Jpn. J. Cancer Res. **91**, 451–457, April 2000

Evaluation of the Interobserver Agreement in the Number of Mitotic Figures of Breast Carcinoma as Simulation of Quality Monitoring in the Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) Protocol

Hitoshi Tsuda,[1,14] Futoshi Akiyama,[2] Masafumi Kurosumi,[3] Goi Sakamoto,[2] Katsushige Yamashiro,[4] Tetsunari Oyama,[5] Takahiro Hasebe,[6] Kaori Kameyama,[7] Tadashi Hasegawa,[1] Shinobu Umemura,[8] Keiichi Honma,[9] Takachika Ozawa,[10] Keiko Sasaki,[11] Hideo Morino[12] and Shozo Ohsumi[13]

[1]Pathology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, [2]Department of Breast Pathology, Cancer Institute, Tokyo 170-0012, [3]Department of Clinical Pathology, Saitama Cancer Center, Saitama 362-0806, [4]Department of Clinical Research, National Sapporo Hospital, Sapporo 003-0804, [5]Second Department of Pathology, Gunma University School of Medicine, Maebashi 371-8511, [6]Pathology Division, National Cancer Center Research Institute East, Kashiwa 277-8577, [7]Division of Diagnostic Pathology, Keio University School of Medicine, Tokyo 160-8582, [8]Department of Pathology, Tokai University School of Medicine, Isehara 259-1193, [9]Department of Clinical Pathology, Niigata Cancer Center Hospital, Niigata 951-8566, [10]Department of Pathology, Hamamatsu Medical Center, Hamamatsu 432-8580, [11]Department of Pathology and Clinical Laboratories, Aichi Cancer Center Hospital, Nagoya 464-8681, [12]Department of Pathology, Labor Welfare Corporation Kansai Rosai Hospital, Amagasaki 660-8511 and [13]Department of Clinical Research, National Shikoku Cancer Center Hospital, Matsuyama 790-0007

In the National Surgical Adjuvant Study for Breast Cancer (NSAS-BC), node-negative breast cancers were divided into higher- and lower-risk groups according to the histopathological nuclear grade given at individual collaborating hospitals, and the higher-risk group was entered into a randomized protocol of adjuvant therapy. Because the nuclear grade was the composite of nuclear atypia and mitotic counts, maintenance of interobserver agreement in mitotic counts was indispensable for the success of the protocol study. Fourteen pathologists participating in the protocol judged whether or not 20 photomicrographs suspected of showing mitotic cancer-cell figures truly showed mitoses. After standardizing the counting method, these pathologists counted the number of mitotic figures per 10 high-power fields of hematoxylin-eosin-stained main-tissue sections of 20 tumors. Areas where mitotic counts were considered to be the most frequent by each pathologist were compared for these tumors. For the judgment of whether the photomicrograph indicated mitosis, the level of interobserver agreement was moderate (κ=0.569). In the observations of 20 tumors, interobserver agreement level of mitotic counts was moderate (κ=0.506), that of nuclear atypia scoring was fair (κ=0.265), and that of nuclear grading was substantial (κ=0.633). The counted area was almost the same among the observers in 9 tumors, split into two areas in 6, and dispersed in 5. Concordance in judgment was achieved in 7 of the first 9 and in all of the third 5, but only in one of the second 6. The cause of discordance was mostly derived from tumor heterogeneity and the difference in the site where mitoses were counted. Interobserver agreement level was considered to be satisfactory, and it was expected that the case entry would be performed appropriately in the protocol study. The selection of the counting area was confirmed to be important for the acquisition of high-level agreement level in mitotic counts.

Key words: Breast cancer — Mitotic figures — Interobserver agreement — Protocol study

The number of mitotic figures is a well-established prognostic factor in primary breast cancer.[1–3] Usually, mitotic counts are combined with nuclear atypia and structural atypia to judge the histological grade of malignancy.[1–3] The quality control of nuclear atypia scoring is difficult because of the non-quantitative nature of atypia scoring. On the other hand, the number of mitotic figures is a quantitative parameter which would make it easier to achieve a high level of interobserver agreement. However, to achieve a high level of interobserver agreement in mitosis counting, attention should be paid to the selection of the counting area and the establishment and standardization of the counting method and criteria for mitotic figures.[4–6]

In the Japan National Surgical Adjuvant Study of Breast Cancer protocol (NSAS-BC), the pathology section was set up to establish histological criteria for assessing high-

[14]To whom correspondence should be addressed.
E-mail: hstsuda@gan2.ncc.go.jp

risk node-negative breast cancers and to standardize the subjective criteria used by collaborating pathologists for nuclear atypia and mitotic counts.[7–9] In the NSAS-BC protocol, node-negative breast cancers were divided into higher- and lower-risk groups according to a histopathological nuclear grade given at individual collaborating hospitals, and the higher-risk group was entered into a randomized protocol of adjuvant therapy.

