Breast Cancer Res Treat (2013) 139:769–777
DOI 10.1007/s10549-013-2576-0

CLINICAL TRIAL

Reducing false-positive biopsies: a pilot study to reduce benign biopsy rates for BI-RADS 4A/B assessments through testing risk stratification and new thresholds for intervention

Chris I. Flowers · Cristina O'Donoghue · Dan Moore · Adeline Goss · Danny Kim · June-Ho Kim · Sjoerd G. Elias · Julia Fridland · Laura J. Esserman

Received: 20 May 2013 / Accepted: 22 May 2013 / Published online: 14 June 2013

Abstract The aim of this study is to evaluate Breast Imaging Reporting and Data Systems (BI-RADS) 4A/B subcategory risk estimates for ductal carcinoma in situ (DCIS) and invasive cancer (IC), determining whether changing the proposed cutoffs to a higher biopsy threshold could safely increase cancer-to-biopsy yields while minimizing false-positive biopsies. A prospective clinical trial was performed to evaluate BI-RADS 4 lesions from women seen in clinic between January 2006 and March 2007. An experienced radiologist prospectively estimated a percent risk-estimate for DCIS and IC. Truth was determined by histopathology or 4-year follow-up negative for malignancy. Risk estimates were used to generate receiver-operating characteristic (ROC) curves. Biopsy rates, cancer-to-biopsy yields, and type of malignancies missed were then calculated across postulated risk thresholds. A total of 124 breast lesions were evaluated from 213 women. An experienced radiologist gave highly accurate risk estimates for IC, DCIS alone, or the combination with an area under ROC curve of 0.91 (95 % CI 0.84–0.99) ($p < 0.001$), 0.81 (95 % CI 0.69–0.93) ($p = 0.011$), and 0.89 (95 % CI 0.83–0.95) ($p < 0.001$), respectively. The cancer-to-biopsy yield was 30 %. Three hypothetical thresholds for intervention were analyzed: (1) DCIS or IC $\geq$ 10 %; (2) DCIS $\geq$ 50 % or IC $\geq$ 10 %; and (3) IC $\geq$ 10 %, which translated to 22, 48, and 56 % of biopsies avoided; cancer-to-biopsy yields of 36, 47, and 46 %; and associated chance of missing an IC of 0, 1, and 2 %, respectively. Expert radiologists estimate risk of IC and DCIS with a high degree of accuracy. Increasing the cut off point for

C. I. Flowers
345 Bayshore Boulevard, Apt 608, Tampa, FL 33606, USA
e-mail: chrisflowers@mac.com

C. O'Donoghue · L. J. Esserman (✉)
University of California at San Francisco, 1600 Divisadero St, PO Box 1710, San Francisco, CA 94115, USA
e-mail: laura.esserman@ucsfmedctr.org

C. O'Donoghue
e-mail: cristina.odonghue@gmail.com

D. Moore
UCSF Biostatistics Division, Box 0560, San Francisco, CA 94107, USA
e-mail: dmoore@biostat.ucsf.edu

A. Goss
University of Pennsylvania, Suite 100 Stemmler Hall, 3450 Hamilton Walk, Philadelphia, PA 19104, USA
e-mail: addie.goss@gmail.com

D. Kim · J.-H. Kim
University of California at San Francisco, San Francisco, CA, USA
e-mail: danny.kim@ucsf.edu

J.-H. Kim
e-mail: juneho.kim@jhmi.edu

S. G. Elias
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, PO Box 85500, 3508GA Utrecht, The Netherlands
e-mail: S.Elias@umcutrecht.nl

J. Fridland
Department of Obstetrics and Gynecology, Harbor-UCLA Medical Center, 1000W. Carson St, Torrance, CA 90509, USA
e-mail: jfridland@obgyn.humc.edu

recommending biopsy, substituting with a short-term follow-up protocol with biopsy if any change, may safely reduce the number of false-positive biopsies.

Keywords False-positive biopsy · Mammography screening · BI-RADS · Biopsy thresholds

Introduction

The diagnostic workup of a suspicious mammographic finding includes taking biopsies to determine whether abnormal findings represent malignancy, with the goal of identifying invasive cancer (IC) at the earliest stage, thereby reducing mortality. There is, however, a considerable variation in the rate of biopsy across the United States as well as internationally, with cancer-to-biopsy yields relatively low, ranging from 22 to 33 % [1–4] and over half a million breast biopsies performed annually [5]. In addition, negative surgical open biopsy rates have been shown to be twice as high in the United States as they are in the United Kingdom, despite similar cancer detection rates [6]. Over a 10-year period, 61 % of women undergoing annual mammographic screening will be called back for an abnormality, and 7–9 % will receive a false-positive biopsy recommendation [7–9]. The negative consequences of benign biopsies include fear, pain, anxiety, direct financial expenses, indirect costs related to work missed, and risk of complications [10–12].

A Breast Imaging Reporting and Data Systems (BI-RADS) 4 assessment is given to lesions that carry a risk of malignancy between 2 and 95 % and, in the United States, most BI-RADS 4 lesions are biopsied (69–95 %) [13, 14]. The 4th Edition (2003) of the BI-RADS guidance chapter provided more refined categories of risk within the BI-RADS Category 4 creating three sub-categories (4A, 4B, and 4C) [15], and the 5th Edition (2013) recommends risk estimates for malignancy (Table 1) [16]. However, there is no distinction between the risk of ductal carcinoma in situ (DCIS) or IC, instead using an overall risk estimate for malignancy. The opportunity of sub-classifying BI-RADS 4 is that low-risk 4A lesions may clinically be evaluated separately and followed rather than immediately biopsied as is done for most 4B and 4C lesions.

