



# The effect of an e-learning module on grading variation of (pre) malignant breast lesions

Carmen van Dooijeweert<sup>1</sup> · Ivette A. G. Deckers<sup>2</sup> · Emma J. de Ruiter<sup>1</sup> · Natalie D. ter Hoeve<sup>1</sup> · Celien P. H. Vreuls<sup>1</sup> · Elsken van der Wall<sup>3</sup> · Paul J. van Diest<sup>1</sup>

Received: 27 February 2020 / Revised: 16 April 2020 / Accepted: 16 April 2020 / Published online: 13 May 2020  
© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2020

## Abstract

Histologic grade is a biomarker that is widely used to guide treatment of invasive breast cancer (IBC) and ductal carcinoma in situ of the breast (DCIS). Yet, currently, substantial grading variation between laboratories and pathologists exists in daily pathology practice. This study was conducted to evaluate whether an e-learning may be a feasible tool to decrease grading variation of (pre)malignant breast lesions. An e-learning module, representing the key-concepts of grading (pre)malignant breast lesions through gold standard digital images, was designed. Pathologists and residents could take part in either or both the separate modules on DCIS and IBC. Variation in grading of a digital set of lesions before and after the e-learning was compared in a fully-crossed study-design. Multiple outcome measures were assessed: inter-rater reliability (IRR) by Light's kappa, the number of images graded unanimously, the number of images with both extreme scores (i.e., grade I and grade III), and the average number of discrepancies from expert-consensus. Participants were included as they completed both the pre- and post-e-learning set (DCIS-module:  $n = 36$ , IBC-module:  $n = 21$ ). For DCIS, all outcome measures improved after e-learning, with the IRR improving from fair (kappa: 0.532) to good (kappa: 0.657). For IBC, all outcome measures for the subcategories tubular differentiation and mitosis improved, with >90% of participants agreeing on almost 90% of the images after the e-learning. In contrast, the IRR for the subcategory of nuclear pleomorphism remained fair (kappa: 0.523 vs. kappa: 0.571). This study shows that an e-learning module, in which pathologists and residents are trained in histologic grading of DCIS and IBC, is a feasible and promising tool to decrease grading variation of (pre)malignant breast lesions. This is highly relevant given the important role of histologic grading in clinical decision making of (pre)malignant breast lesions.

## Introduction

Histologic grade is an important prognostic biomarker, which is widely used to guide breast cancer treatment [1–3]. In the Netherlands, grade currently indicates the need for adjuvant

chemotherapy in nearly a third of breast cancer patients [1, 4]. For ductal carcinoma in situ of the breast (DCIS), grade influences radiotherapy decisions [1, 5] and indicates the need for a sentinel lymph node procedure [1]. Moreover, histologic grade may become the single biomarker that decides whether DCIS patients should or should not be treated, as this is now being investigated by multiple clinical trials [6–9].

Subsequently, accurate, and reproducible grading is of major clinical importance. However, significant inter- and intra-laboratory variation in histologic grading of invasive breast cancer (IBC) and DCIS exists in daily pathology practice [4, 10]. Providing laboratories with feedback reports, in which laboratory-specific case-mix adjusted proportions per grade were benchmarked against other laboratories, resulted in a promising, yet small, decrease in grading variation [11]. As substantial differences in grade-specific proportions were observed between pathologists within individual laboratories [4, 10], we hypothesized that

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41379-020-0556-6>) contains supplementary material, which is available to authorized users.

✉ Paul J. van Diest  
p.j.vandiest@umcutrecht.nl

<sup>1</sup> Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands

<sup>2</sup> Foundation PALGA (the nationwide network and registry of histo- and cytopathology in the Netherlands), Houten, the Netherlands

<sup>3</sup> Department of Medical Oncology, University Medical Center Utrecht, Utrecht, the Netherlands

variation in grading may be primarily explained by differences in grading practices of individual pathologists.

Therefore, we believe that training of pathologists in the assessment of histologic grade could attribute to better synchronization, and thereby a decrease in grading variation of both IBC and DCIS. This is further supported by Elston and Ellis, who emphasize that grading should only be performed by trained pathologists [12]. Yet, in the Netherlands, pathologists or pathology residents are currently not specifically trained in histologic grading of (pre)malignant breast lesions.

