Short-Term Breast Cancer Prediction by Random Periareolar Fine-Needle Aspiration Cytology and the Gail Risk Model

Carol J. Fabian, Bruce F. Kimler, Carola M. Zalles, Jennifer R. Klemp, Sahar Kamel, Sandy Zeiger, Matthew S. Mayo

Background: Biomarkers are needed to refine short-term breast cancer risk estimates from epidemiologic models and to measure response to prevention interventions. The purpose of our study was to determine whether the cytologic appearance of epithelial cells obtained from breast random periareolar fine-needle aspirates or molecular marker expression in these cells was associated with later breast cancer development. Methods: Four hundred eighty women who were eligible on the basis of a family history of breast cancer, prior precancerous biopsy, and/or prior invasive cancer were enrolled in a single-institution, prospective trial. Their risk of breast cancer according to the Gail model was calculated, and random periareolar fine-needle aspiration was performed at study entry. Cells were characterized morphologically and analyzed for DNA aneuploidy by image analysis and for the expression of epidermal growth factor receptor, estrogen receptor, p53 protein, and HER2/NEU protein by immunocytochemistry. All statistical tests are two-sided. Results: At a median follow-up time of 45 months after initial aspiration, 20 women have developed breast cancer (invasive disease in 13 and ductal carcinoma in situ in seven). With the use of multiple logistic regression and Cox proportional hazards analysis, subsequent cancer was predicted by evidence of hyperplasia with atypia in the initial fine-needle aspirate and a 10-year Gail projected probability of developing breast cancer. Although expression of epidermal growth factor receptor, estrogen receptor, p53, and HER2/NEU was statistically significantly associated with hyperplasia with atypia, it did not predict the development of breast cancer in multivariable analysis. Conclusion: Cytomorphology from breast random periareolar fine-needle aspirates can be used with the Gail risk model to identify a cohort of women at very high short-term risk for developing breast cancer. We recommend that cytomorphology be studied for use as a potential surrogate end point in prevention trials. [J Natl Cancer Inst 2000;92:1217–27]

Recent studies (1–3) indicate that breast cancer incidence may be substantially reduced in high-risk cohorts by treatment with tamoxifen, prophylactic oophorectomy, and prophylactic mastectomy. Although these reports are encouraging, preventive drug therapy and prophylactic surgery are expensive and may be associated with side effects. Consequently, high priority continues to be placed on the development of accurate short-term risk models so that women most likely to benefit from preventive therapy can be readily identified. The original Gail model, which predicts risk of in situ and invasive cancer in white women (4), and the subsequent version used in the Breast Cancer Prevention Trials, which predicts risk of invasive cancer by race, have been validated (5) for U.S. women undergoing annual screening (6,7). However, both versions of the model underpredict or overpredict risk for some subsets of women (5,7,8).

Biomarkers that vary with risk and response to prevention interventions have been termed “surrogate end point biomarkers” (9–11). In addition to being biologically and statistically significantly associated with cancer development, surrogate end point biomarkers should be present in a reasonable proportion of at-risk individuals, should be obtainable by minimally invasive procedures, and should be reversible with prevention interventions that have been shown previously to decrease cancer incidence. Ideally, surrogate end point biomarkers should also be amenable to quantitation and standardization (9–11). Once validated, surrogate end point biomarkers can be used to refine risk estimates based on epidemiologic models and to measure response to prevention interventions.

A number of risk biomarkers have been suggested as potential surrogate end point biomarkers. These biomarkers include mammographic breast density (12,13), serum insulin-like growth factor-1 and its binding protein, insulin-like growth factor-binding protein-3 (14–19), serum levels of estradiol and testosterone in postmenopausal women (20), and breast tissue markers (11,21,22). Breast tissue markers have theoretical appeal as being the most directly reflective of underlying neoplastic processes. Proliferative breast disease with and without atypia and lobular carcinoma in situ are associated with increased risk (23–29), and breast cancer incidence in women with atypical hyperplasia or lobular carcinoma in situ is substantially reduced after treatment with tamoxifen (30).

