

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

Carmen van Doijeweert,¹ Paul J van Diest ¹, Inge O Baas,² Elsken van der Wall,² Ivette A Deckers³

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jclinpath-2019-206362>).

¹Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

²Department of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

³Foundation PALGA (the nationwide network and registry of histopathology and cytopathology in the Netherlands), Houten, The Netherlands

Correspondence to

Professor Paul J van Diest, Pathology, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; p.j.vandiest@umcutrecht.nl

Received 4 December 2019
Revised 2 March 2020
Accepted 29 March 2020
Published Online First
10 April 2020



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: van Doijeweert C, van Diest PJ, Baas IO, et al. *J Clin Pathol* 2020;**73**:793–799.

ABSTRACT

Aims Histological grade is widely used to guide the management of invasive breast cancer (IBC). Yet, substantial interlaboratory and intralaboratory grading variations exist in daily pathology practice. To create awareness and to facilitate quality improvement, feedback reports, containing case-mix-adjusted laboratory-specific grades benchmarked against other laboratories, were sent to the individual laboratories by 1 March 2018. We studied the effect of these feedback reports on interlaboratory grading variation up till 1 year later.

Methods Overall, 17 102 synoptic pathology reports of IBC resection specimens from 33 laboratories, obtained between 1 March 2017 and 1 March 2019 were retrieved from the Dutch Pathology Registry (PALGA). An overall deviation score (ODS), representing the sum of deviations from the grade-specific overall proportions, was calculated to compare the absolute deviation for all grades at once. Case-mix correction was performed by two multivariable logistic regression analyses, providing laboratory-specific ORs for high-grade versus low-grade IBC.

Results After feedback, the overall range between laboratories decreased by 3.8%, 6.4% and 6.6% for grades I, II and III, respectively. Though the mean ODS remained similar (13.8% vs 13.7%), the maximum ODS decreased from 34.1% to 29.4%. The range of laboratory-specific ORs decreased by 21.9% for grade III versus grades I–II.

Conclusions An encouraging decrease in grading variation of IBC was observed after laboratory-specific feedback. Nevertheless, the overall grading variation remains substantial. In view of the important role of grading in patient management, it is adamant that not only feedback should be provided on a regular basis but also other interventions, such as additional training, are required.

INTRODUCTION

To date, the histological grade is widely used to guide the therapeutic management of invasive breast cancer (IBC)^{1–4} as it remains one of the most well-established prognostic factors.^{2 3 5} When strictly adhering to the current Dutch guideline,¹ which is similar to the rest of Europe,⁴ grade even indicates the need for adjuvant chemotherapy in

approximately one third of patients with breast cancer in the Netherlands.⁶ Furthermore, grading is used to guide radiotherapy decisions^{1 4 7} and the use of genetic profiling tests.^{1 4 8–10}

Despite its important role in patient management, we previously showed that substantial, and clinically relevant, variation in the grading of IBC exists on a nationwide scale in daily clinical practice in the Netherlands.⁶ Studies in which multiple IBC lesions were graded by several pathologists also showed that reproducibility was no more than moderate.^{11–14} This suggests that patients may be undergraded and overgraded in specific pathology laboratories and or by specific pathologists, which may subsequently result in undertreatment and overtreatment of a substantial number of breast cancer patients.⁶ As this may influence outcome, including exposure to unnecessary toxicity, it is clear that standardised histological grading is of key importance.

The results of our previous study⁶ were sent to the individual laboratories as feedback reports by the nationwide Dutch Pathology Registry (PALGA) to facilitate quality improvement, as auditing and benchmarking improve the quality of breast cancer care.^{15–20} By benchmarking their laboratory-specific proportions per histological grade against other laboratories, pathologists in individual laboratories were enabled to discuss and reflect on their grading practices and could conclude that adaptations were necessary.

This study was conducted to examine the effect of the case-mix-adjusted, laboratory-specific feedback reports on the interlaboratory variation in histological grading of IBC using real-life data from synoptic (structured) pathology reports in the Netherlands.

METHODS

Data source

Data were retrieved from PALGA, the nationwide network and registry of histopathology and cytopathology in the Netherlands, which contains excerpts of all pathology reports from Dutch laboratories, with nationwide coverage since 1991.²¹ All data within the PALGA database are pseudonymised both in the laboratories and by a trusted third party (ZorgTTP, Houten, the Netherlands). All pathology laboratories were anonymised to the researchers

by PALGA in a final step. Laboratories that wanted to receive feedback on the pathologist level (in addition to the overall laboratory feedback) were asked to send their local pathologist information to PALGA as the PALGA database did not contain pathologist information before 2019. All data were retrieved and handled in compliance with the General Data Protection Regulation Act and this study was approved by the scientific and privacy committee of PALGA.

Study population

All synoptic pathology reports of patients with IBC resection specimens in the Netherlands between 1 March 2017 and 1 March 2019 were retrieved from PALGA (n=25 420) (online supplementary 1).

Overall, 38 of 42 Dutch pathology laboratories used the synoptic (PALGA) protocol from 1 March 2017 and onwards. Of these laboratories, we included those that reported at least 50 IBC resection specimens per year.

We excluded all resection specimen reports of patients with neoadjuvant treatment as grading may be influenced by chemotherapy.^{22–24} Furthermore, synchronous IBCs, defined as an ipsilateral lesion within 6 months of the previous IBC resection report, were considered paired measurements of which we only included the first report (online supplementary 1).