Training of pathologists by e-learning seems feasible, as e-learning modules have been shown to decrease grading variation of dysplasia in colorectal adenomas [13] and to improve consistency in the histopathological diagnosis of sessile serrated colorectal lesions [14]. In addition, an e-learning module is easily accessible online for medical professionals throughout the country, or even worldwide, without the need for a live tutor or planned course days [15]. Thus, it enables pathologists to train when and where they want.

The aim of this study was to evaluate whether an e-learning module may be an effective tool to decrease grading variation of (pre)malignant breast lesions (i.e., DCIS and IBC) by studying variation in grading of a digital set of lesions by pathologists before and after the e-learning.

## Materials and methods

### E-learning design

The e-learning module was designed by the co-authors and agreement on its content was reached by three expert breast pathologists (PvD, CV, RG) and one expert breast pathologist's assistant (NtH). The e-learning presents the key-concepts of histologic grading of (pre)malignant breast lesions, including background information, discussion of the specific grading classifications for DCIS and IBC, and an extensive review of these criteria with example images.

The e-learning module consists of two separate modules, one on grading of IBC and one on grading of DCIS. Pathologists and residents could participate in either one or both modules. Separate analyses were conducted for the DCIS- and IBC-module.

All images in the e-learning module were derived from breast cancer cases from daily pathology practice in our institute based on consensus. All patient-related information was removed from all images in the e-learning module to comply with the General Data Protection Regulation.

### Classifications of histologic grading

For grading of DCIS, we used the classification of Holland [16] in our e-learning (Supplementary 1), as this is the

guideline recommended by the Dutch breast cancer guideline, and, as we know from our previous survey among pathologists [10], this seems to be the most widely used classification in daily pathology practice.

The modified Bloom and Richardson guideline (Elston–Ellis modification of Scarff–Bloom–Richardson grading system, also known as the Nottingham grading system) [17, 18] (Supplementary 2) was used in this e-learning module for IBC, as it is the most widely used grading system for IBC and it is globally incorporated in breast cancer guidelines [1, 3, 19, 20]. This classification combines the assessment of cell morphology (nuclear pleomorphism), the proliferation (mitotic count), and differentiation (tubule formation), resulting in a total score and subsequent grade [18].

### Recruitment of participants

Pathologists and pathology residents were invited to participate in this e-learning through the news-bulletin of the Dutch Society for Pathology (NVvP).

### Study design

#### Baseline grading variation

Before they could start the e-learning module, participants were obliged to assign a grade to 52 snapshots of DCIS lesions, as they would in daily practice, to determine the baseline variation in grading between participants for DCIS. As agreed upon by consensus of the expert-panel (PvD, CV, RG, NtH) 16, 19, and 17 of these snapshots were DCIS grades I, II, and III, respectively. Several examples of these snapshots can be found in Supplementary 3. All architectural patterns (solid, cribriform, papillary, and micropapillary) were present in these tests, and some cases included comedonecrosis.

For IBC, participants were obliged to score snapshots per subcategory of the modified Bloom and Richardson guideline [17, 18], i.e. tubular differentiation (19 snapshots; 7 category 1, 5 category 2, 7 category 3), nuclear pleomorphism (19 snapshots; 6 category 1, 6 category 2, 7 category 3), and mitosis (30 mitoses, 20 non-mitoses). With regard to mitoses, participants were asked to state whether they would count 50 figures as a mitosis or not. This was done because we hypothesized that recognizing mitosis is the bottleneck, rather than counting itself. Furthermore, this way, snapshots were sufficient, while whole slides are used for this in daily clinical practice (and pathologists select the field where they count). Several examples of these snapshots can be found in Supplementary 4. These snapshots were primarily derived from invasive ductal carcinoma cases and tubular, cribriform carcinomas for tubule formation, and some invasive lobular

carcinoma cases. Since snapshots for tubule formation, nuclear pleomorphism, and mitoses were taken from different cases, only improvement for individual subcategories could be assessed.

After scoring the pre e-learning set of DCIS lesions and/or all images of the IBC subcategories, participants would start the e-learning itself.

### Post e-learning grading variation

After completion of the e-learning, participants were asked to re-grade the same DCIS lesions, and/or IBC lesions to determine the post e-learning variation in grading between participants. All lesions were presented in a different order than prior to the e-learning. Only after this second round of grading, participants received feedback on their answers.

### Outcome measures and statistical analyses

Variation in grading was measured according to several outcome measures both pre- and post-e-learning. We included scores of participants who completed grading of both the pre- and post-e-learning set of lesions, for DCIS and/or IBC. Hence, a fully-crossed design was used [21].

