Grading Ductal Carcinoma in Situ of the Breast Using an Automated Proliferation Index

Christopher J. Stasik, DO, Marilyn Davis, BS, SCT, Bruce F Kimler, PhD, Fang Fan, MD, PhD, Ivan Damjanov, MD, PhD, Patricia Thomas, MD, Ossama W Tawfik, MD, PhD.

Departments of 1Pathology and Laboratory Medicine, 2Radiation Oncology, University of Kansas Medical Center, Kansas City, KS 66160 USA

Abstract. Tumor grade, size and margin status are the most significant factors in predicting the behavior of ductal carcinoma in-situ (DCIS). The inclusion of necrosis and nuclear grade in the grading of DCIS has demonstrated a fair but suboptimal agreement between pathologists. The grading of DCIS was studied and compared to the Van Nuys (VN) system, by using our newly proposed unifying “nuclear grade + proliferation index (N+P) grading system for invasive carcinomas.

162 DCIS tumors were studied including 49 VN I, 31 VN II, and 82 VN III cases. The VN and N+P systems were compared with each other and correlated with tumor size, ER, PR, p53, Her-2, EGFR, Bcl-2, p27 and p21 status.

The two systems demonstrated similar frequencies for the different grades and an agreement with each other for all of the biomarkers studied. The greatest difference between the two systems was observed for those tumors initially classified as VN II (94% being down-graded to N+P I) and VN III (80% being down-graded to N+P II).

These results suggest that the N+P system, combining nuclear grade with automated MIB-1 count, is a potentially valid and reproducible grading system for both non-invasive and invasive mammary carcinomas. It is automated, less subjective in assessing mitotic activity and necrosis and correlates with other prognostic biomarkers.

Keywords: Ductal carcinoma in situ, Grading, Automation

Introduction

Ductal carcinoma in-situ (DCIS) is a precursor lesion for invasive ductal carcinoma. Tumor size, margin status and grade are the most significant prognostic factors in determining the biologic behavior of DCIS. Although architectural features of DCIS are of value, nuclear grade and necrosis are more useful in predicting local recurrence, prognosis, and invasive transformation as well as in guiding treatment. While there is no universally agreed upon grading system for DCIS, current practice is to grade DCIS on cytonuclear grade in combination with necrosis, and not on the architectural features. The current grading system proposed by the WHO 2003 is a three-tiered system based upon the Van Nuys (VN) grading system (1,2) and that proposed by Scott et al (3), both of which are modifications of the original Lagios classification in 1989 (4). Although the use of necrosis and nuclear grade in the classification scheme of DCIS has demonstrated fair general agreement between pathologists in distinguishing low and high grade DCIS, there is significant disagreement in the grading of the remaining intermediate group. These findings have raised concerns about the system being imprecise in
assessing both parameters (cytornuclear grade and necrosis), allowing examiner subjectivity to influence tumor grade.

Using nuclear grade and MIB-1 (Ki67) automated proliferative activity, we have recently proposed a new grading system for invasive ductal and lobular carcinomas (the nuclear grade plus proliferation [N+P] system) which was comparable to the modified Scarff-Bloom-Richardson (SBR) system in terms of defining prognostically relevant groups (5,6). Replacing the manual mitotic count with an automated MIB-1 count has been shown to be beneficial, providing both standardization and precision. The greatest difference between the two systems was observed for those tumors initially classified as SBR grade II with 28% being “downgraded” to N+P I and 42% being “upgraded” to N+P III. Separation of the overall survival curves was better by the N+P system than by the SBR system. The N+P system appeared to be superior to the SBR system as it correlated better with overall survival for patients with invasive ductal and lobular carcinoma and was less subjective in assessing mitotic activity.

In the current study, we applied our proposed grading system to classify 162 cases of DCIS. The VN and N+P systems were compared with each other and correlated with tumor size, ER, PR, p53, Her-2, EGFR, Bcl-2, p27 and p21 status of each tumor.

Materials and Methods

Patient Cohort: A total of 162 DCIS specimens were examined. Histologic samples of tumors were obtained from core needle biopsy, lumpectomy or mastectomy specimens. A total of 49 VN grade I, 31 VN grade II, and 82 VN grade III cases were evaluated. The retrospective study was approved by the Human Subjects Committee at the University of Kansas Medical Center.

