Interobserver variability between general and expert pathologists during the histopathological assessment of large-core needle and open biopsies of non-palpable breast lesions

H.M. Verkooijen[a,1], J.L. Peterse[b], M.E.I. Schipper[c], E. Buskens[d], J.H.C.L. Hendriks[e], R.M. Pijnappel[f], P.H.M. Peeters[d,*], I.H.M. Borel Rinkes[a], W.P.Th.M. Mali[g], R. Holland[e] on behalf of the COBRA Study Group

[a]*Department of Surgery, University Medical Center Utrecht, The Netherlands*
[b]*Department of Pathology, Antoni van Leeuwenhoek Hospital/Netherlands Cancer Institute, Amsterdam, The Netherlands*
[c]*Department of Pathology, University Medical Center Utrecht, The Netherlands*
[d]*Julius Center for Patient Oriented Research, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands*
[e]*National Expert and Training Center for Breast Cancer Screening, University Medical Center St. Radboud, Nijmegen, The Netherlands*
[f]*Department of Radiology, Martini Hospital Groningen, The Netherlands*
[g]*Department of Radiology, University Medical Center Utrecht, The Netherlands*

Abstract

The purpose of this study was to assess whether general pathologists are able to make as accurate and reproducible a diagnosis on large-core needle biopsies as on open breast biopsy specimens. A total of 688 patients underwent a stereotactic large-core (14G) needle biopsy and subsequent surgical excision of 718 non-palpable breast lesions. Forty-two pathologists from 10 departments of pathology (generalists) made a diagnosis on both the needle and open biopsy specimens. Afterwards, three pathologists and two radiologists with extensive experience in breast pathology (experts) diagnosed all of the biopsy specimens. The general pathologists made a similar histological diagnosis as the experts in 632 (88%) of the needle biopsies and 649 (90%) of the open biopsy specimens. Accordingly, the interobserver agreement for the diagnosis of large-core needle biopsies between the general and experts pathologists was excellent (kappa 0.83) and not significantly different from the interobserver agreement for the diagnosis of open breast biopsies (kappa 0.86). However, many inconsistencies were observed in the category of borderline lesions: only 24% of the large-core needle biopsies and 43% of the open biopsies with an expert diagnosis of 'borderline' were diagnosed similarly by the general pathologists. Additionally, the risk of benign/malignant inconsistencies between general pathologists and experts was approximately 1 in 55 for both needle and open biopsies.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Breast neoplasms; Diagnosis; Breast biopsy; Biopsy; Needle; Interobserver variability; Breast diseases

1. Introduction

Ever since screening mammography was introduced, pathologists are increasingly confronted with the histopathological evaluation of non-palpable breast lesions. Until recently, these lesions were generally diagnosed by means of a needle-localised open breast biopsy. A con-siderable proportion of women undergoing this invasive procedure have benign disease only and therefore less invasive techniques, such as stereotactic large-core needle biopsy, have been introduced to maximise the number of accurate and definitive diagnoses that can be made and allowing a number of open breast biopsies to be avoided.

Accurate and reproducible histopathological assessment of non-palpable breast lesions is of crucial importance when deciding upon optimal treatment choices. Previously, the reproducibility of the histopathological diagnosis of open biopsy specimens was shown to be

---

reasonably good [1]. Because a smaller amount of tissue is obtained, the assessment of large-core needle biopsies might be more complicated. Accordingly, more inconsistencies in the diagnosis of pathologists might occur. We set out to determine the interobserver agreement between general pathologists and an expert panel in a series of patients with non-palpable breast lesions who underwent stereotactic large-core needle biopsies followed by an open breast biopsy.

