PubMed: 25781441]
35. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* 1991;19;403–410. [PubMed: 1757079]
36. 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;167–180. [PubMed: 7831528]
37. 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;1154–1157. [PubMed: 7723550]
38. Lagios MD. Heterogeneity of duct carcinoma in situ (DCIS): Relationship of grade and subtype analysis to local recurrence and risk of invasive transformation. *Cancer Lett* 1995;90;97–102. [PubMed: 7720048]
39. Schuh F, Biazus JV, Resetkova E, Benfica CZ, Ventura Ade 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;10;93. [PubMed: 26159429]
40. Rakha EA, Bennett RL, Coleman D, Pinder SE, Ellis IO, Pathology UKNCCfB. Review of the national external quality assessment (EQA) scheme for breast pathology in the UK. *J Clin Pathol* 2017;70;51–57. [PubMed: 27406051]
41. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Res* 2006;8;R61. [PubMed: 17069663]
42. Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchio C, Reis-Filho JS. Breast cancer precursors revisited: Molecular features and progression pathways. *Histopathology* 2010;57;171–192. [PubMed: 20500230]
43. Pang JM, Gorringer KL, Fox SB. Ductal carcinoma in situ - update on risk assessment and management. *Histopathology* 2016;68;96–109. [PubMed: 26768032]
44. Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD et al. The lower anogenital squamous terminology standardization project for hpv-associated lesions: Background and consensus recommendations from the college of american pathologists and the american society for colposcopy and cervical pathology. *Arch Pathol Lab Med* 2012;136;1266–1297. [PubMed: 22742517]
45. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000;47;251–255. [PubMed: 10896917]
46. Donker M, Litiere 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;4054–4059. [PubMed: 24043739]
47. Rudloff U, Jacks LM, Goldberg JI, Wynveen CA, Brogi E, Patil S et al. Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. *J Clin Oncol* 2010;28;3762–3769. [PubMed: 20625132]

48. Lazzeroni M, Guerrieri-Gonzaga A, Botteri E, Leonardi MC, Rotmensz N, Serrano D et al. Tailoring treatment for ductal intraepithelial neoplasia of the breast according to ki-67 and molecular phenotype. *Br J Cancer* 2013;108;1593–1601. [PubMed: 23579208]
49. Tunon-de-Lara C, André G, Macgrogan G, Dilhuydy JM, Bussi eres JE, Debled M et al. Ductal carcinoma in situ of the breast: Influence of age on diagnostic, therapeutic, and prognostic features. Retrospective study of 812 patients. *Ann Surg Oncol* 2011;18;1372–1379. [PubMed: 21108045]
50. Early Breast Cancer Trialists' Collaborative Group, Correa C, McGale P, Taylor C, Wang Y, Clarke M et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. *J Natl Cancer Inst* 2010;2010;162–177.
51. Nofech-Mozes S, Hanna W, Rakovitch E. Molecular evaluation of breast ductal carcinoma in situ with Oncotype DX DCIS. *Am J Pathol* 2019;189;975–980. [PubMed: 30605628]
52. Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C et al. Whole slide imaging equivalency and efficiency study: Experience at a large academic center. *Mod Pathol* 2019;32;916–928. [PubMed: 30778169]
53. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *Am J Surg Pathol* 2018;42;39–52. [PubMed: 28961557]
54. Harrison BT, Hwang ES, Partridge AH, Thompson AM, Schnitt SJ. Variability in diagnostic threshold for comedo necrosis among breast pathologists: Implications for patient eligibility for active surveillance trials of ductal carcinoma in situ. *Mod Pathol* 2019; doi: 10.1038/s41379-019-0262-4.
55. Toss MS, Miligy I, Al-Kawaz A, Alsleem M, Khout H, Rida PC et al. Prognostic significance of tumor-infiltrating lymphocytes in ductal carcinoma in situ of the breast. *Mod Pathol* 2018; 31; 1226–1236. [PubMed: 29559742]