Unfortunately, a clinical model in which benign breast tissue can be obtained easily and repeatedly over time for risk assessment or for measurement of response to prevention interventions has remained elusive [reviewed in (31)]. Repeated directed biopsies of palpable masses or mammographic abnormalities over

Affiliations of authors: C. J. Fabian, J. R. Klemp, S. Kamel, S. Zeiger (Division of Clinical Oncology, Department of Internal Medicine), B. F. Kimler (Department of Radiation Oncology), M. S. Mayo (Department of Preventive Medicine and Kansas Cancer Institute), University of Kansas Medical Center, Kansas City, KS; C. M. Zalles, Department of Pathology, Hutchinson Hospital, Hutchinson, KS. Correspondence to: Carol J. Fabian, M.D., University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, KS 66160-7820 (e-mail: cfabian@kumc.edu).

See “Notes” following “References.”

© Oxford University Press
an extended time are impractical because the lesion of interest may be removed with the first biopsy. Repeated random biopsies for risk surveillance and/or for measurement of response to a prevention intervention are logistically simpler, but random core biopsy specimens may contain few terminal ductal–lobule units unless the biopsy specimens are obtained from mammographically dense areas (32). Atypical cells in nipple aspirates have been shown to have a prospective association with increased breast cancer risk; however, even in experienced hands, approximately 40% of nipple aspirates are acellular (33, 34). Random periareolar fine-needle aspiration is inexpensive and may be performed repeatedly with minimum morbidity, and the majority of samples are cellular in high-risk premenopausal and perimenopausal women (35–39).

We have reported previously (36, 40, 41) preliminary results from a prospective trial begun in 1989. In that trial, women at increased epidemiologic risk on the basis of family history of breast cancer, a prior breast cancer, or precancerous biopsy underwent random fine-needle aspiration. Cells in the aspirate were assessed for morphologic pattern and the biomarkers DNA ploidy, epidermal growth factor receptor (EGFR), estrogen receptor (ER), p53 protein, and HER2/NEU protein. Women were then followed for breast cancer development. Immunostaining for p53 and HER2/NEU proteins was not performed during the first 2 years of the study. Although the Gail risk was calculated at entry for all subjects, it was not used as an eligibility criterion. Hyperplasia and hyperplasia with atypia in the fine-needle aspirate, as well as the expression of single and multiple biomarkers, were more prevalent in high-risk women than in low-risk control women (38, 42). Prevalence of biomarker expression increased with increasing cytologic abnormality (38, 42). Finally, in a selected group of 224 of these high-risk women for whom measurement of all five biomarkers was attempted or performed, subsequent cancer development (six cancers at a median follow-up of 32 months) was strongly predicted by multiple biomarker expression of the three-test set (EGFR, ER, and p53 protein) and a 10-year Gail risk of developing breast cancer (43).

In this article, we have updated results on 480 high-risk women aspirated during the period from August 1989 through January 1999 and have examined the relationship of the Gail risk, fine-needle aspirate cytology, biomarker expression, and cancer development. We found that both hyperplasia with atypia and the Gail risk are strongly predictive of subsequent cancer development and present a model that characterizes the relationship between the two variables.

SUBJECTS AND METHODS

Eligibility

This study was performed after approval by the University of Kansas Medical Center Human Subjects Committee, in accord with an assurance filed with and approved by the Department of Health and Human Services. Subjects signed an informed consent before each breast aspiration.

Women were self-referred or referred by a physician and must have had at least one of the following major risk factors to be eligible: 1) a family history of breast cancer (in one or more first-degree or in two or more second-degree relatives affected with breast cancer), 2) a prior lymph node-negative breast cancer, or 3) a prior biopsy indicating atypical lobular or ductal hyperplasia or carcinoma in situ. A total of 480 women were eligible for this study. For women with prior lymph node-negative invasive breast cancer, only the contralateral breast was aspirated. Subjects with prior ductal carcinoma in situ (DCIS) treated with radiation or mastectomy were classified as having a prior cancer, and only the contralateral breast was aspirated. Subjects with a small DCIS treated by excision alone were classified as having precancerous disease, and both breasts were aspirated. Ninety-six percent of women were Caucasian, 2.5% were African-American, and 1.5% were other minorities. Women were generally required to be between 30 and 60 years of age. Women younger than 30 years have a low short-term risk of breast cancer (44), and women older than 60 years often have fatty involutional breasts that are unlikely to yield sufficient cells for analysis. Therefore, a woman younger than 30 years could be aspirated only if she were within 10 years of the age at which her youngest relative developed breast cancer or if she had clinical proliferative breast disease. Women older than 60 years could be entered only if they had recent evidence of proliferative breast disease.