Table 1 BI-RADS 4 subcategories

Assessment category	Recommendation	Risk of malignancy (%)
4A: low suspicion	Biopsy should be performed in the absence of clinical contraindication	≥2 to <10
4B: moderate suspicion		≥10 to <50
4C: high suspicion		≥50 to <95

Subcategories of BI-RADS 4 scoring with risk estimates for malignancy included in American College of Radiology's Breast Imaging-Reporting and Data System, BI-RADS: Mammography [16]

This pilot study evaluates the diagnostic accuracy of an experienced radiologist in assigning separate risk estimates for both DCIS and IC. We tested the impact of several different thresholds for biopsy, using risk estimates, to determine if more refined thresholds would safely reduce false-positive biopsies, and whether there is a category of risk for which immediate intervention could be safely replaced with short-term follow-up similar to BI-RADS 3, with little impact on delaying the diagnosis of consequential invasive breast cancer.

Materials and methods

Patients

Data were prospectively collected from a cohort of 213 consecutive female patients that were referred for further evaluation of a breast lesion to the Coordinated Diagnostic Evaluation Program (CDEP) at the Breast Care Center at the University of California, San Francisco, between January 1, 2006 and March 31, 2007. CDEP was initially created as a multidisciplinary program for patients with abnormal mammograms. This study was approved by IRB and was HIPAA compliant. The cohort had a combined 224 lesions. The majority of women (161) were referred to the clinic for a BI-RADS score of 0 or 4. Off the remaining referred women, 7 had a BI-RADS 5 score, 30 had a BI-RADS 1–3 score on a prior mammogram, and 15 were referred for some other suspicious finding without a prior mammogram or BI-RADS score. Women with BI-RADS 1–3 were referred to CDEP because of patient preference, family history, or patient age. Patient demographics and medical history were collected by questionnaire at each examination. The patients had a multidisciplinary assessment and followed standard care.

Radiological assessment

For this study, the 124 BI-RADS 4 or 5 lesions were reviewed by a radiologist with 29 years of mammography experience who was blinded to final outcomes and clinical data. BI-RADS 4 lesions were subcategorized into A through C lesions and prospectively assigned a percent risk-estimate for DCIS and IC separately following guidelines as described in BI-RADS 5th edition (Table 1) [16]. The expert radiologist differentiated risk estimates for lesions based on morphology such as classifying round, coarse, vascular, or punctate microcalcifications as low-

risk as opposed to pleomorphic or fine linear/branching microcalcifications which were classified as higher risk. Linear and segmental distributions were considered higher risk than clustered or regional microcalcifications [3]. If an asymmetric density or mass was associated with calcifications, then a risk of at least 50 % for IC was allocated [19].

Final diagnosis

Pathological findings (core biopsy, fine-needle aspiration, or surgical specimen) at the patient's definitive intervention served as the reference standard. For patients who declined biopsy or surgical intervention, the reference standard was a negative 4-year screening or diagnostic mammogram.

Statistical analysis

Stata and R software programs were used for statistical analysis. Diagnostic accuracy was assessed using receiver-operating characteristic (ROC) curves. ROC curves were generated and area under the curve (AUC) calculated for three groups of risk estimates: (1) risk estimate of IC versus outcome of IC ($n = 124$); (2) risk estimate of DCIS in cases with invasive risk estimates <2 % versus outcome of DCIS ($n = 52$); and (3) highest risk estimate (DCIS or IC) versus outcome of malignancy ($n = 124$). Lesions with risk estimates for both DCIS and IC were categorized into the IC risk estimate groups. We performed a logistic regression of outcome (1 if patient had the outcome, 0 if not) from the radiologist's risk estimate for each of the three outcomes: DCIS, IC, and any malignancy (either DCIS or invasive or both).

Results

Study population

The cohort of 213 consecutive female patients, enrolled from January 2006 to March 2007, had 224 total lesions and most were seen at the CDEP clinic for the evaluation of an abnormal mammogram with a BI-RADS score of 0, 4, or 5. At the time of radiological assessment, priors were reviewed in combination with additional imaging if required. Four lesions did not have imaging for review and 88 lesions were determined to have BI-RADS scores 1, 2, or 3 and were eliminated from this analysis. Of the remaining BI-RADS 4 and 5 lesions, two were excluded because the patient came to CDEP with a pathological diagnosis, and six were excluded because appropriate follow-up data were not available. Analysis was confined to the remaining 116 BI-RADS 4 and 8 BI-RADS 5 lesions; totaling 124 lesions from 108 patients. Figure 1 shows the flow diagram of the lesions included in the study. For our analysis, mean patient age at the CDEP appointment was 54.9 years ± 13.8 (SD). Additional patient characteristics based on the 124 lesions are listed in Table 2.