Per pathology report, we extracted patient characteristics (sex, age and type of surgery) and tumour characteristics (tumour size, histological subtype, histological grade, oestrogen receptor (ER) and progesterone receptor (PR) status, and human epidermal growth factor receptor 2 (HER2) status). Reports with any missing data on one of the patients or tumour characteristics were excluded from further data analysis (online supplementary 1).

Feedback reports

Laboratory-specific feedback reports, regarding the variation in the grading of IBC between 1 January 2013 and 31 December 2016, were sent to the laboratories by PALGA by March 1 2018. These feedback reports showed the laboratory-specific proportions per histological grade, benchmarked against the overall national proportions and the proportions of the other anonymised laboratories. Thereby, laboratories were enabled to discuss and reflect on their grading practice, and perhaps conclude that adaptations were necessary. The general feedback report is available on the PALGA website (in Dutch only).²⁵

Ten laboratories provided (coded) pathologist information for their data, which gave these pathologists the advantage to benchmark their own grading practice against other pathologists in their laboratory and to the national mean. According to the literature, this type of individual feedback is more effective than providing general (laboratory level) data only.^{26–29} Feedback reports were sent to the laboratories by 1 March 2018, which resulted in a prefeedback group of synoptic pathology reports from 1 March 2017 to 1 March 2018, and a postfeedback group of synoptic pathology reports from 2 March 2018 to 1 March 2019.

Histological grading

Histological grading of IBC was determined according to the modified Bloom and Richardson guideline (Elston-Ellis modification),^{30–31} with a score of 1–3 on its three components (tubule formation, nuclear pleomorphism and mitotic count). This results in a total score and subsequent grade (3–5=grade I, 6–7=grade II, 8–9=grade III).

Statistical analysis

Patient and tumour characteristics were summarised and differences between the prefeedback and postfeedback group were tested by means of a χ^2 test for categorical variables and by a non-parametric Mann-Whitney U test for continuous variables.

Overall proportions per grade (I, II and III) were determined before and after the feedback reports and considered the national proportions. The absolute differences from the national proportion per laboratory are presented in bar charts per grade for the prefeedback and postfeedback periods. Laboratories that also received feedback on the pathologist level are indicated by striped bars.

An overall deviation score (ODS) was computed to compare the absolute deviation for all three grades at once. The ODS was calculated by the sum of absolute deviations from the grade-specific national proportions per period (prefeedback and postfeedback). Differences in ODS of individual laboratories before and after feedback were compared using a Wilcoxon signed-rank test.

As a possible way to interpret the type of change in laboratories after feedback, we used multiple definitions of change. First, we arbitrarily defined laboratories with an absolute change of $\leq 2\%$ as 'not shifting'. Second, in the case of an absolute change of $>2\%$, we defined two types of change. Laboratories with a smaller deviation from the overall mean were defined as 'less deviant'. Similarly, laboratories that became more deviant from the overall mean were defined as 'more deviant'.

To compare relative differences among laboratories, we used a logistic regression model, providing ORs and 95% CIs per laboratory. We performed two logistic regression analyses, with different definitions of low-grade and high-grade IBC, as there is no clear binary cut-off in clinical practice. For example, grade III is considered a risk factor (high grade) according to radiotherapy guidelines,^{1–4–7} whereas according to chemotherapy guidelines grades II–III is considered a risk factor (high grade) with possible subsequent therapy consequences.^{1–4} Therefore, in one logistic regression analysis, low-grade IBC was defined as grades I–II and high-grade IBC as grade III, whereas in the other logistic regression analysis, low-grade IBC was defined as grade I and high-grade IBC as grades II–III.

For the choice of reference laboratory, we arbitrarily chose the laboratory best resembling the national distribution with regard to the specific logistic regression analysis. Multivariable logistic regression analyses were performed to correct for differences in case mix. Case-mix variables were selected based on our previous research⁶ and included age, tumour size, type of surgery, histological subtype, HER2 status and hormone receptor status. Hormone receptor status was considered positive when either or both the ER- or PR-receptor status were positive and was taken into account as a binary variable (either positive or negative). According to the current Dutch guideline,¹ the receptor status for ER and PR is considered positive when $\geq 10\%$ of the tumour cells show ER and PR staining by immunohistochemistry (IHC). The overall number of men was too low to take into account in a multivariate model, however, men did not cluster in specific laboratories. To compare differences in the case-mix-adjusted ORs of the individual laboratories, we calculated the positive OR difference (ie, the difference of a laboratory-specific OR to the reference OR of 1.00) both prefeedback and postfeedback and compared the differences of the individual laboratories by a Wilcoxon signed-rank test for both multivariable logistic regression analyses (grade I vs grades II–III and grades I–II vs grade III).

Lastly, the effect of feedback on the pathologist level was tested by comparing the mean ODS before and after feedback, between the laboratories that received feedback both on pathologist and laboratory levels and laboratories that received feedback on the laboratory level only. In addition, type of change between these groups was compared by means of a χ^2 test.

All statistical analyses were performed by using IBM SPSS Statistics V.25.0.0.2. Values <0.05 were considered statistically significant.

RESULTS

Patient, tumour, and laboratory characteristics

A total of 17 102 IBC synoptic resection specimen reports from 16 734 patients were included in our data analysis. For some patients, more than one pathology report was included as this concerns either a bilateral tumour or an ipsilateral tumour that was reported >6 months after the first diagnosis (online supplementary 1). Of the included reports, 8767 were reported before and 8335 were reported after feedback reports were sent to the laboratories by PALGA.