As primary outcome measure, the inter-rater reliability (IRR) was calculated by Light's kappa [22], which is a type of kappa-statistic suitable for a fully-crossed study-design [21]. This overall kappa represents the arithmetic mean of kappa scores for all coder pairs [21, 22]. The 95% confidence interval for kappa was calculated using bootstrapping from 1000 replications. Interpretation of the kappa-statistics was performed according to the proposed interpretation of Cicchetti [23] (K < 0.40: poor, K 0.40–0.59: fair, K 0.60–0.74: good, K 0.75–1.00: excellent). Secondary outcome measures were the number of images scored unanimously (i.e. agreement by 90 and 100% of scoring participants) and the number of single lesions scored as both grades I and III (for DCIS) or both category one and three (for IBC: tubular differentiation, nuclear pleomorphism). Lastly, the average number of discrepancies with grade (DCIS) or sub-score (IBC) from the reference score by the expert panel (PvD, CV, RG, NtH) was determined, overall, and for subgroups of participants (expert breast pathologists, general pathologists, and residents).

All statistical analyses were performed using IBM SPSS Statistics version 25 and R [24].

## Results

### Participants

Thirty-six (DCIS) participants were included in the data-analyses for DCIS, while 21 participants were included in

**Table 1** Characteristics of e-learning participants.

Characteristics	DCIS ( <i>n</i> = 36)	IBC ( <i>n</i> = 21)
<b>Function, <i>n</i> (%)</b>		
Expert breast pathologist	10 (27.8%)	7 (33.3%)
General pathologist	20 (55.6%)	9 (42.9%)
Resident	6 (16.7%)	5 (23.8%)
<b>Years of experience, mean (SD)</b>		
Expert breast pathologist	13.2 (7.3)	13.1 (8.3)
General pathologist	9.5 (6.2)	8.6 (7.5)
Resident	2.7 (1.2)	2.8 (1.3)

the data-analyses for IBC, as they completed both the pre- and post-e-learning grading sets for the separate modules. The majority of participants identified themselves as general pathologist, followed by expert breast pathologists, and residents (Table 1). Expert breast pathologists executed their function for the largest numbers of years, with a mean of >13 years (Table 1).

### DCIS

All outcome measures for DCIS improved after the e-learning (Table 2). Overall agreement (IRR) improved from fair (kappa: 0.532) to good (kappa: 0.657) [23], and the amount of lesions graded unanimously by >90% (33/36) and 100% (36/36) of participants increased after e-learning, with best observed agreement for DCIS grade III. Interestingly, before the e-learning, almost 30% of DCIS images (*n* = 15) were graded as both grades I and III by the different participants, which decreased to <10% (*n* = 5) after the e-learning (Table 2). Lastly, the average number of discrepancies with the reference score by the expert panel decreased for all subgroups of participants.

### Invasive breast cancer

All outcome measures that contribute to grading, except for two outcome measures for nuclear pleomorphism, improved after e-learning (Table 3).

For tubular differentiation, overall agreement (IRR) improved from good (kappa: 0.653) to excellent (kappa: 0.846) [23], while the number of images graded unanimously by >90% (19/21) and 100% (21/21) participants increased as well (from 63.2% to 89.5%, and from 0.0% to 47.4%, respectively). The number of images scored both as category 1 and 3 was notably high before e-learning (78.9%), yet a substantial decrease was observed after e-learning (10.5%). Finally, the average number of discrepancies decreased for all subgroups of participants.

For nuclear pleomorphism, overall agreement (IRR) remained only fair (kappa: 0.523 vs. kappa: 0.571) after the

**Table 2** Variation in grading of ductal carcinoma in situ of the breast before and after e-learning (36 participants).

