Breast-tissue sampling for risk assessment and prevention

C J Fabian¹, B F Kimler², M S Mayo³ and S A Khan⁴

¹Departments of Internal Medicine  
²Radiation Oncology  
³Preventive Medicine and Public Health, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KA 66160, USA  
⁴Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

(Requests for offprints should be addressed to C J Fabian; Email: cfabian@kumc.edu)

Abstract

Breast tissue and duct fluid provide a rich source of biomarkers to both aid in the assessment of short-term risk of developing breast cancer and predict and assess responses to prevention interventions. There are three methods currently being utilized to sample breast tissue in asymptomatic women for risk assessment: nipple-aspirate fluid (NAF), random periareolar fine-needle aspiration (RPFNA) and ductal lavage. Prospective single-institution trials have shown that the presence of atypical cells in NAF fluid or RPFNA specimens is associated with an increased risk of breast cancer. Furthermore, RPFNA-detected atypia has been observed to further stratify risk based on the commonly used Gail risk-assessment model. A prospective trial evaluating risk prediction on the basis of atypical cells in ductal-lavage fluid is ongoing. The ability of other established non-genetic biomarkers (mammographic breast density; serum levels of bioavailable estradiol, testosterone, insulin-like growth factor-1 and its insulin like growth factor binding protein-3) to stratify risk based on the Gail model is as yet incompletely defined. Modulation of breast intra-epithelial neoplasia (i.e. hyperplasia with or without atypia) with or without associated breast-tissue molecular markers, such as proliferation, is currently being used to evaluate response in Phase II chemoprevention trials. RPFNA has been the method most frequently used for Phase II studies of 6–12 months duration. However, ductal lavage, RPFNA and random and directed core needle biopsies are all being utilized in ongoing multi-institutional Phase II studies. The strengths and weaknesses of each method are reviewed.

Use of breast-tissue biomarkers in risk assessment and prevention

Need for risk biomarkers

Over 211 000 women are estimated to develop invasive breast cancer in the USA in 2005 (Jemal et al. 2005). These women will generally undergo some combination of surgery, radiation, antihormone and/or chemotherapy which for many will result in appreciable long-term morbidity (Kuehn et al. 2000, Stanton et al. 2001, Ganz et al. 2002). Despite advances in early detection and treatment, 40 000 women previously diagnosed with invasive breast cancer were predicted to die in 2004 (Jemal et al. 2005). Prevention would be a preferable alternative to treatment of established disease, if those women most likely to benefit from the prevention intervention could be readily identified. Tamoxifen has been identified as a cost-effective intervention for primary risk reduction for asymptomatic women of 35–70 years without prior invasive cancer if they have previously had a biopsy exhibiting atypical ductal hyperplasia (ADH), ductal or lobular carcinoma in situ (DCIS, LCIS), or currently have an estimated 5-year Gail model risk of >1.67% (Fisher et al. 1998, Cuzick et al. 2002, Hershman et al. 2002). Tamoxifen has been recommended by the US Preventive Services Task Force (Kinsinger et al. 2002), the American Society of Clinical Oncology (Chlebowski et al. 1999) and the Canadian Task Force on Preventive Health (Levine et al. 2001) under these circumstances. Yet, despite a relative reduction in cancer incidence of 32–49%, only a minority of high-risk women without a prior diagnosis of DCIS or invasive cancer agree to take it following a recommendation by...
their health-care provider (Port et al. 2001, Vogel et al. 2002, Bober et al. 2004, Tchou et al. 2004). A woman's reluctance to take 5 years of tamoxifen as preventive therapy appears to be based on the fear of side effects coupled with uncertainty of the benefits, particularly if the 5-year Gail model risk of >1.67% is the primary tool used to determine suitability for prevention therapy (Port et al. 2001). The Gail risk model is based on five variables captured as part of the Breast Cancer Detection Project (BCDP): current age, age at menarche, first live birth, number of breast biopsies and number of affected first-degree relatives, as well as a correction factor for atypical hyperplasia if it has been observed in a diagnostic biopsy (Gail et al. 1989; Table 1). The Gail model is simple to use and has been validated for populations undergoing regular screening and an updated version with 5-, 10-, 20- and 30-year risk calculated by race is available on the National Cancer Institute (NCI) website (http://bcra.nci.nih.gov/brc/). Unfortunately, it has only modest discriminatory value for the individual woman, and thus may not be helpful in decision-making with regards to whether to take tamoxifen for prevention (Rockhill et al. 2001). Indeed, Freedman et al. (2003) have suggested that benefit from tamoxifen prevention therapy is likely to accrue to less than 25% of Caucasian women of ages 35–70 identified as high risk on the basis of a 5-year Gail risk of >1.67%. A large number of risk factors are not considered by the Gail model and this may provide partial explanation for its modest individual discriminatory value.

In an attempt to improve the individual discriminatory value of the Gail model, other models such as the Tyrer–Cuzick model have recently been developed which incorporate additional common risk factors including height, weight, presence of affected relatives with ovarian cancer, second-degree relatives with breast cancer, age of affected relatives and whether the relative had unilateral or bilateral breast cancer (Tyrer et al. 2004). The Tyrer–Cuzick model was developed from data obtained from the International Breast Cancer Intervention Study I (IBIS-1) prevention study and may be more relevant to women seeking risk assessment in anticipation of a prevention intervention than the Gail model, which is based on a screening cohort. The discriminatory value of the Tyrer–Cuzick model may be superior to the Gail model for women with a single affected relative (Tyrer et al. 2004).

Although risk models based on historical personal and family history are useful, increasing attention is being given to risk biomarkers that may improve short-term predictive accuracy for the individual woman. Biomarkers may be particularly useful in helping women who are identified as being at increased risk from epidemiologic models make decisions about medical or surgical prevention options. To the extent they can be modulated, biomarkers may also be used to monitor response to prevention interventions and/or predict response to a particular type of intervention. Of particular interest are breast-tissue changes which are highly associated with later cancer development. These changes are currently being utilized to select cohorts and assess response in Phase I and II trials of potential new prevention agents (Boone & Kellogg 1993).

We will review the concept of risk biomarkers with emphasis on those derived from breast tissue and the methods to acquire specimens for the purpose of both risk assessment and prevention.

### Table 1 Risk factors included in the Gail risk model for predicting the development of breast cancer

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Category</th>
<th>Associated relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current age (years)</td>
<td>20–70 (decades)</td>
<td>*</td>
</tr>
<tr>
<td>Age at menarche (years)</td>
<td>≥14, 12–13, ≤12</td>
<td>1.0, 1.10, 1.21</td>
</tr>
<tr>
<td>Number of first-degree relatives</td>
<td>with breast cancer</td>
<td></td>
</tr>
<tr>
<td>Age at first live birth (years)</td>
<td>&lt; 20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20–24</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>25–29/nulliparous</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>≥30</td>
<td>1.55</td>
</tr>
<tr>
<td>Number of biopsies</td>
<td>0, 1, ≥2</td>
<td>1.0, 1.70, 2.88 (age &lt; 50)</td>
</tr>
<tr>
<td>ADH in breast biopsy</td>
<td>Absence, presence</td>
<td>1.0, 1.27, 1.62 (age ≥50)</td>
</tr>
</tbody>
</table>

*Age does not provide an associated relative risk *per se*, but rather provides the starting point (based on Surveillance, Epidemiology and End Results (SEER) data) for calculation of additional risk elements.
Characteristics of ideal risk biomarkers

Characteristics of ideal risk biomarkers include: biologic plausibility, differential expression in low-versus high-risk populations, presence in a reasonable proportion of the high-risk population, association with cancer in prospective studies, expression minimally influenced by normal physiologic processes, the ability to obtain the marker by minimally invasive techniques and an assessment method that provides reproducible results (Kelloff et al. 1996, Boone et al. 1997).

Established risk biomarkers

Deleterious germline mutations in highly penetrant genes such as BRCA1/BRCA2 are strong predictors of breast cancer development but occur in less than 5–10% of women with breast cancer and in only 1% of the general population (Peto et al. 1999, Nathanson et al. 2001, Rebbeck 2002). Common single nucleotide polymorphisms of genes whose protein products are involved in carcinogen and hormone metabolism and/or DNA repair are associated with relative risks of 1.4–2.0; but two and three gene polymorphism combinations may be associated with much higher relative risks (Coughlin & Piper 1999, Feigelson et al. 2001, Pharoah et al. 2002, Comings et al. 2003, Aston et al. 2005). The established risk biomarkers serum-bioavailable estradiol and testosterone in postmenopausal women (Missmer et al. 2004, Tworoger et al. 2005), serum insulin-like growth factor-I (IGF-I) and its binding protein-3 (IGFBP-3) in premenopausal women (Hankinson et al. 1998), mammographic breast density (Boyd & Piper 1999) and breast intra-epithelial neoplasia (Page & Dupont 1990; Table 2) have much broader applicability than germline mutations in tumor-suppressor genes. Further, since they are subject to modulation, these risk biomarkers might be used to monitor change in breast cancer susceptibility from a prevention intervention. Mammographic breast density and intra-epithelial neoplasia are the most attractive risk biomarkers of the potentially modulated markers as they are useful in both pre- and postmenopausal women. However, Tice et al. (2004b) has reported recently that mammographic density adds only modestly to the Gail model in improving discriminatory accuracy.

