Who needs adjuvant systemic treatment? Predicting prognosis in node-negative breast cancer; from bench to bedside

Bueno de Mesquita, J.M.

Publication date
2009

Citation for published version (APA):
The impact of inter-observer variation in pathological assessment of node-negative breast cancer on clinical risk assessment and patient selection for adjuvant systemic treatment

Jolien M Bueno-de-Mesquita, Dimitry SA Nuyten, Jelle Wesseling, Harm van Tinteren, Sabine C Linn, Prof Marc J van de Vijver

Department of Pathology (JMBdM, JW, MJvdV), Department of Radiation Oncology (DSAN), Biometric Department (HvT), and Division of Medical Oncology (SCL), Netherlands Cancer Institute, Amsterdam, the Netherlands Department of Pathology (MJvdV), Amsterdam Medical Centre, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands

Annals of Oncology 2009, in press

Abstract

Introduction: It is well known that there is considerable inter-observer variability in assessment of the pathological parameters that are used to select node-negative breast cancer patients for adjuvant systemic treatment. There is only limited data available in how many patients this leads to differences in treatment decisions.

Methods: Clinical and pathological data of 694 patients <61 years with primary unilateral T1-4N0M0 breast cancer were analysed. Grade, oestrogen receptor (ER), and HER2 status were first assessed locally; subsequent central re-evaluation of these parameters was performed. Clinicopathological low or high risk was assessed using national Dutch guidelines and the Adjuvant! Online (www.adjuvantonline.com).

Results: The local pathological examination was discordant with central review for grade, ER and HER2 in 28% (kappa 0.56; grade 2 tumours 35% discordant), 5% (kappa 0.85), and 4% (kappa 0.81), respectively. If clinical risk was assessed based on Dutch guidelines or Adjuvant! Online, respectively 15% (1 out of 7 patients; kappa 0.70) or 8% (kappa 0.83) of patients would have been assigned to a different clinical risk group.

Conclusion: Inter-observer variation in pathological examination of breast carcinomas results in significant differences in grade, ER status, HER2 status, clinicopathological risk and subsequently in adjuvant systemic treatment advice.
Introduction

In breast cancer, prognostic factors are used to identify patients who are at relatively high risk of developing distant metastases because these patients will potentially benefit most from adjuvant systemic treatment (e.g., chemotherapy, endocrine treatment and targeted therapy); predictive factors are used to select the optimal type of treatment. Currently used prognostic and predictive factors in breast cancer include age, tumour diameter, lymph node status, histologic grade, oestrogen receptor status, progesterone receptor status and HER2 status.1 For decision making on adjuvant systemic treatment most treatment guidelines advice adjuvant systemic treatment for patients with lymph node metastases; for lymph node-negative breast cancer patients, various algorithms based on age, tumour diameter and histologic grade are used.1 Several clinico-pathological guidelines for adjuvant systemic treatment have been developed on the basis of these factors (e.g. St Gallen guidelines, Adjuvant! Online and Nottingham Prognostic Index).2-10 With the exception of age, these factors are determined by pathological examination of the breast cancer specimen. Several studies have shown considerable inter-observer variation in the assessment of these factors between pathologists.11-23 Lack of reproducibility and reliability of pathological examination and laboratory assays affect clinico-pathological risk assessment and potentially adjuvant systemic treatment advice and decision making, and could therefore affect patient outcome.

Concordance between pathologists in assessing grade was found to range from 50% to 85%.11-23 In general, these studies concluded that despite being subject to considerable inter-observer variation, which can be decreased by using standardized protocols, grade was a valuable prognostic indicator.