Outcome measure (total questions = 52)	PRE e-learning	POST e-learning
<b>IRR (Light's kappa (95% confidence interval))</b>	0.532 (0.457–0.591)	0.657 (0.582–0.730)
<b>Cases graded 100% unanimous (36/36)</b>	4 (7.7%)	9 (17.3%)
Grade I ( <i>n</i> = 16)	–	–
Grade II ( <i>n</i> = 19)	–	2 (10.5%)
Grade III ( <i>n</i> = 17)	4 (23.5%)	7 (41.2%)
<b>Cases graded &gt; 90% unanimous (33/36)</b>	13 (25.0%)	24 (46.2%)
Grade I ( <i>n</i> = 16)	–	6 (37.5%)
Grade II ( <i>n</i> = 19)	3 (15.8%)	6 (31.6%)
Grade III ( <i>n</i> = 17)	10 (58.8%)	12 (70.6%)
<b>Cases graded both grades I and III</b>	15 (28.8%)	5 (9.6%)
Grade I ( <i>n</i> = 16)	4 (25.0%)	–
Grade II ( <i>n</i> = 19)	11 (57.9%)	5 (26.3%)
Grade III ( <i>n</i> = 17)	–	–
<b>Average number of discrepancies</b>	12.8 (24.6%)	8.1 (15.6%)
Expert breast pathologists	10.4 (20.0%)	7.6 (14.6%)
General pathologists	12.7 (24.4%)	7.6 (14.6%)
Residents	17.0 (32.7%)	10.8 (20.8%)

e-learning module. The number of images graded 100% unanimous slightly increased (from 15.8 to 26.3%), while the number of images graded unanimously by >90% of participants remained stable (31.6% pre- and post-e-learning). Overall, for both outcome measures, best agreement was observed for nuclear pleomorphism category 3. After e-learning, an increase was observed for the number of images scored as both category 1 and 3 (0.0–15.8%), all of which were deemed category 2 by consensus of the expert-panel. Hence, for these lesions, the same lesion may be treated differently in daily clinical practice, based upon which pathologist graded it, as this is a difference of two points on the overall score, and most probably would lead to a different overall grade (3–5 = grade I, 6–7 = grade II, 8–9 = grade III). Lastly, the average number of discrepancies did show improvement after e-learning (27.1 to 18.8%).

For mitoses, overall agreement (IRR) showed improvement from good (kappa: 0.699) to excellent (kappa: 0.826). After the e-learning, an increase was observed for the number of lesions scored unanimously by >90% and 100% of participants (from 36.0% to 56.0%, and from 74.0% to 86.0%, respectively). The overall number of discrepancies decreased from 10.5% before e-learning to 6.6% after e-learning.

A few example images with low- and high- concordance after e-learning for both DCIS and IBC can be found in Supplementary 5.

## Discussion

This study shows that an e-learning, in which pathologists and residents are trained in histologic grading of DCIS and IBC, is a feasible and promising tool to decrease grading variation, as 18/20 outcome measures improved after e-learning. This is highly relevant considering the substantial grading variation of (pre)malignant breast lesions in current daily clinical practice [4, 10] bearing in mind that grading plays a decisive role in clinical decision making such as the indication for chemotherapy in IBC, and for a sentinel lymph node procedure, and partial breast irradiation in DCIS. Increased consensus with regard to grading will diminish variation in treatment and thereby most likely also patient outcome [4].

To evaluate the effect of our e-learning, we used multiple outcome measures (Tables 2 and 3). The IRR was chosen as outcome measure as this provides a way of quantifying the degree of agreement [21] between the e-learning participants on histologic grading. It is important to acknowledge that kappa-values may be influenced by the choice of the specific kappa statistic, and there is no clear guideline which specific kappa statistic to use for a fully-crossed design with multiple ( $\geq 2$ ) coders and an ordinal outcome measure (grades I–III, category 1–3) [21, 25–27]. Here, we chose Light's kappa [21], which uses the arithmetic mean of the kappa's of all possible coder pairs.

To provide further insights, we assessed the number of lesions scored unanimously (both by all, and by >90% of participants), and the number of lesions which were scored both grades I and III or category 1 and 3, as this outlines where difficulties may and may not lay (specific grades/subcategories).

For DCIS, almost 30% of the pre-e-learning images were scored as both grades I and III (of which the majority was deemed grade II by our expert panel). Although most discrepancies were only one grade apart, this may point out a challenge for current clinical trials, who focus on the safety of active surveillance in grades I and II DCIS patients [6–9], especially since central pathology review is not carried out in all trials [28]. This means that, in current daily pathology practice, grade III DCIS lesions may erroneously be deemed grade II and randomized to active surveillance. This underlines that central pathology review is essential for these trials, and raises the question whether histologic grade, in its current state, should be the single identifying biomarker for low-risk DCIS, as is also pointed out by Cserni et al. [29].

**Table 3** Variation in scoring of the 3 constituents of invasive breast cancer grading before and after e-learning (21 participants).