Breast intra-epithelial neoplasia: the risk biomarker with the closest biologic association with cancer


Proliferative breast disease without atypia (moderate to florid hyperplasia, sclerosing adenosis, papillomas, etc.) is found in approximately 25–30% of diagnostic biopsies and is associated with a 1.4–2.0-fold increase in the relative risk for breast cancer (Dupont & Page 1985, Carter et al. 1988, London et al. 1992, Fitzgibbons et al. 1998, Wang et al. 2004). Higher relative risks associated with proliferative disease without atypia (e.g. 2.0 versus 1.4) may be associated with older age (>50 years) or a positive family history (London et al. 1992, Wang et al. 2004).

Atypical hyperplasia in diagnostic biopsies, whether ductal or lobular, is associated with an approximate 5-fold increase in relative risk without regard to other risk factors (Dupont & Page 1985, Tavassoli & Norris 1990, Page et al. 1991, Dupont et al. 1993). Women with atypia without a positive family history have an approximately 4-fold increase whereas women with a positive family history have an approximately 10-fold increase in their relative risk of breast cancer (Dupont & Page 1985, Dupont et al. 1993).

Table 2 Risk factors and relative risks for breast cancer

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Associated relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum-bioavailable estradiol (postmenopausal; mean age 64 years)</td>
<td>4 (&gt;2.7 vs. &lt; 2.7 pg/ml)</td>
</tr>
<tr>
<td>Serum IGF-I, IGF-I/IGFBP-3 (premenopausal)</td>
<td>2 (upper vs. lower quartile)</td>
</tr>
<tr>
<td>Single nucleotide polymorphisms (SNPs)</td>
<td>1.4–8 single or multiple</td>
</tr>
<tr>
<td>Mammographic breast density</td>
<td>5 (&gt;75 vs. &lt; 5%)</td>
</tr>
</tbody>
</table>
| Breast IEN
  Hyperplasia | 1.4–2 |
  Atypical hyperplasia | 4–5 |
  In situ cancer | 10–20 |

IEN, intra-epithelial neoplasia.

Several investigators, including Wellings & Jensen (1973) and more recently Allred et al. (1998, 2001) and Reis-Filho & Lakhani (2003), have suggested that atypical hyperplasia may arise more commonly from an intermediate lesion called an unfolded lobule (A for ductal, B for lobular) than hyperplasia of the usual type (HUT; Fig. 1). In fact, both atypical hyperplasia and HUT may both arise from unfolded lobules (Wellings & Jensen 1973, Allred et al. 1998, 2001). These unfolded lobules are characterized by increased cellularity and proliferation with distention of the terminal lobule duct unit.

ADH often shares molecular and genetic changes with DCIS as assessed by immunostaining (Boecker et al. 2002) or mRNA gene profiles (Ma et al. 2003). The frequency of loss of heterozygosity at at least one locus occurs in approximately 50% of ADH lesions and somewhat less frequently for HUT lesions (O’Connell et al. 1998). The most prevalent chromosomal losses are at 16q and 17p for both HUT and atypical hyperplasia, similar to what is observed for DCIS (Lakhani et al. 1995, Amari et al. 1999, Gong et al. 2001). Comparative genomic hybridization also indicates patterns of similar chromosomal gains and losses for non-invasive and invasive lobular cancer. Of particular importance is the loss of 16q, which contains E-cadherin, a tumor-suppressor gene involved in cell adhesion and cell-cycle regulation. E-cadherin is expressed in normal cells, but is lost in LCIS and invasive lobular cancer (reviewed in Reis-Filho & Lakhani 2003).

Although ADH is reported in 5% or less of diagnostic biopsies, it has been reported in 9% of autopsy specimens from average-risk women (Nielsen et al. 1987) and 39% of prophylactic mastectomy specimens from high-risk women (Hoogerbrugge et al. 2003). In the series by Hoogerbrugge et al. (2003), 57% of women with a family history consistent with that of a mutation in BRCA1 and/or BRCA2 had atypical ductal or lobular lesions and/or in situ cancer and these lesions were often multifocal or multicentric. Most women at increased risk for breast cancer by virtue of family history or other factors have never

**Figure 1** Histological continuum of the neoplastic progression towards invasive breast cancer. The relative risks (RR) refer to the increased likelihood of developing invasive cancer.
had a diagnostic biopsy. The question then becomes, how might we best detect intra-epithelial neoplasia, particularly atypical hyperplasia, via non-diagnostic tissue sampling? Further, do morphologic changes suggestive of intra-epithelial neoplasia detected as part of non-diagnostic tissue sampling carry similar predictive weight as those found in diagnostic biopsies performed following an abnormal exam or breast-imaging procedure?

**Methods of detecting breast intra-epithelial neoplasia for risk assessment**

Core biopsy, nipple aspiration for collection of nipple-aspirate fluid (NAF), ductal lavage and random periareolar fine-needle aspiration (RPFNA) are all being utilized to collect breast epithelial tissue for risk assessment in asymptomatic women without suspicious lesions on physical exam or mammography. Morphology has been prospectively correlated with later breast cancer development for only two of these methods: NAF and RPFNA.

**NAF**

NAF is generally collected following 5–10 min of manual massage, with or without the use of a Sartorius-type breast pump. Warming of the breast with a heating pad and scrubbing the nipple to dislodge keratin plugs are also often advocated (Sartorius et al. 1977). Ability to produce NAF is influenced by cohort selection and the number of attempts (Sauter et al. 1996, Klein et al. 2001, Wrensch et al. 2001, King et al. 2004). Approximately 80% of women are reported to produce NAF after five or more attempts (Sauter et al. 1996, King et al. 2004). NAF production has been reported in 39–66% of women without regard to risk (Wrensch et al. 1992, 2001), and 50–95% of high-risk women (Sauter et al. 1997, Dooley et al. 2001, Antill et al. 2004, Kurian et al. 2004, Sharma et al. 2004).

Young age (30–50 years), prior lactation and non-Asian ethnicity are positively associated with the ability to produce NAF (Wrensch et al. 1992). In an update of their original series, women producing NAF exhibiting proliferative epithelium with or without atypia had a 2.4–2.8-fold risk of breast cancer compared with those who did not produce NAF (Wrensch et al. 1992). In an update of their original series, women producing NAF exhibiting proliferative epithelium with or without atypia had a 2.4–2.8-fold risk of breast cancer compared with those who did not produce NAF with a median follow-up time of 21 years (Wrensch et al. 2001). Tice et al. (2004a) recently reported that adding NAF cytology to the Gail risk model improved model fit in a cohort of 6904 women with 100 000 patient years of follow-up. The relative incidence for the highest quintile compared with the lowest was 3.2 for the Gail model and 5.3 for the model including NAF cytology. There was no significant interaction with age.

In summary, both NAF production and NAF cytomorphology have been associated with elevated risk in prospective trials and NAF is easy and inexpensive to collect. There is preliminary evidence that NAF cytomorphology may also stratify risk based on the Gail model. Unfortunately, up to 50% of high-risk women fail to produce NAF, and up to 73% of NAF samples have insufficient cells for morphologic assessment (Dooley et al. 2001, Sharma et al. 2004, Francescatti et al. 2004). Given these limitations, several investigators have turned to molecular analysis.
of NAF fluid including hormone levels (Elia et al. 2002, Chatterton et al. 2004), proteomic patterns (Sauter et al. 2004, Alexander et al. 2004) and gene methylation (Evron et al. 2001, Krassenstein et al. 2004). Others have sought more reliable methods of obtaining epithelial cells from breast tissue.

RPFNA

A second method of non-lesion directed tissue sampling is RPFNA. This technique is based on the premise that if there are widespread proliferative changes within the breast, then there is an appreciable chance that these changes might be detected by random tissue sampling. The rationale is supported by the multifocal, multicentric proliferative changes observed in autopsy series (Bhathal et al. 1985, Nielsen et al. 1987) as well in prophylactic mastectomy series from high-risk women (Hoogerbrugge et al. 2003). Rather than assessing specific ducts that produce NAF, RPFNA attempts to detect a field change. Presumably those individuals who have atypia which can be detected by random tissue sampling would have the highest density of precancerous lesions within the breast tissue and a higher short-term risk of breast cancer than those women in whom atypia was not detected by this technique.