All patients originated from 33 of 42 Dutch pathology laboratories, as 4 laboratories did not implement synoptic reporting between 1 March 2017 and 1 March 2019 and 5 laboratories graded <50 IBC lesions within the synoptic PALGA protocol per period (prefeedback and/or postfeedback). The number of synoptic IBC reports per laboratory ranged from 64 to 613 (median 239) in the year before the feedback reports, whereas the number of synoptic pathology reports per laboratory in the year after the feedback reports ranged from 52 to 637 (median 207). Characteristics of all included IBC resection specimen reports are listed in table 1.

The overall mean age (SD) at diagnosis was 63.2 (11.9) years and patients were primarily female individuals (99.2%). Breast-conserving surgery was performed in approximately two thirds of patients (68.2%). The majority of tumours were of ductal (not otherwise specified) subtype (78.2%), with a positive ER/PR status (89.8%), whereas only a small minority of tumours had a positive HER2 status (7.9%). Most characteristics, including age, sex, tumour size, type of surgery and histological subtype, we similarly distributed prefeedback and postfeedback. A minimal but significant increase of hormone-receptor-positive tumours was observed after the feedbacks reports, whereas a significant decrease was observed for HER2-positive tumours ($p=0.010$).

Overall national proportions for IBC grades I, II and III were, respectively, 30.5%, 49.5% and 20.0% before the feedback reports, whereas IBC grades I, II and III were reported in, respectively, 32.0%, 49.2% and 18.8% after the feedback reports ($p=0.048$).

Interlaboratory differences in histological grading

After feedback, the total range between laboratories decreased for all grades; 3.8% for grade I (from 17.5%–45.5% to 17.3%–41.5%), 6.4% for grade II (from 34.3%–64.5% to 35.0%–58.8%) and 6.6% for grade III (from 10.9%–37.1% to 9.9%–29.5%) (figure 1).

The mean overall ODS remained similar after feedback (13.8 vs 13.7%), which is also reflected by the similar ODS of individual laboratories (Wilcoxon signed-rank test, $p=0.955$). The maximum ODS, however, decreased from 34.1% to 29.4% (figure 2). Overall, 11 (33.3%), 13 (39.4%) and 16 (48.5%) of laboratories showed no shift ($\leq 2\%$) after feedback for grades I, II and III (table 2).

Table 1 Characteristics of the 17 102 included invasive breast cancer resection specimen reports from the PALGA database between 1 March 2017 and 1 March 2019

	Total (N=17 102)	PRE (N=8767)	POST (N=8335)
Histological grade, n (%)			
Grade I	5337 (31.2)	2672 (30.5)	2665 (32.0)
Grade II	8445 (49.4)	4344 (49.5)	4101 (49.2)
Grade III	3320 (19.4)	1751 (20.0)	1569 (18.8)
Age (years)*	63.2 (11.9)	63.0 (11.9)	63.4 (11.9)
Sex, n (%)			
Female	16 971 (99.2)	8962 (99.1)	8279 (99.3)
Male	131 (0.8)	75 (0.9)	56 (0.3)
Tumour size (cm)*	1.8 (1.3)	1.8 (1.3)	1.7 (1.3)
Type of surgery, n (%)			
Mastectomy	5433 (31.8)	2844 (32.4)	2589 (31.1)
Breast conserving	11 669 (68.2)	5923 (67.6)	5746 (68.9)
Histological subtype, n (%)			
Ductal	13 373 (78.2)	6886 (78.5)	6487 (77.8)
Lobular	2353 (13.8)	1172 (13.4)	1181 (14.2)
Other	1376 (8.0)	709 (8.1)	667 (8.0)
ER/PR status, n (%)			
Negative	1740 (10.2)	939 (10.7)	801 (9.6)
Positive	15 362 (89.8)	7828 (89.3)	7534 (90.4)
HER2 status, n (%)			
Negative	15 753 (92.1)	8030 (91.6)	7723 (92.7)
Positive	1349 (7.9)	737 (8.4)	612 (7.3)

PRE includes cases before, and POST includes cases after sending feedback reports to individual laboratories.

*Mean (SD).

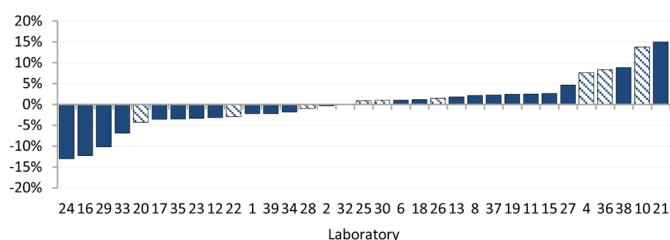
ER/PR, oestrogen receptor/progesterone receptor; HER2, human epidermal growth factor receptor 2.

Among laboratories that shifted $>2\%$ after feedback, the number of laboratories that became more deviant was similar to the number of laboratories that became less deviant after feedback (30.3% vs 36.4% for grade I, 30.3% vs 30.3% for grade II and 24.2% vs 27.3% for grade III, respectively) (table 2).