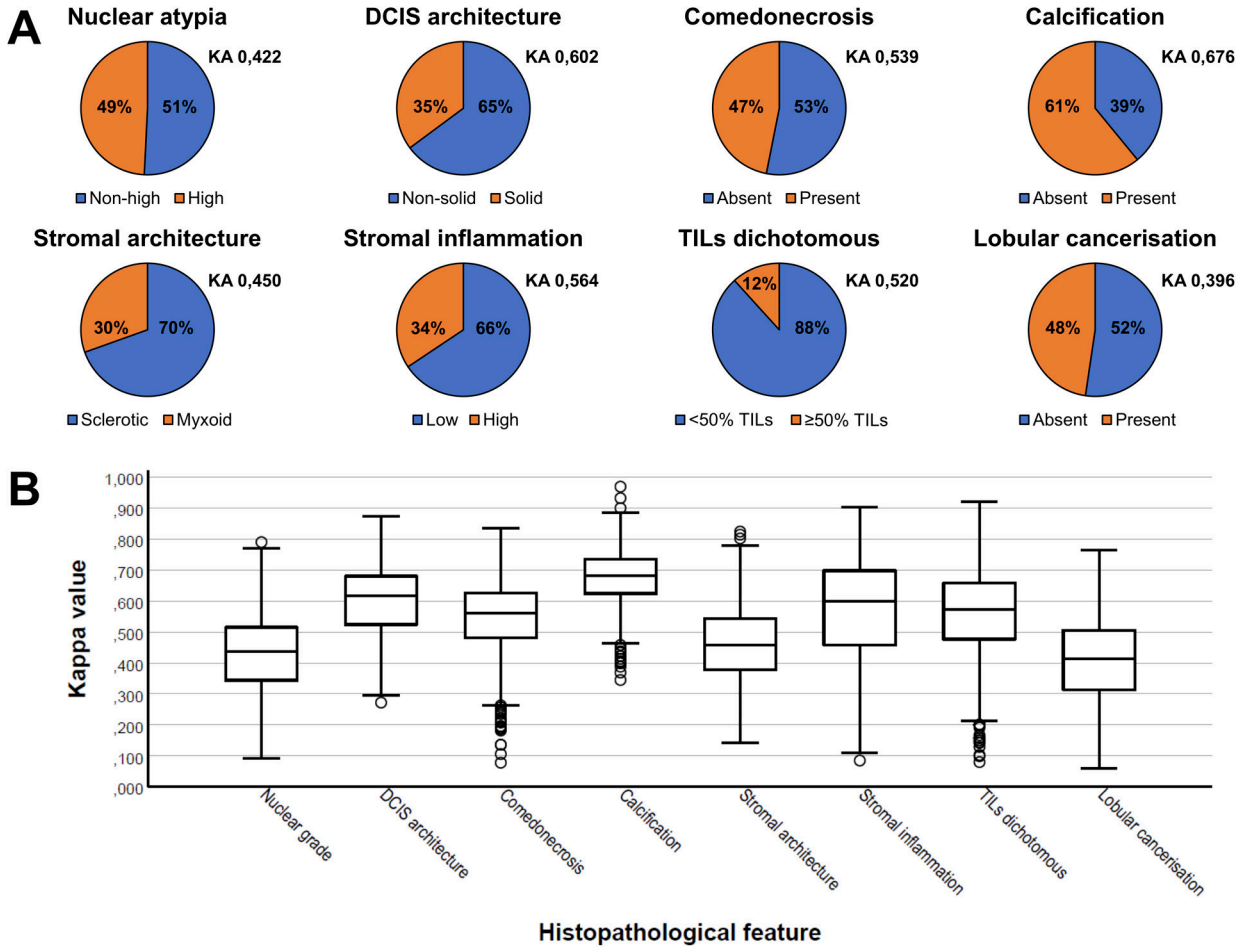


Figure 1. (A) Pie charts illustrating the histopathological assessment of the ‘average pathologist’ (Px) as a percentage, based on the arithmetic mean of all pathologists’ scores per dichotomously assessed histopathological feature. KA signifies the Krippendorff’s alpha statistic per feature. (B) Box-and-whisker plots illustrating the differences in the distribution in Cohen’s kappa values per pathologist duo per histopathological feature. Circles represent outliers. DCIS: ductal carcinoma *in situ*. TILs: tumor-infiltrating lymphocytes.