Tubular differentiation (total cases = 19)	PRE e-learning	POST e-learning
<b>IRR (Light's kappa (95% confidence interval))</b>	0.653 (0.529–0.739)	0.846 (0.745–0.918)
<b>Cases graded 100% unanimous (21/21)</b>	–	9 (47.4%)
Category 1 (n = 7)	–	4 (21.1%)
Category 2 (n = 5)	–	2 (10.5%)
Category 3 (n = 7)	–	3 (15.8%)
<b>Cases graded &gt;90% unanimous (19/21)</b>	12 (63.2%)	17 (89.5%)
Category 1 (n = 7)	5 (71.4%)	6 (85.7%)
Category 2 (n = 5)	2 (40.0%)	4 (80.0%)
Category 3 (n = 7)	5 (71.4%)	7 (100.0%)
<b>Cases scored both category 1 and 3</b>	15 (78.9%)	2 (10.5%)
Category 1 (n = 7)	5 (26.3%)	1 (14.3%)
Category 2 (n = 5)	4 (24.1%)	–
Category 3 (n = 7)	6 (31.6%)	1 (14.3%)
<b>Average number of discrepancies</b>	3.0 (15.5%)	1.0 (5.3%)
Expert breast pathologists	4.6 (24.1%)	1.3 (6.8%)
General pathologists	1.8 (9.4%)	0.8 (4.1%)
Residents	2.8 (14.7%)	1.0 (5.3%)
Nuclear pleomorphism (total cases = 19)	Pre e-learning	Post e-learning
<b>IRR (Light's kappa (95% confidence interval))</b>	0.523 (0.372–0.6314)	0.571 (0.443–0.672)
<b>Cases graded 100% unanimous (21/21)</b>	3 (15.8%)	5 (26.3%)
Category 1 (n = 6)	–	–
Category 2 (n = 6)	–	–
Category 3 (n = 7)	3 (42.9%)	5 (71.4%)
<b>Cases graded &gt;90% unanimous (19/21)</b>	6 (31.6%)	6 (31.6%)
Category 1 (n = 6)	1 (16.7%)	–
Category 2 (n = 6)	1 (16.7%)	1 (16.7%)
Category 3 (n = 7)	4 (57.1%)	5 (71.4%)
<b>Cases scored both category 1 and 3</b>	–	3 (15.8%)
Category 1 (n = 6)	–	–
Category 2 (n = 6)	–	3 (50.0%)
Category 3 (n = 7)	–	–
<b>Average number of discrepancies</b>	5.1 (27.1%)	3.6 (18.8%)
Expert breast pathologists	6.3 (33.1%)	4.0 (21.1%)
General pathologists	4.4 (23.2%)	2.8 (14.7%)
Residents	4.0 (21.1%)	3.8 (20.0%)
Mitoses (total lesions = 50)	PRE e-learning	POST e-learning
<b>IRR (Light's kappa (95% confidence interval))</b>	0.699 (0.586–0.785)	0.826 (0.746–0.891)
<b>Cases graded 100% unanimous (21/21)</b>	18 (36.0%)	28 (56.0%)
Mitosis (n = 30)	10 (33.3%)	18 (60.0%)
Non-mitosis (n = 20)	8 (40.0%)	10 (50.0%)
<b>Cases graded &gt;90% unanimous (19/21)</b>	37 (74.0%)	43 (86.0%)

**Table 3** (continued)

Mitoses (total lesions = 50)	PRE e-learning	POST e-learning
Mitosis (n = 30)	21 (70.0%)	25 (83.3%)
Non-mitosis (n = 20)	16 (80.0%)	18 (90.0%)
<b>Average number of discrepancies</b>	5.2 (10.5%)	3.3 (6.6%)
Expert breast pathologists	4.6 (9.1%)	3.9 (7.7%)
General pathologists	4.7 (9.3%)	2.1 (4.2%)
Residents	7.2 (14.4%)	4.2 (8.2%)

For IBC, most difficulties were observed for scoring nuclear pleomorphism, which is in line with previous studies [4, 30–33], and may be due to the fact that scoring of this subcategory is least quantitative, leaving most room for variation in interpretation. It should also be mentioned that we used snapshots of lesions (of very similar magnification), which makes interpreting nuclear pleomorphism somewhat more difficult, especially since comparison to other cells (for example epithelium) by zooming out was not possible. In contrast, for tubular differentiation and mitoses, >90% of participants agreed on almost 90% of the questions after the e-learning, indicating that for these two constituents of grade the learning effect is much greater.