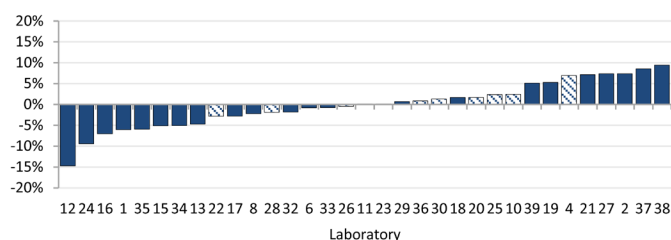
For the multivariate logistic regression analysis of grade III versus grades I–II, laboratory 30 had the lowest mean deviation from the national proportion before and after feedback for grade III (0.2%) and was chosen as a reference laboratory. Before feedback, adjusted ORs ranged from 0.37 (95% CI: 0.21 to 0.67) to 2.15 (95% CI: 1.26 to 3.67). After feedback, the range of adjusted ORs decreased from 0.37 (95% CI: 0.20 to 0.68) to 1.76 (95% CI: 1.01 to 3.07) (figure 3A). Consequently, the absolute overall OR range decreased by 21.9% from 1.78 to 1.39. Positive OR differences of the individual laboratories did not significantly differ (Wilcoxon signed-rank test, $p=0.702$). For the multivariate logistic regression analysis of grades II–III versus grade I, laboratory 32 had the lowest mean deviation from the national proportion before and after feedback for grade I (0.9%) and was chosen as a reference laboratory. Before feedback, adjusted ORs ranged from 0.48 (95% CI: 0.29 to 0.77) to 2.00 (95% CI: 1.10 to 3.65), resulting in the absolute overall OR range of 1.52. After feedback, the range of adjusted ORs slightly increased (10.5%), that is, from 0.42 (95% CI: 0.26 to 0.67) to 2.10 (95% CI: 1.24 to 3.58) with corresponding absolute overall OR range of 1.68 (figure 3B). Positive OR differences of the individual laboratories did not significantly differ (Wilcoxon signed-rank test, $p=0.640$).

Original research

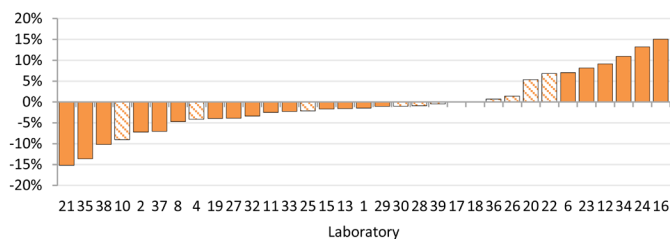
A: Before feedback reports: grade I



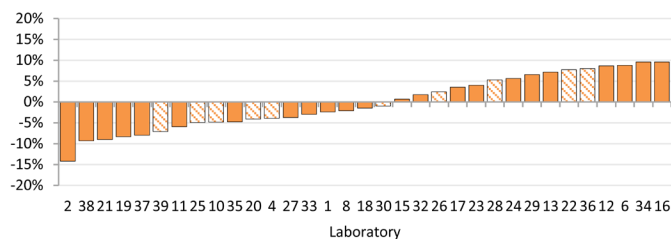
D: After feedback reports: grade I



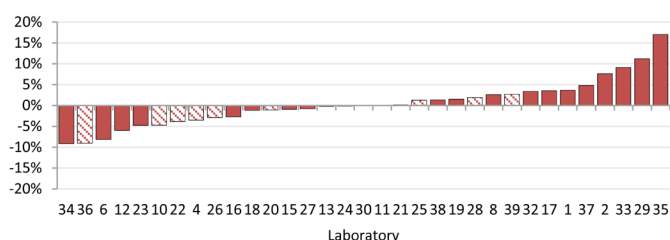
B: Before feedback reports: grade II



E: After feedback reports: grade II



C: Before feedback reports: grade III



F: After feedback reports: grade III

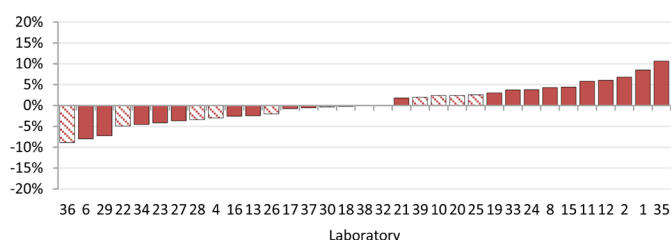


Figure 1 Interlaboratory (n=33) variation in histological grading of invasive breast cancer before (A–C) (n=8767) and after feedback reports (D–F) (n=8335). Percentages per laboratory show the absolute deviation from the national proportion per histological grade for grade I (A+D), grade II (B+E) and grade III (C+F). Laboratory numbers for all subfigures (A–F) correspond. All laboratories are ranged from lower (negative values) to higher proportions (positive values). Striped bars indicate laboratories that received feedback on the pathologist level (n=10).

Feedback on the pathologist level

Ten of the included laboratories received feedback both on laboratory and on pathologist levels (figures 1 and 2; striped bars). Although the mean prefeedback ODS of these laboratories was lower (10.7%) than the mean prefeedback ODS of laboratories

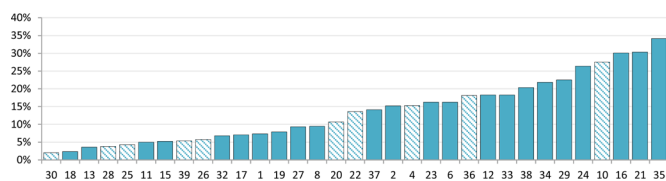
that only received feedback on the laboratory level (15.1%), both groups did not show noteworthy changes after feedback (10.7% and 15.0%, respectively). Furthermore, the type of change in laboratories after feedback did not significantly differ for both groups (table 2). Yet, a significantly higher proportion of laboratories that received feedback on the pathologist level showed no shift after feedback for grade I, whereas a similar pattern was observed for grade III ($p=0.103$), whereas this was not observed for grade II.