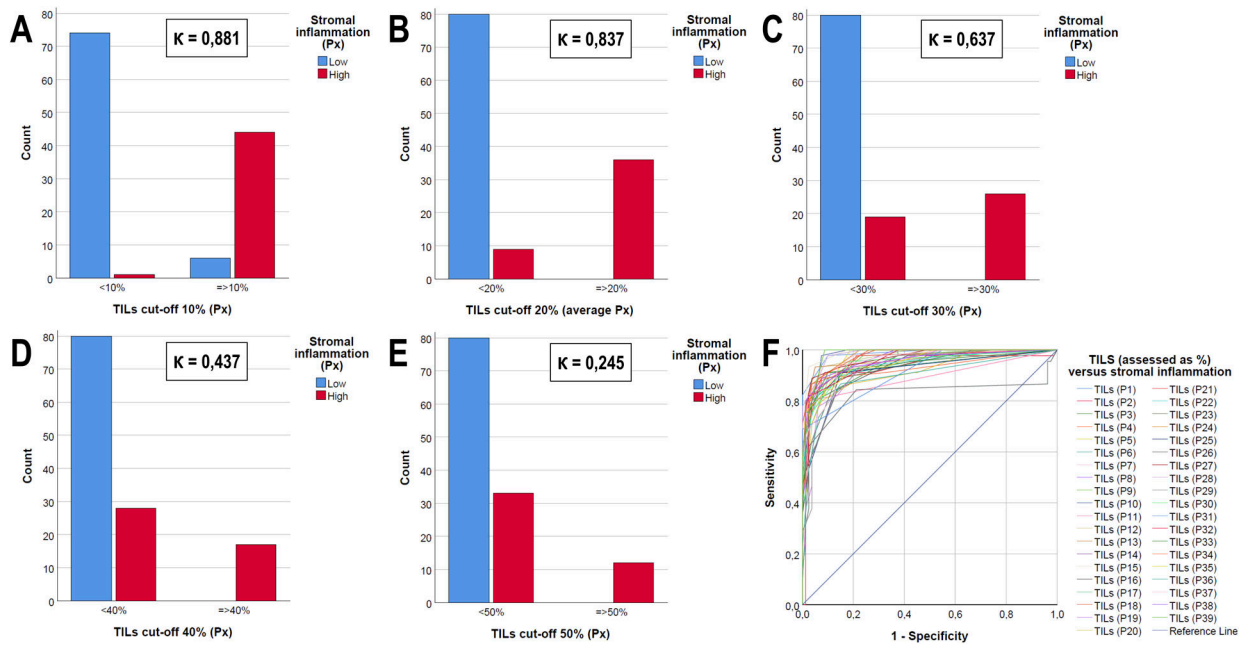


Figure 2.

Bar charts (A-E) and ROC-curve (F) illustrating the interrelationship between semi-quantitatively assessed stromal inflammation *versus* dichotomized tumor-infiltrating lymphocytes (TILs) assessed as a percentage in ductal carcinoma in situ (DCIS). The agreement between both parameters depends upon the cut-off used: Cohen's kappa values (K) are mentioned for a threshold of 10% (A), 20% (B), 30% (C), 40% (D) or 50% (E). A ROC-curve (F) illustrates this interrelationship for each pathologist separately.