2. Patients and methods

This study is part of the multi-institutional COBRA study (COre Biopsy after RAdiological localisation), which has been described elsewhere in Ref. [2]. Briefly, the aim of the COBRA study was to compare the diagnostic accuracy of stereotactic large-core needle biopsy with that of open breast biopsy for the diagnosis of non-palpable breast disease. The study was executed according to the Helsinki Declaration. The Dutch National Insurance Council and all local Institutional Review Boards approved the study protocol. Between April 1997 and February 2000, 826 consecutive patients with 871 non-palpable breast lesions after having given their informed consent, underwent a stereotactic large-core needle biopsy followed by a needle-localised open breast biopsy. Patients were referred from 19 hospitals to undergo a stereotactic large-core needle biopsy in one of five centres (University Medical Center, Utrecht, Antoni van Leeuwenhoek Hospital, Amsterdam, Dr Daniel den Hoed Clinic, Rotterdam, Bosch Medicentrum Den Bosch and Martini Hospital, Groningen). Biopsies were performed according to a standardised protocol using prone biopsy tables (Fisher Imaging Denver CO and Lorad Stereoguide Danbury CT) and a 14-gauge core needle long throw (2.2-cm excursion) automated biopsy device (C.R. Bard Inc., Covington, GA, USA). At least five needle biopsy specimens were obtained per lesion. For lesions consisting of microcalcifications, we X-rayed the biopsy specimens to ascertain the presence of calcifications. A total of 42 pathologists working from 10 different pathology laboratories performed the initial histopathological assessment of the large-core needle biopsy specimens in a routine setting. From now on, we will refer to this group as the 'general pathologists'. Based on the initial diagnosis, a further diagnostic or therapeutic strategy was planned. If large-core needle biopsy revealed invasive breast cancer, definitive surgery (breast-conserving therapy or mastectomy with axillary dissection or sentinel node biopsy) was offered. Where the large-core needle biopsy revealed ductal carcinoma *in situ* (DCIS), breast-conserving therapy or mastectomy (in cases of extensive DCIS) was performed. According to the study protocol, all patients with a non-malignant large-core needle biopsy result underwent a diagnostic needle-localised open breast biopsy. The same group of general pathologists diagnosed all of the surgical specimens (unblinded for the core biopsy diagnosis).

An expert panel consisting of three expert pathologists and two expert radiologists reviewed 718 non-palpable lesions detected in 688 consecutive patients (82% of all lesions enrolled in the COBRA study). Prior to reviewing the histopathological slides, the panel radiologists examined the mammograms and, where available, the ultrasound images, spot compression and magnification views of the non-palpable lesions. They informed the expert pathologists on the mammographic aspects of the non-palpable lesions. Taking into account this radiological information, the pathologists first assessed the large-core needle biopsy slides and thereafter the surgical specimens. During this review process, the panel was blinded for the diagnoses established by the general pathologists.

2.1. (Statistical) analysis

The histopathological diagnoses of the general pathologists and the review panel were classified into five categories: invasive breast cancer, DCIS, borderline lesions, benign lesions and normal breast tissue. The reason for this division is that in daily practice these five diagnostic categories have different treatment consequences. Normal breast tissue was defined as biopsy specimens containing only normal breast tissue. This diagnosis does not explain the mammographic abnormality and is therefore always a reason for additional diagnostic interventions [2]. The category benign was used for the cases in which a benign lesion was found which explained the mammographic findings, e.g. fibroadenoma as a substrate for a density or fat necrosis with dystrophic calcifications explaining mammographic microcalcifications. The term borderline lesion was used in cases of atypical hyperplasia (ductal or lobular) or lobular carcinoma *in situ*. The presence of a borderline lesion in open breast biopsies indicates an elevated risk of breast cancer [3–5]. Previous studies have shown that finding a borderline lesion in large-core needle biopsies is an indication for open breast biopsy because in 23–50% of the cases the excision biopsy will yield malignancy [2,6–10]. In cases of ductal carcinoma *in situ* or invasive breast cancer, surgery is required, including axillary dissection or sentinel node biopsy in cases of invasive breast cancer.

We compared the interobserver agreement between the general pathologists and experts for the histopathological assessment of large-core needle biopsies, as well as for the open breast biopsies. We hypothesised that the small amount of diagnostic material obtained by means of the large-core needle biopsy would lead to a less reproducible histopathological assessment. Accordingly,

we expected more inconsistencies between the general pathologists and experts for the assessment of the large-core needle biopsies than for the assessment of the open biopsy specimens. The level of interobserver agreement was calculated using a kappa statistic [11]. The kappa statistic includes a correction for the level of agreement expected by chance alone. Values of kappa near the maximum of 1 indicate perfect agreement. Values higher than 0.40 are considered acceptable and higher than 0.75 are considered excellent [12]. Kappa statistics and 95% confidence intervals (CIs) were calculated using Agree 6.0 (Groningen, The Netherlands).