DISCUSSION

Using nationwide data from structured (synoptic) pathology reports, we studied case-mix-adjusted, laboratory-specific feedback reports as an intervention to decrease interlaboratory variation in histological grading of IBC. This study shows an encouraging decrease in nationwide grading variation after sending feedback reports to individual laboratories, reflected by a decrease in absolute range of grade-specific proportions after feedback for all grades (I–III), the decrease of the maximum ODS and the range of laboratory-specific ORs showing a notable decrease of 21.9% for grade III versus grades I–II. Overall, this shows that most deviant laboratories became less deviant, whereas the overall mean ODS and positive OR differences of individual laboratories did not significantly differ.

The primary aim of the laboratory-specific feedback reports was to create awareness among pathologists by highlighting that grading variation in current clinical practice is substantial and

A: Before feedback reports: overall deviation score (ODS)



B: After feedback reports: overall deviation score (ODS)

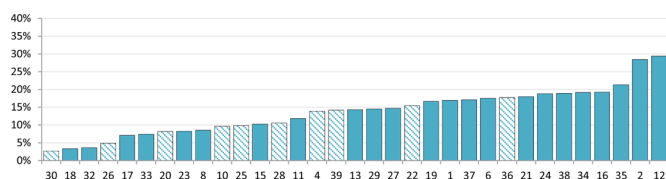


Figure 2 Interlaboratory (n=33) variation in histological grading of invasive breast cancer before (A) (n=8767) and after feedback reports (B) (n=8335). Each bar represents the ODS per laboratory. Laboratory numbers for (A) and (B) correspond. All laboratories are ranged from lower to higher ODS. Striped bars indicate laboratories that received feedback on the pathologist level (n=10). ODS, overall deviation score.

Table 2 Type of absolute change in the 33 included laboratories after feedback per histological grade, with differentiation between laboratories that received feedback on the pathologist level (n=10) and laboratories that only received feedback on the laboratory level (n=23)

Type of change	Total laboratories (N=33)	Laboratories with feedback on the pathologist level (N=10)	Laboratories without feedback on the pathologist level (N=23)	P value
Grade I, n (%)				
No shift (≤2%)	11 (33.3)	6 (60.0)	5 (21.7)	0.032*
Shift (>2%)				
Less deviant	10 (30.3)	3 (30.0)	7 (30.4)	0.190†
More deviant	12 (36.4)	1 (10.0)	11 (47.8)	
Grade II, n (%)				
No shift (≤2%)	13 (39.4)	4 (40.0)	9 (39.1)	0.963*
Shift (>2%)				
Less deviant	10 (30.3)	3 (30.0)	7 (30.4)	1.000†
More deviant	10 (30.3)	3 (30.0)	7 (30.4)	
Grade III, n (%)				
No shift (≤2%)	16 (48.5)	7 (70.0)	9 (39.1)	0.103*
Shift (>2%)				
Less deviant	8 (24.2)	1 (10.0)	7 (30.4)	0.600†
More deviant	9 (27.3)	2 (20.0)	7 (30.4)	

*P value for shift vs no shift between laboratories with and without feedback on the pathologist level.

†P value for type of change when laboratories shifted $>2\%$ after feedback between laboratories with and without feedback on the pathologist level.

improvement is warranted. It is important to stress that the aim of the feedback reports was not to just simply make 'higher' grading pathologists grade their tumours lower and vice versa. The awareness that the feedback reports created enabled pathologists to discuss how they grade with other pathologists. Furthermore, they could perhaps conclude that they interpret the guideline differently or less strictly than other pathologists. In addition to inciting a dialogue between pathologists, we also hope that our previous paper⁶ opens the dialogue between pathologists and oncologists. As we have previously shown, grade determines whether patients will get chemotherapy in approximately 30% of patients with breast cancer,⁶ thus, awareness of grading variation is also very important to oncologists. Moreover, one could also think of peer consultation in these cases, where grade determines whether a specific therapy is indicated.

Data included in this study were from synoptic pathology reports only, as currently over 80% of IBC resection specimens are reported this way.³² Moreover, besides increased overall completeness of pathology reports,^{33,34} it has recently been shown that synoptic reporting also improves patient care.³⁴ Besides advantages in patient care, easy data extraction from synoptic pathology reports also enables the assembly of nationwide laboratory-specific feedback reports on any chosen biomarker (histological grade or hormone receptor and/or HER2 status³⁵). We believe that these feedback reports are an important first step towards the improvement of breast cancer care by creating insight and awareness in the variation of biomarker assessment, which is supported by the results of this study.

Thirty-eight of the current 42 pathology laboratories in the Netherlands implemented synoptic reporting between 1 March