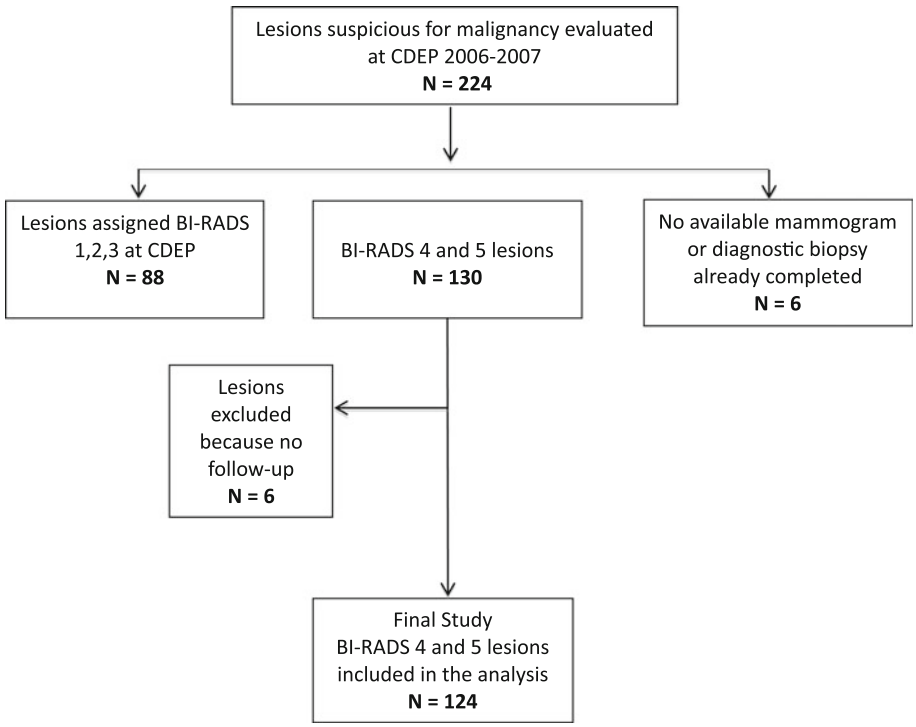

Fig. 1 BI-RADS 4 and 5 lesions included in the final analysis. In 213 consecutive female patients, 224 lesions were evaluated. Eighty-eight lesions were excluded because they were read as BI-RADS 1-3. Two lesions had final pathological diagnosis at presentation and four lesions had no imaging available. Off BI-RADS 4 and 5 lesions, six were excluded because they did not have follow-up. The remaining 124 lesions were used for analysis

Table 2 Demographic and clinical characteristics of the study population

Factor	Lesions[b] ($N = 124$)	%
Age (years)		
<40	17	14
40–49	27	22
50–59	35	28
60–69	26	21
70–79	13	10
>79	6	5
Menopausal status		
Premenopausal	44	35
Perimenopausal	4	3
Postmenopausal	76	61
Family history of breast cancer		
Yes	46	37
No	77	62
Missing	1	1
Presenting with symptoms (lump, pain, discharge)		
Yes	38	31
No	86	69
Reason for referral to CDEP[a]		
Abnormal screening mammogram	81	65
Abnormal diagnostic mammogram	98	79
Other abnormal index imaging	5	4
No abnormal index imaging	5	4
CDEP BI-RADS score		
BI-RADS 4	116	94
BI-RADS 5	8	6

[a] Multiple selections are possible; percentages do not add to 100 %

[b] Lesions map to characteristics of 108 patients based on the frequency of lesions

Final diagnosis

Off the 124 lesions, definitive diagnosis was obtained via biopsy or surgical excision for 115 lesions, yielding a biopsy rate of 93 % for BI-RADS 4 and 5 lesions at the CDEP. For 7 % ($n = 9$) of the 124 lesions, benign outcome was confirmed by a negative mammogram at 4 years. Off these nine lesions, two had been recommended for biopsy at CDEP, six were downgraded to BI-RADS 3, and one with 'tea-cup' calcification had been recommended for routine screening.

Thirty-five of the 115 biopsied lesions were found to have IC or DCIS, yielding an overall cancer-to-biopsy yield at CDEP of 30 %. Pathological diagnosis was determined by a core biopsy ($n = 16$), fine-needle aspiration ($n = 12$), and excisional biopsy ($n = 7$). Twenty-three of these lesions were IC, 12 of which also had accompanying DCIS. Twelve lesions were diagnosed as DCIS alone. Eighty-nine lesions (72 %) were found to be benign: 80 by biopsy and 9 by 4-year follow-up. A total of eight benign lesions (6 %) were found to be high-risk lesions such as atypical hyperplasia. Final pathological diagnoses of the benign and malignant lesions are described in Table 3.