For each category separately, we calculated the proportion of expert diagnoses that were similarly diagnosed by the general pathologist to identify the categories with the poorest agreement between the general pathologist and experts. We took the expert diagnoses as the reference standard, because we considered these diagnoses to be the most accurate; i.e. the expert pathologists had extensive experience and special interest in the histopathological assessment of small breast lesions and were assisted by two experienced radiologists and were therefore able to make a diagnosis under optimal conditions. We used the Chi square test to compare the proportions (Statistical Package for Social Sciences (SPSS 9.0).

3. Results

The characteristics of the patients and the non-palpable lesions are presented in Table 1. In Table 2, the histopathological diagnoses of general pathologists and experts, based on the large-core needle biopsies, are presented. In 632 of the 718 lesions (88%), the general pathologists established an identical histopathological diagnosis on the large-core needle biopsy specimens as the experts. After correction for chance, the level of interobserver agreement appeared to be 'excellent' (kappa 0.83 (95% CI 0.78–0.88) [12].

In Table 3, the histopathological diagnoses based on the open biopsy specimens are presented. In 649 of the

Table 1
Characteristics of 688 patients with 718 non-palpable breast lesions

Characteristic	N (%)
Age (in years, range between parentheses)	58 (29–85)
Patients referred by national screening programmes	406 (59)
Radiology	
Microcalcifications	348 (48)
Mass (spiculated or well defined)	258 (36)
Microcalcifications with tissue distortion	108 (15)
Architectural distortion	22 (3)
Focal asymmetry	2 (0.3)

718 open biopsies (90%), the general pathologists reached the same histopathological diagnosis as the experts. The level of interobserver agreement for the open biopsy specimens was also 'excellent' (kappa 0.86 (95% CI 0.81–0.91) and not significantly different from the interobserver agreement of the large-core needle biopsy specimens.

Table 4 illustrates that most discrepancies between the general and experts pathologists were observed in the borderline and normal breast tissue categories. This was the case for large-core needle biopsies, as well as for open biopsies. Only 24% of the large-core needle biopsies and 43% of the open biopsies with an expert diagnosis of 'borderline' were diagnosed similarly by the general pathologists. Fifty percent of the large-core needle biopsies and 35% of the open biopsies with an expert diagnosis of 'normal breast tissue' were diagnosed similarly by the general pathologists. When the experts categorised the non-palpable lesions as benign, DCIS or invasive cancer, a similar diagnosis was made by the general pathologists in most cases (for both types of biopsy). None of the observed differences in the inconsistencies between the assessment of needle and open biopsies were statistically significant.

The general pathologists assessed one large-core needle biopsy, diagnosed as invasive breast cancer by the experts, as benign. In addition, the experts diagnosed seven needle biopsies as DCIS, while the general pathologists evaluated these samples as benign. These discrepancies would have led to different management strategies in these patients and may therefore have affected the prognosis of the patient. The same phenomenon was observed in the group of open biopsy specimens. Here, two lesions diagnosed as invasive cancer and nine lesions diagnosed as DCIS by the experts were not recognised as such by the general pathologists. The opposite also occurred: general pathologists diagnosed one large-core needle biopsy lesion as invasive cancer and four as DCIS, while the experts diagnosed these lesions as benign. Similarly, the general pathologists diagnosed three open biopsy specimens as DCIS, while the experts evaluated them as benign. Thus, the risk of benign/malignant inconsistencies between the general pathologists and the experts was approximately 1 in 55 for both the needle and open biopsies.

4. Discussion

This study demonstrated that the level of interobserver agreement between general and expert pathologists for the diagnosis of large-core needle biopsies is not significantly different from the level of interobserver agreement for the diagnosis of open breast biopsies. In addition, the kappa value of 0.83 indicated an 'excellent' interobserver agreement [12]. Therefore, we can conclude

Table 2
Histological diagnoses based on large-core needle biopsies made by routine pathologists (vertically) and experts (horizontally)

Histological diagnosis (routine pathologists)	Histological diagnosis (experts)					Total
	Normal breast tissue	Benign	Borderline	DCIS	Invasive breastcancer	
Normal breast tissue	13	11	2			26
Benign	11	197	19	7	1	235
Borderline	2	8	8	5	1	24
DCIS		4	5	153	3	165
Invasive breast cancer		1		6	261	268
Total	26	221	34	171	266	718

DCIS, ductal carcinoma *in situ*.