Because nuclear grade was the composite of nuclear atypia and mitotic counts, maintenance of interobserver agreement in mitotic counts was indispensable for the success of the protocol study. In the present study, 14 pathologists participating in the NSAS-BC protocol individually judged the number of mitotic counts for 20 tumors using the main hematoxylin-eosin (HE)-stained tissue sections. Interobserver agreement level in the scoring of mitotic counts was estimated by κ statistics for the simulation of quality monitoring in the NSAS-BC protocol.

MATERIALS AND METHODS

Criteria for mitotic figures During interphase, the DNA is doubled in preparation for cell division. During prophase, the nuclear envelope breaks down, and a spindle forms between the two centrioles. At metaphase, the chromosomes align at the equator of the cell, and as the anaphase begins, the duplicated chromosomes (called chromatids) are separated. At telophase, the chromosomes reach the mitotic poles, and the cell begins to pinch in. At each pole are the same number and type of chromosomes as were present in the cell before it divided.[10]

The criteria for mitotic figures are identical with those by van Diest *et al*., as follows: 1. The nuclear membrane is absent, and cells have passed the prophase; 2. Condensed chromosomes are present, either clotted (late prophase to prometaphase), in a plane (metaphase/anaphase), or in separate clots (telophase).[5]

Twenty photomicrographs suspected of showing mitotic figures of breast cancer cells were prepared. Fourteen pathologists independently judged whether or not these photomicrographs truly showed mitoses. These pathologists are responsible for the diagnosis of breast cancers in collaborating hospitals where more than 40 patients with node-negative breast cancers received surgical therapy in 1998.

Preparation of main tumor sections Twenty cases of primary breast cancer were selected randomly from routinely processed archival cases between 1990 and 1994. Five micrometer-thick HE-stained tissue sections of the 20 breast carcinomas were prepared at the National Cancer Center Research Institute, Tokyo and later distributed to 14 collaborating pathologists.

Counting procedure We also utilized the method by van Diest *et al*.[5] The area for scoring the counted mitoses was selected according to the following procedure. First, after screening all the tumor fields by means of a $\times 10$ or $\times 20$ objective lens, two or three parts with high density of cancer cells were selected, and the number of mitotic figures was counted at these parts. Second, among these high-density areas, one area with the largest number of mitotic figures was marked in ink. This area was approximately 0.5 mm$\times$0.5 mm in width. Third, in this area with the largest number of mitoses, about 50 visual fields around the points having the largest number of mitoses were examined. Using a $\times 40$ objective lens, a score of 1 was given if <5 mitotic figures were observed, of 2 if 5–10 mitotic figures were observed, and of 3 if >10 mitotic figures were observed. Only invasive components were counted. The counting was performed slowly and accurately. Although there are many categorizations for the number of mitotic figures, we adopted this criterion because the categorization above correlated well with the prognosis of Japanese patients with node-negative breast carcinoma.[7]

Mitotic counts were adjusted to "field number" 20 to standardize the number of mitoses among hospitals according to the property of the eyepieces of the light microscopes used by individual pathologists (Table I).

Scoring of nuclear atypia and nuclear grade The nuclear atypia was scored as 1, 2 and 3 when the degree of atypia was low, intermediate and high, respectively.[7] The nuclear grade was scored as 1, 2 and 3 when the sum of scores for the nuclear atypia and those for mitotic counts were 2–3, 4 and 5–6, respectively.[7]

Statistical analysis The degree of interobserver agreement for the score of mitotic counts was tested using the generalized κ test for more than two observers from the viewpoint of categorical data.[11–13] In accordance with the criteria of Landis and Koch, the κ statistics were divided into several scales to determine the strength of agreement.[14]

Table I. Adjustment of the Criteria for Mitotic Counts according to the Property of the Eyepieces of a Light Microscope