Discriminative ability of DCIS and IC risk estimation by an experienced breast radiologist

Risk estimates for IC only, DCIS only, and both IC and DCIS were above zero in 43 % ($n = 54$), 40 % ($n = 53$), and 17 % ($n = 17$), respectively. The ROC curve assessing the risk estimate for IC versus outcome of IC ($n = 124$) had an AUC of 0.91 (95 % CI 0.84–0.99) (Fig. 2a). The ROC curve assessing the risk estimate for DCIS alone was

Table 3 Final histopathological diagnosis of study lesions

Malignant lesions	Lesions ($N = 35$)	%
Invasive breast cancer	23	66
Metastatic breast cancer	1	3
IDC (5 Grade 1, 10 Grade 2, 2 Grade NOS)	17	49
ILC (2 Grade 2, 1 Grade NOS)	3	9
Malignant myoepithelioma	1	3
Unspecified IC	1	3
DCIS (3 Grade 1, 6 Grade 2, 3 Grade 3)	12	34
Benign lesions	Lesions ($N = 80$)	%
Acellular amorphous debris	1	1
Atypical ductal hyperplasia[a]	3	4
Atypical Lobular Hyperplasia[a]	2	3
Apocrine metaplasia	4	5
Benign	8	10
Benign microcalcifications	4	5
Cyst	5	6
Duct ectasia	2	3
Fibroadenoma	17	21
Fibroadipose tissue	1	1
Fibrocystic changes	15	19
Fibrosis	6	8
LCIS[a]	3	4
Lobular hyperplasia	1	1
Proteinaceous debris and neutrophils	1	1
Sclerosing adenosis	3	4
Stromal hyperplasia	1	1
Usual ductal hyperplasia	3	4
No cancer at 2-year follow-up[b]	Lesions ($N = 9$)	

NOS not otherwise specified

[a] High-risk lesions associated with an increased risk of breast cancer

[b] Nine lesions were followed without a biopsy and had no cancer at 2-year follow-up

analyzed in those lesions with <2 % estimated risk of IC ($n = 52$); the AUC was found to be 0.81 (95 % CI 0.69–0.93) (Fig. 2b). A third ROC curve ($n = 124$) was generated to compare the expert reader's highest risk estimate (for DCIS or IC) with an outcome of malignancy (DCIS or IC); the AUC was found to be 0.89 (95 % CI 0.83–0.95) (Fig. 2c).

Hypothetical biopsy thresholds: resultant biopsy rates, cancer-to-biopsy yields

The radiologist's risk estimates were stratified into BI-RADS risk categories and the types of lesions that were found for each category were identified and characterized (Table 4). Several possible scenarios for new biopsy thresholds are listed in Table 5 with defined categories for biopsy. Lesions whose risk estimates fall between the current thresholds (risk >2 % for either DCIS or IC) and the new proposed biopsy thresholds would be recommended for a 6-month follow-up and a subsequent biopsy if a change is noted. Effects of raising the biopsy threshold are shown in terms of malignant lesions missed and benign biopsies avoided in Table 5 and Fig. 3. Under current clinical guidelines for BI-RADS 4 and 5, all lesions in this analysis would have been biopsied with a cancer-to-biopsy yield of 28 %. The actual biopsies performed during CDEP generated a cancer-to-biopsy yield of 30 %.

Figure 3 shows the fraction of biopsies and the consequent cancer-to-biopsy yield for the three newly proposed biopsy thresholds chosen from Table 5 and designated as scenarios #1–3. The newly proposed biopsy thresholds are described in the following three scenarios. Scenario (1) If only lesions with risk estimates above 10 % for either DCIS or IC were recommended for biopsy, the cancer-to-biopsy yield would be 36 and 22 % of current biopsies would be avoided. No malignancies would be missed. (2) If only lesions with DCIS risk estimates above 50 % for DCIS or above 10 % for IC were recommended for biopsy, the cancer-to-biopsy yield would be 47 and 48 % of biopsies would be avoided. The consequence was one stage 1A IC missed (IDC, grade 2, 3 mm) and four DCIS lesions (including one grade 3 DCIS) missed. The IC would likely be picked up at 6-month follow-up if change was observed. (3) If the biopsy thresholds were increased such that lesions were biopsied only if they received IC risk estimates greater than or equal to 10 %, the cancer-to-biopsy yield would be 46 %, and 56 % of biopsies would be avoided. However, the diagnosis of two ICs (IDC, grade 2, 3 mm; IDC, grade 1, 2.5 mm) and eight DCIS lesions (including two grade 3 DCIS) would have been missed and postponed by 6 months or more. Focusing only on diagnosing IC, the percent of lesions where diagnosis would be postponed would be 0 % (scenario #1), 1 % (scenario #2), and 2 % (scenario #3). The total DCIS and IC lesions that would have been recommended for a 6-month follow-up were 5 of the 124 lesions (4 %) in scenario #2, and 10 lesions (8 %)