Skolnick et al. (1990) performed four-quadrant FNA on first-degree relatives of cancer patients and compared these aspirates to age-matched controls without affected family members. Cytologic evidence of proliferative breast disease with or without atypia was observed in 35% of high-risk women compared with 13% of controls. Fabian et al. (1994, 2000) used a modification of this technique. Instead of four-quadrant aspirates, two sites per breast were aspirated approximately 1 cm from the nipple areolar complex in both the upper-outer and upper-inner quadrants. Buffered lidocaine was used to anesthetize the skin and deeper subcutaneous tissue. Utilizing a 1.5 inch 21-gauge needle and a 12 cc syringe prewetted with RPMI, four or five aspirations were performed through each of the anesthetized areas. To reduce risk of bleeding and hematoma formation, women are asked to discontinue non-steroidal anti-inflammatory drugs, vitamin E or fish oil products 3 weeks prior to the procedure. Currently the majority of women are also offered vitamin K (10 mg) for 3 days prior to the procedure. Cold packs are applied to the breasts for approximately 10 min after the aspirations and then the breasts and chest wall are bound firmly with a soft gauze for several hours. Women are then instructed to wear a tight-fitting sports bra for several days. Severe hematoma formation requiring surgical evacuation and/or infection requiring oral antibiotics occurred in fewer than 1% of aspiration visits (Fabian et al. 2000). RPFNA produces minimal discomfort with a median reported pain score of 1 on a 0–10 scale (Chamberlain et al. 2003).

Although the procedure is called random as it is not directed towards a palpable mass or lesion detected by breast imaging, areas in which some resistance is encountered with the tip of the needle are sampled preferentially. Material from all aspiration sites is pooled in a single 15 cc tube and processed for cytomorphology and biomarkers. In our original series (Fabian et al. 2000), material was expressed into RPMI and processed via a Millipore™ filter on to slides (Barrett & King 1976). Since 1999, RPFNA specimen processing has been modified such that material is expressed directly into 10 cc of a modified Cytolyt™ fixative (9 cc of Cytolyt™ plus 1 cc of 10% neutral buffered formalin). Cells remain in the modified Cytolyt™ for 24–48 h on a test-tube rocker prior to
transfer to Preservercyt™. ThinPrep™ slides are then made according to standard instructions provided by Cytyc. Generally four slides are made: one for cytomorphology, with the remainder reserved for other biomarkers. The addition of formalin is useful in preserving estrogen receptor (ER) and preventing cellular degeneration if cells are exposed to extreme temperatures during shipment as part of multicenter collaborations. The number of epithelial cells obtained is related directly to the cytomorphology pattern observed. For the RPFNA procedure, we categorize cell number for each slide as <10, 10–99, 100–499, 500–999, 1000–5000 and >5000. In general, non-proliferative specimens have 100–499 cells per slide and it is possible to make only one or two slides. Specimens with hyperplasia have a median of 1000–5000 cells/slide and it is generally possible to make three or four slides per aspiration setting. Women with atypia have a median of >5000 cells/slide and it is almost always possible to make four or more slides per aspiration (C J Fabian et al., unpublished observations).

A cohort of 480 women with a median age of 44 years and a median 10-year Gail risk of 4% underwent an initial RPFNA and were asked to return for a follow-up RPFNA 6–12 months later. 82% returned for the follow-up RPFNA. Results from the first and second aspiration were combined for a baseline data set and subjects were followed for cancer development. 94% of subjects had adequate cytology for morphologic assessment from the initial aspiration. Utilizing the combined baseline dataset, 30% exhibited non-proliferative cytology, 49% hyperplasia and 21% hyperplasia with atypia. Considering only the initial aspiration, 12% were considered to have hyperplasia with atypia (Zalles et al. 1995, Fabian et al. 2000). 60% of the women were premenopausal. Premenopausal and postmenopausal women on hormone-replacement therapy (HRT) had a higher prevalence of RPFNA atypia than postmenopausal women not on HRT (P=0.001; Fabian et al. 2000). At a median follow-up time of 45 months, women with baseline hyperplasia with atypia were more likely to have developed DCIS and/or invasive cancer than women without atypia (Fig. 3). Further, women with 10-year Gail risks above the median of 4% (corresponding roughly to a 5-year Gail risk of 1.7%) could be stratified into very high and moderately high risk on the basis of RPFNA atypia (Fig. 4). Women with both RPFNA atypia and 10-year Gail risks of >4% had a 15% incidence of DCIS and/or invasive cancer at 3 years, whereas women with a 10-year Gail estimate of <4% had a 4% incidence of DCIS or invasive cancer within 3 years. For the entire cohort, both 10-year Gail risk and RPFNA atypia were predictive of cancer development. For women premenopausal at the time of study entry, RPFNA atypia and prior precancerous diagnostic biopsy (atypical hyperplasia, LCIS) were predictive of subsequent breast cancer development of DCIS or invasive cancer (P=0.044). Although this subcohort analysis must be viewed with caution, it is possible that RPFNA atypia may be a more sensitive risk predictor in premenopausal than postmenopausal women.

Proliferative breast disease is a continuum with overlapping morphologic features. Thus, the substantial

Figure 3 The risk for subsequent development of breast cancer is predicted by cytologic findings from RPFNA of women at high risk for the development of breast cancer. The incidence of cancer was 3.9% at a median follow-up time of 45 months (tick marks indicate censored subjects). The risk of breast cancer was statistically significantly higher in the 102 women with evidence of cytologic atypia in their initial RPFNA specimen than in the 378 women without evidence of atypia. Adapted from Fabian et al. (2000).
intra- and inter-observer variance described previously in the interpretation of both cytologically and histologically prepared specimens is not surprising (Rosai 1991, Schnitt et al. 1992, Sidawy et al. 1998). Using a single experienced cytopathologist and predefined criteria for non-proliferative specimens, hyperplasia or hyperplasia with atypia (Zalles et al. 1995), intra-observer variance was approximately 25% in our RPFNA series (Fabian et al. 2000, 2002). Masood et al. (1990) has developed a semiquantitative scoring index in which six cytologic characteristics are assigned 1-4 points depending on the degree of abnormality observed. Although there is overlap, non-proliferative samples generally score in the 6–10 range, hyperplasia 11–14 and hyperplasia with atypia 15–18 (Masood et al. 1990). Use of this index allows identification of samples that are borderline between hyperplasia and atypia (e.g. a score of 14) and may also reduce interpretive variance. Intra-observer variance was reduced from 25% with traditional categorical descriptors to 16% with the Masood index system when significant variance was considered to be a change in the index score of three or more points (Fabian et al. 2002).

In 1996, at a National Cancer Institute Conference, a Uniform Approach for diagnostic fine-needle aspiration biopsies (FNABs) was adopted (Uniform Approach 1997). Five categories were recognized: (1) unsatisfactory/insufficient cellularity; (2) benign; (3) atypical/indeterminate; (4) suspicious, probably malignant and (5) malignant. It was suggested that diagnostic FNABs falling into the atypical/indeterminate category be followed by surgical biopsy (Uniform Approach 1997). In a series reported by Boerner et al. (1999), 5% of diagnostic FNABs were atypical/indeterminate and cancer was found in approximately half of these specimens at follow-up excisional biopsy. At the present time it is unknown whether either the Masood scoring index or the Uniform Approach criteria, when applied to RPFNA specimens, would result in less inter- and intra-observer variance; nor whether it would provide superior or inferior predictive ability for development of breast cancer.

In summary, RPFNA utilizing the technique developed by Fabian et al. in which four or five aspirates are taken from each of two anesthetized sites per breast is associated with 94% cytomorphologic evaluability in a high-risk cohort where age is predominately 30–60 years. A random single aspiration from the upper-outer quadrant is not likely to produce the same results (only 60% morphologic evaluability reported; Khan et al. 1998). Further, cytologic evidence of atypia confers a 5-fold increase in risk compared with the absence of evidence of atypia and allows stratification of women with elevated Gail risk into high and very high categories. Although more invasive than NAF, the procedure may be performed comfortably and supply costs are modest. The primary drawback to this procedure is that the location of marked atypia, if observed, is unknown.

**Ductal lavage**

Ductal lavage is an extension of the NAF technique. In this procedure, NAF-producing ducts are cannulated with a microcatheter, saline or other physiologic solution is infused, the breast is massaged and the ductal...
lavage effluent is collected and expressed into a tube of fixative. In the multi-institution study published by Dooley et al. (2001), the ductal-lavage effluent was expressed into tubes of Cytolyt™ and mailed to a central processing location. The liquid fixative/cell mixture was then poured through a Millipore™ filter system, and cells captured on a filter paper that must be transferred subsequently to a glass side and dissolved with chloroform or other suitable solvent. This is a very efficient system for maximum cell capture but nuclear morphology can be suboptimal if the filter is not completely dissolved.

The multicenter study indicated that NAF production was possible in 83% of 500 eligible women from a high-risk cohort (57% of whom had a contralateral breast cancer and 39% with a 5-year Gail risk of >1.7%). 92% of women with NAF production underwent successful duct cannulation. Adequate samples for cytomorphic assessment (>10 cells) were obtained from 78% of women who underwent successful duct cannulation. Thus, 60% of women presenting for breast-tissue-based risk assessment produced NAF, underwent successful cannulation and had evaluable epithelial cells in their lavage specimen (Dooley et al. 2001).