A limitation of this study may be that participants graded the same digital set of lesions after finishing the e-learning. Participants could go through the e-learning (and both tests) at their own pace, even in multiple sessions, so it is unknown what the “wash-out time” was between the pre- and post-test. Therefore, it cannot be ruled out that, although the lesions were presented in a different order, participants “recognized” some pictures. However, we would like to emphasize that participants only received extensive feedback after the final test. In addition, grading the same set of lesions, was deemed the best way to observe the effect of the e-learning on grading variation as this enables comparing “baseline” variation before e-learning with grading variation after e-learning. Grading a different set of lesions after the e-learning would not enable us to distinguish between an effect of the e-learning or simply because grading of the post-e-learning set may have been “easier” or “more difficult”.

Furthermore, as complete consensus was not always obtained after e-learning, the images chosen based on “true grade/category” consensus of our expert panel may still not be the perfect examples. Nonetheless, we would like to emphasize that the average number of discrepancies from this “true grade/category” decreased after e-learning for both DCIS and all subcategories of IBC. Therefore, if this type of training by e-learning is to be enrolled on a larger scale, one could think of a larger consensus panel of expert



breast pathologists, especially since, for tubular differentiation and nuclear pleomorphism, most discrepancies from our expert panel were observed among this subgroup of participants.

A final limitation may be our decision to use mitosis identification in snapshots, rather than the selection of high power fields and actual mitosis counting. This was done because of practical reasons, however, we would like to emphasize that this indeed leaves out selection of high power fields, which may be a challenge, especially in a heterogeneous tumor, leaving room for grading variation within this subcategory.

Supported by the results of this study, we believe that training of pathologists in the assessment of histologic grade could further attribute to better synchronization and thereby a decrease in grading variation of both IBC and DCIS. An e-learning module as described in this study, in combination with monitoring variation in histologic grading of IBC and DCIS in daily clinical practice by laboratory- and pathologist specific feedback, may be considered by the Dutch Society of Pathology as one of the pathways for those pathologists that want to qualify as breast pathologist. Likewise, pathologists who participate in the Dutch national colorectal cancer screening program are already obliged to participate in an e-learning. In addition, the use of peer-learning programs for training and quality was identified as one of the strategies to improve patient access to high-quality oncologic pathology by Nass et al. [34]. Lastly, the e-learning may well be implemented in training programs of pathology residents.

While both laboratory-specific feedback and training of pathologists by e-learning may be promising tools to decrease grading variation, we believe that it is also important to emphasize that histologic grade is not a fact, nor the truth, but merely a model or tool that consists of a statistically computed set of cut-off values in a spectrum of histopathologic features to classify expected tumor behavior. Yet, these cut-offs are often used as hard criterion by clinicians in clinical decision making. For example, according to the Dutch breast cancer guideline, for a large group of patients ( $\geq 35$  years, N0-, HER2-, 1.1–2 cm) [1],  $\geq$ grade II indicates that they are eligible for chemotherapy, whereas grade I does not. Yet, the majority of patients ( $>75\%$ ) has a score on the switch point of grades, that is scores 5 or 6, and scores 7 or 8 [4]. Thus, for these patients, the difference of only one point on any of the subcategories of the Bloom and Richardson classification may already alter their overall histologic grade, and thus may have different therapeutic implications. Therefore, we believe that awareness, and understanding of the difficulties of histologic grading, among clinicians is crucial as well to improve clinical decision making for patients. Furthermore, in the current era of shared-decision making, these difficulties should also be discussed with patients, where relevant.

In conclusion, this study shows that an e-learning module, in which pathologists and residents are trained in histologic grading of DCIS and IBC, is a feasible and promising tool to decrease grading variation of (pre) malignant breast lesions. This is highly relevant given the important role of histologic grading in clinical decision making of DCIS and IBC whereby an influence on outcome cannot be ruled out.

**Acknowledgements** The authors thank all pathologists and residents that participated in this study. We also kindly acknowledge prof. dr. Roel Goldschmeding (RG) for participating in the expert panel of the e-learning module.