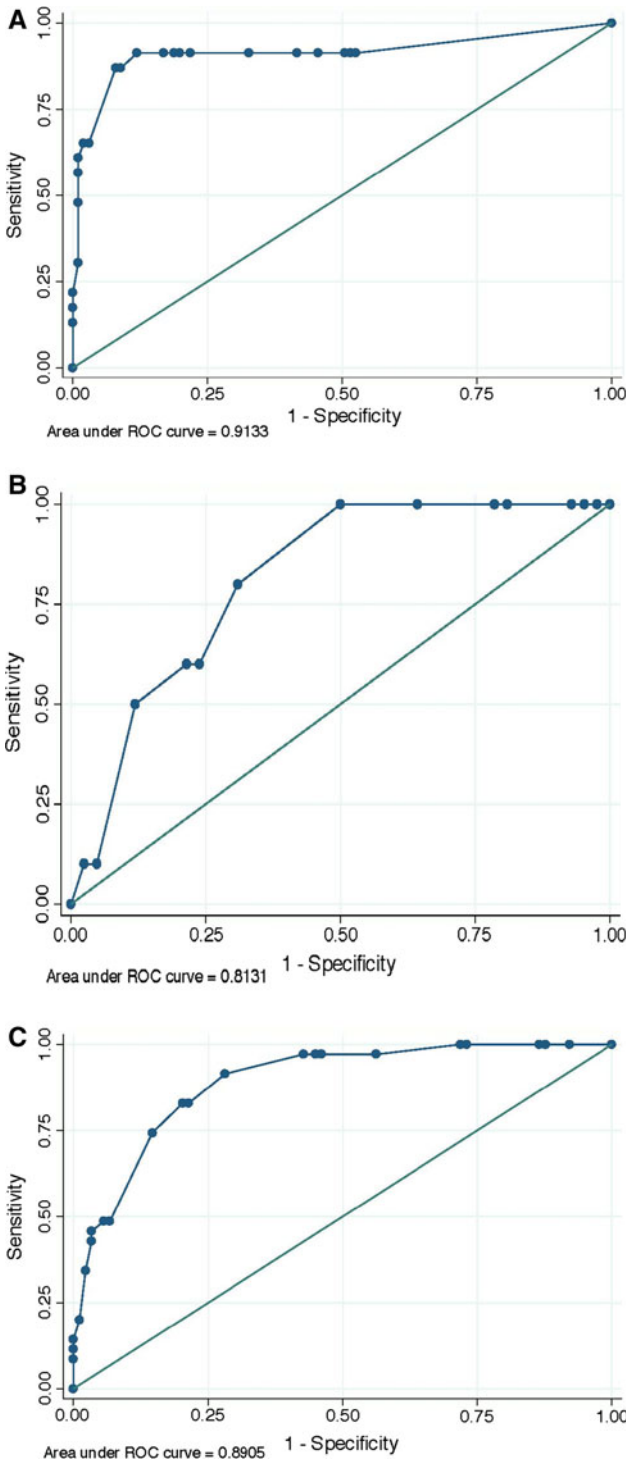

Fig. 2 ROC curve comparing the expert radiologist's risk estimates for **a** IC versus IC outcomes ($n = 124$) with an AUC of 0.91 (95 % CI 0.84–0.99), **b** DCIS versus DCIS outcome ($n = 53$) with an AUC of 0.81 (95 % CI 0.69–0.93), **c** DCIS or IC with outcome of malignancy (DCIS or IC) ($n = 124$) with an AUC of 0.89 (95 % CI 0.83–0.95)

Table 4 Malignancies identified within risk estimate categories

Corresponding BI-RADS category	Risk estimate range (%)		No. lesions	Final diagnosis			
	DCIS	IDC		Inv Ca	DCIS alone		Benign
					Gr. 1–2	Gr. 3	
DCIS-4A	2–9	<2	8	0	0		8
					0	0	
DCIS-4B	10–49	<2	33	1[a]	4		28
					3	1	
DCIS-4C	50–94	<2	12	1[b]	4		7
					3	1	
Inv Ca-4A	Any RE	2–9	18	0	0		18
					0	0	
Inv Ca-4B	Any RE	10–49	25	1	3		21
					2	1	
Inv Ca-4C	Any RE	50–94	23	15	1		7
					1	0	
Inv Ca-5	Any RE	95–100	5	5	0		0
					0	0	
Total			124	23	12		89

Data are number of lesions with risk estimates for DCIS or IC as percentages. Data to the right are number of lesions

Inv Ca invasive cancer, *Gr* grade

[a] IDC, grade 2, 3 mm (Stage 1A)

[b] IDC, grade 1, 2.5 mm (Stage 1A)

Table 5 The effect of increasing biopsy thresholds on biopsy rates, cancer-to-biopsy yields, and malignancies missed

Biopsy recommendation: (scenarios in Fig. 3a)	Lesions biopsied *N* (%)	Biopsies avoided *N* (%)	Number with IC or DCIS N	Ca/Bx yield (PPV) (%)	Number with DCIS rec 6-month F/U *N* (%)	Number with high-grade DCIS rec 6-month F/U N (%)	Number with IC rec 6-month F/U *N* (%)	Total number with IC or DCIS rec 6-month F/U N (%)
All BI-RADS 4 & 5 cases (current guidelines)	124 (100)	0 (0)	35	28	0 (0)	0 (0)	0 (0)	0 (0)
10 % ≤ DCIS or 2 % ≤ IC RE	116 (94)	8 (6)	35	30	0 (0)	0 (0)	0 (0)	0 (0)
10 % ≤ DCIS or 10 % ≤ IC RE (scenario #1)	97 (78)	27 (22)	35	36	0 (0)	0 (0)	0 (0)	0 (0)
50 % ≤ DCIS or 2 % ≤ IC RE	83 (67)	41 (33)	30	36	4 (3)	1 (1)	1 (1)	5 (4)
50 % ≤ DCIS or 10 % ≤ IC RE (scenario #2)	64 (52)	60 (48)	30	47	4 (3)	1 (1)	1 (1)	5 (4)
2 % ≤ IC RE	73 (59)	51 (41)	25	34	8 (6)	2 (2)	2 (2)	10 (8)
10 % ≤ IC RE (scenario #3)	54 (44)	70 (56)	25	46	8 (6)	2 (2)	2 (2)	10 (8)