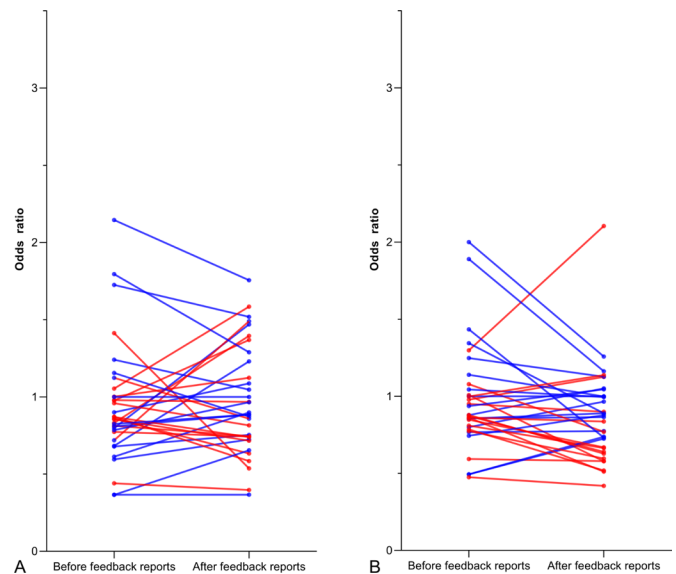


Figure 3 Case-mix-adjusted ORs per laboratory before and after the feedback reports were calculated by multivariate logistic regression analyses: (A) grade III vs grade I–II, (B) grade II–III vs grade I. ORs are adjusted for age, tumour size, type of surgery, histological subtype, hormone receptor status and HER2 status. Each laboratory is represented by two dots (before and after feedback) connected by one line. The colour of the line and dots shows whether the OR after feedback shifted towards an OR of 1.00, and thus became less deviant (blue) or the OR after feedback shifted away from an OR of 1.00, and thus became more deviant (red) from the reference laboratory. HER2, human epidermal growth factor receptor 2.

2017 and 1 March 2019. Five of these 38 laboratories were, nevertheless, excluded from further data analyses as they synoptically graded <50 IBC in either the prefeedback or postfeedback period. Two laboratories likely started using the protocol somewhere in the prefeedback period (<50 reports) as their synoptic IBC report number increased considerably (>230) during the postfeedback period. Two other laboratories had low synoptic IBC report numbers in general (30–60 per period) and the fifth laboratory stopped reporting synoptically (425 prefeedback, 0 postfeedback) for unknown reasons.

Although it seems that some laboratories (or pathologists) grade only few IBC cases annually, it is important to emphasise that these pathologists may still report IBC resection specimens narratively (ie, outside the synoptic PALGA protocol), thus they may grade more IBC cases in clinical practice than our data may suggest. In addition, we previously showed that both laboratories that grade few and many IBC within the synoptic protocol may report significantly deviant proportions per grade.⁶

However, if IBC resection specimen numbers from our study are the true numbers per laboratory, one could argue the desirability of laboratories assessing <50 IBC resection specimens annually, which comes down to <1 IBC resection specimen per week. With the current situation of substantial nationwide inter-laboratory grading variation, and the potential clinical consequences in mind,⁶ one could argue that grading may be only be undertaken by trained, or maybe even only by expert breast pathologists. This should be the subject of future research.

Overall, grades I, II and III IBC were observed in 31.2%, 49.4% and 19.4% of all IBC resection specimens, respectively. This is in line with our previous study (2013–2016) and other studies,^{36–41} although percentages vary between the different

studies. The distribution of grades did vary significantly between the prefeedback and postfeedback period ($p=0.048$); a slight increase of grade I IBC (30.5% vs 32.0%) and a slight decrease of grade III IBC (20.0% vs 18.8%) was apparent, whereas grade II IBC remained relatively stable (49.5% vs 49.2%). This shift in distribution may be initiated by the feedback reports, however, it may also reflect a true change or random variation in the breast cancer population. Either way, this made it more difficult to show significant deviations towards the mean.

In addition, we found relatively low numbers of HER2-positive IBC reports (7.9%) as compared with generally adopted numbers of 15%–20%.⁴² Furthermore, we found relatively high numbers of almost 90% ER-positive tumours, whereas approximately 15% of breast cancers are usually reported as triple negative.⁴³ Both findings are likely due to the fact that we excluded patients who received neoadjuvant treatment, which is the preferred initial approach in patients with HER2-positive and triple-negative breast cancer.⁴⁴

We analysed the data in an absolute and a relative manner, comparing laboratories both to the national proportion and a reference laboratory. Overall, the mean ODS and changes within individual laboratories were non-significant (ODS, positive OR differences). However, all analyses did show a decrease in the extremes (absolute range per grade, maximum ODS and the 21.9% decrease in ORs for grade III vs grades I–II) after feedback. In addition, the slight increase (10.5%) of the absolute range of OR before and after feedback for grades II–III versus I seems to be mainly caused by a single extremely deviating laboratory (#12), as this laboratory became significantly more deviant in the same direction after feedback. An impressive decrease in the absolute OR range of 45.0% can be observed for the remainder of the laboratories. Overall, this shows that most deviant laboratories became less deviant, whereas the majority of the other laboratories remained stable. We therefore believe that these results show an encouraging decrease in breast cancer grading variation after feedback.

Interestingly, we found no differences between laboratories that only received feedback on the laboratory level and laboratories that additionally received feedback on the pathologist level. This may be due to the multistep, semiobjective way of the grading according to the modified Bloom and Richardson guideline,^{30–31} with scores on the three components of grading, resulting in a total score and subsequent grade. This makes it more objective and thus more robust to direct influences on the overall grade. In this light, it may be interesting to reflect on the three different subcategories (tubular differentiation, nuclear pleomorphism, mitotic count) at the pathologist level. In addition, although the mean ODS did not change in both groups, the mean ODS of laboratories that also received feedback on the pathologist level was notably lower than the mean ODS of laboratories that received feedback on the laboratory level only. Hence at the starting point, laboratories with feedback on the pathologist level were already less deviant from the overall mean, which may have influenced their urgency to adjust their grading practices. This may be reflected by the significantly higher proportion of laboratories showing no shift for grade I among laboratories that received feedback on the pathologist level as compared with laboratories that received feedback on the laboratory level only. Lastly, the relatively low number of laboratories that received feedback on the pathologist level ($n=10$), makes it difficult to draw definite conclusions. However, literature does suggest that feedback is more effective when individual rather than general data are provided.^{26–29}

As with all interventions implemented in an uncontrolled environment, such as daily clinical practice, the observed decrease in grading variation can probably not solely be attributed to the feedback reports. However, within the time frame of our previous study⁶ and this study (31 December 2016 to 1 March 2019), no guideline changes and no major other interventions or events on a nationwide scale took place.