"Field number"	Mitotic counts per 10 HPF using a 40× objective lens			Type of light microscope
	1 point	2 points	3 points	
18	0–3	4–8	9≤	Nikon CWF 10× Nikon CFD 10×
20	0–4	5–10	11≤	Olympus WHK 10× Nikon CFWN 10×
22	0–5	6–12	13≤	Olympus WH 10× Nikon CFI 10×
26	0–7	8–17	18<	Olympus SWH 10×
26.5	0–7	8–17	18<	Olympus SWHK 10× Nikon CFUWN 10× Nikon CFIUW 10×

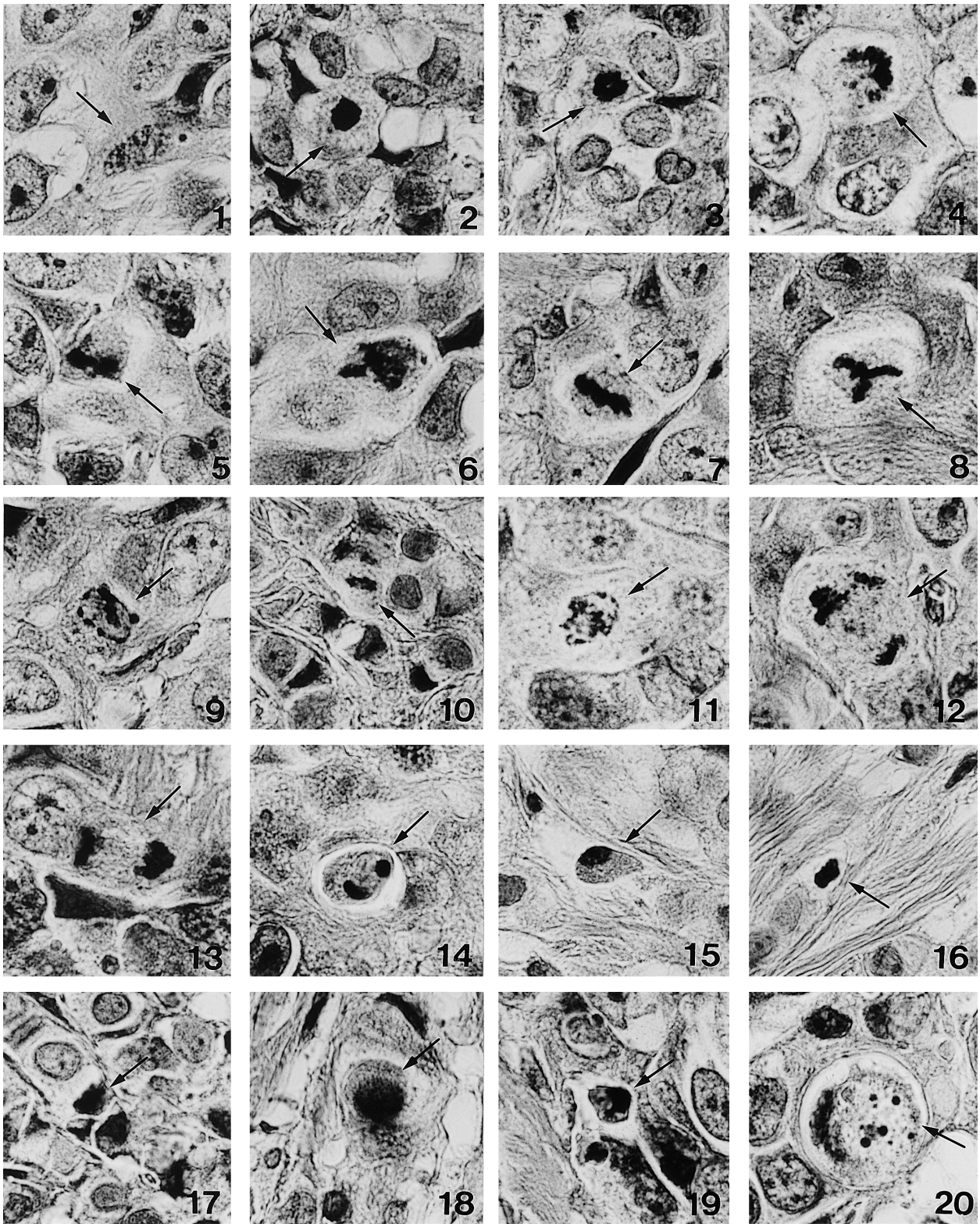

Fig. 1. Twenty photomicrographs which were suspected of showing mitotic figures of cancer cells (arrows). Photo #1 was taken as early prophase, #2–6 and #11 as late prophase to prometaphase, #7, 8, 10 as metaphase, #12–13 as metaphase, and others as non-mitoses. The distribution of judgments by 14 pathologists is shown in Table II.