Unless otherwise stated, data are the number of lesions with percentages in parentheses. The denominator is 124 lesions for all percentages except for the positive predictive value of the cancer to biopsy yield

RE risk estimate, *IC* invasive cancer, *Ca/Bx Yield* cancer to biopsy yield, *Rec* recommend, *Mo* month

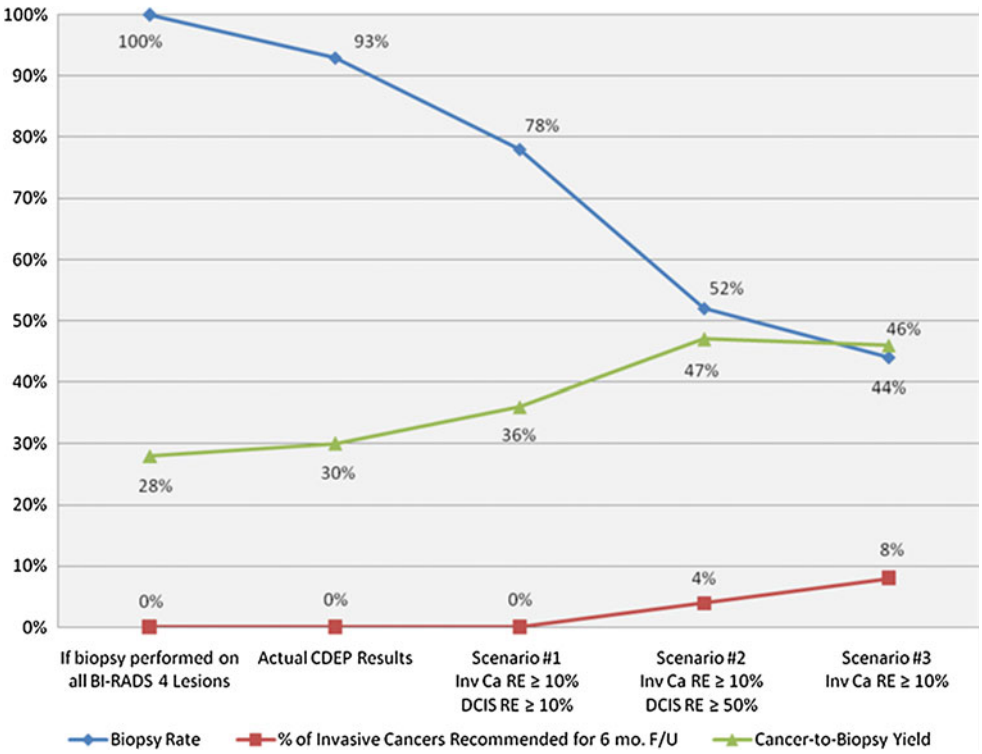

Fig. 3 Comparing biopsy threshold scenarios. The fraction of biopsies and the consequent cancer-to-biopsy yield of current guidelines, biopsy 100 %; the CDEP results; and three hypothetical biopsy threshold scenarios. The lesions recommending for 6-month follow-up in the scenarios on the *x*-axis are one IC (3 mm, ER+, low-grade invasive ductal carcinoma [IDC]) in scenario #2 and two ICs in scenario #3. If we consider high-grade DCIS, scenario #2 recommended one case and scenario #3 recommended two cases for 6-month follow-up

in scenario #3. Table 5 shows greater detail of the IC and DCIS cases that would have received a 6-month follow-up option instead of biopsy.

Discussion

Current criticisms of mammography screening programs include concern about overdiagnosis, the generation of false positives, and associated biopsies [17, 18]. While overdiagnosis is controversial for IC, there is a developing disquiet about overtreatment of DCIS, especially low-to-intermediate grade DCIS [19]. It is likely that majority of these lesions would not progress to IC, and if they do progress, the risk is 5–15 years from the original time of detection. High-grade DCIS may be associated with a higher risk of developing invasive breast cancer, and this risk is usually within 5 years of diagnosis. However, most DCIS is treated like invasive disease. Increasing awareness of the potential for overtreatment is leading to a reconsideration of the approach to DCIS, especially for low-to-intermediate grade DCIS, and a shift to explore chemoprevention as an alternative. Certainly, there is no urgency to detect such lesions. This pilot study was designed to determine whether it was possible to give a separate risk estimate for DCIS and IC and to lay the groundwork for predicting biologic type. Increasingly, we understand that breast cancer is a heterogeneous collection of diseases, where the tempo of disease ranges from indolent to aggressive. The results of this study show that an experienced radiologist can accurately provide risk estimates for both DCIS and IC. Revising thresholds for biopsy demonstrates that there is only a very low risk of delaying diagnosis, and the lesions for which diagnosis is delayed appear to be those with more indolent behavior. If these risk estimates can be validated by a larger study, then they could be used to place some calcifications in the BI-RADS 4A, 3, or even 2 categories and generate new biopsy threshold recommendations.

Given the substantial number of biopsies performed for benign lesions we sought to identify potential new thresholds to refine biopsy recommendations and to optimize management. Experienced mammographers are likely to be highly accurate in their ability to assign more refined risk estimates of both DCIS and IC, as set by BI-RADS subcategories. These categories can be used to assign new biopsy thresholds that may result in safely avoiding many benign biopsies, which is bourn out by some units reporting a high PPV for malignancy.