The most important predictive factors that are used for treatment decision making in breast cancer are oestrogen receptor (ER) status and HER2 status. Also for these predictive factors, considerable inter-observer variation for the assessment results has been observed ranging from 11% to 28%.24-26 In part these differences are due to methodological differences in the staining technique used, and in part they are due to differences in the interpretation of the staining results.22-23 Intra-laboratory and inter-laboratory variation in ER assay methodology and reporting of results have varied considerably since the ER test was first introduced in clinical practice in the 1970s.27

None of the studies investigating the inter-observer variation of prognostic and predictive markers have directly evaluated the impact on clinico-pathological risk assessment and patient selection for adjuvant systemic treatment. Therefore, the aim of the present study was to assess the effect of inter-observer variation in pathological examination of node-negative breast cancer on the selection of patients for adjuvant systemic treatment using various clinico-pathological guidelines (i.e., national Dutch CBO guidelines, St Gallen guidelines, Nottingham Prognostic Index, and Adjuvant! Online).
Methods

Study design
For this retrospective study, the clinical and pathological data from two consecutive patient series diagnosed between 1996 and 2006 were analysed. The first series consisted of 123 patients of a validation study of the 70-gene prognosis signature performed by Bueno-de-Mesquita et al.; hereafter referred to as the ‘validation-series’.28 The second series is a sub-set of 585 patients from the previously reported RASTER-study (ISRCTN71917916); hereafter referred to as ‘the RASTER-series’.29 Clinical and pathological data retrieved from the patient files for the 14 patients for whom tumours samples could not be retrieved did not significantly differ from the 694 patients analysed for this study (data not shown). This study was part of the RASTER study. Institutional approval for the RASTER study was obtained from the Institutional Review Board of the Netherlands Cancer Institute.29

Patients

Validation-series
The “validation-series” consisted of a consecutive series of patients younger than 55 years of age who had received adequate local therapy for early stage breast cancer, defined as node-negative, tumour diameter < 5 cm (pT1-2). Patients in this series were treated at two Dutch hospitals, the Netherlands Cancer Institute (NKI; Amsterdam) and the Reinier de Graaf Hospital (RdGG; Delft) between 1996 and 1999. Details on the validation-series have been reported previously.28 Paraffin embedded tumour blocks with a representative tumour sample could not be retrieved in 1 out of 123 (0.8%) patients in this series.

RASTER-series
Details on the patient series of the RASTER-study (ISRCTN 71917916). have been reported previously.29 For this study only patients who were eligible according to inclusion and exclusion criteria from the RASTER-study were included in the analysis (N=585). Patients were diagnosed with node-negative breast cancer between 2004 and 2006 at 16 Dutch hospitals. Paraffin embedded tumour blocks with a representative tumour sample could not be retrieved in 13 out of 585 (2%) patients in this series.

Local pathological examination
H&E stained sections were assessed at the pathology department of the participating hospitals by local pathologists for both series. In the RASTER-series two hospitals utilised the same pathology department therefore the original histopathological examination was performed at 15 pathology departments. Histological tumour grade according to the Elston & Ellis method30, oestrogen receptor (ER), progesterone receptor (PR), and HER2 status were determined according to locally used methods. According to Dutch guidelines, the oestrogen and progesterone receptor were considered positive if at least 10% of tumour cells stained positive using an
Impact of inter-observer variation

immunohistochemical assay. Samples were deemed HER2-positive if the score was 3+ in immunohistochemical assay (defined as uniform intense membrane staining of >10% of invasive tumour cells). If the score was 2+ in immunohistochemical assay (defined as uniform membrane staining of >10% of invasive tumour cells with a moderate intensity of staining) and a fluorescent in-situ hybridisation result (FISH) or chromogenic in situ hybridisation (CISH) result was available, the FISH/CISH result (positive or negative) was used. For FISH, a ratio of HER2 to centromer chromosome 17 ratio more than 2.0 was considered positive; for a CISH a result of more than five HER2 gene copies per nucleus was considered positive. If FISH/CISH was not available, a score of 2+ was considered negative for HER2. A score of 0 (defined as weak uniform or partial membrane staining of >10% of invasive tumour cells) was always considered negative for HER2.