Table II. Judgment by 14 Pathologists Concerning Whether 20 Photomicrographs Suspected of Mitotic Cancer-cell Figures Were Considered as Mitosis or Not

Photo No.	Number of pathologists		
	Judgment		
	Mitosis	Not mitosis	Unknown
1	0	13	1
2	13	0	1
3	14	0	0
4	13	1	0
5	9	3	2
6	12	0	2
7	14	0	0
8	14	0	0
9	2	11	1
10	13	0	1
11	6	7	1
12	14	0	0
13	14	0	0
14	5	6	3
15	1	11	2
16	1	9	4
17	3	8	3
18	2	10	2
19	0	14	0
20	0	12	2

Table III. Mitotic Count Judgments for 20 Tumors by 14 Pathologists Using Main Tumor Tissue Sections

Tumor No.	Number of pathologists		
	Number of mitotic figures per 10 HPF		
	<5	5–9	>9
1	7	7	0
2	14	0	0
3	0	1	13
4	13	1	0
5	10	4	0
6	9	5	0
7	0	6	8
8	14	0	0
9	1	4	9
10	0	0	14
11	10	4	0
12	14	0	0
13	0	2	12
14	9	5	0
15	13	1	0
16	0	2	12
17	8	5	1
18	12	2	0
19	4	9	1
20[a)]	13	0	0

a) One pathologist did not give a score.

The most probable reason for interobserver disagreement in mitotic counts would be the difference in the counting area. For the evaluation of concordance in counting an area on each tumor tissue section among observers, the area was marked in ink and the scores of mitotic counts given were compared.

RESULTS

In Fig. 1, photomicrograph #1 had been taken as early prophase, #2–6 and 11 as late prophase to prometaphase, #7, 8, 10, and 11 as metaphase, #12–13 as anaphase, and #9 and 14–20 as non-mitosis by one of the pathologists (H.T.). Most observers agreed with that interpretation for photos #1–10, #12, #13, and #15–20. Judgments were divided for #11 and 14. Number 11 was considered as a degenerated or pyknotic nucleus rather than prometaphase or metaphase by nearly half of the observers. In the case of #14, it was difficult to decide whether it was an anaphase cell or simply a pyknotic cell. For the judgments as to whether the photomicrograph indicated mitosis, the level of interobserver agreement was moderate (κ=0.569) (Table II, Fig. 1).

In the study of interobserver agreement using HE-stained tissue sections, the categorization of mitotic counts was concordant among 10 or more observers for tumors #2–5, 8, 10–13, 15, 16, 18, and 20 (Table III). In particular, there was total agreement concerning tumors #2, 8, 10, 12, and 20. On the other hand, the concordance was lower in tumors #1, 6, 7, 9, 14, 17, and 19. In the observations of these 20 tumors, the agreement level was moderate (κ=0.506).

To reveal the reason for this inconsistency in the categorization of mitotic counts, the counted area was compared among observers for each tumor. In nine tumors, observers largely agreed on the counted area (group 1). In six tumors, observers split into two groups concerning the counted area (group 2), and in the other five, observers split into three or more groups (group 3). Concordance in judgment was achieved in 7 of the 9 tumors (group 1) and in all five of the third 5 (group 3), but only in one of the second 6 (group 2). In Fig. 2, representative cases of groups 1, 2, and 3 are schematically drawn. In tumor #6 in group 2, all four observers who counted the *b* area gave a score of 1, whereas two of six observers who counted the *a* area gave a score of 2, and the other three observers who counted area *c* or *d* also gave a score of 2 (Fig. 2A). In tumor #7 in group 2, six of eight observers who counted the *a* area gave a score of 3, whereas four of five observers who counted the *b* area gave a score of 2 (Fig. 2B). This