## Compliance with ethical standards

**Conflict of interest** All authors declare that there is no conflict of interest. This research was funded by the Quality Foundation of the Dutch Association of Medical Specialists (SKMS). The funding source had no role in the design and conduct of this study, nor did it have a role in the review or approval of the manuscript and the decision to submit the manuscript for publication.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. The Netherlands Comprehensive Cancer Organisation (IKNL). Oncoline: Breast Cancer Guideline, 2017. <https://www.oncoline.nl/borstkanker>.
2. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol.* 2008;26:3153–8.
3. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res.* 2010;12:207.
4. van Doonijeweert C, van Diest PJ, Willems SM, Kuijpers C, 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.* 2020;146:769–80.
5. Smith BD, Bellon JR, Blitzblau R, Freedman G, Haffty B, Hahn C, et al. Radiation therapy for the whole breast: Executive summary of an American Society for Radiation Oncology (ASTRO) evidence-based guideline. *Pr Radiat Oncol.* 2018;8:145–52.
6. 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;51:1497–510.
7. 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;51:2296–303.
8. Groen EJ, Elshof LE, Visser LL, Rutgers EJT, Winter-Warnars HAO, Lips EH, et al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). *Breast.* 2017;31:274–83.
9. Youngwirth LM, Boughey JC, Hwang ES. Surgery versus monitoring and endocrine therapy for low-risk DCIS: The COMET Trial. *Bull Am Coll Surg.* 2017;102:62–3.

10. van Doonijeweert 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;174:479–88.
11. van Doonijeweert C, van Diest PJ, Baas IO, van der Wall E, Deckers IA. Variation in breast cancer grading: the effect of creating awareness through laboratory-specific and pathologist-specific feedback reports in 16 734 patients with breast cancer. *J Clin Pathol.* 2020. <https://doi.org/10.1136/jclinpath-2019-206362>.
12. 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.* 2002;41:154–61.
13. Madani A, Kuijpers C, Sluijter CE, Von der Thusen JH, Grunberg K, Lemmens V, et al. Decrease of variation in the grading of dysplasia in colorectal adenomas with a national e-learning module. *Histopathology.* 2019;74:925–32.
14. IJspeert JE, Madani A, Overbeek LI, Dekker E, Nagtegaal ID. Implementation of an e-learning module improves consistency in the histopathological diagnosis of sessile serrated lesions within a nationwide population screening programme. *Histopathology.* 2017;70:929–37.
15. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. *Acad Med.* 2006;81:207–12.
16. 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;11:167–80.
17. 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;11:359–77.
18. 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;19:403–10.
19. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol.* 2007;25:5287–312.
20. Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. *J Clin Oncol.* 2002;20:3628–36.
21. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol.* 2012;8:23–34.
22. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull.* 1971;76:365–77.
23. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6:284–90.
24. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. <https://www.R-project.org>.
25. Warrens MJ. Inequalities between multi-rater kappas. *Adv Data Anal Classif.* 2010;4:271–86.
26. Berry KJ, Johnston JE, Mielke PW Jr. Weighted kappa for multiple raters. *Percept Mot Skills.* 2008;107:837–48.
27. Conger AJ. Integration and generalization of kappas for multiple raters. *Psychol Bull.* 1980;88:322–8.
28. 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.* 2020;146:1189–97.
29. Cserni G, Sejbén A. Grading ductal carcinoma in situ (DCIS) of the breast—what’s wrong with it? *Pathol Oncol Res.* 2019. <https://doi.org/10.1007/s12253-019-00760-8> (Epub ahead of print).
30. Boiesen P, Bendahl PO, Anagnostaki L, Domanski H, Holm E, Idvall I, et al. Histologic grading in breast cancer—reproducibility between seven pathologic departments. *South Sweden Breast Cancer Group. Acta Oncol.* 2000;39:41–5.
31. Frierson HF Jr., Wolber RA, Berean KW, Franquemont DW, Gaffey MJ, Boyd JC, et al. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am J Clin Pathol.* 1995;103:195–8.
32. Italian Network for Quality Assurance of Tumour Biomarkers (INQAT). Quality control for histological grading in breast cancer: an Italian experience. *Pathologica.* 2005;97:1–6.
33. Meyer JS, Alvarez C, Milikowski C, Olson N, Russo I, Russo J, et al. Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Mod Pathol.* 2005;18:1067–78.
34. Nass SJ, Cohen MB, Nayar R, Zutter MM, Balogh EP, Schilsky RL, et al. Improving cancer diagnosis and care: patient access to high-quality oncologic pathology. *Oncologist.* 2019;24:1287–90.