The number of epithelial cells was estimated by counting the number of cell clusters and multiplying the number of clusters by the average number of cells in a cluster. Adequate epithelial cells for a morphologic designation (>10 cells) were obtained in NAF from 27% (111/417) of women versus 78% (299/383) of women undergoing successful duct cannulation. The median number of epithelial cells from evaluable NAF specimens was 120 (range 10–74 300) versus 4000 (range 24–143 000) or 13 500 (range 43–492 000) depending upon which microcatheter was used for ductal lavage.

Morphologic assessment was also performed centrally by two expert cytopathologists using modified Uniform Approach criteria (Uniform Approach 1997). Morphology was reported as insufficient, benign, mild atypia, marked atypia or malignant. 8% of the 500 eligible subjects had atypia by NAF compared with 18% by ductal lavage. Most cases of atypia were mild and inter-observer variance between two cytopathologists was reported as 11% utilizing the modified Uniform Approach criteria. Concordance between NAF and ductal-lavage cytomorphology was poor. Half the women with atypia in their NAF specimens had ductal-lavage specimens interpreted as benign. Three-quarters of atypical lavage specimens were associated with benign or acellular NAF specimens (Dooley et al. 2001).

Both NAF and ductal-lavage procedures were reported as well tolerated with a median pain score of 8 mm for NAF and 24 mm for ductal lavage on a 0–100 mm visual analogue scale. However, 28% of subjects underwent the procedure in the operating room under general anesthesia. In the multi-institutional series, sterile technique was used for ductal lavage. Subsequently, Francescatti et al. (2004) have reported on a series of 114 subjects undergoing lavage using aseptic but not sterile technique: no infections were noted. Similar to the Dooley study, the mean age was 52 years, mean 5-year Gail risk was 3.1%, and 39% had contralateral breast cancer. This group found that 56% of subjects presenting for risk assessment via ductal lavage had cytologically evaluable results. Reasons for non-evaluability included lack of NAF production (23%), inability to cannulate a NAF-producing duct (5%) or insufficient cells in the effluent (16%). The 57% cytologic evaluable rate for all women presenting for study is similar to the 60% rate reported by Dooley et al. (2001).

There are two mechanical challenges during ductal lavage which are responsible for a relatively modest portion of cases of cytologic inevaluability. These are passage of the catheter through the nipple sphincter and successful navigation through the lactiferous sinus into a duct without piercing the wall of the duct. The group at Northwestern has suggested several modifications aimed at sphincter relaxation and/or increasing patient comfort, which, in their experience, increase the rate of successful duct cannulation. These include use of nitroglycerin paste to relax the sphincter and subcutaneous nipple block with lidocaine (Golewale et al. 2003). Far more frequent causes for cytologic inevaluability are the inability to produce NAF to guide catheter placement and lack of epithelial cells. Several investigators highly skilled in performing ductal lavage have been able to cannulate non-NAF-producing ducts. They report that NAF-non-producing ducts are often cellular, particularly if there are other NAF-producing ducts within the breast (Cazzaniga et al. 2003, Love & King 2004). Whether this is an approach which can be transferred to less-experienced clinicians remains to be seen.

Other investigators have not been able to reproduce the yields of high epithelial cells reported in the original multi-institutional study by Dooley et al. (2001), even when cannulating only NAF-producing ducts. Currently, the ThinPrep™ technique described for RPFN is also used to process specimens obtained by ductal lavage. This processing technique is associated with improved nuclear morphology compared
with the previous Millipore™ system even though cellularity may be reduced. Experienced investigators report an average cell yield of 5000 cells per duct successfully lavaged (Khan et al. 2002).

An advantage of lavage over RPFNA is that investigation of ducts producing atypical or frankly malignant cells can be investigated via ductoscopy. In a series by Noga et al. (2002) mild to marked ductal lavage atypia was found in 42 ducts from 68 patients with pathologic nipple discharge. The majority (71%) of ducts with atypia were found to have intraductal papilloma and only 5.7% were found to have cancer. Pleomorph spindle-shaped cells which may be confused with atypical proliferative lesions may be a result of uneven or incomplete fixation. These fixation artifacts are probably secondary to the saline lavage solution. Use of a more isotonic fluid such as lactated Ringers or Plasmolyte™ for lavage and not allowing the Cytolyt™ lavage fluid ratio to exceed 3 may reduce fixation artifact.

The potential for early detection of breast cancer sets ductal lavage apart from other minimally invasive techniques for risk assessment. Khan et al. (2004) studied 44 breasts from 39 women, 38 of which had histologic evidence of cancer (although one had only lobular carcinoma in situ). Mean age was 50 years. 87% of breasts with cancer produced NAF. In only 5/38 (13%) of cancerous breasts were markedly atypical or malignant cells observed and in only 16/38 breasts were mildly or markedly atypical cells observed (Khan et al. 2004). Thus, in the study reported by Khan et al., the sensitivity for cancer detection was 13–42% depending on whether mild or marked atypia is used as a threshold. In a second study of women with suspicious microcalcifications undergoing core needle biopsy, NAF was obtained in six of 10 breasts with DCIS, but the DCIS-containing ducts yielded fluid in only one woman (Khan S A et al. 2005). The disappointing low sensitivity for detection of cancer with ductal lavage may in part be due to ductal anatomy and distribution of cancer. Going & Moffat (2004) have demonstrated that a minority of ducts drain the majority of breast-tissue volume. Further, only approximately one-third of ducts would be readily accessible by ductal lavage or ductoscopy: the rest taper to a minute orifice and some do not communicate with the skin surface. Badve et al. (2003) reviewed 801 mastectomies performed for DCIS or invasive cancer and found nipple and central duct involvement in only 22% of cases.

The sensitivity of ductal lavage for cancer detection is lower than that of mammography (61–81%) in a young screening population (Humphrey et al. 2002, Carney et al. 2003) or breast magnetic resonance imaging (79%) in a high-risk population (Kriege et al. 2004). However, the sensitivity of ductal lavage may be similar to that of mammography (33%) in a young high-risk population (Kriege et al. 2004).

Women with ductal-lavage atypia are presumed to be at increased risk of breast cancer based on the elevated risk observed for women with atypical cells in NAF or RPFNA specimens (Vogel 2004). However, the impact of ductal-lavage-detected atypia on the short-term risk for breast cancer is presently unknown since there was no follow-up of participants in the multicenter study reported by Dooley et al. (2001). A prospective trial is currently underway at multiple centers in the US in which women at increased risk for breast cancer will undergo ductal lavage at 6-month intervals over a 3-year period and will be followed for clinical breast cancer development. In this study, cytomorphology will be assessed at the individual participating institutions rather than through a central review board. Several published reviews suggest that women with ductal-lavage-detected atypia should be offered standard risk-reduction options such as tamoxifen (Morrow et al. 2002, O'Shaughnessy et al. 2002). Women with moderate to marked atypia may undergo ductoscopy; however, reimbursement for either ductal lavage or ductoscopy by third-party carriers is variable. Given the low specificity for cancer, removal of breast tissue on the basis of ductal-lavage atypia alone in the absence of a suspicious lesion on ductoscopy or breast-imaging modalities is discouraged (Morrow et al. 2002).

In summary, ductal-lavage cytomorphology (atypia) is currently being utilized for clinical risk stratification although the magnitude of risk conferred by ductal-lavage atypia has yet to be defined. Ductal lavage produces evaluable material for cytomorphology in 56–60% of women presenting for lavage when production of NAF guides attempts at duct cannulation, and 78% of women with a successful duct cannulation. Lavage is reported as well tolerated. Costs for procedure-related materials are substantially higher than NAF and RPFNA. Ductal lavage has low sensitivity for cancer detection and should not be used for that purpose. However, ductal-lavage atypia can be further investigated by ductoscopy.

Core needle biopsy

Core needle biopsy holds the promise of better architectural definition and large numbers of epithelial cells for study; but non-lesion-directed core biopsies
as a method of harvesting tissue for risk assessment and prevention trials have had mixed results. Mansoor et al. (2000) reported predominately atrophic terminal lobular duct units in 11-gauge core needle biopsies of normal breast tissue adjacent to benign lesions requiring stereotactic biopsy. In this series, the median number of normal cores per patients was two (range 1–7). Non-atrophic terminal lobule duct units were present in only 47% of patients. Postmenopausal women on HRT and women with dense heterogenous parenchyma were most likely to have non-atrophic terminal lobule duct units (Mansoor et al. 2000). To date, prevention trials using core needle biopsy as the sampling technique have not accrued subjects at a rapid rate although the procedure is described as well-tolerated (Harper-Wynne et al. 2002, Mohsin et al. 2003, Palomares et al. 2004). A 60–90% success rate has been described for obtaining adequate tissue at both the baseline and follow-up core biopsy (Harper-Wynne et al. 2002, Mohsin et al. 2003, Palomares et al. 2004, Stearns et al. 2004). The relative risks for non-directed core biopsy findings of hyperplasia with or without atypia have yet to be defined.

Which procedure is superior for obtaining epithelial cells for risk assessment?

This is a complicated question without a simple answer and will to a great extent depend on the skill set of the health-care professional performing the tissue sampling and the available resources. For this comparison, we have not included random core biopsy as there is minimal experience with this technique used in this fashion at the present time.