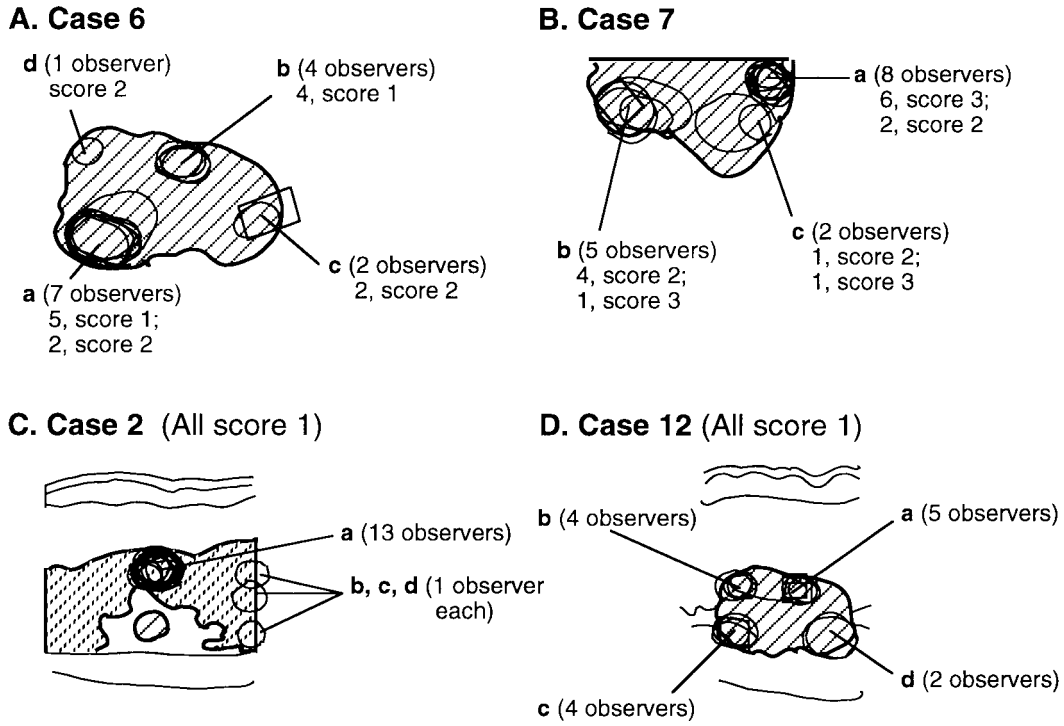

Fig. 2. Schematic presentation of the cut surface of four breast cancers. The area where mitotic figures are counted is indicated. In A (case 6) and B (case 7), counted areas largely split into two (*a* and *b*), and the score of mitotic figures tended to differ between areas. In C (case 2), observers showed total agreement concerning the counted area, whereas in D (case 12), observers split into 4 groups. However, in C and D, the score of mitotic counts was almost identical among observers.

tendency was common among the cases of group 2. In contrast, in tumor #2 in group 1 and tumor #12 in group 3, the category of mitotic counts was almost identical among observers regardless of the counted area (Fig. 2, C and D). The discordance was mostly derived from tumor heterogeneity and the difference in the site where mitoses were counted.

In these 20 tumors, interobserver agreement level of the scoring of nuclear atypia was fair (κ=0.265) (Table IV). The interobserver agreement level of nuclear grading, the composite of nuclear atypia and the number of mitotic figures, was judged to be substantial (κ=0.633) (Table V). In 11 tumors (#1, 4–6, 8, 11, 14, 15, and 17–19), the score of nuclear grade given by observers was split into grade 1 and grades 2–3. The level of discordant grading, which was estimated by dividing the sum of non-modal observations for the above 11 tumors by all observations for 20 tumors, was 12.0% (31 of 259).