In this study of 124 lesions, 2 hypothetical thresholds for biopsy, scenario #1 and #2, seem most promising. In this pilot study, a biopsy threshold of $\geq$ 10 % DCIS or $\geq$ 10 % IC risk (scenario #1) avoids 22 % of biopsies with a cancer-to-biopsy yield of 36 % without delaying diagnosis for any malignant lesions. A biopsy threshold of $\geq$ 50 % DCIS or $\geq$ 10 % IC risk (scenario #2) results in avoiding 48 % of biopsies, a cancer-to-biopsy yield of 47 %, but postpones diagnosing one IC and four non-invasive (DCIS)

lesions. However, the IC was only 3 mm and low grade, and it is highly likely to have been identified 6 months later, still as stage 1a or 1b, with little consequence.

The introduction of mammographic screening led to a significant rise in detection of DCIS, which has become a target for screening [20]. The question is whether only high-grade DCIS should be a focus of early detection. DCIS now accounts for 20–30 % of all "malignant" diagnoses of breast cancer, almost entirely from screening. Yet after removal of approximately 60,000 DCIS cases annually for over 10 years, there has not been a concomitant drop in IC, suggesting that many of these lesions would not necessarily progress to IC if left undetected [17]. Although the natural history of DCIS is unknown, autopsy data indicate the existence of a reservoir of DCIS in the population that is never diagnosed and never attains clinical relevance [21, 22]. The consequence of delayed diagnosis of DCIS is likely to be negligible. In addition, there is a great value in risk stratifying low and intermediate versus high-grade DCIS. Low-grade DCIS lesions have an uncertain risk for progression to IC, as our understanding of the natural history of these lesions is poor. After excision, the risk of an IC developing ranges from 5 to 30 % over a period of 2–15 years after excision [23–25]. If a high-grade DCIS progresses, it will do so over a period of 2–5 years [26].

BI-RADS 4, with a wide range of risk of malignancy from 2 to 95 %, does not differentiate between a low-grade DCIS, which may never have clinical significance [27], and a consequential IC [17]. BI-RADS 4 includes many patients who do not have malignant or even high-risk lesions. The results demonstrate that with new biopsy thresholds, the United States can decrease biopsies performed for benign lesions to approach the cancer-to-biopsy yield rates of other countries such as Sweden (30–47 %) [28] and the United Kingdom (50–64 %) [6, 29]. Focusing on diagnosing IC or high-grade DCIS lesions may be one way to arrive at a threshold that lowers false positives while maintaining sensitivity for ICs.

BI-RADS 4 lesions that are ultimately benign are those that are most frequently thought to have a risk of non-high-grade DCIS. There is a fear of missing associated IC, although that is usually only identified in conjunction with the high-grade DCIS lesions that have fairly characteristic appearances on mammogram and are usually assigned a >50 % chance of being DCIS. The fact that DCIS is not an emergency, and does not require urgent intervention, may allow us to consider recommending a 6-month follow-up instead of biopsy. This is unlikely to have an impact on survival, even if the lesion ultimately is diagnosed as DCIS.

The fear of missing cancers is a potent driver of excess biopsies. Although controversial, there is increasing support for the view that some proportion of screen-detected cancers are slow-growing low-risk tumors, with indolent behavior [30–32]. A delay of 6 months in the detection of such lesions is unlikely to cause harm. The challenge is to distinguish benign and slow-growing lesions from those where there is an urgent need for resolution, recommending a short-term follow-up for the former and biopsy for the latter. For many low-risk radiographic findings, evidence of growth over time may help sort out which lesions require biopsy: IC lesions will progress and change over 6 months of observation and can be detected in a timely manner at an early stage.

The findings of this study demonstrate the potential clinical utility of experienced radiologists, providing separate risk estimates for DCIS and IC. In this small study, lesions that might have been missed and recommended for a 6-month follow-up were likely to have little if any immediate risk if diagnosis is delayed for 6 months. Biopsy thresholds also give radiologists and clinicians the justification and support for allowing disease dynamics to determine what is consequential and worthy of bringing to clinical attention [33].

This study has several weaknesses. First, as a pilot study, we have a small number of cases, which may not be representative. Also, intervention decisions may not be properly made at 6-month follow-up and diagnosis may be delayed. Women may not fully understand risk and the importance of follow-up. Lastly, only one experienced radiologist generated the risk estimates for this study. Our experienced radiologist's predictive ability may not be representative of other radiologists. We are in the process of validating this hypothesis in a reader study of 750 BI-RADS 4 and 5 lesions across five University of California Medical Centers as part of the Athena Breast Health Network. This will test academic radiologists of varying experience. If validated, we also plan to extend our study to community radiologists to show that this can work outside of academia.

This study suggests that using a biopsy threshold of risk estimates ≥ 50 % for DCIS and ≥ 10 % for IC may effectively and potentially safely improve cancer-to-biopsy yields. It was only intended as a pilot study to explore and validate new thresholds for biopsy, and has subsequently led to the above reader study. Management of lower risk lesions with a 6-month observation may increase patient anxiety and may only postpone biopsy, but may also enable us to safely observe specific lesions and decrease the biopsy rate. If validated, it will be necessary for clinicians to educate their patients about the safety of observation and communicate the plan for follow-up.