Table 3 gives the relative strengths and weakness of the three procedures. NAF harvest is clearly the least time-consuming, requires the least training to perform, is non-invasive, cheap and associated with the least discomfort, but it also is the least likely to provide evaluable epithelial cells. RPFNA occupies the middle ground in time to perform, training and expense. It is minimally invasive but comfortable for most patients with a median pain score of 1 on a numeric assessment scoring system of 0–10 (Chamberlain et al. 2003). It is the procedure most likely, in our experience (Zalles et al. 2003), to produce evaluable epithelial cells. Ductal lavage probably requires the most time to perform, the most training to master, especially for non-surgeons, and is associated with the greatest expense for procedure associated materials. The likelihood of obtaining epithelial cells for evaluation is intermediate between NAF and RPFNA. The magnitude of risk elevation has not been defined for ductal-lavage atypia as it has for NAF and RPFNA atypia. Nonetheless, a number of practitioners are using ductal-lavage atypia in the Gail model in a similar fashion to ADH in diagnostic biopsies, which results in an increase of the relative risk by 1.82 (Vogel et al. 2002, Vogel 2004). Ductal lavage was associated with a median pain score of 24 mm on a visual analogue scale (0–100 mm) and is also well-tolerated by most women. In preliminary studies with a head-to-head comparison of ductal lavage and RPFNA, RPFNA appeared to be associated with the least discomfort (Arun et al. 2003, Chamberlain et al. 2003). Ozanne & Esserman (2004) have suggested recently that use of RPFNA atypia to select which individuals with a 5-year Gail risk of ≥1.7% should receive tamoxifen is much more cost effective than Gail risk alone in terms of dollars spent per life-year saved. Unless the total cost of ductal lavage can be reduced to $350 per procedure, ductal lavage is no more cost effective than giving tamoxifen to all women with a Gail risk of ≥1.7%. Even at the low cost of $350 per procedure, ductal-lavage atypia is still less cost effective than RPFNA atypia in their model.

Are risk-eligible women with evidence of atypia more likely to undertake a prevention intervention?

Although this question has yet to be directly addressed, Bober et al. (2004) reported that risk-eligible post-menopausal women who had a prior abnormal biopsy including atypical hyperplasia were significantly more likely to accept prevention treatment with tamoxifen or participation in the STAR trial of tamoxifen versus raloxifene than women without a history of abnormal biopsy. Having a first-degree relative with breast cancer and/or a biopsy per se did not predict acceptance of prevention drug treatment. However, women with a history of an abnormal biopsy were much more likely

**Table 3** Comparison of NAF, ductal lavage (DL) or RPFNA for risk assessment and prevention

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasiveness</td>
<td>NAF</td>
</tr>
<tr>
<td>Discomfort (pain score 0–10)</td>
<td>DL</td>
</tr>
<tr>
<td>Supply cost (US $)</td>
<td>RPFNA</td>
</tr>
<tr>
<td>Time (min)</td>
<td>15</td>
</tr>
<tr>
<td>Easy to learn?</td>
<td>++</td>
</tr>
<tr>
<td>Adequate cytology (%)</td>
<td>23–66</td>
</tr>
<tr>
<td>Reproducibility* (%)</td>
<td>? ~60</td>
</tr>
</tbody>
</table>

*(little change at 3–12 months in absence of intervention).*
to perceive that their physician was advising them to undergo a prevention intervention than women with a family history alone (Bober et al. 2004). On multivariable analysis, only perceived recommendation by a physician, not atypia, was significant in predicting uptake of prevention drug therapy. A second study of factors affecting tamoxifen acceptance among high-risk women found that a history of ADH or LCIS is the strongest determinant of willingness to take tamoxifen (Tchou et al. 2004). Conversely, Didwania et al. (2003) reported that only 2/11 subjects with mild ductal-lavage atypia were influenced by the results to take tamoxifen.

**Potential breast-tissue molecular risk biomarkers**

Considering the inter- and intra-observer variance observed with cytomorphology for both ductal lavage and diagnostic and/or RPFN A, there is a great deal of interest in supplementing morphologic interpretations with molecular markers (Fabian et al. 2002, Ljung et al. 2004, Gornstein et al. 2004, Sneige 2004). Simple assessment of ploidy has been accomplished for both RPFN A and ductal-lavage samples (Fabian et al. 2000, Sauter et al. 2004). Sauter et al. (2004) noted higher frequencies of aneuploidy and hypertetraploidy in ductoscopy lavage specimens from women known to have breast cancer.

Genetic markers of allelic imbalance such as loss of heterozygosity and comparative genomic hybridization suggest a close relationship between atypical duct hyperplasia, DCIS and invasive cancer (O’Connell et al. 1998, Amari et al. 1999, Allred et al. 2001, Gong et al. 2001). Comparative genomic hybridization studies further indicate that well-differentiated DCIS and poorly differentiated DCIS are distinct genetic entities separately evolving into low- and high-grade invasive cancer (Buerger et al. 1999, 2001). These types of study also suggest that ductal and lobular cancers appear to evolve from different precursor lesions (Reis-Filho & Lakhani 2003).

Gene-expression profiling provides an estimate of the relative abundance of a particular gene compared with a reference sample. RNA is isolated, reverse transcribed to cDNA, labeled with a fluorescent dye and hybridized to a microarray. In the past, mRNA expression profiling has required appreciable amounts of fresh or frozen tissue which made the study of precancerous lesions difficult (Das & Singal 2002). However, laser-assisted microdissection, RNA linear amplification techniques (Van Gelder et al. 1990, Zhao et al. 2002) and specialized processing (Baunoch et al. 2003, Ma et al. 2003) allow mRNA expression profiling or quantitative real-time PCR to be performed on discrete lesions from formalin-fixed paraffin-embedded tissue obtained from core needle biopsies or fine-needle aspirations (Ellis et al. 2002, Fabian et al. 2003). Using mRNA expression profiling, Ma et al. (2003) provided additional evidence that ADH is a genetically advanced precancerous lesion and that ADH and DCIS are direct precursors of invasive ductal cancer. Relatively few genetic differences were found between ADH, DCIS and invasive cancer in the same breast, although appreciable differences were found between low- and high-grade in situ and invasive cancers from different individuals. Genes whose expression increased between DCIS and invasive cancer were related to proliferation and cell-cycle regulation (Ma et al. 2003). Based on studies such as these, some investigators have hypothesized that ADH is a committed precursor lesion whose molecular phenotype may predict the type of later in situ or invasive cancer (Jeffrey & Pollack 2003). It follows that markers of allelic imbalance or gene expression profiling might be utilized to supplement morphologic interpretations in the identification of high-risk lesions such as atypical hyperplasia.

Many early changes in carcinogenesis may be at the translational rather than at the transcriptional level, which could be theoretically useful in identifying women with non-proliferative or proliferative changes at high risk of developing more-advanced precancerous lesions and eventually cancer. Further, gene-specific levels of mRNA and their protein products do not necessarily correlate, indicating the importance of post-transcriptional influences (Gygi et al. 1999, Celis et al. 2000). Several proteomic technologies are available including two-dimensional gel analysis, surface-enhanced laser desorption/ionization-time of flight (SELDI-TOF) and sandwich antigen capture or direct immunoassays. Investigators have used both antibody arrays and two-dimensional gel profiling to study differences between normal and malignant ductal and lobular units. A large number of proteins differentially expressed in normal and malignant tissue were identified (Czerwenka et al. 2001, Hudelist et al. 2004). These included proteins involved in cellular trafficking and cytoskeletal and extracellular matrix regulation (including e1F), cell signaling and apoptosis (including Rho, 14-3-3 proteins and proteins involved in epidermal growth factor receptor (EGFR) phosphorylation; Fish et al. 1995, Wulfkuhle et al. 2002, Hudelist et al. 2004). A disadvantage of current proteomics techniques is they must be performed on fresh or frozen tissue/fluid that has not been fixed in
formalin. This limits proteomic analysis on diagnostic biopsies, especially for multi-institutional trials.

Proteomics pattern assessment is readily performed on NAF using the SELDI technique. Pawleutz et al. (2001) observed different protein profiles in NAF fluid in women with and without breast cancer. Interestingly, Pawlik et al. (2004) have reported differences in protein patterns in NAF samples in healthy women versus those with early-stage breast cancer, but no significant differences between the involved and uninvolved breasts. Sauter et al. (2002) found five differentially expressed protein profiles present in over 75% of women with invasive cancer but <10% of NAF samples from normal women.

