

## Ductal Lavage Is an Inefficient Method of Biomarker Measurement in High-Risk Women

Seema A. Khan,<sup>1</sup> Heather A. Lankes,<sup>2</sup> Deepa B. Patil,<sup>3</sup> Michele Bryk,<sup>1</sup> Nanjiang Hou,<sup>2</sup> David Ivancic,<sup>1</sup> Ritu Nayar,<sup>3</sup> Shahla Masood<sup>4</sup> and Alfred Rademaker<sup>2</sup>

### Abstract

Effective methods of serial epithelial sampling to measure breast-specific biomarkers will aid the rapid evaluation of new preventive interventions. We report here a proof-of-principle phase 2 study to assess the utility of ductal lavage (DL) to measure biomarkers of tamoxifen action. We enrolled women with a 5-year breast cancer risk estimate >1.6% or the unaffected breast of women with T<sub>1a</sub> or T<sub>1b</sub> breast cancer. After entry DL, participants chose tamoxifen or observation and underwent repeat DL 6 months later. Samples were processed for cytology and immunohistochemistry for estrogen receptor  $\alpha$ , Ki-67, and cyclooxygenase-2. Of 182 women recruited, 115 (63%) underwent entry and repeat DL; 85 (47%) had sufficient cells for analysis from  $\geq 1$  duct at both time points; in 78 (43%), cells were sufficient from  $\geq 1$  matched ducts. Forty-six women chose observation and 39 chose tamoxifen. We observed greater reductions in the tamoxifen group than in the observation group for Ki-67 (adjusted  $P = 0.03$ ) and estrogen receptor  $\alpha$  (adjusted  $P = 0.07$ ), but not in cyclooxygenase-2 (adjusted  $P = 0.4$ ) labeling. Cytologic findings showed a trend toward improvement in the tamoxifen group compared with the observation group. Interobserver variability for cytologic diagnosis between two observers showed good agreement ( $\kappa = 0.44$ ). Using DL, we observed the expected changes in tamoxifen-related biomarkers; however, poor reproducibility of biomarkers in the observation group, the 53% attrition rate of subjects from recruitment to biomarker analyses, and the expense of DL are significant barriers to the use of this procedure for biomarker assessment over time.

Ductal lavage (DL) is a minimally invasive technique that allows sampling of breast ductal epithelium in healthy high-risk women with a significantly better cell yield compared with nipple aspirate fluid (1–3). This allows the possibility of monitoring response to prevention agents by serial sampling of epithelial cells, using biomarkers relevant to the agent being tested. Thus, DL is a potentially attractive tool for assessing the effects of chemoprevention agents (1).

To establish the principle that biomarkers of chemoprevention agents can be monitored using DL, we initiated a phase 2, nonrandomized trial using tamoxifen, the gold-standard

breast cancer chemoprevention agent (4). We reasoned that the efficacy of tamoxifen in the prevention of estrogen receptor (ER)-positive breast cancer has been clearly shown and a study using tamoxifen as the preventive intervention should enable the efficient evaluation of DL as a tool for repeat epithelial sampling in phase 2 prevention studies. Furthermore, because tamoxifen does not prevent all breast cancer cases, if markers of tamoxifen efficacy could be identified it may be possible to target therapy to those most likely to benefit from it.

We now report the final results of this trial, having recruited 182 women, of whom 85 were evaluated for biomarker results at two time points. We present a comparative analysis of women who accepted tamoxifen therapy following baseline DL and those who declined. In addition, we address an aspect of DL that adds to the complexity and cost of this technique of biomarker monitoring; that is, whether there is an advantage to analyzing each duct separately or whether averaging duct samples from the same woman provides similar information. We have therefore reported data for individual ducts, as well as for women (an average of all ducts lavaged during a single procedure). We report here our findings on cytology, as well as cell number and other immunohistochemical biomarkers [ER $\alpha$ , Ki-67, and cyclooxygenase 2 (COX-2)], for the 39 women in the tamoxifen group and 46 in the observation group.

**Authors' Affiliations:** Departments of <sup>1</sup>Surgery, <sup>2</sup>Preventive Medicine, and <sup>3</sup>Pathology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; and <sup>4</sup>Department of Pathology, University of Florida College of Medicine, Jacksonville, Florida  
Received 6/25/2008; revised 10/24/2008; accepted 10/30/2008; published OnlineFirst 2/17/09.

**Grant support:** Bluhm Family Program for Breast Cancer Early Detection and Prevention and NIH/National Cancer Institute P50 CA89018-02 (Avon Progress for Patients Supplement). H.A. Lankes was supported by National Cancer Institute grant R25 CA100600.

**Requests for reprints:** Seema A. Khan, Department of Surgery, Northwestern University, 303 East Superior Street, Lurie 4-133, Chicago, IL 60611. Phone: 312-503-2112; Fax: 312-503-2555; E-mail: skhan@nmh.org.

©2009 American Association for Cancer Research.

doi:10.1158/1940-6207.CAPR-08-0119

## Materials and Methods

### Study design

Subjects were recruited between March 1, 2003, and March 31, 2006, from the Bluhm Family Program for Breast Cancer Early Detection and Prevention and the Lynn Sage Breast Center at Northwestern Memorial Hospital, Chicago, IL. The study was approved by the institutional review board of Northwestern University, and all participants signed a document of informed consent.