Central histopathology review
For each tumour in the study, one paraffin block containing a representative part of the tumour was collected at the Netherlands Cancer Institute. An H&E stained section was used to assess histological grade according to the Elston & Ellis method, and oestrogen receptor, progesterone receptor and HER2 status (assessed by immunohistochemistry) were determined by three experienced breast pathologists (MJvdV, JLP and JW). In case of discordance between the original pathological examination and central review with regard to grade, the examination was performed again by a panel of experienced breast pathologists who then agreed on a final grade. For the validation-series this panel consisted of MJvdV, JPL and JW and for the RASTER-series the panel consisted of MJvdV and JW. MJvdV performed the central review in 75% (523/694) of the patients, JLP and JW in 12% (86/694) and 12% (85/694) of patients, respectively.

Immunohistochemical staining was performed on 3μm thick paraffin embedded tumour sections after microwave antigen retrieval (citrate buffer, pH 6.0 for 15 minutes) using commercially available monoclonal mouse antibodies to ER (1D5+6F11, 1:200 dilution; NeoMarkers, LabVision, Fremont, CA), monoclonal mouse antibodies to PR (PR-1, 1:400 dilution; ImmunoLogic, Duiven, the Netherlands) and monoclonal rabbit antibodies to HER2 (SP332, 1:3000 dilution; NeoMarkers, LabVision), and using the Power Vision detection kit (ImmunoLogic), and the Stainer Lab Vision 2D (LabVision).

For determination of oestrogen receptor, progesterone receptor and HER 2 status the same criteria were used as for the local assessment. However, in case of an HER2 2+ score in the absence of locally performed FISH or CISH, CISH was performed on 5μm paraffin embedded tumour sections. Sections were heated for 90 minutes at 90 minutes at 75°C in an incubator, de-paraffinised in fresh xylene twice for 10 minutes each time, dehydrated in three changes of absolute ethanol for 5 minutes and air-dried. Pre-treatment, hybridisation and post-hybridisation washes were performed according to the manufacturer's description using Zymed SpoT-Light® HER2 CISH™ Kit. If less than 6 copies per tumour cell were present, a tumour was considered HER2 negative; if 6 or more copies were present, a tumour was considered HER2 amplified.
Patient and tumour characteristics of the local and centrally reviewed histopathological findings (N=694) are summarized in table 1. In situ hybridisation was available in 71% (29/41; 9 FISH and 20 CISH) of samples with HER2 2+ score after central review, the in situ hybridisation result was positive in 13 (4 FISH/9 CISH) and negative in 16 samples (5 FISH/11CISH). Due to technical problems we were not able to obtain a FISH or CISH result in 12 samples; these samples (2+ in immunohistochemical assay) were considered negative for HER2.

Table 1: Patient and tumour characteristics, based on local and central pathological examination