DISCUSSION

To achieve a high level of interobserver agreement, the quality of tissue processing and the selection of the counting area are crucial.[5] In the present study, the quality of tissue sections was uniform because all sections were processed in an identical institute. Because of the difference in the number of available staff and the technical level among hospitals, equalization of the quality of tissue specimens is not easy. This problem cannot be solved by the personal effort of pathologists alone. Nonetheless, the quality of tissue specimens could be improved if individuals exercised the utmost care in all ways. For example, resected specimens can be fixed better, especially in the neighboring area of the cut surface of the tumor, when the specimens are directly soaked in an ample volume of formalin in Tupperware and the cut surface of the tumor is down but not fastened to the plate. On dissection, at least one piece of well-fixed tumor should be processed for pathological diagnosis. Delay in fixation was shown to influence the mitotic activity of cancer cells in strictly designed experiments.[15] However, this subject is still controversial because the interval between the resection and starting fixation did not influence the number of mitotic counts in a significant way in recent studies using specimens for diagnosis.[7, 16]

Fourteen (70%) of the 20 tumors examined in the present study were considered to be rather uniform in mitotic activities throughout the tumor tissue. However, in

Table IV. Nuclear Atypia Judgments for 20 Tumors by 13 Pathologists Using Main Tumor Tissue Sections[a)]

Tumor No.	Modal nuclear atypia score	Number of pathologists		
		Nuclear atypia score		
		1	2	3
1	2	2	11	0
2	2	5	8	0
3	3	0	3	10
4	2	4	9	0
5	2	4	9	0
6	2	3	10	0
7	3>2	0	6	7
8	2	2	10	1
9	2>3	0	7	6
10	3	0	0	13
11	2	2	11	0
12	2	3	10	0
13	3	0	5	8
14	2	1	12	0
15	2	2	11	0
16	2	0	11	2
17	2	3	10	0
18	2	2	11	0
19	2	0	10	3
20	1=2	6	6	0

a) One pathologist did not give a nuclear atypia score. For tumor 20, another pathologist did not give a score.

Table V. Nuclear-grade Judgments for 20 Tumors by 13 Pathologists Using Main Tumor Tissue Sections[a)]

Tumor No.	Modal nuclear-grade score	Number of pathologists		
		Nuclear atypia score		
		1	2	3
1	1>2	7	6	0
2	1	13	0	0
3	3	0	1	12
4	1	12	1	0
5	1	10	3	0
6	1	10	3	0
7	3	0	2	11
8	1	12	1	0
9	3	0	4	9
10	3	0	0	13
11	1	10	3	0
12	1	13	0	0
13	3	0	1	12
14	1	9	4	0
15	1	12	1	0
16	3	0	2	11
17	1	8	5	0
18	1	12	1	0
19	2	2	10	1
20	1	12	0	0

a) One pathologist did not give a nuclear atypia score. For tumor 20, another pathologist did not give a score.

six tumors, differences in the counting area gave rise to interobserver discordance in the scoring of mitotic counts. Slow and accurate selection of the counting area was confirmed to be important not only to evaluate correctly the status of mitotic activities in the tumor, but also to achieve high-level interobserver agreement.

The low κ value (0.265) for the scoring of nuclear atypia did not mean a truly low agreement level. The modal nuclear atypia score was 2 in 16 tumors and 3 in only 4 tumors, but no tumor showed the modal score of nuclear atypia 1. This kind of bias to nuclear atypia score 2 would have caused an apparently low level of κ. On the other hand, the κ value for nuclear grading was very high and suggested that the grading was performed with an adequate interobserver agreement level. In the NSAS-BC protocol, node-negative breast cancers of nuclear grade 1 are to be excluded, but cases of nuclear grade 2 and 3 are considered to be eligible for entry. There was a possibility that 11 tumors (#1, 4–6, 8, 11, 14, 15, and 17–19) were erroneously entered into the protocol or erroneously excluded by a small number of observers. According to the previous study, interobserver disagreement would be derived either from the intermediate nature of the tumor itself or from erroneous scoring.[8)] Tumor #1 is an example of such a situation. Because nearly half of the 31 discordant gradings would also contain a disagreement about tumor nature, the approximate true level of erroneous judgments would be about 6.0%.

This study simulated the quality management of mitotic count scoring in the NSAS-BC protocol. Because 38 hospitals are participating in the protocol study and the area of examination of mitotic counts is selected not from one section but from multiple ones per tumor, the agreement level would become lower in the quality monitoring. The NSAS-BC pathology section has held slide conference sessions twice a year since 1996 for the standardization of nuclear atypia criteria among collaborating pathologists.[8, 9)] A color atlas for the reference of daily nuclear atypia scoring was edited and distributed to collaborators in 1997 and 1999 (revised version). These activities should result in the maintenance and improvement of the agreement level of nuclear atypia scoring.