Table 3
Histological diagnoses based on open breast biopsies made by routine pathologists (vertically) and experts (horizontally)

Histological diagnosis (routine pathologists)	Histological diagnosis (experts)					Total
	Normal breast tissue	Benign	Borderline	DCIS	Invasive breast cancer	
Normal breast tissue	6	4				10
Benign	11	207	11	9	2	240
Borderline		11	10	4		25
DCIS		3	2	139	5	149
Invasive breast cancer				7	287	294
Total	17	225	23	159	294	718

that large-core needle biopsies are equally difficult to assess as open biopsies and that general pathologists are able to evaluate large-core needle biopsy specimens of non-palpable breast lesions with a high accuracy.

The reproducibility in the histopathological assessment of the large-core needle biopsies of non-palpable breast lesions has not been previously studied. Palli and colleagues studied the interobserver variability for the diagnosis of surgical specimens of non-palpable lesions among 16 pathologists [13]. They reported an overall kappa score of 0.72. However, the 81 lesions assessed in this study were selected, resulting in a higher proportion of 'difficult' lesions (11% borderline lesions compared with only 5% in our study). Similarly, a considerable

Table 4
Proportion of lesions, classified according to expert diagnosis, that was similarly diagnosed by routine pathologists

Expert diagnosis	Proportion of lesions diagnosed similarly by the routine pathologists		
	Core biopsies	Excision biopsies	*P* value
Normal	50% (13/26)	35% (6/17)	NS
Benign	89% (197/221)	92% (207/225)	NS
Borderline	24% (8/34)	43% (10/23)	NS
DCIS	89% (153/171)	87% (139/159)	NS
Invasive	98% (261/266)	98% (287/294)	NS

DCIS, ductal carcinoma *in situ*; NS, non-significant.

interobserver variability for the diagnosis of open breast biopsies of non-palpable lesions among six pathologists with extensive experience in breast pathology was reported by Schnitt and colleagues [14]. However, they selected only proliferative breast lesions (epithelial hyperplasia without atypia, atypical hyperplasia or DCIS) and therefore a larger interobserver variability was to be expected. We decided to compare the differences between the histopathological assessment of large-core needle and open biopsies in a group of non-selected, 'daily practice' lesions. The prevalence of borderline lesions in the present study was relatively low; i.e. 5% of the large-core needle biopsies and 3% of the open biopsies were diagnosed as borderline.

Like previous investigators, we observed low levels of interobserver agreement in the category of borderline lesions [1,13,14]. Although the presence of a borderline lesion in an open biopsy specimen implies an elevated risk for developing breast cancer [3,4], it is not usually followed by further diagnostic intervention. Therefore, inconsistencies in evaluating open biopsies as either benign or borderline have relatively little impact. However, the presence of a borderline lesion in a large-core needle biopsy, will always entail an excisional biopsy. Accordingly, inconsistencies in evaluating large-core needle biopsies as borderline or benign will lead to unwanted differences in the treatment of individual patients. Therefore, with the advent of large-core needle

biopsy techniques, consistency in reporting borderline lesions becomes increasingly important.

Whenever a diagnosis of normal breast tissue is obtained, it is very likely that the biopsy procedure has not been executed properly. In cases of large-core needle biopsy, the biopsy needle has probably not penetrated the lesion. In cases of an open breast biopsy, the needle localisation, surgical excision or histopathological sampling has probably been executed incorrectly. Previously, it was demonstrated that in cases of a large-core needle biopsy yielding only normal breast tissue, the risk of the presence of malignancy is 16% [2]. Therefore, this finding always demands a re-biopsy or an open biopsy. Similarly, when an open biopsy yields only normal breast tissue, repeat mammography and sometimes repeat a biopsy is warranted. Therefore, also in this category, consistency in the histopathological reporting between pathologists is very important.

High levels of interobserver agreement were observed for the categories 'benign', 'DCIS' and 'invasive cancer'. Nevertheless, there were a few major disagreements between the general pathologists and experts in these categories, i.e. the experts diagnosing invasive cancer while the general pathologists fail to identify the lesion as such and vice versa. These major discrepancies might have serious impact on the therapeutic decisions made for individual patients, ultimately affecting prognosis (i.e. failing to diagnose malignancies) and morbidity (diagnosing benign lesions as malignant and performing unnecessary surgery). The incidence of these major discrepancies was low and was similar in the large-core needle biopsy specimens and open breast biopsies.