All women were required to have a mammogram interpreted as not suspicious for breast cancer within 12 months before aspiration, plus a breast examination of the breast that the needle aspiration was performed; the contralateral breast was aspirated as normal or not sufficiently abnormal to warrant a diagnostic biopsy. However, many of these women had dense breasts upon mammographic examination and diffusely nodular breasts upon clinical examination.

Gail Risk Calculations

The projected probability of developing in situ or invasive breast cancer at 10, 20, and 30 years was calculated at entry for each high-risk woman, according to the model developed by Gail et al. (4). The Gail model uses age at menarche, age at first live birth, current age, number of affected first-degree relatives, number of breast biopsies, and whether or not atypical hyperplasia was present for risk calculations. The Gail risk calculations were not used to determine eligibility because the model was not designed to assess risk for women with some known major risk factors, such as lobular carcinoma in situ or prior invasive breast cancer, or women with a strong family history of breast cancer but without an affected first-degree relative. For women with prior lobular carcinoma in situ or DCIS, the assumption was made that they had underlying atypical hyperplasia, even if that was not specifically noted on a pathology report. For women who had undergone lumpectomy plus radiation therapy or mastectomy for treatment of their breast cancer, only the contralateral breast was aspirated and considered at risk. In other words, any prior biopsy findings of atypical hyperplasia or carcinoma in situ in the involved breast did not contribute to the calculated risk of developing cancer in the other breast. However, any biopsies of the involved breast before definitive management were retained for the calculation.

Aspiration

Women were cautioned not to take aspirin or a nonsteroidal anti-inflammatory agent for at least 3 weeks before aspiration to minimize hematoma formation. pH-buffered 1% lidocaine was used as local anesthetic (one part sodium bicarbonate to three parts lidocaine). Buffered lidocaine was used to decrease local discomfort and cellular distortion. A 1.5-inch 21-gauge needle attached to a 10- to 12-mL syringe, pretwisted with tissue culture medium, was used for the aspiration. The needle was positioned immediately adjacent to the areola at approximately 3 o’clock and later at 9 o’clock, and the position was varied slightly to avoid superficial blood vessels. As mentioned above, women with prior DCIS or invasive breast cancer had only the contralateral breast aspirated. Tissue was probed deeply to sample the terminal lobular–duct unit, where most cancers are thought to arise (45). Typically, eight to 10 aspirations were performed per breast; half were done through the upper outer quadrant site and half through the upper inner quadrant site. This number might have been adjusted down if bleeding was encountered at the site or upward if it was apparent that little material was being aspirated. Before being processed, all cells from the aspirations were pooled in 5 mL of ice-cold RPMI-1640 medium. The vast majority of aspirations reported herein were conducted by one physician (C. J. Fabian), who was also responsible for training four other individuals who have performed aspirations over the years.

After the aspiration was completed, cold packs were applied to the aspiration sites for approximately 10 minutes, the breasts and chest wall were firmly bound in gauze, and a tight-fitting sports bra was worn over the gauze wrap to minimize breast movement and to decrease the chance of later hematoma formation. Women were instructed to wear a spandex sports bra for 4 or 5 days. Women expressing anxiety about the procedure were offered oral lorazepam before aspiration, but most women did not take it after the initial aspiration. The procedure itself produced little discomfort for most women. However, some bruising occurred in most individuals and often produced some temporary discomfort that was usually relieved with acetaminophen. Severe hematoma formation requiring surgical evacuation or infection requiring oral antibiotics oc-
Women were then asked to return in 6 months for another aspiration. Eighty-two percent of women received a second aspiration, with 47% having their second aspiration within 6 months. Because of scheduling difficulties (e.g., holidays and vacations, synchronizing with menstrual cycle, and subject request), the interval between the first and the second aspiration was delayed to 7–12 months in 25% of the subjects and to 13–21 months in 10%. Eighty-eight high-risk women (18%) had only a single aspiration. Sixteen of these women (3%) had actually had a second aspiration, but the results had not been interpreted by the analysis cutoff date. Forty-one (9%) were aspirated in the first 2 years of the study, when less emphasis was placed on repeat aspirations; 31 (6%) simply chose not to return. Their data were included in the analysis. To minimize sampling variance, results from the first and from the second aspirations were pooled to make up the initial dataset, with the more abnormal findings being used. Twenty-nine percent of women had more abnormal findings on the second aspiration than on the first, including 6% for whom no cytology reading was possible on the first aspiration. Consideration of only the results from the first aspiration did not alter the conclusions reached when pooling was used.