ACKNOWLEDGMENTS

This study was supported in part by a sponsorship grant from Taiho Pharmaceutical Co., Tokyo as a related project of the NSAS-BC protocol, which is conducted by Dr. S. Akashi-Tanaka, Dr. T. Fukutomi, and Dr. T. Watanabe, National Cancer Center Hospital, Tokyo. We thank Ms. Y. Yamauchi and Ms. T. Takarabe for their excellent technical assistance.

(Received November 1, 1999/Revised January 18, 2000/ Accepted January 26, 2000)

REFERENCES

1) Bloom, H. J. G. and Richardson, W. W. Histological grading and prognosis in breast cancer. *Br. J. Cancer*, **11**, 359–377 (1957).
2) Contesso, G., Mouriesse, H., Friedman, S., Genin, D., Sarrazin, D. and Rouesse, J. The importance of histologic grade in long-term prognosis of breast cancer: a study of 1,010 patients, uniformly treated at the Institut Gustave-Roussy. *J. Clin. Oncol*., **5**, 1378–1386 (1987).
3) Elston, C. W. Grading of invasive carcinoma of the breast. *In* "Diagnostic Histopathology of the Breast," ed. D. L. Page and T. V. Anderson, pp. 300–311 (1987). Churchill Livingstone, New York.
4) Simpson, J. F., Dutt, P. L. and Page, D. L. Expression of mitoses per thousand cells and cell density in breast carcinomas: a proposal. *Hum. Pathol*., **23**, 608–611 (1992).
5) van Diest, P. J., Baak, J. P. A., Matze-Cok, P., Wisse-Brekelmans, E. C. M., van Galen, C. M., Kurver, P. H. J., Bellot, S. M., Fijnheer, J., van Gorp, L. H. M., Kwee, W. S., Los, J., Peterse, J. L., Ruitenberg, H. M., Schapers, R. F. M., Schipper, M. E. I., Somsen, J. G., Willig, A. W. P. M. and Ariens, A. T. Reproducibility of mitosis counting in 2,469 breast-cancer specimens: results from the multicenter morphometric mammary carcinoma project. *Hum. Pathol*., **23**, 603–607 (1992).
6) Hilsenbeck, S. G. and Allred, D. C. Improved methods of estimating mitotic activity in solid tumors. *Hum. Pathol*., **23**, 601–602 (1992).
7) Tsuda, H., Akiyama, F., Kurosumi, M., Sakamoto, G. and Watanabe, T., for the NSAS-BC pathology section. Establishment of histological criteria for high-risk node-negative breast carcinoma in a randomized clinical trial of adjuvant therapy. *Jpn. J. Clin. Oncol*., **28**, 486–491 (1998).
8) Tsuda, H., Akiyama, F., Kurosumi, M., Sakamoto, G., Watanabe, T. and NSAS-BC collaborating pathologists. The efficacy and limitations of repeated slide conferences for improving interobserver agreement when judging nuclear atypia of breast cancer. *Jpn. J. Clin. Oncol*., **29**, 68–73 (1999).
9) Tsuda, H., Akiyama, F., Kurosumi, M., Sakamoto, G. and Watanabe, T., for the NSAS-BC pathology section. Monitoring of interobserver agreement in nuclear atypia scoring of node-negative breast carcinomas judged at individual collaborating hospitals in the National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) protocol. *Jpn. J. Clin. Oncol*., **29**, 413–420 (1999).
10) Gilbert, S. F. "Developmental Biology," 3rd Ed., pp. 3–32 (1991). Sinauer Associates, Sunderland.
11) Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull*., **76**, 378–382 (1971).
12) Light, R. J. Measures of response agreement for qualitative data: some generalizations and alterations. *Psychol. Bull*., **76**, 365–377 (1971).
13) Fleiss, J. L. "Statistical Methods for Rates and Proportions," pp. 212–236 (1981). John Wiley & Sons, New York.
14) Landis, J. R. and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174 (1977).
15) Græm, N. and Helwig-Larsen, K. Mitotic activity and delay in fixation of tumour tissue. *Acta Pathol. Microbiol. Scand. A*, **87**, 375–378 (1979).
16) Donhuijsen, K., Schmidt, U., Hirche, H., van Beuningen, D. and Budach, V. Changes in mitotic rate and cell cycle fractions caused by delayed fixation. *Hum. Pathol*., **21**, 709–714 (1990).