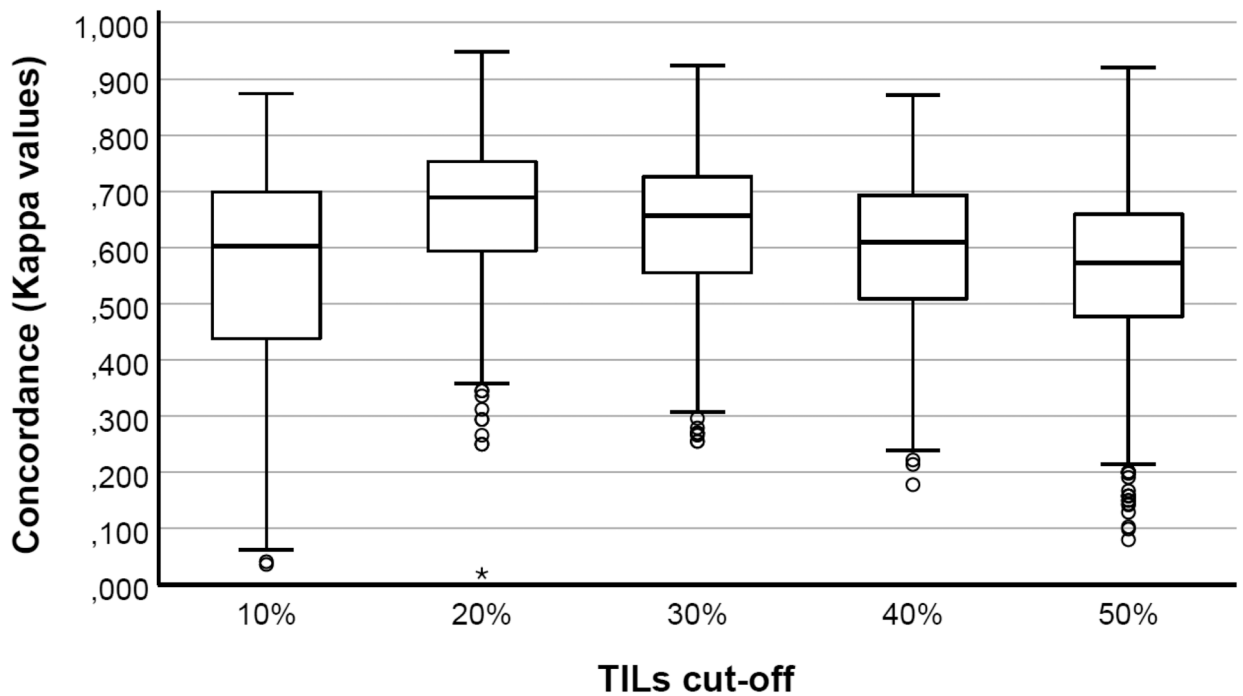


Figure 3.

Box-and-whisker plots illustrating the differences in the distribution in Cohen's kappa values per pathologist duo per cut-off for dichotomized tumor-infiltrating lymphocytes (TILs). The 20% and 50% cut-off for tumor-infiltrating lymphocytes assessed as a percentage showed the lowest and highest inter-observer variability, respectively. Circles represent outliers; asterisks represent extremes.

Table 1.

Distribution of the answers of the 39 participating pathologists regarding potential confounders that might influence the degree of inter-observer variability.

Multiple choice question	n (%)
In what kind of laboratory do you work?	
Academic	25 (64)
Non-academic	12 (31)
Both academic and non-academic	2(5)
What kind of microscope do you use for daily practice?	
Conventional light microscope	36 (92)
Digital and conventional light microscopy	3 (8)
How much of your time is dedicated to breast pathology?	
< 20% (< 1 day per week)	7 (18)
20 and < 40% (between 1 and 2 days per week)	10 (26)
40 and < 60% (between 2 and 3 days per week)	4 (10)
60 and < 80% (between 3 and 4 days per week)	8 (20)
80% (> 4 days)	10 (26)
Which grade do you mention in case of heterogeneity?	
Highest grade	35 (90)
Predominant grade	2 (5)
Both highest and predominant grade	2 (5)
Which grading system do you use for ductal carcinoma in situ grading?	
ASCO/CAP	13 (33)
Holland classification	0 (0)
Lagios classification	0 (0)
Pinder classification	2 (5)
Van Nuys classification	4 (10)
WHO classification	16 (41)
Other	0 (0)
Van Nuys + WHO classification	2 (5)
Van Nuys + ASCO/CAP classification	1 (3)
No answer	1 (3)
Do you routinely report tumor-infiltrating lymphocytes?	
Always mentioned in report	2 (5)
Sometimes mentioned in report	18 (46)
Never mentioned in report	18 (46)
No answer	1 (3)
Do you routinely report lobular cancerization?	
Always mentioned in report	10 (26)
Sometimes mentioned in report	19 (49)
Never mentioned in report	9 (23)
No answer	1 (3)

Multiple choice question	n (%)
Do you routinely report calcifications?	
Always mentioned in report	36 (92)
Sometimes mentioned in report	2 (6)
Never mentioned in report	0 (0)
No answer	1 (3)
Do you routinely report comedonecrosis?	
Always mentioned in report	35 (90)
Sometimes mentioned in report	4 (10)
Never mentioned in report	0 (0)
Do you routinely report ductal carcinoma in situ architecture?	
Always mentioned in report	29 (74)
Sometimes mentioned in report	8 (21)
Never mentioned in report	2 (5)

ASCO/CAP: American Society of Clinical Oncology/College of American Pathologists; WHO: world health organisation

Absolute agreement among pathologists regarding the evaluation of eight different histopathological features in ductal carcinoma in situ of the breast.