**Assays**

The initial tissue processing was begun in the clinic room where the aspiration was performed. Aspirated material was expressed from the syringe directly into a 5-mL tube of ice-cold tissue culture medium (RPMI-1640 medium supplemented with 10% fetal calf serum) in an ice bath. Cells remained in tissue culture medium at ice temperature for up to 2 hours before fixation. Cells were pooled from the right and left breasts before aliquots were taken for individual assays. We pooled cells because our primary goal was to detect a field effect, which might be useful for short-term cancer prediction. Our goal was not to detect a cancer in an individual breast or direct a diagnostic procedure. Specimens were processed first by removing fibrin clots, centrifuging at 2000 rpm for 10 minutes. One third of the specimen was used for cytology. The remaining two thirds of the specimen was divided equally for the other four or five assays, so that the slide for cytology had at least twice the material of any of the other slides.

Cytology preparations were filtered through a 25-mm Millipore filter (5-μm pore size) rather than smeared onto a slide, to reduce cellular distortion and to maximize cell yield. Cytology specimens were stained immediately and interpreted within a few weeks of aspiration by the cytopathologist (C. M. Zalles). Cytology was classified as either nonproliferative (which included normal as well as apocrine metaplasia) or proliferative (epithelial hyperplasia), with or without atypia according to previously published criteria (41). To avoid controversy and confusion resulting from use of the term “atypical hyperplasia” to describe cytologic rather than histologic changes (46–48), we have used the term “hyperplasia with atypia.” We use this term to describe both lobular and ductal hyperplasia with atypia, and it is synonymous with what we previously termed “proliferative breast disease with atypia” (41), in keeping with the new standard terminology (49) for diagnostic fine-needle aspirations. However, because these are not diagnostic aspirations and we are hesitant to ascribe the label “disease” to proliferative changes observed in random fine-needle aspirates, we have decided to simply label these changes “hyperplasia with atypia.” Our criteria for atypia, however, have not changed since the study was initiated in 1989. Specifically, specimens with hyperplasia with atypia exhibited a generally high cellular yield with multiple clusters of epithelial cells arranged in complex sheets and clusters with partial layering or overriding of cells, with loss of cohesion, loss of polarity, nuclear hyperchromasia, and distinct nucleoli. Myoepithelial cells could be identified as present but were usually rare (41). Although the number of cells was not specifically counted, there were usually at least 1000 epithelial cells per cytology slide. The minimum for interpretation was 10 cells.

For Feulgen’s staining of DNA and immunocytochemistry, lysing buffer was applied to the specimen, followed by centrifugation at 2000 rpm for 10 minutes at 4 °C and resuspension in RPMI-1640 medium. The specimen was distributed over a slide and centrifuged at room temperature in a cytocentrifuge (Shandon Inc., Pittsburgh, PA) for 4 minutes at 1250 rpm. Specimens were immediately fixed in acetone for assay of p53 and HER2/NEU or sequentially fixed in 10% buffered formalin, methanol, and acetone for assay of EGFR and ER. Slides were stored in freezing medium at −20 °C until they were stained. All procedures, including interpretation of cytology and immunocytochemistry assays, were typically completed within 4–6 weeks of the aspiration. DNA ploidy determinations were performed by image analysis (CAS; Becton-Dickinson, Mountainview, CA). Specimens containing 10% or more of cells with a DNA index of less than 0.85 or greater than 1.15 were considered aneuploid. PAH240 and antibody 3 (Oncogene Research Products, Cambridge, MA) were the antibodies used to detect p53 and HER2/NEU, respectively. Although PAH240 has been considered specific for mutant p53 protein, nuclear and cytoplasmic staining has been observed in normal cells and may be due to the method of fixation, especially under conditions of oxidative stress (50,51). EGFR was determined with clone F7 anti-EGFR antibody (Sigma Chemical Co., St. Louis, MO), and ER was determined with an ER-ICA (Immunocytochemical Assay) kit (Abbott Laboratories, Abbott Park, IL). DNA ploidy assessment was attempted in 84% of the 480 women, EGFR assessment was attempted in 96%, ER assessment was attempted in 93%, p53 assessment was attempted in 85%, and HER2/NEU assessment was attempted in 76%. Overall, 77% of the 480 women had a successful test for DNA ploidy, 83% had a successful test for EGFR, 81% had a successful test for ER, 66% had a successful test for p53, and 59% had a successful test for HER2/NEU. For ploidy or immunocytochemistry interpretation, a minimum of 25 epithelial cells were required. However, most slides had more than 100 epithelial cells. Readers were instructed to score at least 100 epithelial cells from the most positive area of the slide. For Feulgen and immunocytochemical markers were scored negative 0 to +, representing a range from no evidence of expression to intense expression equivalent to the level of expression in positive control slides. Specimens with 10% or more of scored ductal cells staining with an intensity of 2+ or greater (EGFR, p53, and HER2/NEU) were interpreted as showing evidence of expression (52,53). Provided the above criteria were met, EGFR and HER2/NEU samples were considered to be positive if there was cytoplasmic or membrane staining. p53 samples were considered to be positive if nuclear staining was observed, even if some cytoplasmic staining was also noted. Because ER staining tends to be weak in cytospin preparations (54), ER was considered to be expressed if a staining intensity of 1+ or greater in 10% or more of scored ductal cells was observed.