Potential subjects completed a questionnaire regarding breast cancer risk factors at their first visit and a 5-year breast cancer risk estimate was calculated using statistical models (5, 6). Eligible women ages 35 to 60 years were at increased risk for breast cancer (5-year risk estimate of >1.6%) or had completed local therapy for unilateral estrogen receptor–positive duct carcinoma *in situ* or invasive breast cancer <1 cm in size and did not require chemotherapy. Baseline DL was done by one of two operators (S.A.K. or M.B.). In premenopausal women, the DL procedure was not timed to a specific menstrual phase, but information regarding the last menstrual period and the onset of the next period was recorded. Participants were counseled regarding the cytologic findings from the DL samples, and the risks and benefits of tamoxifen therapy, and were allowed to choose tamoxifen therapy or observation. Subjects underwent repeat lavage 6 months later. As much as possible, follow-up DL in premenopausal women was done in the same phase of the cycle as the first DL procedure.

### Ductal lavage procedure

The details of the DL procedure have been described previously (1). Briefly, this involved application of a topical anesthetic cream, warming and massage of the breast, use of an aspirator cup (Cytoc Corp.) to elicit nipple aspirate fluid, and cannulation of fluid yielding and (when possible) non–fluid yielding ducts using a microcatheter (Cytoc Corp.) and a physiologic buffer solution (Plasmalyte; Baxter Healthcare Corporation). The lavage effluent was collected in Cytolyte (Cytoc Corp.). The location of the lavaged duct(s) was noted on an 8 × 8 nipple grid, and the duct orifice was identified with a knotted piece of prolene suture. The 12 o'clock axis of the areola was marked with a skin pen, and the nipple was photographed. When women returned for repeat DL, every attempt was made to cannulate ducts that had been previously lavaged (matched ducts). New fluid yielding ducts were also lavaged if any were identified.

### Analysis of cells

Samples were processed as previously described (1). Briefly, the lavage effluent was centrifuged and the cell pellet resuspended in 20 mL of Preservcvt (Cytoc Corp.) solution. A ThinPrep slide (Cytoc Corp.) was prepared from one fourth of the cell suspension and stained using the Papanicolaou technique (7). Cytology was evaluated by two observers (R.N. and S.M.) and categorized as insufficient cellular material for cytologic diagnosis, benign, mild atypia, severe atypia, or malignant, as described previously, using the consensus criteria developed for interpretation of DL samples (3). Discordant diagnoses were resolved for the final analysis by joint review. In the analysis by woman, the worst cytologic diagnosis was used; that is, if two ducts showed benign cytology but a third duct showed mild atypia, the woman was designated as having mild atypia for that procedure. A second aliquot was used for immunohistochemical evaluation of ER $\alpha$ , using heat antigen retrieval (DakoCytomation) and a 1:200 dilution of Clone SP1 (Lab Vision Corporation). A third aliquot was used for double immunohistochemical staining of the nuclear protein Ki-67 and the cytoplasmic protein COX-2 (1). Mouse monoclonal antibodies Ki-67, clone MIB-1 (DakoCytomation) diluted 1:200, and COX-2 (Cayman Chemical Company) diluted 1:100 were used sequentially, followed by color development (3,3'-diaminobenzidine for ER- $\alpha$  and COX-2 and Vector red for Ki-67). The validity of the double-staining procedure was established in MCF-7 cells by using single-label staining

for Ki-67 and COX-2 compared with the results of dual staining. There was an excellent correspondence of the fraction of labeled cells in single- and dual-stained slides over three separate experiments. The labeling index (LI) for each marker was calculated by counting positively and negatively stained epithelial cells (average of 1,000 cells per slide). A colocalization index (colocalization LI) was calculated as cells expressing both COX-2 and Ki-67. The intraobserver variability was assessed by blinded repetition of counts on 20 random immunohistochemistry slides each for ER- $\alpha$  and Ki-67–COX-2.

A minimum of 100 cells per slide (average 1,000 cells) was counted and the LI for each marker was calculated per duct as the number of positive cells divided by the total epithelial cell number. For each duct, the number of epithelial cells on all slides was summed to generate the total cell number per duct. The total cell number per woman was the sum of epithelial cells from all ducts lavaged during a single procedure; a woman had sufficient cellularity for analysis if she had at least one cellular duct at both time points. Total epithelial cell number was categorized as insufficient (<100 cells), borderline (100–399 cells), sufficient (400–999 cells), moderate (1,000–4,999 cells), and abundant (5,000 or more cells). The biomarker indices were generated per woman as the sum of the number of positive cells from all ducts divided by the sum of the number of epithelial cells from all ducts, expressed as a percentage.

### Statistical analysis

Statistical analyses were done both by woman and by matched duct. Subject age and 5-year breast cancer risk estimates were compared between the tamoxifen and observation groups using the independent sample *t* test. Race and menopausal status was compared between the groups using Fisher's exact test. Cell yield and cytology were categorized into five ordinal categories before analysis. Ordinal data were compared between the baseline and 6-month lavage using the weighted  $\kappa$  statistic and 95% confidence interval. Spearman correlations and accompanying *t* test for zero correlation were used to correlate baseline and 6-month markers measured in the same woman. Median biomarker levels were compared between the groups using the Wilcoxon rank sum test. When analyses were done by matched duct, the Wilcoxon rank-sum test for clustered samples was used (8). Cytology was compared between groups using the Fisher's exact test by woman and by duct. When adjusting for multiple ducts within a woman, cytology was analyzed using the test for proportions accounting for such clustering as described by Donner and Klar (9). Patients with missing data were excluded from the individual analysis for which the data was missing.