<table>
<thead>
<tr>
<th></th>
<th>Original pathological examination</th>
<th>Central pathological review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total n (%)</td>
<td>Validation n (%)</td>
</tr>
<tr>
<td>Number of patients</td>
<td>694</td>
<td>122</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 35 years</td>
<td>44 (6%)</td>
<td>12 (19%)</td>
</tr>
<tr>
<td>36-45 years</td>
<td>88 (14%)</td>
<td>10 (8%)</td>
</tr>
<tr>
<td>41-50 years</td>
<td>130 (19%)</td>
<td>20 (15%)</td>
</tr>
<tr>
<td>46-50 years</td>
<td>225 (32%)</td>
<td>40 (33%)</td>
</tr>
<tr>
<td>51-65 years</td>
<td>183 (26%)</td>
<td>40 (33%)</td>
</tr>
<tr>
<td>&gt; 65 years</td>
<td>44 (6%)</td>
<td>44 (6%)</td>
</tr>
<tr>
<td>Tumour size (pT,M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT1 (&lt;20 mm)</td>
<td>400 (71%)</td>
<td>75 (18%)</td>
</tr>
<tr>
<td>pT2 (≥20-50 mm)</td>
<td>200 (29%)</td>
<td>47 (24%)</td>
</tr>
<tr>
<td>pT3 (≥50 mm)</td>
<td>1 (0.1%)</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>Histological tumour type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal</td>
<td>566 (82%)</td>
<td>100 (89%)</td>
</tr>
<tr>
<td>Lobular</td>
<td>67 (10%)</td>
<td>16 (13%)</td>
</tr>
<tr>
<td>Other</td>
<td>56 (8%)</td>
<td>7 (6%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3 (4%)</td>
<td></td>
</tr>
<tr>
<td>Histologic grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (good)</td>
<td>144 (21%)</td>
<td>46 (13%)</td>
</tr>
<tr>
<td>2 (intermediate)</td>
<td>331 (48%)</td>
<td>80 (6%)</td>
</tr>
<tr>
<td>3 (poor)</td>
<td>217 (31%)</td>
<td>49 (22%)</td>
</tr>
<tr>
<td>Missing</td>
<td>2 (3%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Oestrogen-receptor status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>140 (20%)</td>
<td>33 (21%)</td>
</tr>
<tr>
<td>Positive</td>
<td>554 (80%)</td>
<td>89 (13%)</td>
</tr>
<tr>
<td>Missing</td>
<td>9 (1%)</td>
<td></td>
</tr>
<tr>
<td>Progesterone-receptor status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>214 (31%)</td>
<td>42 (37%)</td>
</tr>
<tr>
<td>Positive</td>
<td>470 (69%)</td>
<td>71 (13%)</td>
</tr>
<tr>
<td>Missing</td>
<td>30 (5%)</td>
<td>9 (1%)</td>
</tr>
<tr>
<td>HER2-receptor status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>513 (78%)</td>
<td>48 (19%)</td>
</tr>
<tr>
<td>Positive</td>
<td>80 (13%)</td>
<td>14 (23%)</td>
</tr>
<tr>
<td>Missing</td>
<td>101 (16%)</td>
<td>40 (41%)</td>
</tr>
</tbody>
</table>

*pT1N0M0 pathological Tumour, Nodes, and Metastases staging system, pT pathological T stage. Percentages in tables and tests may not add up to 100% or add up over 100% due to rounding off. The ‘...’ in cells mean 0.

No significant difference was present in patient and locally assessed tumour characteristics between patients in whom a representative tumour sample was and was not available (data not shown). Mean age was 47 years (range 26-60; SD 7), and the mean tumour diameter was 17 mm (range 2-80; SD 8).
National Dutch CBO guidelines (2004 version) for adjuvant systemic treatment
The risk assessment using clinico-pathological factors is referred to as ‘clinical risk’.
Clinical risk assessment (high versus low) using the national Dutch CBO guidelines (2004 version; formulated by the Dutch Institute for Healthcare Improvement (CBO); www.oncoline.nl) was based on the following factors: age, lymph node status, tumour size (pT), and histologic grade (Elston & Ellis method)30. Patients with a clinical high risk tumour will benefit most from adjuvant systemic treatment (e.g. chemotherapy, endocrine therapy or both; Oxford Overview) and therefore adjuvant systemic treatment is advised in these patients.33,34 The use of the Dutch ‘CBO-guidelines’ has been described in detail by Bueno-de-Mesquita et al.29

International adjuvant systemic treatment guidelines
International guidelines used in this study to assess the clinical risk and need for adjuvant systemic treatment were: St Gallen-guidelines4, Nottingham Prognostic Index (NPI)5-8, and Adjuvant! Online (www.adjuvantonline.com, version 8.0).9,10 A moderate or high clinical risk was considered to be an indication for adjuvant systemic treatment.
For patients included in the validation-series the St Gallen guidelines published in 199835 were used as described in Bueno-de-Mesquita et al.28 For patients of the RASTER-series the St Gallen guidelines published in 2005 were used.4
The Nottingham Prognostic Index (NPI) computes a score using the following algorithm: 0.2*size (cm) + grade + nodal status.5-8
The Adjuvant! Online Software calculates a 10-year survival probability based on the patient’s age, tumour size, histological tumour grade, oestrogen receptor status and nodal status. A low clinical risk was defined as patients with a 10-years survival probability of at least 90%. The application of the Adjuvant! Online Software was conform the use of this guideline as described in Bueno-de-Mesquita et al.29 The Adjuvant! Online Software does not give an advice, but returns risk estimates and the calculated benefits of adjuvant systemic treatment.9,10