Another method of assessing gene/protein function is with methylation-specific PCR. DNA methylation is an important process in epigenetic cellular memory that restricts or permits differential gene expression in descendant cells (Widschwendter & Jones 2002a, 2002b). This may be particularly important for evaluating the function of tumor-suppressor genes whose expression is attenuated or lost during the initiation or promotional phases of breast carcinogenesis (Das & Singal 2004). The tumor-suppressor genes HIC-1, RASSFIA and 14-3-3σ are methylated in a substantial proportion of cases of hyperplasia with and without atypia (Fujii et al. 1998, Ferguson et al. 2000, Umbricht et al. 2001, Lehmann et al. 2002). RASSFIA is involved with the regulation of cell-cycle progression via inhibition of cyclin D1 accumulation and promotion of apoptosis. Low-frequency promoter methylation of GSTP, CDH, BRCA1, p16 and RARβ2 have all been observed in benign breast tissue (Troch et al. 2003, Bae et al. 2004, Bean et al. 2005). These abnormalities allow escape from normal senescence and apoptosis and an extended period of susceptibility to proliferation and carcinogenic influences (Widschwendter et al. 1997, Huschtscha et al. 1998, Nuovo et al. 1999, Tlsty et al. 2001, Neumeister et al. 2002, Holst et al. 2003). Methylation abnormalities such as these may be detected from fixed microdissected or whole-slide scrapings from RPFNA (Troch et al. 2003, Sukumar et al. 2004) or ductal-lavage samples (Fackler et al. 2004).

The increased proportion of cells expressing ER and/or the proliferation marker Ki-67 may signal the transition from non-proliferative to proliferative epithelium. Proliferation in terminal lobular duct units varies with age, menopausal status and phase of menstrual cycle; and is highest for premenopausal women in the luteal phase of the cycle (Soderqvist et al. 1997, Potten et al. 1998). In the normal breast epithelium, proliferation, as measured by Ki-67, is positively correlated with serum progesterone levels but not serum estradiol, prolactin, bioavailable testosterone, androstenedione or IGF-I (Soderqvist et al. 1997). For premenopausal women, the proportion of normal breast epithelial cells expressing Ki-67 expression has been reported as approximately 1% in the follicular phase and 2–3% in the luteal phase (Soderqvist et al. 1997, Shoker et al. 1999). For postmenopausal women, the proportion of breast epithelial cells expressing Ki-67 is less than 1% (Shoker et al. 1999). The proportion of epithelial cells expressing Ki-67 increases in hyperplasia (>1%) and hyperplasia with atypia (2–5%) in both histologic and RPFNA specimens (Shoker et al. 1999, Allred et al. 2001, Khan Q J et al. 2005).

The proportion of breast epithelial cells expressing ER also varies with age and menopausal status as well as cell-cycle phase. The proportion of cells expressing ER is lowest in the luteal phase of the menstrual cycle and highest in postmenopausal women. ER has been reported to average 20% in the follicular portion of the cycle and 0–5% in the luteal portion of the cycle in normal lobules from premenopausal women and 18–40% in normal lobules from postmenopausal women (Soderqvist et al. 1993, Markopoulos et al. 1998, Khan et al. 1999, Shoker et al. 1999). The proportion of cells expressing PR in non-proliferative breast tissue is generally greater than those expressing ER and progesterone receptor (PR) expression does not vary significantly over the cycle (Soderqvist et al. 1993, Khan et al. 1999). In hyperplasia the proportion of ER-positive cells increases to 45% or greater and 90% for ADH (Shoker et al. 1999, Allred et al. 2001).

In a cross-sectional study of women with known outcome, Khan et al. (1999) found that an increased level of ER relative to PR in benign biopsies was associated with increased risk of breast cancer relative to those in whom the proportion of cells staining positive for PR was greater than or equal to those staining positive for ER (Khan et al. 1999).

In normal, non-proliferative breast tissue, ER-positive cells rarely express proliferation antigens; rather, proliferation is observed in adjacent ER-negative cells, which respond to paracrine influences from their ER-positive neighbors (Clarke et al. 1997). Studies in normal premenopausal breast tissue showed that only 0.01% of cells are dual-labeled for both ER and Ki-67 (Clarke et al. 1997, Shoker et al. 1999). Despite a lower overall percentage of epithelial cells expressing Ki-67, Shoker et al. (1999) reported that the proportion of dual-labeled cells was 4–20-fold higher in postmenopausal than premenopausal women. The proportion of epithelial cells expressing Ki-67 or dual-labeled for both Ki-67 and ER shows a progressive
increase between hyperplasia ADH and carcinoma in situ (Shoker et al. 1999). A negative association between Ki-67 and ER expression is maintained in hyperplasia of the usual type but this is lost in ADH. This would indicate a lack of suppression of ER expression as cells enter the cell cycle in atypical hyperplasia (Shoker et al. 1999). An increase in the proportion of cells staining individually as well as dually for Ki-67 and ER with progression to atypia may indicate a shift from paracrine to autocrine control of proliferation.

**Combining morphologic and molecular markers for risk stratification**

Molecular markers have the potential to further stratify risk prediction based on epidemiologic models and breast-tissue morphology although few prospective studies have been performed. Immunocytochemical expression of ER, p53, EGFR and HER-2 was assessed in cytospin preparations along with morphology from Millipore™ preparations in high-risk women in our prospective study (Fabian et al. 2000). Cytoplasmic and/or membrane staining of 10% or more of ductal cells (>2+ intensity) was considered as evidence of expression for both EGFR and HER-2. Given the use of cytospin preparations, lack of immediate fixation in a formalin fixative and/or antigen retrieval of ER expression was likely underestimated and thus >1+ intensity staining for ER in 10% or more of ductal cells was considered evidence of expression.

All single markers were strongly predictive of cytologic atypia ($P < 0.001$ in univariate analysis) and multiple marker expression of the three-biomarker set EGFR, ER and p53 was strongly predictive of atypia in multivariable analysis ($P < 0.001$). ER was the only single molecular marker predictive of cancer development ($P=0.048$) in a univariate analysis, although multiple markers in the three-set test (EGFR, ER and p53) were strongly predictive in univariate analysis for cancer development and for time to cancer development ($P=0.021$ and 0.003, respectively). Neither single nor multiple biomarkers (p53, EGFR, ER) were predictive for DCIS or invasive cancer in a multivariable analysis when RPFNA cytomorphology, 10-year Gail risk and prior diagnosis of LCIS or atypical hyperplasia in a diagnostic biopsy were included in the equation (Fabian et al. 2000). RPFNA hyperplasia with atypia, 10-year Gail risk and prior LCIS or atypical hyperplasia in a diagnostic biopsy were all predictive for cancer development (Fabian et al. 2000).

Quantitative PCR may be readily performed for 6–12 biomarkers on cellular material available from one of the four slides generally made from a RPFNA (Petroff et al. 2004). Quantitative PCR may provide for more accurate and reproducible assessments than immunocytochemistry especially for biomarkers expressed primarily in the membrane or cytoplasm. Quantitative PCR may also allow for measurements of more biomarkers in small samples than can be performed with immunocytochemistry. There is, however, poor correlation with focally expressed markers such as Ki-67, or those in which protein stabilization not elevated mRNA levels give rise to enhanced expression (Ginestier et al. 2002).

Methylation-specific PCR to determine loss of expression of tumor-suppressor genes is being utilized to help identify cancer in markedly atypical cytology specimens, and currently is being studied to determine whether it might identify proliferative disease likely to progress to cancer. Methylation-specific PCR can easily be performed on cells available from RPFNA or ductal-lavage samples (Evron et al. 2001, Fackler et al. 2004, Krassenstein et al. 2004, Moore et al. 2004). A new technique called quantitative multiplex methylation-specific PCR (QM-MSP) allows quantitative assessment of the extent of methylation of several genes (Fackler et al. 2004).

Chromosomal alterations in ductal lavage specimens matching those in corresponding cancers measured by comparative genomic hybridization or fluorescent in situ hybridization matching those in corresponding tumors have been observed in women who have breast cancer regardless of whether atypical or malignant cytomorphology is present (Adduci et al. 2004).

Whether one or more of these molecular techniques improves prediction based on epidemiologic models combined with cytomorphology with or without mammographic breast density needs to be addressed in a prospective trial.

**Use of breast-tissue biomarkers in prevention**

Biomarkers highly associated with short-term risk and subject to modulation may be used to select cohorts and measure a response to a prevention intervention. Biomarkers used to measure a response to an intervention are called surrogate endpoint biomarkers or SEBs (Kelloff et al. 1994, 1996). Important properties of an SEB are similar to those of risk biomarkers discussed above. In addition to risk-marker properties,
favorable modulation of the SEB by established prevention interventions should be associated with reduced cancer risk. Risk biomarkers which are currently being used to assess response to a prevention intervention in Phase I and II trials are (1) the ratio of serum IGF-I to its binding protein IGFBP-3, (2) serum-bioavailable estradiol and sex hormone-binding globulin, (3) mammographic breast density and (4) breast intra-epithelial neoplasia (hyperplasia with and without atypia) and associated changes within breast intra-epithelial neoplasia, such as proliferation (Table 4).

Tamoxifen, an established prevention drug, has been shown to favorably modulate all of the above biomarkers except serum hormones (Brisson et al. 2000, Chang et al. 2000, Bonanni et al. 2001, Tan-Chiu et al. 2003, Cuzick et al. 2004). Individuals who did not exhibit favorable biomarker modulation in response to tamoxifen do not necessarily go on to develop breast cancer (Tan-Chiu et al. 2003, Cuzick et al. 2004). In fact, the observed benefit from drug treatment is greater than would be expected from the extent of modulation of many biomarkers, particularly mammographic density (Cuzick et al. 2004).