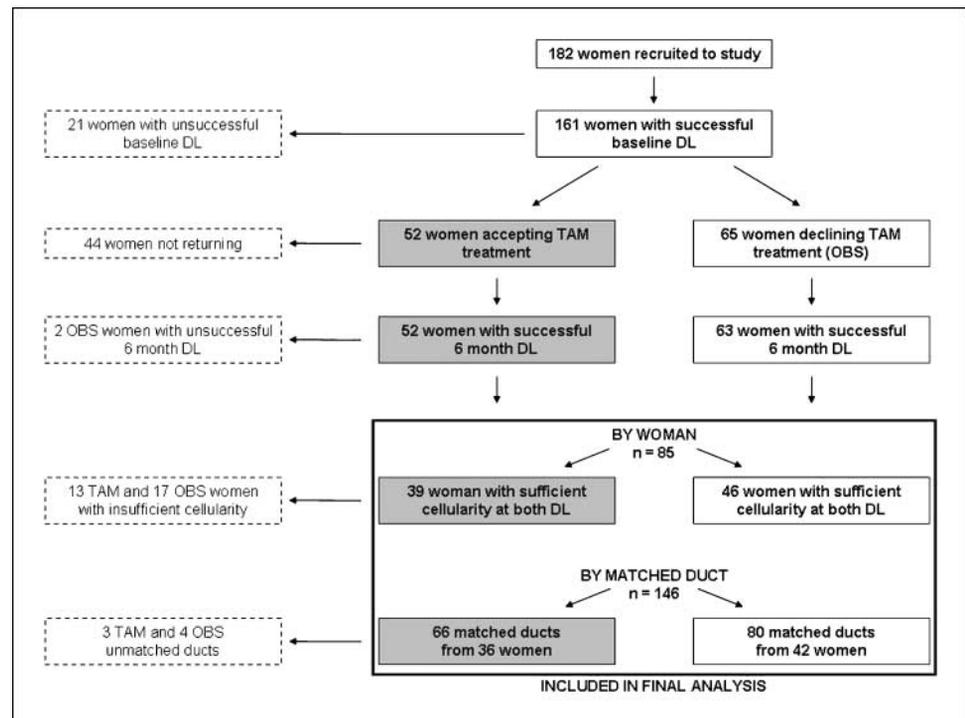
## Results

### Patient demographics

The success of DL procedures on the 182 women recruited to the study is presented in Fig. 1. Of the total study population, 161 women (88%) had at least one duct lavaged at baseline; 117 women (64%) had sufficient cellularity for analysis at baseline. Women who underwent baseline DL were asked to return for a repeat lavage at 6 months. Of 161 women, 44 (27%) did not return for reasons including difficulty with travel and work schedule; only 4 women stated that their unwillingness to return was due to pain of the procedure. Tamoxifen acceptance by the 117 women with sufficient cellularity for analysis at baseline was good: 52 (44%) chose tamoxifen treatment and the remaining 65 (56%) chose observation.

In the subsequent analyses in this report, the "by woman" analyses focus on the 85 women with sufficient cellularity at both time points (i.e.,  $\geq 100$  cells in at least one duct at both time points), of whom 39 accepted tamoxifen therapy and 46 declined. In 78 of these 85 women, a total of 146 ducts

Fig. 1. Study outline.



could be recannulated and lavaged again at the 6-month time point (designated “matched” ducts). These are included in the “by matched duct” analyses because the lavaged ducts were identical between baseline and repeat lavage. In the remaining 7 women, the ducts lavaged at baseline and 6 months were different, and therefore these 7 women are not included in the “matched duct” analysis. The median number of ducts lavaged at each time point was similar in the tamoxifen and observation groups, as was the number of ducts yielding more than 100 epithelial cells on at least one slide. The demographic characteristics of the 85 women with sufficient cellularity for analysis are presented in Table 1, which also shows the number of lavaged ducts and the biomarker values at baseline for the whole study population and the two analytic groups.

### Menstrual cycle phase

Forty-four women were premenopausal at study entry and were menstruating regularly. Baseline DL was done in the follicular phase in 24 women and in the luteal phase in 20 women (Table 2). Our study design stipulated that the second DL procedure be done in the same phase as the entry procedure; however, despite all attempts to achieve this, the phase at the two time points was discordant in 12 women because some women who were premenopausal at entry developed irregular periods or ceased menstruation during the 6 months following study entry (Table 1). This was more frequent in the tamoxifen group ( $P = 0.01$ ).

Cytology, cell number, and biomarker expression at baseline lavage were examined descriptively by menstrual cycle phase in the 85 women and 209 ducts lavaged at baseline (Table 2). The frequency of cytologic atypia was similar across menstrual phase and was observed in 54% and 45% of women in the follicular and luteal phases, respectively, and was lower in postmenopausal women (29%). When an-

alyzed by duct, the proportion of samples exhibiting cytologic atypia was similar between follicular phase, luteal phase, and postmenopausal women (34%, 20%, and 26%, respectively, Fisher's exact  $P = 0.97$ ). Cell number (when summed across ducts) was higher in follicular phase than in luteal phase samples and was similar between the luteal phase and postmenopausal samples. In agreement with published studies, there was a trend for ER $\alpha$  LI to be lower, in women who underwent DL during the luteal phase, compared with the follicular phase (10–12). There was no difference in the Ki-67 or COX-2 LI between women undergoing DL in the luteal phase compared with the follicular phase. By woman, cellular atypia was positively correlated with epithelial cell number ( $P < 0.0001$ ), as was cellular atypia with Ki-67 LI ( $P = 0.01$ ). By duct, epithelial cell number was positively correlated with atypia ( $P < 0.0001$ ).