Statistical analysis
Analyses were performed using the statistical package SPSS (SPSS for Windows, Release 15.0.1. (22-11-2007); Chicago: SPSS Inc.). The differences in patient and tumour characteristics between the two patient series were tested using the Pearson Chi-Square test. In case of ordinal variables (age, pT-stage of TNM, histological grade and nodal status) with more than two groups, we tested for trends (using Cochran-Armitage test). A significant finding was defined as a P-value smaller than 0.05. The level of agreement between the original pathological examination and the central pathological review was expressed by means of a Cohen’s kappa. Level of agreement between clinical risk assessment based on original pathological examination and central pathological review, and its impact on adjuvant systemic treatment advice, was also expressed by means of a Cohen’s kappa. A kappa of one indicates perfect agreement, where a kappa of zero indicates no agreement.
Who needs adjuvant systematic treatment

Results

Inter-observer variation between local pathological evaluation and central review

The results of the original pathological examinations were discordant with the results of the central review for histologic tumour type (ductal, lobular or other) in 12% (82/687) of patients (kappa 0.56), grade in 28% (195/689) of patients (kappa 0.56), oestrogen receptor in 5% (33/685) of patients (kappa 0.85), progesterone receptor in 12% (83/679) of patients (kappa 0.72), and HER2 status in 4% (25/589) of patients (kappa 0.81; table 2).

Table 2: Original histopathology examination of paraffin embedded tumour samples versus central histopathology review for histological tumour type (A), grade (B), oestrogen receptor (C), progesterone receptor (D) and HER2 status (E). In cells with a grey background the central review does not correspond with the original examination.
Impact of inter-observer variation

The discordance between the original pathological evaluation and the central review has been analysed for separate prognostic factors. Four percent (21/564; 95% CI 2-5%; table 2) of tumours originally classified as “ductal type” were scored as lobular or another non-ductal type histology at central review and 21% (14/67; 95% CI 11-31%) of the tumours classified as lobular type were classified as ductal or another special histology at central review. Grade 2 tumours were most sensitive to inter-observer variation (35% (115/328) discordance; 95% CI 30-40%) compared to grade 1 (25% (36/144) discordance; 95% CI 18-32%) and grade 3 tumours (20% (44/217) discordance; 95% CI 15-26%; table 2B).

Inter-observer variation in the assessment of oestrogen receptor status was higher in tumours originally assessed as oestrogen receptor negative (14% (19/139) were re-classified as ER positive; 95% CI 8-19%) compared to oestrogen receptor positive tumours (3% (14/546) were re-classified as ER negative; 95% CI 1-4%). Assessment of originally progesterone receptor negative tumours were more sensitive to inter-observer variation (16% (33/212) were re-classified as PR positive; 95% CI 11-20%) at central review compared to tumours originally assessed as progesterone positive (11% (50/467) were re-classified as PR negative; 95% CI 8-14%). Tumours that were originally assessed as HER2 positive were scored as HER2 negative at central review in 21% of cases (17/80; 95% CI 12-30%); tumours that were originally assessed as HER2 negative were assessed as HER2 positive at central review in 2% of cases (8/509; 95% CI 0.5-3%).

Effect of discordance of histopathological evaluation on clinical risk assessment

When the clinicopathological risk factors were used to assess recurrence risk according to the Dutch CBO guidelines, there was discordance in the risk assessment (high versus low) between the original evaluation and the central review 15% (102/689) of patients (kappa 0.70; table 3A). Clinical CBO risk based on centrally reviewed data resulted in a change from high risk to low risk in 9% (61/689) of patients; based on the local evaluation adjuvant systemic treatment was advised to these patients, but based on the central review results no adjuvant systemic treatment would have been advised. On the other hand, for 6% (41/689) of the patients the local assessment was a low clinical risk, but based on the central review results, a high risk would be assigned and adjuvant systemic treatment would have been advised. This implies that original clinical CBO high risk was more sensitive to inter-observer variation (19% (61/319; 95% CI 15-23%) at central review than CBO low risk (11% (41/370); 95% CI 8-14%).