Sampling breast tissue for SEB also provides the potential to assess biomarkers predictive of the response to a particular class of agents (e.g. ER for selective estrogen receptor modulators (SERMs)). This would allow appropriate matching of an individual with an intervention to which she is most likely to respond assuming the mechanism of action is known and predictive biomarkers have been identified (Paik et al. 2004). Repeated sampling also allows for assessment of markers which may provide evidence that the individual is in fact benefiting from drug treatment, i.e. reduction in proliferation or improvement in abnormal morphology.

**Clinical models for Phase I trials**

The toxicity profile is generally well known for most drugs being considered as potential prevention agents. As even minimal side effects are often unacceptable when a drug is to be given to a healthy women over a prolonged period, Phase I prevention trials focus on establishing the lowest dose at which a drug modulates a risk and/or a mechanism-of-action biomarker (Kelloff et al. 1994, Fabian et al. 1998, 2004a, 2005). Phase IA trials explore the effects of dose on several biomarkers. Phase IB trials are usually placebo-controlled and confirm that a given drug dose modulates a biomarker reliably (Fig. 5).

At present, the most popular Phase IA model is the so-called presurgical model in which women with a DCIS or a small invasive cancer who have undergone a diagnostic core biopsy are randomized between one of several drug doses in the interval (generally 2–4 weeks) between biopsy and re-excision, lumpectomy or mastectomy. A no-treatment control (either randomized or non-randomized) may be used as well to determine the effect of biopsy and other influences (such as stopping HRT) on the biomarkers being assessed. Generally, a proliferation biomarker such as Ki-67 is used as the primary endpoint and 8–12 subjects are entered at each dose level (Fabian et al. 2004a, 2005). In Phase IB, the dose(s) from Phase IA which have shown favorable modulation in 80% of subjects is chosen and subjects are randomized to study
drug or placebo (Decensi et al. 2003, Fabian et al. 2004a). Restriction of entry to a relatively homogenous population (e.g. similar menopause status and grade) may reduce the wide variation seen in Ki-67 and thus the number of subjects needed. For example, for a postmenopausal cohort not previously on HRT with non-high-grade tumors and a median Ki-67 of 10% and a standard deviation of 9%, a 50% reduction in Ki-67 could be detected with 40 evaluable subjects in the treatment group and 20 subjects in the placebo group (Fabian et al. 2004a).

Figure 5 Schematic of Phase IA presurgical model for breast cancer chemoprevention trials. Study agent is administered in the interval between diagnostic biopsy and planned surgical re-excision.

Clinical models for Phase II trials

Phase II trials are generally randomized, double-blind, placebo-controlled studies in which high-risk subjects are enrolled for 6–12 months. However, in cases in which the effect size of the agent on the primary endpoint biomarker is uncertain, a single-arm pilot study may be very useful. Primary response endpoints most often employed include modulation of morphology in high-risk women with baseline intra-epithelial neoplasia, modulation of proliferation or modulation of mammographic breast density (Boyd et al. 1997, Fabian et al. 2002, Harper-Wynne et al. 2002). Serum IGF-I/IGFBP-3 may be employed as well (Bonnani et al. 2001, Decensi et al. 2004). These SEBs are measured at baseline and study completion (Fig. 6).

In the US, Phase II trials using breast-tissue biomarkers as the primary endpoint have to date only been completed successfully in a reasonable time frame using RPFNA as the tissue-sampling method, although trials with ductal lavage and serial random and ultrasound-guided biopsy are ongoing.

Fabian et al. (2002) reported accrual of 119 subjects in 23 months to a single-institution trial of 6 months of α-difluromethylornithine versus placebo. Approximately four women underwent aspiration for every woman placed on study as the study required demonstration of hyperplasia or hyperplasia with atypia at baseline, sufficient cells for three other biomarker studies, as well as a number of other medical parameters including a reasonably normal auditory evaluation at baseline. Mean age of entrants was 46 years. Considering only eligibility based on cytologic morphology, approximately one in two women aspirated would have been eligible. 96% of subjects completed the study including a follow-up RPFNA 6 months after randomization.

The main endpoint of the study was improvement in cytologic morphology which was similar for both the α-difluromethylornithine- and placebo-treated groups. Samples from 28% of subjects randomized to the placebo were interpreted as showing improvement using the traditional categories of non-proliferative, hyperplasia and hyperplasia with atypia. 18% of samples from placebo-treated subjects were interpreted as showing improvement when improvement was defined as reduction of 3 or more Masood score points. Average baseline Masood score was 13.5 with a mean decrease of 0.46 ± 2.5 points in the placebo group. Although this amount of variation in the placebo group is within the range of discordance reported for benign breast specimens (Sidawy et al. 1998) with the modest sample size (119 subjects) utilized for this trial,
a 60% improvement in cytologic category and/or a mean reduction of 2 Masood score points (from an average baseline of 13.5) would be needed to detect a significant difference relative to placebo with an 80% power and a type I error rate of 5%. Alternatively, the sample size could be increased to detect smaller differences. Two models have been proposed in this regard which would facilitate sample-size estimation and statistical analysis. These have been termed the prevention-of-progression and reversal-of-atypia models (O’Shaughnessy et al. 2002, Fabian et al. 2005).

In the prevention-of-progression model, women without atypia but with hyperplasia (traditional criteria) are randomized to placebo or drug for 6–12 months. The primary endpoint is hyperplasia with atypia in the follow-up RPFNA sample (Fig. 7). Even if up to 25% of subjects randomized to placebo showed atypia in the follow-up sample, a 50% reduction in the incidence of hyperplasia with atypia could be detected in 335 subjects, with 80% power and a 5% type I error rate (Fabian et al. 2005). Approximately 670 high-risk subjects would have to be screened by RPNA for this type of trial design.

A reversal of the hyperplasia-with-atypia model in which all subjects had atypia at entry would require two or three times as many screening aspirations but fewer subjects entered on trial (Fig. 7). As few as 130 subjects would need to be entered into the treatment portion to detect a 50% reduction in fine-needle aspiration atypia after 6–12 months and 280 to detect a 33% reduction. Corresponding numbers of women who would need to be screened with an aspiration are 650 and 1400. It is unknown at present whether hyperplasia with atypia can be reversed after 6–12 months, and what effect failure to detect atypia on follow-up RPFNA might have on subsequent cancer incidence. The Phase II reversal of atypia model at present is probably best employed for drugs/regimens predicted to have a strong apoptotic effects (such as withdrawal of estrogen).

Because of the large number of subjects required for modulation of morphology for either the prevention-of-progression or reversal-of-atypia design, pilots and smaller Phase II studies often utilize modulation of proliferation as the primary response indicator. Ki-67/MIB-1 is an attractive proliferation marker because of its reproducibility (Keshgegian & Cnaan 1995, Biesterfeld et al. 1998). The proportion of cells expressing MIB-1 varies with age, menopause status and phase of the menstrual cycle as well as morphologic abnormality. Utilizing RPFNA cytology specimens (obtained in the follicular phase for premenopausal women), we have observed that median Ki-67 in premenopausal women with hyperplasia is 2% and 1% for postmenopausal women. Further, women with hyperplasia without atypia had a median Ki-67 of 1.1% versus 2.8% for hyperplasia with atypia (Khan Q J et al. 2005). Similar values for Ki-67 have been reported for ductal lavage specimens from high-risk women (Cazzaniga et al. 2003).
proportion of cells expressing Ki-67 it is necessary to count at least 500 ductal cells and 1000 would be optimum. Several small chemoprevention studies with modulation of Ki-67 in RPFNA cytology specimens as the main endpoint are ongoing (Fig. 8), and one using letrozole on postmenopausal women on HRT for control of menopausal symptoms shows evidence of modulation of Ki-67 despite continuing HRT during letrozole treatment (Fabian et al. 2004b).

Detection of a 50% reduction in the proportion of cells expressing Ki-67 would require approximately 50 subjects in each arm, assuming a median Ki-67 of 3.5% at baseline and a standard deviation of a similar value, with 80% power and a 5% type I error rate. Detecting a 33% reduction in proliferation would require approximately 200 subjects (Fabian et al. 2005).

These are theoretical estimates as studies assessing the variability of Ki-67 in postmenopausal women or premenopausal women in the same phase of the menstrual cycle over 6–12 months have not yet been performed.