### Reproducibility of cellular parameters in the observation group

Cytologic diagnosis; the number of epithelial cells obtained; and the ER $\alpha$ , Ki-67, and COX-2 LIs were compared across time points in the observation group to assess the reproducibility of these measures (Table 3). There was a decline in the total number of epithelial cells obtained between the baseline and 6-month lavage. There was good correlation between the ER $\alpha$  LI at both time points by woman ( $r = 0.49$ ,  $P = 0.0008$ ) and by matched duct ( $r = 0.25$ ,  $P = 0.04$ ). The correlation was borderline for COX-2 by woman ( $r = 0.31$ ,  $P = 0.05$ ) and significant by matched duct ( $r = 0.27$ ,  $P = 0.03$ ). The correlation between Ki-67 LI at both time points was nonsignificant by woman ( $r = 0.18$ ,  $P = 0.26$ ) and by matched duct ( $r = 0.10$ ,  $P = 0.40$ ), as shown in Table 3. The  $\kappa$  statistic (95% confidence interval) for agreement of cytologic diagnosis between time points in the observation group was 0.10 (–0.15–0.35) by woman and 0.05 (–0.16–0.26) by matched duct.

**Choice of tamoxifen therapy**

Of the 85 women who had sufficient cells for analysis at two time points, 39 accepted tamoxifen therapy and 46 declined. Tamoxifen users were not significantly older than nonusers (mean age 51 years versus 49 years,  $P = 0.136$ ). The fraction of tamoxifen users who showed cytologic atypia at baseline DL (15 of 39, 39%) was similar to those who declined tamoxifen (18 of 46, 39%). A diagnosis of DCIS or T<sub>1a</sub> or T<sub>1b</sub> carcinoma in the contralateral breast was more frequent among tamoxifen users than in the observation group (26% versus 11%) but not significantly so. Of those participants who did

not have a history of breast cancer, the mean 5-year breast cancer risk estimate was 3.1 for the tamoxifen users, compared with 3.0 for those who declined ( $P = 0.625$ ).

**Effect of tamoxifen therapy on cellular parameters**

In the 85 women (39 on tamoxifen and 46 who declined tamoxifen), the fraction of tamoxifen users who showed cytologic atypia at baseline DL (16 of 39, 41%) was similar to those who declined tamoxifen (18 of 46, 39%). In the tamoxifen group, 36% of women and 23% of matched ducts showed improvement in cytology, whereas in the observation group

**Table 1.** Subject demographics

	All women (n = 85)	Tamoxifen (n = 39)	Observation (n = 46)	P*
	Mean (Range)	Mean (Range)	Mean (Range)	
Age (y)	50 (33-64)	51 (42-63)	49 (33-64)	0.10
5-y risk estimate	3.0 (0.6-6.9)	3.1 (1.7-6.9)	3.0 (0.6-5.6)	0.64
	n (%)	n (%)	n (%)	P†
Cancer diagnosis				
DCIS	6 (7)	5 (13)	1 (2)	0.09
Invasive‡	9 (11)	5 (13)	4 (9)	0.73
Race				
White	69 (81)	35 (90)	34 (74)	0.09
Other	16 (19)	4 (10)	12 (26)	
Menopausal status (b/6)§				
Pre/pre	32 (38)	10 (26)	22 (48)	0.04
Post/post	41 (48)	19 (49)	22 (48)	0.99
Pre/post	12 (14)	10 (26)	2 (4)	0.01
	Median (mean)	Median (mean)	Median (mean)	P¶
Biomarker				
ER	24.59 (25.26)	27.98 (29.23)	19.88 (21.88)	0.002
Ki67	0.11 (0.59)	0.32 (0.57)	0.11 (0.50)	0.30
Cox2	40.00 (42.43)	46.15 (44.97)	36.64 (38.14)	0.10
Baseline				
No. lavaged ducts per woman				
All	3.0 (2.98)	2.0 (2.72)	3.0 (3.2)	0.05
Cell counts ≥100	2.0 (2.46)	2.0 (2.44)	2.0 (2.48)	0.47
Biomarker				
ER	22.17 (22.88)	23.86 (24.25)	20.14 (21.43)	0.18
Ki67	0.00 (0.31)	0.00 (0.20)	0.08 (0.42)	0.25
Cox2	38.91 (39.25)	35.09 (38.88)	31.67 (35.51)	0.71
6 mo				
Median lavaged ducts per woman				
All	3.0 (3.04)	3.0 (2.92)	3.0 (3.13)	0.23
Cell counts ≥ 100	2.0 (2.29)	2.0 (2.18)	2.0 (2.39)	0.13

\*P value from two-sided independent sample *t* test.

†P value from two-sided Fisher's exact test.

‡T<sub>1a/b</sub> carcinoma in contralateral breast.

§Self-reported at baseline (b) and 6 mo (6) DL.

¶P value from Wilcoxon signed-rank test.