Using Adjuvant Online, 8% (54/689) of patients (kappa 0.83; missing 4) would have had a different clinical risk and adjuvant systemic treatment advice after central review (5% (32/689) high to low risk, and 3% (22/689) low to high risk; table 3B). Originally clinical low risk (10%; 22/229; 95% CI 6-13%) based on Adjuvant Online was equally sensitive to inter-observer variation as originally high risk (7% (32/461); 95% CI 5-9%) at central review.
According to St Gallen guidelines, 12% of patients (kappa 0.58) would have been given a different clinical risk and treatment advice after central review (7% high to low risk, and 5% low to high risk; table 3C). Originally clinical low risk (29%; 33/113; 95% CI 21-38%) tumours based on St Gallen guidelines were more sensitive to inter-observer variation than originally high risk (9%; 51/575; 95% CI 7-11%). Based on the NPI, 14% (95/689; missing 5) of patients (kappa 0.72) would have been given a different clinical risk and treatment advice after central review (7% originally high to low risk, and 7% originally low to high risk; table 3D). Originally high risk tumours based on the NPI were equally susceptible to inter-observer variation (16% (48/300); 95% CI 12-20%) as originally low risk patients (12% (47/389); 95% CI 9-15%). The impact of inter-observer variation for clinico-pathological risk assessment was most apparent in histological grade compared to the other factors evaluated. Especially grade 2 tumours were sensitive to inter-observer variation compared to grade 1 and 3 tumours as kappa’s differed substantially (table 4).
The findings of our study show that inter-observer variation in pathological examination of breast carcinomas results in substantial differences in clinico-pathological risk assessment and subsequently in adjuvant systemic treatment advice. Histologic tumour grading was most sensitive to inter-observer variation among pathologists. Central review changed the histologic tumour grade in 28% of patients. Especially, grade 2 tumours were susceptible (35%) to inter-observer variation. In other studies, the concordance between pathologists in assessing grade was found to range from 50% to 85% which is in line with our results (72%).11-23

Central review changed the clinical risk assessment and adjuvant systemic treatment advice of 102 out of 689 patients (15% or approximately 1 in 7 patients) if Dutch CBO guidelines (2004 version) were used. For 61 of these patients (9%) the central review resulted in a change from a high to a low risk and as a consequence the treatment advice to undergo adjuvant systemic treatment would change into the advice not to undergo adjuvant systemic treatment. For 41 (6%) of the patients adjuvant systemic treatment would have been advised after central review, but was not advised based on the original evaluation. Clinical CBO low risk tumours (based on local assessment) were less sensitive to inter-observer variation (11%; 1 in 10 patients) than high risk tumours (19%; 1 in 5 patients). If Adjuvant Online was used, clinical risk and adjuvant systemic treatment advice would change in 1 in 12 patients (8%). The discordance in clinical risk and treatment advice based on the St Gallen guidelines and NPI was respectively, 12% (1 in 8) and 14% (1 in 7) of patients. Dutch CBO guidelines, St Gallen and Nottingham Prognostic Index were more sensitive to inter-observer variation than Adjuvant Online! suggesting that tumour diameter and/or age are more heavily weighed for risk assessment by Adjuvant Online!, than by other guidelines. It is of note that St Gallen low risk tumours were most sensitive to inter-observer variation (29%; 1 in 3 patients) compared to St Gallen high risk tumours (9%; 1 in 11 patients) and compared to other guidelines. This is probably