Criteria for the Van Nuys Grading System: The histologic parameters evaluated using the VN system consist of nuclear grade and the presence of comedo necrosis (Table 1). Nuclear grade was scored conventionally from 1 to 3. Nuclear grade 1 is defined as having small, monotonous cells with nuclei equivalent to 1-1.5 times the diameter of a red blood cell (RBC), diffuse chromatin and inapparent nucleoli. Mitoses are rare to absent. Nuclear grade 2 is characterized by nuclei equivalent to 1.5-2 times the diameter of a RBC, coarse chromatin, infrequent nucleoli and rare mitotic activity. Nuclear grade 3 exhibits nuclei with diameters >2.5 RBC equivalents, pleomorphism with irregular nuclear contour, coarse clumped and vesicular chromatin, one or more prominent nucleoli and frequent mitotic figures. Comedo necrosis is defined as recognizable necrotic cells with karyorrhectic or pyknotic nuclear fragments and loss of nuclear detail accumulating centrally in the duct. Final VN grades are based on the combination of nuclear grade and the presence or absence of comedo-type necrosis (Table 1).

Criteria for the Nuclear grade and Proliferation index (N+P) Grading System: The N+P grading system is a three-tiered system, evaluating two features: nuclear pleomorphism and an automated MIB-1 count. Nuclear grade was scored conventionally from 1 to 3 as per the VN system. The proportion of cells positive for MIB-1 staining was evaluated using image analysis and expressed as a percentage using automated methods. Tumors were assigned to one of three groups based on the proportion of MIB-1-positive cells identified: 0-9%, 10-25%, and >25%. These cutpoints were selected based on our previous experience in grading invasive ductal and lobular carcinomas (5,6). N+P grade I was defined as having neither nuclear grade 3 nor MIB-1 expression >25%; N+P grade II was defined as having either nuclear grade 3 or MIB-1 expression >25%; and N+P grade III was defined as having both nuclear grade 3 and MIB-1 expression >25% (Table 1).

Figure 1 shows screen images from the Automated Cellular Imaging System (ACISTM) (San Juan Capistrano, CA) program depicting the process of selecting tissue areas to be evaluated for MIB-1 expression in representative N+P grades I, II and III DCIS lesions.
Immunohistochemical Studies: At diagnosis, tissue blocks containing the most representative and well-preserved tumor areas had been selected for immunohistochemical (IHC) studies. IHC analyses for ER, PR, Her-2, MIB-1, p53, epidermal growth factor receptor (EGFR), bcl-2, p27, and p21 and ploidy analysis were performed on all specimens tissue fixed with 10% neutral buffered formalin. Her-2 antibody was detected using the HercepTest (DAKO, Carpinteria, CA). The individual antibodies, vendor, titration titer, time of titration, epitope retrieval method and method of detection that had been used are shown in Table 2.

Positive IHC reactions had been defined as a dark brown reaction on the cell membrane for Her-2 and EGFR, positive nuclear staining for ER, PR, MIB-1, p53, p27, p21 and positive cytoplasmic staining for bcl-2; with areas of high-density immunostaining.
selected for image analysis or manual scoring. For proliferation index (PI) of MIB-1, the percentage of nuclei with immunopositivity had been determined using the PI program of first the CAS (Cell Analysis System) 200 image analyzer (Bacus Laboratory, Chicago, IL) prior to 2001 and then the Automated Cellular Imaging System (ACISTM) (San Juan Capistrano, CA). For ER, PR, and p53, both the CAS-200 and ACISTM systems had been used for automated counts. Manual microscopy had been used to score tumor staining with antibodies to EGFR, p27, p21 and bcl-2. Her-2 staining had been quantified using a score of 0 or 1+ to indicate no staining, and 2+ or 3+ to denote positive staining, per the scoring instructions included in the HercepTest kit. Results had been validated using the Her-2 scoring system of the ACIS system and by FISH. A 10% cutoff was used to define positivity for MIB-1, ER, PR, EGFR and Bcl-2. Positive staining was defined as >5% for p53, >50% for p27 and >1% for p21.

### Results

The N+P system, in comparison with the VN system was correlated with a variety of histological and prognostic markers. A total of 162 patients with DCIS were analyzed. 47% of the cases were assigned the same grade by both the VN and N+P grading systems. Forty-nine VN I cases were analyzed, the majority of which (45 cases) remained grade I by the N+P system. Thirty-one patients had VN II DCIS lesions. When graded according to N+P criteria, only 2 patients remained grade II. The remaining 29 patients (94%) were downgraded to N+P grade I. Eighty-two patients presented with VN III DCIS. Seventeen of these patients remained grade III by the N+P system while 65 patients (40%) were downgraded to N+P grade II (Table 3).