**Table 2.** Summary of cellular markers by menstrual status at baseline DL

	Premenopausal (n = 44)		Postmenopausal (n = 41)
	Follicular (n = 24)	Luteal (n = 20)	
	Median (IQ range)	Median (IQ range)	Median (IQ range)
ER LI	25 (18-38)	17 (13-29)	25 (19-34)
Ki-67 LI	0.14 (0.04-1.07)	0.17 (0.04-0.58)	0.11 (0-0.50)
COX-2 LI	44 (29-53)	44 (21-49)	37 (30-60)
Cell no.	21,782 (13,782-36,816)	11,635 (6,796-23,074)	12,087 (6,518-22,713)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Cytology			
ICMD	1 (4)	0 (0)	2 (5)
Benign	10 (42)	11 (55)	27 (66)
Atypia	13 (54)	9 (45)	12 (29)
	Premenopausal (n = 120)		Postmenopausal (n = 89)
	Follicular (n = 71)	Luteal (n = 49)	
	Median (IQ range)	Median (IQ range)	Median (IQ range)
ER LI	24 (16-35)	17 (10-25)	25 (17-34)
Ki-67 LI	0.09 (0-0.97)	0.09 (0-0.66)	0.10 (0-0.47)
COX-2 LI	40 (28-56)	34 (18-46)	40 (28-60)
Cell no.	6,531 (4,831-11,599)	4,655 (1,841-9,217)	6,366 (2,920-11,875)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Cytology			
ICMD	2 (3)	0 (0)	4 (4)
Benign	45 (63)	39 (80)	62 (70)
Atypia	24 (34)	10 (20)	23 (26)

Abbreviations: IQ, interquartile range; ICMD, insufficient cellular material for diagnosis.

improved cytology was observed in 22% of women and 18% of matched ducts. There was also a deterioration in cytology findings, from benign to mild atypia; in the tamoxifen group, this occurred in 13% of women and 9% of matched ducts, compared with worsened cytology in 17% of women and 15% of matched ducts in the observation group. Thus, improvement was more frequent than worsening of cytology in the tamoxifen group than in the observation group, but these differences were not statistically significant.

Reductions in ER $\alpha$ , Ki-67, and COX-2 from baseline to 6 months were observed in the tamoxifen group and were significant for Ki-67 by woman ( $P = 0.04$ ) and by matched duct ( $P = 0.002$ ). When the analyses were adjusted for multiple ducts within women, the differences in Ki-67 were still significant ( $P = 0.03$ ; Table 4). However, these findings should still be considered tentative because the fraction of Ki-67-positive cells was very low. In the tamoxifen group, median Ki-67 LI

went from 0.32 to 0, and in the observation group from 0.11 to 0.08. Additionally, there was a larger proportion of women in the tamoxifen group who had a history of early breast cancer. The decrease in ER $\alpha$  was also greater in the tamoxifen group, but was of borderline significance when the analyses were adjusted for multiple ducts per woman ( $P = 0.07$ ). For COX-2, there were no significant differences by woman or by matched duct.

#### Interobserver correlations of cytologic diagnosis

We performed a comparative analysis of cytologic diagnosis to assess the reproducibility of results between two experienced cytopathologists (R.N. and S.M.). The results presented in Table 5 include all samples (baseline and 6 months) and indicate that there was good correlation between the two observers. Of the 306 ducts analyzed by both observers (R.N. and S.M.), cytologic diagnosis was concordant in 205 (67%) of

samples: 144 were diagnosed with benign cytology by both observers, 20 ducts were diagnosed with mild atypia, and 41 had insufficient cellular material for diagnosis; however, 101 samples (33%) were given discordant diagnoses of insufficient, benign, or mild atypia. The weighted  $\kappa$  statistic was 0.44 (95% confidence interval, 0.36-0.53).

## Discussion

We have previously reported that DL samples can be used for the measurement of biomarkers and that sufficient cellularity for analysis at the initial lavage was obtained from 70% of the women (1). Here, we report the results of serial monitoring of breast epithelium, comparing women treated with tamoxifen therapy with women who chose observation. We chose a 6-month interval for the assessment of biomarker modulation because this is frequently employed in phase 2 prevention studies (13, 14); moreover, a demonstration of biomarker stability at the 6-month time point would be a good starting point for future studies. Additionally, we intended to assess the modulation of cytologic atypia by tamoxifen and it seemed unlikely that a shorter interval of therapy would achieve this. A longer interval may have affected compliance for return lavage and would have increased the duration and therefore the expense of the study.

We found that sufficient numbers of epithelial cells for biomarker analysis were obtained in only 47% (85 of 182) of women at both time points due to attrition of the study population at several levels (see Fig. 1). In the tamoxifen group, we saw the expected declines in the widely used biomarkers that we chose to study, selected for their expected response to tamoxifen therapy (cytologic atypia, ER $\alpha$ , and Ki-67); however, the attrition of the sample size and the resulting decrease in power, along with a high level of variability in cytologic and immunohistochemical parameters in the observation group, rendered it difficult to identify statistically significant changes

between the tamoxifen and observation groups. We saw trends toward improvement in cytologic findings and ER $\alpha$  expression. The Ki-67 labeling index was substantially lower in our DL samples than the ~2% value reported for normal breast samples obtained by random fine needle aspiration or core biopsy. This may be related to the fact that the luminal cells that are presumably exfoliated and collected during a DL procedure have lower proliferation rates. However, we saw a significant decline in Ki-67 LIs between the tamoxifen and observation groups by woman ( $P = 0.04$ ) and by matched duct ( $P = 0.002$ ) and after adjustment for multiple ducts per subject ( $P = 0.03$ ). This was despite poor reproducibility of Ki-67 LI across the two time points, suggesting a strong effect of tamoxifen on cell proliferation. The poor reproducibility of immunohistochemical biomarkers may, in part, be attributed to the fact that our cellularity threshold for inclusion of immunohistochemistry slides was >100 epithelial cells, in contrast to studies using random fine needle aspiration samples where Fabian et al. have used a threshold of >500 epithelial cells. The higher threshold is more feasible for random FNA material, where 16 to 20 needle passes from two breasts are pooled, than for DL samples that are typically handled as separate samples for each individual duct. This consideration was partly what drove us to examine the question of whether maintaining separate duct samples has any advantage over pooling them (see below). Nevertheless, the mean number of cells on immunohistochemistry slides in our study was 1,000, and only 9% were in the 100 to 500 cell range.