Two reviewers independently scored all immunocytochemistry slides. Intraobserver variance was 4% as reported previously (38). One pathologist (C. M. Zalles) reviewed all cytology slides. Periodically, the pathologist was asked to blindly review slide sets originally interpreted over the prior 12–18 months and to reassess cytologic category interpretation as nonproliferative, hyperplasia, or hyperplasia with atypia. For cytology, intraobserver subcategory variance has ranged from as low as eight (8%) in 100 to 26 (24%) in 110; sets that contained the greatest proportion of slides with hyperplasia with or without atypia were associated with the greatest intraobserver variance. Currently, all cytology slides are reviewed twice in a blinded fashion. If there is intraobserver variance, then slides are assessed a third time, and the two similar readings are listed as the designation.

**Follow-up**

After the initial cytologic and biomarker characterization, high-risk subjects were followed for development of invasive or in situ breast cancer. Most subjects returned for repeat aspirations every 1–3 years, depending on their cytology results. Subjects with three aspirations exhibiting nonproliferative cytology were told that they did not need to return for aspiration but would need to continue yearly screening and follow-up. Periodically, questionnaires were mailed to subjects who had not been seen in the clinic for more than a year. Attempts were made to contact by telephone any subject who did not return a questionnaire. Any self-reporting of breast cancer development was followed up by a request for medical records and a pathology report from the attending physician. Where possible, slides and/or tissue blocks were obtained and reviewed.

**Statistical Analysis**

Statistical analysis was performed with the computer program package SPSS for Windows (Release 9.0, SPSS, Inc., Chicago, IL). Contingency table analyses and two-sided Student’s t tests were used to compare subsets within the high-risk group and to calculate P values. Specifically, distributions of cytologic characterization versus menopausal status, use of hormone replacement ther-
apy (HRT), and risk factors at entry were analyzed by contingency tables. Continuous variables (e.g., age at entry and the Gail risk) were analyzed by Student's t tests. The association of epidemiologic risk factors and molecular markers in cells from fine-needle aspirates with a finding of hyperplasia with atypia was analyzed by contingency table analysis and logistic regression. Univariate analysis of eventual development of DCIS or invasive cancer as a function of epidemiologic risk factors and cytomorphology and molecular markers in cells from fine-needle aspirates was performed by contingency table analysis and two-sided Student's t tests. Univariate analysis of time to development of DCIS or invasive cancer was analyzed by the Kaplan–Meier survival analysis. Subjects who died of any cause, who developed a new primary non-breast cancer or metastatic breast cancer, or who underwent bilateral prophylactic mastectomy were censored at the time of the event. We compared categorical variables by the log-rank test and continuous variables by Cox proportional hazards analysis.

Exploratory models to estimate the joint effect of variables on breast cancer development and time to breast cancer development were desired. A stepwise multiple logistic regression was used to assess what variables may have a joint effect on breast cancer development. Explanatory factors used in this procedure were race, age at menarche, age at first live birth, age at entry in the study, menopausal status, use of HRT, family history of breast cancer, precancerous mastopathy in a prior biopsy, prior breast cancer, multiple risk categories (two or more of the previous three items), a 10-year Gail risk adjusted for presence of atypical hyperplasia in a prior biopsy, cytologic evidence of epithelial hyperplasia with atypia in random fine-needle aspirates, and expression of multiple biomarkers. Similarly, stepwise Cox proportional hazards regression was used to assess what variables may have a joint effect on time to breast cancer development. Exploratory factors used in this procedure were the same as those used in the stepwise multiple logistic regression. Stepwise regression techniques provide an efficient method for finding a small set of important covariates from a large number of potential covariates. Stepwise regression is an important tool in exploratory data analysis, even given the fact that statistical significance levels may be distorted (55). All P values are from two-sided tests.