Although feedback reports were sent by 1 March 2018, it is very likely that they were discussed in the laboratories somewhere in the week(s) thereafter, for example, in a regular staff meeting. Hence, this means that the actual postfeedback period may have started somewhat later than 1 March 2018, which could have clouded the effect of the feedback reports. Hence, the actual effect of the feedback reports may be even greater.

Despite the encouraging decrease in nationwide grading variation, we also showed that grading variation remains substantial. Besides continuous monitoring and benchmarking²⁶ of histological grading (and other crucial biomarkers like ER, PR and HER2), which is already being considered by PALGA and the Dutch Society of Pathology, future research might focus on developing an e-learning module to train pathologists and residents in histological grading of IBC in a standardised way to further decrease grading variation. This is underlined by Elston and Ellis, who state that grading of IBC should only be undertaken by specifically trained pathologists.⁴⁵

CONCLUSIONS

An encouraging decrease in nationwide Dutch grading variation of IBCs was observed after feedback. As feedback reports were sent to the laboratories for the first time, this was not (yet) a closed quality loop. Therefore, although our results are encouraging, the full potential of these feedback reports is still unknown. As overall grading variation remains substantial, it seems worthwhile to monitor this by continuing with feedback reports. Closing the quality loop and further training of pathologists, for example, by e-learning, may help to further decrease grading variation and improve clinical decision making and thereby the outcome of our patients.

Take home messages

- ▶ Substantial, and clinically relevant, grading variation of invasive breast cancer exists in daily pathology practice.
- ▶ To create awareness and to facilitate quality improvement, feedback reports, containing case-mix-adjusted laboratory-specific grades benchmarked against other laboratories, were sent to the individual pathology laboratories by 1 March 2018.
- ▶ An encouraging decrease in nationwide Dutch grading variation was observed within the year after laboratory-specific feedback reports.

Handling editor Cheok Soon Lee.

Contributors CvD, PjvD, IOB, EvdW, IAD: involved in the design of the study; read and revised the paper, and agreed with the final version of the paper. CD and IAD: data collection. CvD: analysed data; wrote the paper. CvD, PjvD, EvdW, IAD: data interpretation. PjvD: obtained funding.

Funding This research was funded by the Quality Foundation of the Dutch Association of Medical Specialists (SKMS).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Our data are derived from PALGA, the nationwide network and registry of histo- and cytopathology in the Netherlands. Data are available upon reasonable request by researchers. More information can be found at: <https://www.palga.nl/en/data-requests/researchers.html>.