When we designed our study in 2002 to 2003, there was significant interest in cytologic findings in DL samples. Cytologic atypia had been identified as a potential surrogate end point in phase 2 prevention studies (15) and although its reversibility had never been shown, we postulated that using DL to sample epithelial cells from the same ductal tree over time,

**Table 3.** Reproducibility of cellular markers at baseline and 6-mo DL samples in the observation group

### By woman ( $n = 46$ )

	ER LI	Ki-67 LI	COX-2 LI	Cell no:
Women	44	41	41	46
Baseline median	20	0.11	37	13324
6-mo median	20	0.08	32	8118
Spearman correlation	0.49	0.18	0.31	0.43
$P$	0.0008	0.26	0.05	0.003

### By Matched duct ( $n = 80$ )

	ER LI	Ki-67 LI	COX-2 LI	Cell no:
Matched ducts	71	71	71	80
Baseline median	20	0.08	36	6599
6-mo median	20	0	36	3406
Spearman correlation	0.25	0.10	0.27	0.08
$P$	0.04	0.40	0.03	0.47

NOTE:  $r$ , Spearman correlation coefficient.

**Table 4.** Tamoxifen-related biomarker changes in DL samples**Median Difference\* by woman (n = 85)**

	ER LI	Ki-67 LI	COX-2 LI	Cytology	
	Median (IQ range)	Median (IQ range)	Median (IQ range)	Improved	Worsened
Tamoxifen	-1.85 (-13.68-5.09)	-0.16 (-0.65-0.00)	-7.10 (-24.47-8.04)	36% (14/39)	13% (5/39)
n	37	36	36		
Observation	-0.62 (-6.52-4.09)	0.00 (-0.12-0.21)	-2.35 (-14.90-8.96)	22% (10/46)	17% (8/46)
n	44	41	41		
Difference	-1.23	-0.16	-4.75		
P <sup>†</sup>	0.25	0.04	0.29	0.23 <sup>‡</sup>	0.76 <sup>‡</sup>

**Median Difference\* by matched duct (n = 146)**

	ER LI	Ki-67 LI	COX-2 LI	Cytology	
	Median (IQ range)	Median (IQ range)	Median (IQ range)	Improved	Worsened
Tamoxifen	-7.06 (-16.42-6.63)	-0.11 (-0.43-0.00)	-2.97 (-24.27-8.01)	23% (15/66)	9% (6/66)
n	58	59	59		
Observation	-0.49 (-8.74-8.32)	0.00 (-0.20-0.30)	-3.02 (-13.33-8.94)	18% (14/80)	15% (12/80)
n	71	71	71		
Difference	-6.57	-0.11	0.05		
P <sup>†</sup>	0.05	0.002	0.40	0.53 <sup>‡</sup>	0.32 <sup>‡</sup>
Adjusted P <sup>§</sup>	0.07	0.03	0.40	0.45 <sup>  </sup>	0.27 <sup>  </sup>

\*Six-month lavage minus baseline lavage.

<sup>†</sup>Wilcoxon rank-sum test comparing medians between observation and tamoxifen groups.<sup>‡</sup>Fisher's exact test comparing proportions between observation and tamoxifen groups.<sup>§</sup>Wilcoxon rank-sum test comparing medians between observation and tamoxifen groups, adjusting for multiple ducts within women.<sup>||</sup>Donner and Klar test comparing proportions between observation and tamoxifen groups, adjusting for multiple ducts within women.

it may be feasible to show reversal of cytologic atypia in a given duct. We did not restrict entry to women with atypical samples at baseline because we wanted to assess both aspects of variability in cytology: change from atypia to benign and from benign to atypical. We report here that cytologic findings from the same duct were variable over time; however, in the tamoxifen group, changes were more likely in the direction of improvement from mild atypia to benign cytology (44% improvement) than in the observation group (16% improvement; Table 4). Although we did not see a statistically significant difference in the cytologic improvement between tamoxifen and observation groups, the relationships between atypia and other parameters were similar to those observed in previous studies of DL (16) and random fine needle aspiration (17) in healthy, high-risk women. Thus, epithelial cell number was positively correlated with atypia ( $P < 0.0001$ ), as was cellular atypia with Ki-67 LI ( $P = 0.01$ ). In addition, as previously reported (10, 18), the ER $\alpha$  LI was lower in women who underwent DL during the luteal phase when compared with women undergoing DL in the follicular phase. There has been some discussion regarding the possibility that mild atypia in DL samples may be nonspecific and related to menstrual cycle changes. We did not find any specific differences in cytologic atypia rates by menstrual cycle phase or menopausal status.