**RESULTS**

**Subject Demographics**

Four hundred eighty high-risk women were entered during the period from August 1989 through January 1999. Sixty-three percent were eligible because of a family history of breast cancer, 12% were eligible because of a prior precancerous biopsy (atypical hyperplasia or carcinoma in situ), 11% had prior breast cancer, and 14% belonged to two or more of the above risk categories. Three hundred sixty-three women (76%) had a family history of breast cancer, including 302 (63%) in the family history risk category and 61 (13%) in the multiple risk category. Of the total of 363 women classified as having a family history of breast cancer, 335 (92%) had at least one first-degree relative with breast cancer and 28 (8%) had no first-degree relatives with breast cancer but had two or more second-degree relatives with breast cancer. The median age at entry was 44 years, and 60% (286 of 480) were premenopausal. Thirty-seven percent of the 194 postmenopausal women were on HRT at entry. The median 10-year Gail predicted probability of breast cancer development was 4.0%.

**Cytology and Risk Factor Associations**

Thirty percent of the cohort had nonproliferative breast cytology, 49% had epithelial hyperplasia, and 21% had hyperplasia with atypia in the initial fine-needle aspirate (Table 1). The incidence of hyperplasia with atypia was higher in premenopausal women on HRT than in postmenopausal women not on HRT (P = .001). Women with cytologic evidence of hyperplasia with atypia had higher mean 10-year Gail risk values (5.0% versus 6.9%; P < .001; Table 2). Women who entered the study with a history of breast cancer were older (median age = 48 years) and less likely (P = .001) to have proliferative cytology, either with or without atypia. Women who entered the study with a prior precancerous biopsy or who belonged to multiple risk categories at study entry were more likely to exhibit hyperplasia with atypia, but statistical significance was not reached (Table 2).

**Cytology and Biomarker Associations**

Prevalence of individual molecular marker expression, in women for whom the specific test was attempted, was 24% for DNA aneuploidy, 24% for ER, 27% for p53, 26% for HER2/NEU, and 42% for EGFR. With the exception of DNA aneuploidy, there was an increase in the prevalence of individual

<table>
<thead>
<tr>
<th>Variable*</th>
<th>P from univariate analysis†</th>
<th>P from multivariable analysis‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemiologic risk factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-y Gail risk</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Prior precancerous biopsy</td>
<td>.063</td>
<td></td>
</tr>
<tr>
<td>Multiple risk categories at entry§</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>On HRT at entry</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>Molecular marker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR expression</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>p53 expression</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>ER expression</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>HER2/NEU expression</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Any expression (five tests)</td>
<td>&lt;.01</td>
<td>.035</td>
</tr>
<tr>
<td>Multiple expression (five tests)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Any expression in three-test set (EGFR, ER, and p53)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Multiple expression in three-test set (EGFR, ER, and p53)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*HRT = hormone replacement therapy; EGFR = epidermal growth factor receptor; ER = estrogen receptor.
†χ² or Fisher's exact test, where appropriate for categorical variables; Student’s t test for continuous variables. All P values are from two-sided tests.
‡Logistic regression (using different blocks of input variables as noted).
§Risk categories at entry included family history of breast cancer, prior precancerous biopsy, and prior breast cancer.

*HRT = hormone replacement therapy.
bimarker expression with increasing cytologic abnormality (Fig. 1). The prevalence of multiple biomarker expression (regardless of which specific biomarkers) was also increased with increasing cytologic abnormality. Sixty-four percent of the women with hyperplasia with atypia expressed multiple biomarkers, whereas only 18% of the women with nonproliferative cytology expressed multiple biomarkers ($P < .001$). Expression of individual markers (EGFR, ER, p53, and HER2/NEU) was associated with atypia by univariate analysis (Table 2). Similarly, any expression and multiple expression in the complete five-biomarker set or the three-biomarker set (EGFR, ER, and p53) were associated with atypia by univariate analysis ($P < .001$). By multivariable analysis, only multiple expression within the three-biomarker set was strongly associated with atypical cytology ($P < .001$), regardless of whether the epidemiologic risk factors (e.g., the Gail risk, on HRT at entry) were added to the analysis; i.e., the Gail risk did not contribute to the equation (Table 2).