### Statistical Analysis:

Overall frequencies and percentages were summarized for tumor grade by both the N+P and VN systems, MIB-1, ER, PR, p53, p21, p27, EGFR, Bcl-2, Her-2, and tumor size. The frequencies of each variable, stratified by the grading system, were calculated and their relationships with each grading system were evaluated using the chi-square test. Kappa coefficient was used to determine the agreement between the N+P and VN grading systems. Summaries of the biomarkers by the N+P system stratified by the VN system are also given. Differences in distribution of positive expression of the biomarkers across the three grades were examined by Chi-square test. Differences in the level of biomarker expression and tumor size between grades were analyzed by non-parametric Kruskal-Wallis test.

### Table 2. Protocols for immunohistochemistry*

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Vendor</th>
<th>Titer</th>
<th>Time</th>
<th>Epitope Retrieval</th>
<th>Method of Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>BioCare</td>
<td>1:1000</td>
<td>30 min</td>
<td>BioCare Nuclear Decloaker</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>p53</td>
<td>Biogenex</td>
<td>1:60</td>
<td>30 min</td>
<td>Citrate, pH6</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>PR</td>
<td>Dako</td>
<td>1:5000</td>
<td>30 min</td>
<td>Citrate, pH6</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>EGFR</td>
<td>Zymed</td>
<td>1:20</td>
<td>30 min</td>
<td>Proteinase K, 10’</td>
<td>LSAB+ (Dako)</td>
</tr>
<tr>
<td>bcl-2</td>
<td>Dako</td>
<td>1:100</td>
<td>30 min</td>
<td>BioCare Reveal</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>MIB-1</td>
<td>Dako</td>
<td>1:1000</td>
<td>30 min</td>
<td>BioCare Reveal</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>HercepTest</td>
<td>Dako</td>
<td>P.D. Her2/neu</td>
<td>30 min</td>
<td>Per Kit Instructions</td>
<td>Kit Components (K5204, Dako)</td>
</tr>
<tr>
<td>p27</td>
<td>Dako</td>
<td>1:500</td>
<td>30 min</td>
<td>BioCare Nuclear Decloaker</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
<tr>
<td>p21</td>
<td>Dako</td>
<td>1:100</td>
<td>30 min</td>
<td>Citrate, pH6</td>
<td>Envision + LP, mouse (Dako)</td>
</tr>
</tbody>
</table>

*All stains were performed using the Dako Autostainer per manufacturer procedures. Buffer used was Tris-buffered saline with Tween; visualization with Dab+ (Dako)
Biomarkers evaluated in our study included MIB-1, ER, PR, EGFR, Her-2, Bcl-2, p53, p27 and p21. Data regarding MIB-1, ER, and PR staining was available for all 162 tumors. Of the remaining biomarkers, not all were available for all tumors.

The two grading systems demonstrated similar frequencies for expression across the different histologic grades and a general agreement with each other for all of the biomarkers studied (Table 4, Figure 2). Staining with ER, PR and Her-2 antibodies followed expected trends with both grading systems (Table 4 and Figure 2). Statistically significant associations between N+P grade and other biomarkers were identified, including EGFR (p=0.027), Bcl-2 (p=0.001), and p53 (p=0.001). Staining with p27 and p21 did not demonstrate differences between grades by either the VN or N+P systems (Table 4, Figure 2).

Alternatively, when the median level of expression of the various biomarkers was correlated with VN

<table>
<thead>
<tr>
<th>Variable (cutpoint for positivity)</th>
<th>N</th>
<th>Percent Positive for Van Nuys Grade Categories</th>
<th>P-Value *</th>
<th>Percent Positive for N+P Grade Categories</th>
<th>P-Value *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>%MIB-1 (Ki-67) (10%)</td>
<td>162</td>
<td>29%</td>
<td>42%</td>
<td>73%</td>
<td>0.001</td>
</tr>
<tr>
<td>% ER (10%)</td>
<td>162</td>
<td>100%</td>
<td>84%</td>
<td>57%</td>
<td>0.001</td>
</tr>
<tr>
<td>% PR (10%)</td>
<td>162</td>
<td>82%</td>
<td>68%</td>
<td>38%</td>
<td>0.001</td>
</tr>
<tr>
<td>% EGFR (10%)</td>
<td>145</td>
<td>4%</td>
<td>18%</td>
<td>25%</td>
<td>0.016</td>
</tr>
<tr>
<td>Her-2 (2.0)</td>
<td>118</td>
<td>10%</td>
<td>12%</td>
<td>35%</td>
<td>0.006</td>
</tr>
<tr>
<td>% Bcl-2 (10%)</td>
<td>161</td>
<td>96%</td>
<td>87%</td>
<td>62%</td>
<td>0.001</td>
</tr>
<tr>
<td>% p53 (5%)</td>
<td>160</td>
<td>10%</td>
<td>19%</td>
<td>43%</td>
<td>0.001</td>
</tr>
<tr>
<td>% p27 (50%)</td>
<td>26</td>
<td>67%</td>
<td>50%</td>
<td>55%</td>
<td>0.78</td>
</tr>
<tr>
<td>% p21 (1%)</td>
<td>26</td>
<td>56%</td>
<td>67%</td>
<td>73%</td>
<td>0.72</td>
</tr>
</tbody>
</table>