Because the interpretation of cytology has a subjective component, we included a blinded review of cytologic diagnosis by an expert reference cytopathologist in our study design. This analysis showed that the variability of cytologic diagnosis is

**Table 5.** Interobserver variability of cytologic diagnosis for 306 ducts analyzed by two observers

First observer (R.N.)	Second observer (S.M.)				Total
	IMCD	Benign	Mild atypia	Severe atypia	
ICMD	41	32	6	0	79
Benign	3	144	41	1	189
Mild atypia	0	16	20	0	36
Severe atypia	0	0	2	0	2
Total	44	192	69	1	306

NOTE: Shaded cells depict samples with concordant diagnoses by two observers. Weighted  $\kappa$  for all ducts: Good (0.44; 95% CI, 0.36-0.53).

only partly related to interobserver variability of interpretation, as suggested by the moderate  $\kappa$  statistic of 0.44. This is very similar to our previous findings in women who underwent DL before mastectomy (1–3). Discordance in interpretation was mostly attributable to samples that exhibited benign or mildly atypical cytology. Similar results were documented in a premastectomy DL study (19) where mild atypia constituted one third of samples and was the most challenging and least reproducible diagnostic category. In another study of DL reproducibility, 14 of 69 women entered into the study underwent two DL procedures and the reproducibility of cellularity and cytology was very similar to that observed in our study (20).

We included premenopausal and postmenopausal women in this study because tamoxifen is effective in both groups. By design, we did not mandate the menstrual cycle phase for the entry lavage because the optimal menstrual cycle phase for breast epithelial sampling in studies of risk and prevention is not known. Some investigators perform such sampling only in follicular phase (21) and others do not time sampling by phase (22). We have found in a previous case-control study using breast biopsy samples that biomarker differences may be more marked in the luteal phase than in the follicular phase (18). Additionally, fixing the sampling in a specific phase adds to the difficulty of scheduling and recruitment. We reasoned, therefore, that it would be optimal to allow the first sample to be collected in either follicular or luteal phase, but we would limit within-person variability by asking women to return for the repeat procedure in the same phase of the cycle as the initial lavage. We would then be able to explore the advantages of sampling in one phase over the other and apply the results to the design of future studies. However, women on tamoxifen developed irregular periods or ceased menstruation during the course of the study significantly more frequently, producing some variation in the endocrine status of women at the two time points, which may contribute to the variability in the DL findings. Our experience highlights the importance of uniformity in menstrual phase and menopause status, and the difficulty of prevention trials designed in an age group that straddles menopause. Although the age group of 40 to 60 years is the most appropriate for testing prevention agents in terms of breast cancer risk and motivation, the changing endocrine environment presents a challenge.

Although DL allows for the repeated sampling of the breast epithelium from the same duct over time, with an expected improvement in the reproducibility of biomarker findings, the analysis of several samples per procedure adds to the expense of the procedure, the expense of laboratory assays, and the complexity of statistical analysis. Our study is the first to assess the advantage of analyzing each duct separately versus averaging all duct samples from a single DL procedure. The descriptive analysis of cellular markers at baseline was similar whether examined by woman or by duct (Table 2). Furthermore, analyses of the reproducibility of cellular markers at both time points in the observation group showed no advantage for the by matched

duct analysis (Table 3), and the tamoxifen-related changes in biomarkers showed only a marginal advantage when analyzed by woman or by matched duct (Table 4). We conclude from this experience that pooling DL samples from different ducts would increase the efficiency of this sampling method.

We experienced high attrition rate in the study subjects related to a variety of factors; these included the inability to perform baseline DL ( $n = 21$ ), failure to return for a second lavage ( $n = 44$ ), inability to perform repeat lavage on subjects who returned ( $n = 2$ ), insufficient cellular material for analysis from both lavage procedures ( $n = 30$ ), and unmatched ducts yielding sufficient cells for analysis in ( $n = 7$ ). Thus, of 182 women recruited to the study, 85 (47%) had successful DL at both time points with sufficient cellular material for analysis, and 78 (43%) had matched ducts with sufficient cellular material for analysis. The largest sources of subject attrition were failure to return (due to travel distance and work schedules) for the second DL procedure (44 of 182, 24%) and insufficient cells at two time points (30 of 182, 16%). Our results are somewhat better than those reported in a previous smaller study of DL at two time points (22). In that study, a total of 67 women were recruited, 22 (32%) did not return for the second procedure, and DL could be repeated 6 months later on 19 women (28%).

Two previous small studies have compared DL to random fine needle aspiration at a single time point in very similar study populations and have found DL to have low cellular yield (23, 24). In the larger of these (23), 86 women were recruited. DL could be performed in 38 ducts and samples adequate for cytologic assessment were obtained in 27 of 86 (31%) of subjects. We used a higher threshold for adequacy (100 rather than 10 epithelial cells) and found a higher proportion of adequate samples at baseline (128 of 182 women, 70%) than in these studies; however, the high attrition rate discussed above, some of which is related to poorer cell yield at the second DL procedure, prevents us from endorsing DL as an improvement over existing tools for biomarker assessment in healthy high-risk women because we do not see any improvement in reproducibility of biomarkers when the analyses are restricted specifically to ducts that were recannulated at two time points.

In summary, we observed the expected trends in tamoxifen-related biomarkers using DL for breast epithelial sampling of the healthy, high-risk breast. However, a 53% attrition rate of subjects from recruitment to biomarker analyses, the expense of the catheter, the time required for the procedure, and the analysis of multiple samples per woman at each time point renders DL an extremely expensive method of breast epithelial sampling. This high cost and the variability of findings over time in the observation group means that this procedure is of questionable utility for biomarker assessment over time in high-risk women.

### Disclosure of Potential Conflicts of Interest

Portions of this work were presented by the first author as an oral abstract at the 2007 American Society of Clinical Oncology Annual Meeting, June 1 to 5, 2007, in Chicago, IL.