This pilot study found that following risk-based biopsy thresholds for BI-RADS 4 lesions by recommending a 6-month follow-up for the lowest-risk lesions, reclassifying to a BI-RADS 3 equivalent, may safely reduce biopsy rates and increase cancer-to-biopsy yields. These thresholds are not meant to be the definitive standards for biopsy but rather a starting point to move forward to determine what

thresholds best improve cancer-to-biopsy yields while avoiding a delay in diagnosis for consequential invasive lesions.

Acknowledgments Dr. Elias was supported by a Dutch Cancer Society fellowship and a René Vogels Foundation travel grant. Dr. O'Donoghue was supported by the Eleanor B. Pillsbury Fellowship at the University of Illinois Chicago Medical Center.

Conflict of interest None of the authors have any conflicts of interest.

References

1. Esserman L et al (2002) Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 94(5):369–375
2. Kerlikowske K et al (2011) Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med 155(8):493–502
3. Bent CK et al (2010) The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. AJR Am J Roentgenol 194(5):1378–1383
4. Weaver DL et al (2006) Pathologic findings from the Breast Cancer Surveillance Consortium: population-based outcomes in women undergoing biopsy after screening mammography. Cancer 106(4):732–742
5. Thorsen C, Eklund M, Ozanne E, Esserman LJ (2010) The Number of Screening Breast Biopsies Performed in the United States in 2010. In: American College of Surgeons 98th Clinical Congress 2012, Chicago, IL, USA
6. Smith-Bindman R et al (2003) Comparison of screening mammography in the United States and the United Kingdom. JAMA J Am Med Assoc 290(16):2129–2137
7. Hubbard RA et al (2011) Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med 155(8):481–492
8. Blanchard K et al (2006) Long-term risk of false-positive screening results and subsequent biopsy as a function of mammography use. Radiology 240(2):335–342
9. Baker SG, Kramer BS (2008) Estimating the cumulative risk of a false-positive under a regimen involving various types of cancer screening tests. J Med Screen 15(1):18–22
10. Brewer NT, Salz T, Lillie SE (2007) Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 146(7):502–510
11. Zagouri F et al (2008) Pain in different methods of breast biopsy: emphasis on vacuum-assisted breast biopsy. Breast 17(1):71–75
12. Yazici B et al (2006) Scar formation after stereotactic vacuum-assisted core biopsy of benign breast lesions. Clin Radiol 61(7):619–624
13. Geller BM et al (2006) Improving the concordance of mammography assessment and management recommendations. Radiology 241(1):67–75
14. Poplack SP et al (2000) Mammography in 53,803 women from the New Hampshire mammography network. Radiology 217(3):832–840
15. D'Orsi CJ, Bassett L, Berg WA et al. (2003) Breast imaging reporting and data system, BI-RADS: mammography, 4th edn. American College of Radiology, Reston, VA
16. Breast Imaging Reporting and Data System, BI-RADS: Mammography, 5th edn. Reston, VA: American College of Radiology (in press)
17. Esserman L, Shieh Y, Thompson I (2009) Rethinking screening for breast cancer and prostate cancer. JAMA 302(15):1685–1692
18. Bleyer A, Welch HG (2012) Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 367(21):1998–2005
19. Ozanne EM et al (2011) Characterizing the impact of 25 years of DCIS treatment. Breast Cancer Res Treat 129(1):165–173
20. Evans AJ, Blanks RG (2002) Should breast screening programmes limit their detection of ductal carcinoma in situ? Clin Radiol 57(12):1086–1089
21. Jorgensen KJ, Gotzsche PC (2009) Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 339:b2587
22. Kettritz U (2008) Modern concepts of ductal carcinoma in situ (DCIS) and its diagnosis through percutaneous biopsy. Eur Radiol 18(2):343–350
23. Page DL et al (1995) Continued local recurrence of carcinoma 15–25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer 76(7): 1197–1200
24. Sanders ME et al (2005) The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer 103(12): 2481–2484
25. Hughes LL et al (2009) Local excision alone without irradiation for ductal carcinoma in situ of the breast: a trial of the Eastern Cooperative Oncology Group. J Clin Oncol 27(32):5319–5324
26. Collins LC et al (2005) Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. Cancer 103(9):1778–1784
27. Jensen RA, Page DL (2003) Ductal carcinoma in situ of the breast: impact of pathology on therapeutic decisions. Am J Surg Pathol 27(6):828–831
28. Andersson I, Janzon L, Sigfusson BF (1985) Mammographic breast cancer screening—a randomized trial in Malmo, Sweden. Maturitas 7(1):21–29
29. Consolidated Guidance on Standards for the NHS Breast Screening Programme (2005) NHSBSP Publication No 60 (Version 2), April 2005. NHS Cancer Screening Programmes, Sheffield
30. Kalager M et al (2012) Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med 156(7):491–499
31. Welch HG, Black WC (2010) Overdiagnosis in cancer. J Natl Cancer Inst 102(9):605–613
32. Nelson HD et al. (2009) Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med 151(10):727–737, W237–42
33. Esserman L, Thompson I (2010) Solving the overdiagnosis dilemma. J Natl Cancer Inst 102(9):582–583