* Chi-square test
Semi-automated method for grading breast cancer

...and N+P grade, there was an apparent significant difference between N+P grade III tumors for most of the biomarkers studied (Figure 2A and B). As compared to the VN system, N+P grade III tumors had significantly lower ER, PR, Bcl-2 and p27 expression (Figure 2A) and significantly higher MIB-1, p53 and p21 expression (Figure 2B).

Tumor size did not correlate with DCIS grade by either the VN or the N+P system. The median sizes for grade I, II and III were 0.6 cm, 0.8 cm and 1.0 cm respectively for the VN system; and 0.65 cm, 1.0 cm and 1.05 cm respectively for the N+P system. These differences were not statistically significant, (p=0.20 for the VN system and 0.34 for the N+P system).

Discussion

Ductal carcinoma in-situ has come to be regarded as an extremely heterogeneous disease, both in terms of its morphologic appearance as well as its clinical behavior. Prior to the advent of mammography, the diagnosis of DCIS was made only infrequently. As imaging techniques have evolved and public awareness has grown, the incidence of diagnosed DCIS has increased dramatically (1,2). At a fundamental level, DCIS is a non-invasive proliferation of atypical epithelial cells originating from the terminal duct lobular unit of the breast. The degree of cellular atypia is highly variable. Subtle forms which blur the line between hyperplasia and neoplasia are common as are more overt lesions that loudly declare their capacity for invasive transformation. In light of such variability, it comes as no surprise that there exists significant uncertainty regarding how to translate the spectrum of DCIS morphology into useful clinical information.

Multiple studies have attempted to correlate the histologic features of DCIS with risk of progression to invasive carcinoma. Numerous classification systems have been described (1,7,8,9), most of which incorporate some combination of nuclear grade, architecture and morphology in an attempt to separate DCIS into high, intermediate and low grades. In North America, the most widely used of these systems is the VN classification (1). The VN classification incorporates nuclear grade and the presence or absence of comedo necrosis to segregate DCIS lesions into three groups.

To be clinically useful, a classification system must be reproducible. All of the currently proposed grading systems rely upon subjective criteria such as nuclear grade and the extent of necrosis. High nuclear grade has been associated with local recurrence (4,10) and is an important feature rightly incorporated into the majority of DCIS classification schemes. Some systems have shown improved reproducibility when nuclear
grade and necrosis are given an intermediate category (3), although other studies have shown disparity between pathologists trying to discern between three gradations (11,12). When specific criteria are defined, pathologists are able to be quite reproducible in the identification of high versus low nuclear grade lesions (13). Clearly, if a DCIS score is to be used in order to make treatment-related decisions, a system that is highly reproducible is required.

The VN classification has provided pathologists and clinicians with a prognostically meaningful approach to the diagnosis of DCIS. Problems exist, however, due to the reliance upon subjective criteria that are open to individual interpretation and variability. While the simplicity of using a single feature (i.e., comedo necrosis or grade 3 nuclei) to indicate a high grade DCIS lesion is attractive, the diagnosis of comedo necrosis in particular is not always straight-forward and the extent to which its presence in an otherwise low-grade lesion should influence the pathologist is not well-defined.

Of the various classification systems published, the VN system has been shown to have the lowest inter-observer variability when strict criteria defining comedo necrosis are applied (14). Without such criteria, inconsistencies are seen in the determination of the presence or absence of comedo necrosis (15). While pathologists are quite consistent when it comes to identifying high nuclear grade, any inconsistency determining comedo necrosis (particularly in the absence of grade 3 nuclei) has a significant impact on the final VN grade.

In most studies which evaluate inter-observer reproducibility of the features used in DCIS grading, participants rarely differ by more than one grade, possibly indicating that a significant contributor to the problem is the subjectivity of the individual examiners. A 1997 consensus conference analyzed the problems of DCIS classification, ultimately failing to endorse any single classification system (16). The importance of nuclear grade was recognized and it was recommended that DCIS be stratified primarily on that basis with mention of necrosis, cell polarization and architectural pattern in the final report (17).