Although the overall consistency between the general and expert pathologists was good, the occurrence of clinically relevant discrepancies leaves room for improvement. We think that a high diagnostic accuracy can be achieved in a general setting by establishing the multidisciplinary evaluation of every large-core needle biopsy and open biopsy procedure. A team consisting of a radiologist, a surgeon and a pathologist (in complicated cases preferably more than one) needs to carefully correlate every histopathological diagnosis to the mammographic findings and only if the two correlate should the diagnosis be accepted. We recommend that whenever a pathologist has reservations over the precision of a histological diagnosis, (s)he should not hesitate to consult a colleague with extensive expertise in breast pathology.

5. COBRA Study Group

A.P.E. Besnard, I.H.M. Borel Rinkes, E. Buskens, W.A.H. Gelderman, A.D. Groote, J.H.C.L. Hendriks, R. Holland, P.K.H. Hut, H.J. de Koning, W.P.Th.M. Mali, M. Oudkerk, P.H.M. Peeters, J.L. Peterse, R.M. Pijnappel, E.J.T. Rutgers, M.E.I. Schipper, P. Schipper, J.G. van den Tweel, H.M. Verkooijen, Th.J.M.V. van Vroonhoven, P.F.G.M. van Waes, Th. Wiggers

References

1. Sloane JP, Ellman R, Anderson TJ, *et al.* Consistency of histopathological reporting of breast lesions detected by screening: findings of the U.K. National External Quality Assessment (EQA) Scheme. *Eur J Cancer* 1994, **30**, 1414–1419.
2. Verkooijen HM. Diagnostic accuracy of stereotactic large-core needle biopsy for nonpalpable breast disease: results of a multicenter prospective study with 95% surgical confirmation. *Int J Cancer* 2002, **99**, 853–859.
3. Marshall LM, Hunter DJ, Connolly JL, *et al.* Risk of breast cancer associated with atypical hyperplasia of lobular and ductal types. *Cancer Epidemiol Biomarkers Prev* 1997, **6**, 297–301.
4. Page DL, Jensen RA, Simpson JF. Premalignant and malignant disease of the breast: the roles of the pathologist. *Mod Pathol* 1998, **11**, 120–128.
5. Tavassoli FA. Ductal carcinoma in situ: introduction of the concept of ductal intraepithelial neoplasia. *Mod Pathol* 1998, **11**, 140–154.
6. Acheson MB, Patton RG, Howisey RL, Lane RF, Morgan A. Histologic correlation of image-guided core biopsy with excisional biopsy of nonpalpable breast lesions. *Arch Surg* 1997, **132**, 815–818.
7. Jackman RJ, Nowels KW, Rodriguez-Soto J, Marzoni FA, Finkelstein SI, Shephard MJ. Stereotactic, automated, large-core needle biopsy of nonpalpable breast lesions: false-negative and histologic underestimate rates after long-term follow up. *Radiology* 1999, **219**, 799–805.
8. Liberman L, Cohen MA, Dershaw DD, Abramson AF, Hann LE, Rosen PP. Atypical ductal hyperplasia diagnosed at stereotaxic core biopsy of breast lesions; An indication for surgical biopsy. *AJR* 1995, **164**, 1111–1113.
9. Brown TA, Wall JW, Christensen ED, *et al.* Atypical hyperplasia in the era of stereotactic core needle biopsy. *J Surg Onco* 1998, **67**, 168–173.
10. Gadzala DE, Cederbom GJ, Bolton JS, *et al.* Appropriate management of atypical ductal hyperplasia diagnosed by stereotactic core needle breast biopsy. *Ann Surg Oncol* 1997, **4**, 283–286.
11. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology. A basic science for clinical medicine. Boston, Little, Brown and Company, 1985, 17–45.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977, **33**, 159–174.
13. Palli D, Galli M, Bianchi S, *et al.* Reproducibility of histological diagnosis of breast lesions: results of a panel in Italy. *Eur J Cancer* 1996, **32**, 603–607.
14. Schnitt SJ, Connolly JL, Tavassoli FA, *et al.* Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. *Am J Surg Pathol* 1992, **16**, 1133–1143.