ORCID iD

Paul J van Diest <http://orcid.org/0000-0003-0658-2745>

REFERENCES

- The Netherlands Comprehensive Cancer Organisation (IKNL). Oncoline: breast cancer guideline, 2017. Available: <https://www.oncoline.nl/borstkanker> [Accessed 1 Oct 2019].
- Rakha EA, El-Sayed ME, Lee AHS, *et al.* Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol* 2008;26:3153–8.
- Rakha EA, Reis-Filho JS, Baehner F, *et al.* Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 2010;12:207.
- Curigliano G, Burstein HJ, Winer EP, *et al.* De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen international expert consensus conference on the primary therapy of early breast cancer 2017. *Ann Oncol* 2017;28:1700–12.
- Metzger-Filho O, Ferreira AR, Jeselsohn R, *et al.* Mixed invasive ductal and lobular carcinoma of the breast: prognosis and the importance of histologic grade. *Oncologist* 2019;24:e441–9.
- van Doijeweert C, van Diest PJ, Willems SM, *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.
- Smith BD, Bellon JR, Blitzblau R, *et al.* Radiation therapy for the whole breast: Executive summary of an American Society for radiation oncology (ASTRO) evidence-based guideline. *Pract Radiat Oncol* 2018;8:145–52.
- Cardoso F, van't Veer LJ, Bogaerts J, *et al.* 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016;375:717–29.
- Krop I, Ismaila N, Andre F, *et al.* Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of clinical oncology clinical practice guideline focused update. *J Clin Oncol* 2017;35:2838–47.
- Senkus E, Kyriakides S, Ohno S, *et al.* Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015;26:v8–30.
- Boiesen P, Bendahl PO, Anagnostaki L, *et al.* Histologic grading in breast cancer--reproducibility between seven pathology departments. South Sweden Breast Cancer Group. *Acta Oncol* 2000;39:41–5.
- Frierson HF, Wolber RA, Berean KW, *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.
- Italian Network for Quality Assurance of Tumour Biomarkers (INQAT) Group. Quality control for histological grading in breast cancer: an Italian experience. *Pathologica* 2005;97:1–6.
- Meyer JS, Alvarez C, Milikowski C, *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.
- van Dam PA, Verkerken L, Hauspy J, *et al.* Benchmarking and audit of breast units improves quality of care. *Facts Views Vis Obgyn* 2013;5:26–32.
- van Bommel ACM, Spronk PER, Vrancken Peeters M-JTFD, *et al.* Clinical auditing as an instrument for quality improvement in breast cancer care in the Netherlands: the National NABON breast cancer audit. *J Surg Oncol* 2017;115:243–9.
- Srigley J, Lankshear S, Brierley J, *et al.* Closing the quality loop: facilitating improvement in oncology practice through timely access to clinical performance indicators. *J Oncol Pract* 2013;9:e255–61.
- Sacerdote C, Bordon R, Pitarella S, *et al.* Compliance with clinical practice guidelines for breast cancer treatment: a population-based study of quality-of-care indicators in Italy. *BMC Health Serv Res* 2013;13:28.
- Laronga C, Gray JE, Siegel EM, *et al.* Florida initiative for quality cancer care: improvements in breast cancer quality indicators during a 3-year interval. *J Am Coll Surg* 2014;219:638–45.
- Veerbeek L, van der Geest L, Wouters M, *et al.* Enhancing the quality of care for patients with breast cancer: seven years of experience with a Dutch auditing system. *Eur J Surg Oncol* 2011;37:714–8.
- Casparie M, Tiebosch ATMG, Burger G, *et al.* Pathology databanking and biobanking in the Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 2007;29:19–24.
- Adams AL, Eltoum I, Krontiras H, *et al.* The effect of neoadjuvant chemotherapy on histologic grade, hormone receptor status, and HER2/neu status in breast carcinoma. *Breast J* 2008;14:141–6.
- Yin H-fang, Wang Y-hong, Qin X-q, Yin HF, Wang YH, *et al.* [Effect of neoadjuvant chemotherapy on histologic grade and expression of biological markers in breast cancer]. *Zhonghua Zhong Liu Za Zhi* 2009;31:858–62.
- Zheng S, Zhang B-lin, Xiao T, *et al.* [Comparison of histopathologic changes and expression of biomarkers in breast carcinoma before and after neoadjuvant chemotherapy]. *Zhonghua Bing Li Xue Za Zhi* 2011;40:465–70.
- Deckers IAG (PALGA Foundation). Feedback variation in grading of invasive breast cancer and DCIS, 2018. Available: https://www.palga.nl/assets/uploads/Spiegelinfo/PALGA%20rapport%20-%20Terugkoppeling%20variatie%20in%20gradering%20mammacarcinoom_algemeen.pdf [Accessed 1 Oct 2019].
- Brehaud JC, Colquhoun HL, Eva KW, *et al.* Practice feedback interventions: 15 suggestions for optimizing effectiveness. *Ann Intern Med* 2016;164:435–41.
- Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;119:254–84.
- Ilgel DR, Fisher CD, Taylor MS. Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology* 1979;64:349–71.
- Archer-Kath J, Johnson DW, Johnson RT. Individual versus group feedback in cooperative groups. *J Soc Psychol* 1994;134:681–94.
- 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.
- 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.
- PALGA Foundation. Annual report 2018 (in Dutch), 2019. Available: <https://www.palga.nl/assets/uploads/Jaarverslag/2018%20Palga%20web.pdf> [Accessed 1 Oct 2019].
- Sluiter CE, van Lonkhuijzen LRCW, van Slooten H-J, *et al.* The effects of implementing synoptic pathology reporting in cancer diagnosis: a systematic review. *Virchows Arch* 2016;468:639–49.
- Sluiter CE, van Workum F, Wiggers T, *et al.* Improvement of care in patients with colorectal cancer: influence of the introduction of standardized structured reporting for pathology. *JCO Clin Cancer Inform* 2019;3:1–12.
- van Doijeweert C, Deckers IAG, Baas IO, *et al.* Hormone- and HER2-receptor assessment in 33,046 breast cancer patients: a nationwide comparison of positivity rates between pathology laboratories in the Netherlands. *Breast Cancer Res Treat* 2019;175:487–97.
- Blamey RW, Hornmark-Stenstam B, Ball G, *et al.* ONCOPOOL - a European database for 16,944 cases of breast cancer. *Eur J Cancer* 2010;46:56–71.
- Kiaer HW, Laenkholm A-V, Nielsen BB, *et al.* Classical pathological variables recorded in the Danish breast cancer Cooperative group's register 1978–2006. *Acta Oncol* 2008;47:778–83.
- Møller H, Henson K, Lüchtenborg M, *et al.* Short-Term breast cancer survival in relation to ethnicity, stage, grade and receptor status: national cohort study in England. *Br J Cancer* 2016;115:1408–15.
- Puig-Vives M, Sánchez MJ, Sánchez-Cantalejo J, *et al.* Distribution and prognosis of molecular breast cancer subtypes defined by immunohistochemical biomarkers in a Spanish population-based study. *Gynecol Oncol* 2013;130:609–14.
- Seneviratne S, Lawrenson R, Scott N, *et al.* Breast cancer biology and ethnic disparities in breast cancer mortality in New Zealand: a cohort study. *PLoS One* 2015;10:e0123523.
- Sun J, Chen C, Wei W, *et al.* Associations and indications of Ki67 expression with clinicopathological parameters and molecular subtypes in invasive breast cancer: a population-based study. *Oncol Lett* 2015;10:1741–8.
- Piccatt-Gebhart MJ, Procter M, Leyland-Jones B, *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;353:1659–72.
- Denkert C, Liedtke C, Tutt A, *et al.* Molecular alterations in triple-negative breast cancer-The road to new treatment strategies. *Lancet* 2017;389:2430–42.
- Burstein HJ, Curigliano G, Loibl S, *et al.* Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen international consensus guidelines for the primary therapy of early breast cancer 2019. *Ann Oncol* 2019;30:1541–57.
- 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.