Any grading system for DCIS must have clinical utility and prognostic meaning. Predicting the biologic behavior of DCIS is extremely difficult and much effort has been put into the investigation of factors which may indicate a capacity for invasive transformation. To date, multiple IHC biomarkers have been identified and correlated with each other as well as histologic features. The advent of targeted therapy towards some of these markers has revolutionized the treatment of breast cancer and added IHC to the list of prognostically important testing to be done. Expression of steroid hormone receptors (ER and PR) has been associated with low nuclear grade DCIS (18) and decreased proliferative activity (19). Conversely, EGFR and Her-2 over-expression have been associated with high nuclear grade lesions (19). The proliferation index (Ki-67) has been evaluated by a number of investigators and found to be associated with high nuclear grade (18-22), “extensive” necrosis (22), lack of ER and PR positivity (21), and p53 overexpression (19,21). High Ki-67 has also been positively correlated with the presence of comedo necrosis (21). In contrast, low Ki-67 staining is associated with ER expression (19), absence of p53 overexpression (21) and lower nuclear grade (DCIS or LCIS) (21).

Previously, our group developed and demonstrated the utility of the N+P system for grading invasive ductal carcinoma (5). Our data showed that our system is superior to SBR grading with improved patient stratification into grades one through three, good correlation with immunohistochemical biomarkers and is prognostic for overall survival. Furthermore, the N+P system was later validated as an excellent grading system for invasive lobular carcinoma (6). The new system was shown to decrease the element of subjectivity for assessing mitotic activity in invasive lobular carcinoma and appeared to be superior to the SBR system in predicting patient survival (6).
In the current study, the N+P system is compared with the VN classification for DCIS. The VN system was chosen for comparison because it is the most widely used grading system in North America and has been shown to have the least inter-observer variation (11,14,15). The trend observed towards lower N+P grades compared to VN grades in disparate cases may reflect the bias seen when using potentially subjective criteria (like comedo necrosis) for DCIS grading. Although nuclear grade rather than mitotic activity is heavily weighed in the VN system, automated MIB-1 measurement combined with nuclear grade appears to be a better (and more objective) measure than comedo necrosis for DCIS grading. This technique results in improved patient stratification, reproducibility and maintains the expected relationships with immunohistochemical biomarkers of prognostic relevance.

With the advent of targeted pharmacotherapy such as Tamoxifen and Herceptin, as well as other therapies under investigation like EGFR tyrosine kinase inhibitors, it is clear that the prognosis of DCIS is complex to predict and is linked to more than just morphologic features. The N+P system, grade for grade, shows similar frequencies of expression of relevant biomarkers with the VN system when evaluating ER, PR, Her-2, EGFR, Bcl-2, p53, p27 and p21. Furthermore when the median level of expression of each biomarker was correlated with tumor grades by the VN and N+P system, there was a significantly better correlation between N+P grade and level of expression for most of the markers, especially for N+P grade III tumors. These findings suggest a better representation of the newly proposed grading system with tumor biology.

In summary, the N+P system is a reproducible method for grading DCIS which is comparable to the VN system and correlates well with prognostically and therapeutically important immunohistochemical biomarkers. By using only cancer grades and the MIB-1 proliferation index, we have developed a valid grading system consisting of one automated, objective measurement (MIB-1) and one subjective criteria (nuclear grade) that has been shown to be reproducible. In addition to DCIS grading, the N+P grading system is also valid for grading invasive ductal and lobular carcinoma (5,6). Future research is needed to assess the reproducibility of these methods. In addition, future studies correlating DCIS grading by the N+P system with disease progression and survival are also needed to further validate this system. This might require collaboration with other groups and would require large numbers of patients with long follow ups. The unification of invasive and in-situ grading of breast lesions under a single umbrella would have multiple benefits. The use of one system would facilitate research into the biology and treatment of the disease as well as the meaningful collection of statistical data. Pathologists would have a means for grading lobular carcinoma, and the juggling of multiple grading systems would be a thing of the past. The confusion surrounding the grading of breast cancer would be decreased and the grading itself would be more consistent. This translates to improved communication between the pathologist and oncologist and, ultimately, improved patient care.

References

8. Poller DN, Silverstein MJ, Galea M, Locker AP, Elston CW, Blamey RW, Ellis IO. Ideas in pathology. Ductal carcinoma in-