Table 2.

Histopathological feature		All 39 pathologist agreed		At least 38 of 39 pathologists agreed	
		Cases n (%)	Absolute agreement n (%)	Cases n (%)	Absolute agreement n (%)
Nuclear atypia	<i>Non-high</i>	5 (3.9)	8 (6)	10 (7.8)	20 (16)
	<i>High</i>	3 (2.3)		10 (7.8)	
Ductal carcinoma in situ architecture	<i>Non-solid</i>	29 (22.7)	35 (27)	44 (34.4)	56 (44)
	<i>Solid</i>	6 (4.7)		12 (9.4)	
Comedonecrosis	<i>Absent</i>	13 (10.2)	21 (16)	23 (18.0)	40 (31)
	<i>Present</i>	8 (6.3)		17 (13.3)	
Calcification	<i>Absent</i>	7 (5.5)	35 (27)	14 (10.9)	53 (41)
	<i>Present</i>	28 (21.9)		39 (30.5)	
Stromal architecture	<i>Sclerotic</i>	14 (10.9)	17 (13)	23 (18.0)	30 (23)
	<i>Myxoid</i>	3 (2.3)		7 (5.5)	
Stromal inflammation	<i>Low</i>	28 (21.9)	31 (24)	37 (28.9)	44 (34)
	<i>High</i>	3 (2.3)		7 (5.5)	
Tumor-infiltrating lymphocytes dichotomous	<i><50%</i>	75 (58.6)	75 (59)	76 (59.4)	90 (70)
	<i>50%</i>	0 (0)		14 (10.9)	
Lobular cancerization	<i>Absent</i>	6 (4.7)	9 (7)	16 (12.5)	21 (16)
	<i>Present</i>	3 (2.3)		5 (3.9)	

Table 3.

Descriptive statistics for the Cohen’s kappa values among 39 pathologists per histopathological feature of ductal carcinoma in situ of the breast.

Histopathological feature	Mean	SD	Minimum	Maximum	IQR	P5	P10	P25	P50	P75	P90	P95
Nuclear grade	0.430	0.124	0.091	0.790	0.171	0.207	0.272	0.344	0.437	0.515	0.586	0.627
Ductal carcinoma in situ architecture	0.600	0.012	0.272	0.873	0.156	0.391	0.444	0.524	0.617	0.680	0.727	0.766
Comedonecrosis	0.544	0.120	0.078	0.836	0.145	0.310	0.387	0.481	0.562	0.626	0.677	0.703
Calcification	0.675	0.094	0.345	0.969	0.110	0.510	0.554	0.624	0.682	0.734	0.791	0.817
Stromal architecture	0.461	0.122	0.142	0.824	0.166	0.256	0.309	0.377	0.458	0.543	0.622	0.673
Stromal inflammation	0.567	0.168	0.085	0.903	0.240	0.249	0.322	0.458	0.600	0.698	0.766	0.795
Lobular cancerization	0.403	0.130	0.059	0.765	0.192	0.172	0.218	0.314	0.414	0.506	0.564	0.598
Combined risk score	0.473	0.144	-0.079	0.881	0.204	0.235	0.280	0.373	0.480	0.577	0.653	0.692
Tumor-infiltrating lymphocytes dichotomous (10% cut-off)	0.553	0.186	0.036	0.874	0.261	0.188	0.261	0.438	0.603	0.699	0.750	0.776
Tumor-infiltrating lymphocytes dichotomous (20% cut-off)	0.668	0.121	0.020	0.948	0.159	0.434	0.505	0.594	0.689	0.753	0.809	0.836
Tumor-infiltrating lymphocytes dichotomous (30% cut-off)	0.637	0.124	0.255	0.924	0.171	0.405	0.465	0.555	0.656	0.726	0.781	0.811
Tumor-infiltrating lymphocytes dichotomous (40% cut-off)	0.595	0.132	0.178	0.871	0.184	0.351	0.401	0.509	0.610	0.693	0.752	0.788
Tumor-infiltrating lymphocytes dichotomous (50% cut-off)	0.556	0.148	0.080	0.920	0.182	0.246	0.353	0.477	0.573	0.659	0.725	0.761

IQR: interquartile range; P: percentile; P50: fiftieth percentile = median; SD: standard deviation.