### References

1. Bhandare D, Nayar R, Bryk M, et al. Endocrine biomarkers in ductal lavage samples from women at high risk for breast cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14:2620–7.
2. Khan SA, Wolfman JA, Segal L, et al. Ductal lavage findings in women with mammographic microcalcifications undergoing biopsy. *Ann Surg* 2005;12:689–96.
3. Khan SA, Wiley EL, Rodriguez N, et al. Ductal lavage findings in women with known breast cancer undergoing mastectomy. *J Natl Cancer Inst* 2004;96:1510–7.

4. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;90:1371–88.
5. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually [see comments]. *J Natl Cancer Inst* 1989;81:1879–86.
6. Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer* 1994;73:643–51.
7. Marshall PN. Papanicolaou staining—a review. *Microsc Acta* 1983;87:233–43.
8. Rosner B, Glynn RJ, Lee ML. Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach. *Biometrics* 2003;59:1089–98.
9. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994;140:279–89.
10. Battersby S, Robertson BJ, Anderson TJ, King RJB, McPherson K. Influence of menstrual cycle, parity and oral contraceptive use on steroid hormone receptors in normal breast. *Br J Cancer* 1992;4:601–7.
11. Boyd M, Hildebrandt RH, Bartow SA. Expression of the estrogen receptor gene in developing and adult human breast. *Breast Cancer Res Treat* 1996;37:243–51.
12. Williams G, Anderson E, Howell A, et al. Oral contraceptives (OCP) use increases proliferation and decreases oestrogen receptor content of epithelial cells in the normal human breast. *Int J Cancer* 1991;48:206–10.
13. Fabian CJ, Kimler BF, Brady DA, et al. A phase II breast cancer chemoprevention trial of oral  $\alpha$ -difluoromethylornithine: breast tissue, imaging, and serum and urine biomarkers. *Clin Cancer Res* 2002;8:3105–17.
14. Fabian CJ, Kimler BF, Zalles CM, et al. Reduction in proliferation with six months of letrozole in women on hormone replacement therapy. *Breast Cancer Res Treat* 2007;73:643–51.
15. O'Shaughnessy JA, Kelloff GJ, Gordon GB, et al. Treatment and prevention of intraepithelial neoplasia: an important target for accelerated new agent development. *Clin Cancer Res* 2002;8:314–46.
16. Cazzaniga M, Severi G, Casadio C, Chiapparini L, Veronesi U, Decensi A. Atypia and Ki-67 expression from ductal lavage in women at different risk for breast cancer. *Cancer Epidemiol Biomarkers Prev* 2006;15:1311–5.
17. Khan QJ, Kimler BF, Clark J, Metheny T, Zalles CM, Fabian CJ. Ki-67 expression in benign breast ductal cells obtained by random periareolar fine needle aspiration. *Cancer Epidemiol Biomarkers Prev* 2005;14:786–9.
18. Khan SA, Rogers MA, Khurana KK, Meguid MM, Numann PJ. Estrogen receptor expression in benign breast epithelium and breast cancer risk [see comments]. *J Natl Cancer Inst* 1998;90:37–42.
19. Brogi E, Robson M, Panageas KS, Casadio C, Ljung BM, Montgomery L. Ductal lavage in patients undergoing mastectomy for mammary carcinoma: a correlative study. *Cancer* 2003;98:2170–6.
20. Visvanathan K, Santor D, Ali SZ, Hong IS, Davidson NE, Helzlsouer KJ. The importance of cytologic intrarater and interrater reproducibility: the case of ductal lavage. *Cancer Epidemiol Biomarkers Prev* 2006;15:2553–6.
21. Fabian CJ, Kimler BF, Zalles CM, et al. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. *J Natl Cancer Inst* 2000;92:1217–27.
22. Visvanathan K, Santor D, Ali SZ, et al. The reliability of nipple aspirate and ductal lavage in women at increased risk for breast cancer—a potential tool for breast cancer risk assessment and biomarker evaluation. *Cancer Epidemiol Biomarkers Prev* 2007;16:950–5.
23. Arun B, Valero V, Logan C, et al. Comparison of ductal lavage and random periareolar fine needle aspiration as tissue acquisition methods in early breast cancer prevention trials. *Clin Cancer Res* 2007;13:4943–8.
24. Zalles CM, Kimler BF, Simonsen M, Clark JL, Metheny T, Fabian CJ. Comparison of cytomorphology in specimens obtained by random periareolar fine needle aspiration and ductal lavage from women at high risk for development of breast cancer. *Breast Cancer Res Treat* 2006;97:191–7.

# Cancer Prevention Research

## Ductal Lavage Is an Inefficient Method of Biomarker Measurement in High-Risk Women

Seema A. Khan, Heather A. Lankes, Deepa B. Patil, et al.

*Cancer Prev Res* 2009;2:265-273. Published OnlineFirst February 17, 2009.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1940-6207.CAPR-08-0119](https://doi.org/10.1158/1940-6207.CAPR-08-0119)

**Cited articles** This article cites 24 articles, 8 of which you can access for free at:  
<http://cancerpreventionresearch.aacrjournals.org/content/2/3/265.full#ref-list-1>

**Citing articles** This article has been cited by 1 HighWire-hosted articles. Access the articles at:  
<http://cancerpreventionresearch.aacrjournals.org/content/2/3/265.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerpreventionresearch.aacrjournals.org/content/2/3/265>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.