---

### Table 4: Impact (clinical relevance) of inter-observer variation on clinical risk assessment

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Grade</th>
<th>total</th>
<th>Discordance based on initial examination vs. central review</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>CBO</td>
<td>grade 2</td>
<td>328</td>
<td>69</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>grade 1 &amp; 3</td>
<td>361</td>
<td>33</td>
<td>9%</td>
</tr>
<tr>
<td>Missing: 5 (1%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AO</td>
<td>grade 2</td>
<td>328</td>
<td>36</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>grade 1 &amp; 3</td>
<td>361</td>
<td>18</td>
<td>5%</td>
</tr>
<tr>
<td>Missing: 5 (1%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St Gallen</td>
<td>grade 2</td>
<td>327</td>
<td>42</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>grade 1 &amp; 3</td>
<td>361</td>
<td>42</td>
<td>12%</td>
</tr>
<tr>
<td>Missing: 5 (1%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPI</td>
<td>grade 2</td>
<td>328</td>
<td>56</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>grade 1 &amp; 3</td>
<td>361</td>
<td>39</td>
<td>11%</td>
</tr>
<tr>
<td>Missing: 5 (1%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Grade 2 tumours are always intermediate/high risk tumours based on the St Gallen guidelines

---

### Discussion

The findings of our study show that inter-observer variation in pathological examination of breast carcinomas results in substantial differences in clinico-pathological risk assessment and subsequently in adjuvant systemic treatment advice. Histologic tumour grading was most sensitive to inter-observer variation among pathologists. Central review changed the histologic tumour grade in 28% of patients. Especially, grade 2 tumours were susceptible (35%) to inter-observer variation. In other studies, the concordance between pathologists in assessing grade was found to range from 50% to 85% which is in line with our results (72%).11-23

Central review changed the clinical risk assessment and adjuvant systemic treatment advice of 102 out of 689 patients (15% or approximately 1 in 7 patients) if Dutch CBO guidelines (2004 version) were used. For 61 of these patients (9%) the central review resulted in a change from a high to a low risk and as a consequence the treatment advice to undergo adjuvant systemic treatment would change into the advice not to undergo adjuvant systemic treatment. For 41 (6%) of the patients adjuvant systemic treatment would have been advised after central review, but was not advised based on the original evaluation. Clinical CBO low risk tumours (based on local assessment) were less sensitive to inter-observer variation (11%; 1 in 10 patients) than high risk tumours (19%; 1 in 5 patients). If Adjuvant Online was used, clinical risk and adjuvant systemic treatment advice would change in 1 in 12 patients (8%). The discordance in clinical risk and treatment advice based on the St Gallen guidelines and NPI was respectively, 12% (1 in 8) and 14% (1 in 7) of patients. Dutch CBO guidelines, St Gallen and Nottingham Prognostic Index were more sensitive to inter-observer variation than Adjuvant Online! suggesting that tumour diameter and/or age are more heavily weighed for risk assessment by Adjuvant Online!, than by other guidelines. It is of note that St Gallen low risk tumours were most sensitive to inter-observer variation (29%; 1 in 3 patients) compared to St Gallen high risk tumours (9%; 1 in 11 patients) and compared to other guidelines. This is probably
due to the fact that grade 2 tumours are intermediate/high risk tumours according to the St Gallen guidelines, while grade 2 tumours were more sensitive to inter-observer variation than grade 1 or 3 tumours.

The distribution of grade and frequency of oestrogen receptor, progesterone receptor and HER2 positive tumours was similar to that of other populations, indicating that patients included in this study were representative of the general breast cancer population.

We found a false positive HER2 assessment in 21% of the cases, a number that is very similar to that reported in other series. Paik et al postulated that a bias is introduced by local pathologists in small volume laboratories who overestimate HER2 positive and thereby may cause a high degree of discordance. This supports the importance of using a high volume laboratory, experienced pathologists and rigorous application of standardized procedures for HER2 testing.

Paik et al. studied the degree of agreement in the assignment of tumour grade among three pathologists as a pre-specified secondary objective in a larger study in which they validated the prognostic 21-gene recurrence score. This 21-gene recurrence score quantifies the likelihood of distant recurrence in oestrogen receptor positive node-negative tamoxifen treated breast cancer. They observed that the concordance among pathologists for poorly differentiated tumours was moderate (kappa 0.61) and for well-differentiated and moderately differentiated grades was low (kappa 0.23 and 0.36, respectively). They concluded that a possible advantage of the recurrence score or comparable gene expression profiles over tumour grading in clinical practice is that it is less susceptible to inter-observer variability and therefore more reproducible.