Ductal lavage is an attractive alternative tissue-sampling method for Phase II prevention trials if reproducibility of cytomorphology or biomarkers is superior to that of RPFNA. In a preliminary analysis of an ongoing Phase II study of women undergoing repeat ductal lavage after 6 months of either tamoxifen or no intervention, 230/266 ducts (86%) could be recannulated 6 months later. Of the women who had at least one duct recannulated at a second lavage procedure, 75% had sufficient cells for cytologic diagnosis at both time points. There was no fall-off in cell yield or overall success of the procedure after 6 months of tamoxifen, although there was a trend towards declining nipple-fluid yield with longer durations of tamoxifen use (Bhandare et al. 2004). Reduction in incidence of cytologic atypia was seen in both the tamoxifen and no-intervention groups, with no significant difference between groups. Further, since NAF production and a successful ductal lavage can be expected in only approximately 60% of high-risk subjects, a larger number of subjects may have to be screened in order to meet the accrual goals, and supplies for ductal lavage are considerably more expensive than for RPFNA. Finally, it is possible that a successful drug intervention may reduce NAF production, making it less likely to obtain an adequate follow-up ductal-lavage specimen. The pros and cons of using of RPFNA, ductal lavage and NAF in prevention studies are detailed in Table 3.

As it is not clear at present which technique is most likely to give the most satisfactory results, before and after a chemoprevention intervention, the NCI is sponsoring a multi-institutional trial comparing ductal lavage and RPFNA in high-risk premenopausal women before and after 12 months of celecoxib (400 mg) twice daily for 12 months. Women eligible for the screening phase undergo a NAF attempt and, if successful, ductal lavage during the follicular phase of the menstrual cycle. Women who do not produce NAF and/or whose lavage is unsuccessful also have an RPFNA the same day. Four slides are made for both the ductal lavage and RPFNA procedure for cytomorphology, Ki-67, ER and cyclo-oxygenase 2 (COX-2). In order to be eligible for the treatment portion women must have >1000 ductal cells on the cytomorphology slide.
Table 5 Clinical trials of breast cancer prevention using breast-tissue-based biomarker modulation

<table>
<thead>
<tr>
<th>Study cohort/condition</th>
<th>Study agent</th>
<th>Tissue sampling techniques</th>
<th>Institution (protocol chair, if known)</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk for development of breast cancer</td>
<td>Tamoxifen + 4-HPR</td>
<td>NAF, 4-quadrant FNAB, rBX</td>
<td>National Cancer Institute</td>
<td>Completed</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Tamoxifen</td>
<td>NAF, DL</td>
<td>Northwestern Univ (Seema Khan)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Green tea extract (Polyphenon E) + aspirin</td>
<td>NAF, DL, RPFNA</td>
<td>Univ California Los Angeles (Mai Brooks)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Celecoxib vs. placebo</td>
<td>FNAB, rBX</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Celecoxib</td>
<td>RPFNA, DL</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Recruitment Completed</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Letrozole</td>
<td>RPFNA</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Recruitment Completed</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Tamoxifen</td>
<td>rBX</td>
<td>National Cancer Institute (Joann Zujewski)</td>
<td>Completed (Conley et al. 2000)</td>
</tr>
<tr>
<td>High risk for development of invasive breast cancer</td>
<td>Tamoxifen</td>
<td>rBX, FNAB</td>
<td>Univ of Texas Southwestern Medical Center (David Euhus)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>High risk for development of breast cancer</td>
<td>Celecoxib vs. Placebo</td>
<td>rBX</td>
<td>SouthWest Oncology Group (Powel Brown)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>High genetic risk for development of breast cancer</td>
<td>Bexarotene</td>
<td>rBX</td>
<td>Baylor College Medicine (Richard Elledge)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Tansideex</td>
<td>dxBx-Surg</td>
<td>Whithington Hospital, Manchester</td>
<td>DeFriend et al. 1994</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Tamoxifen</td>
<td>FNAB</td>
<td>Royal Marsden Hospital</td>
<td>Makris et al. 1998</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Fenretinide + tamoxifen</td>
<td>dxBx-Surg, FNAB</td>
<td>M.D. Anderson Cancer Center (Eva Singletary)</td>
<td>Completed (Singletary et al. 2000)</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Idoxifene</td>
<td>dxBx-Surg</td>
<td>Royal Marsden Hospital</td>
<td>Dowsett et al. 2000</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Raloxifene</td>
<td>dxBx-Surg</td>
<td>Royal Marsden Hospital</td>
<td>Dowsett et al. 2001</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Anastrozole</td>
<td>dxBx-Surg</td>
<td>Haukeland University Hospital</td>
<td>Geisler et al. 2001</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Faslodex vs. Tamoxifen</td>
<td>dxBx-Surg</td>
<td>Nottingham City Hospital</td>
<td>Robertson et al. 2001</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Fulvestrant</td>
<td>dxBx-Surg</td>
<td>South Manchester University Hospital</td>
<td>Bandel et al. 2002</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Letrozole</td>
<td>dxBx-Surg</td>
<td>Royal Marsden Hospital</td>
<td>Harper-Wynne et al. 2002</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Faslodex vs. Placebo</td>
<td>dxBx-Surg</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Completed Fabian et al. 2004b</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Celecoxib vs. Placebo</td>
<td>dxBx-Surg</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Recruitment Completed</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Tamoxifen (three doses)</td>
<td>dxBx-Surg</td>
<td>European Institute of Oncology (Andreas Decensi)</td>
<td>Decensi et al. 2003; Kisanga et al. 2004</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Isoflavone supplements</td>
<td>dxBx-Surg</td>
<td>Univ California Los Angeles (Mai Brooks)</td>
<td>Completed (Sartipour et al. 2004)</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Tamoxifen or no treatment</td>
<td>dxBx-Surg</td>
<td>Georgetown University (Vered Stearns)</td>
<td>Stearns et al. 2004</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Anastrozole</td>
<td>dxBx-Surg</td>
<td>Western General Hospital, Edinburgh</td>
<td>Dixon et al. 2004</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Celecoxib vs. placebo</td>
<td>dxBx-Surg</td>
<td>Univ of Kansas (Carol Fabian)</td>
<td>Completed</td>
</tr>
<tr>
<td>Newly diagnosed breast cancer</td>
<td>Celecoxib</td>
<td>dxBx-Surg</td>
<td>Memorial Sloan-Kettering Cancer Center (Eliza Port)</td>
<td>Recruiting</td>
</tr>
<tr>
<td>Premenopausal women at high risk for development of breast cancer</td>
<td>Deslorelin + estradiol + testosterone</td>
<td>Unspecified</td>
<td>Univ Southern California (Michael Press)</td>
<td>Recruiting</td>
</tr>
</tbody>
</table>

DL, ductal lavage; dxBx-Surg, pre-surgical model with study agent in interval between core needle biopsy diagnosis and surgical re-excision (lumpectomy or mastectomy); FNAB, fine-needle aspiration biopsy; 4-HPR, N- (4-hydroxyphenyl) retinamide; NAF, nipple-aspiration fluid; rBX, random or image-guided core needle biopsy; RPFNA, random periareolar fine-needle aspiration.

>500 ductal cells on the Ki-67 slide with >1% of cells showing evidence of proliferation and >100 ductal cells for ER and COX-2. Although the primary endpoint is modulation of proliferation by RPFNA, significant secondary endpoints include comparison of screening eligibility, adequacy of follow-up specimens, differences in cytomorphology and immunohistochemistry assessment, and procedure tolerance.
A number of Phase II clinical prevention trials are currently ongoing utilizing RPFNA or ductal lavage or core biopsy as the tissue-sampling technique and these are listed in Table 5.

Summary

RPFNA, ductal lavage and NAF are all commonly used methods to obtain breast tissue for risk stratification as well as for response monitoring in pilot and Phase II prevention trials. Both NAF and RPFNA atypia are associated with increased risk of subsequent breast cancer in prospective trials of average and high-risk women, respectively. Currently, the magnitude of increase in relative risk for detected atypia is known for only RPFNA (5-fold) and NAF (approximately 2.8-fold) but will probably be at least as high for ductal-lavage-detected atypia as for NAF-detected atypia. The most efficient method of obtaining epithelial cells for evaluation at a reasonable cost is RPFNA. Considerable issues remain regarding the reproducibility of morphologic assessment across a wide range of settings. Research is ongoing for methods to improve morphologic assessment reproducibility, including the addition of molecular tools. Therefore, breast-tissue-based risk stratification is still best performed within the context of a clinical trial.

RPFNA, ductal lavage and core biopsy are being utilized for tissue sampling in pilot and Phase II prevention studies in which modulation of morphology and/or molecular markers are used as the primary response endpoints following 1–12 months of study drug or placebo. It is not clear at present which method of repeated tissue sampling is the most cost-effective for prevention studies, but an NCI-sponsored study comparing RPFNA and ductal lavage has nearly completed accrual and will address this question directly.

Acknowledgments

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References


normal breast epithelium of women with breast cancer displays an aberrant response to estradiol. *Cancer Epidemiology Biomarkers and Prevention* 25 867–872.


Nuovo GJ, Plaia TW, Belinsky SA, Baylin SB & Herman JG 1999 In situ detection of the hypermethylation-induced inactivation of the p16 gene as an early event in oncogenesis. PNAS 96 12754–12759.


b biomarkers as surrogate endpoints for activity. *Clinical Cancer Research* 10 8332–8340.


Umbricht CB, Evron E, Gabrielson E, Ferguson A, Marks J & Sukumar S 2001 Hypermethylation of 14-13-3 sigma (stratifin) is an early event in breast cancer. *Oncogene* 20 3348–3353.