The inter-observer variation was higher in grade 2 tumours than in grade 1 and 3 tumours, subsequently the impact of inter-observer variation on clinico-pathological risk assessment, regardless of what guideline was used, was higher in grade 2 tumours compared to grade 1 and 3 tumours. Gene expression analyses performed by Sotiriou et al. showed that gene expression in grade 2 tumours resembles that of grade 1 tumours in half of the cases and resembles gene expression of grade 3 tumours in the other half. This observation supports our conclusion that grade 2 tumours are a mixed tumour group and hard to reproducibly classify.

Viale et al. reported the results of a central pathological review of the oestrogen and progesterone receptor status in patients included in the BIG I-98 trial. In this trial, only postmenopausal women with oestrogen and/or progesterone receptor positive tumours were included. The central review confirmed 97% as ER positive and 82% as PR positive, implying an inter-observer variation of 3% (95% CI 3-4%) and 18% (95% CI 17-19%) for a change in hormone receptor status from positive to negative. Our study is in line with these results as the inter-observer variation for ER assessment was 5% (95% CI 3-6%) for the change from positive to negative. The inter-observer variation for PR assessment was lower in our study (12%; 95% CI 10-15%).
Conclusion

We conclude that inter-observer variation in pathological examination of breast carcinomas results in considerable differences in clinico-pathological risk assessment and subsequently in adjuvant systemic treatment advice. Rigorous quality control programs should guarantee the reliability of pathological assessment of breast cancer.

Contributors

JM Bueno-de-Mesquita, SC Linn, and MJ Van de Vijver were responsible for the study design. JM Bueno-de-Mesquita coordinated and performed the study. JM Bueno-de-Mesquita and DSA Nuyten took part in the data collection. JM Bueno-de-Mesquita, MJ van de Vijver and J Wesseling performed the revision. JM Bueno-de-Mesquita and H van Tinteren performed the statistical data analysis. JM Bueno-de-Mesquita, DSA Nuyten, SC Linn, MJ Van de Vijver took part in data interpretation and manuscript writing. All authors were involved in reviewing the manuscript. The corresponding author (Marc J van de Vijver) had access to all data in the study and he had final responsibility for the decision to submit for publication.

Role of the Funding Source

This study was part of the RASTER study (ISRCTN71917916). The RASTER study was financially supported by the Dutch Health Care Insurance Board. The funding source had no role in the study design, data collection, data analysis, data interpretation, in writing the report, or in the decision to submit for publication.

Conflict of Interest Statement

None of the investigators and authors of this study have a conflict of interest to declare.

Acknowledgements

We are indebted to the women who participated in the RASTER-study as well as doctors, nurses, and data managers from the participating hospitals in the Netherlands who enrolled patients in the RASTER-study and the Validation-study. We also are indebted to pathologists of the following Dutch hospitals: Aysis Care Group (Arnhem), Amphia Hospital (Breda), Amstelland Hospital (Amstelveen), Atrium Medical Center, (Heerlen), Canisius-Wilhelmina Hospital (Nijmegen), Deventer Hospital (Deventer), Diakonessenhuis (Utrecht), Gelre Hospitals (Apeldoorn), Maxima Medical Center & PAMM (Eindhoven /Veldhoven), Medisch Spectrum Twente (Enschede), Stichting Pathologie Oost Nederland (Enschede), Nij Smellinghe Hospital (Drachten), Netherlands Cancer Institute (Amsterdam), Onze Lieve Vrouwe Gasthuis (Amsterdam), Reinier de Graaf Hospital (Delft), Rijnland Hospital (Leiderdorp), and Spaarne Hospital (Hoofddorp).
Who needs adjuvant systematic treatment

References