**Predictors of Cancer Development**

At a median follow-up time of 45 months from initial aspiration, 20 of the 480 women have developed cancer (invasive cancer in 13 of them and DCIS in seven). For these 20 women, the median age at entry was not statistically significantly different from that of the women with no cancer detected (46 years versus 44 years). Fourteen cancers developed in the 286 women who were premenopausal at entry, and six cancers developed in the 194 women who were postmenopausal at entry. The median age at the time of diagnosis was 51 years, and the median interval between study entry and cancer diagnosis was 19 months (range = 1–101 months). Cancers were detected in all risk subcategories (Table 3). A Kaplan–Meier hazard analysis estimates that 3.9% of our high-risk cohort will develop DCIS or invasive cancer at a median follow-up time of 45 months or approximately 1% per year (Fig. 2).

Univariate analyses indicated that subsequent cancer development was associated with a Gail risk of developing breast cancer, prior precancerous biopsy, belonging to multiple risk categories at entry, hyperplasia with atypia in the random fine-needle aspirates, ER expression, or expression of multiple markers from the three-biomarker set (EGFR, ER, and p53) (Table 4). The effect of cytologic evidence of hyperplasia with atypia in random fine-needle aspirates on subsequent cancer development is shown in Fig. 2.

Stepwise logistic regression was performed with cancer detection as the dependent variable (see above). Of the variables listed in Table 4, hyperplasia with atypia and 10-year Gail risk of developing breast cancer were the only two predictors of cancer development to enter the logistic regression model. Similarly, with the use of stepwise Cox proportional hazards regression, both the Gail risk of developing breast cancer and atypical cytology in random fine-needle aspirates contribute to risk prediction according to the following hazard function equation:

$$h(t) = [\text{ho}(t)]e^{B1*\text{Gail Risk} + B2*\text{hyperplasia with atypia}}$$

where $\text{ho}(t)$ is an arbitrary and unspecified baseline hazard function at time $t$, $B1$ is a coefficient that reflects change in the hazard rate with increasing value of the Gail risk-projected probability of developing cancer at 10 years, and $B2$ is a coefficient that reflects change in the hazard rate when hyperplasia with atypia is present. Specifically, a 1% increase in the 10-year Gail risk value increases the relative risk by a factor of
1.093 (95% confidence interval [CI] = 1.03–1.16), and atypical cytology increases the relative risk by a factor of 5.02 (95% CI = 2.01–12.56). Although it is accepted that stepwise procedures may lead to errant measures of variability and thus to statistical significance, they are still considered valid tools for exploratory analyses (55). In our analyses, given the fact that the same covariates entered the model in logistic and Cox regression analyses and given the extremeness of the nominal P values, these results may be considered as robust.

Combining the two predictors identified above appears to markedly enhance short-term predictive ability. For ease of graphical representation, we divided the women into three distinct groups by the Gail risk and hyperplasia with atypia (Fig. 3). Women with both a 10-year Gail risk above the median and hyperplasia with atypia (14% of our cohort) had a 15% incidence of breast cancer development within the first 3 years. Women with a 10-year Gail risk above the median but no hyperplasia with atypia (37% of our cohort) had an observed breast cancer incidence of 4% within the initial 3 years. Women with a 10-year Gail risk below the median of 4% had no cancers detected in the initial 3 years of follow-up, regardless of their cytology test results. Only two of the 235 women in this latter group have developed cancer, and these cancers were detected 7.3 and 8.4 years after the initial aspiration.

When the total cohort is divided by menopausal status at entry, cancer development in premenopausal women is predicted by hyperplasia with atypia and prior precancerous biopsy (P <.001) in a multivariable analysis but not by the Gail risk. For postmenopausal women, even though there were only six cases of cancer, cancer was marginally predicted by the Gail risk (P = .044) but not by hyperplasia with atypia (Table 5). Given the small numbers of cancers, the subcohort analysis by menopause status must be interpreted with caution.

**DISCUSSION**

Our study suggests that two readily available tools (i.e., the Gail risk model and cytomorphology of cells in random periareolar fine-needle aspirates) may be used together to define a cohort at very high short-term risk for DCIS and invasive breast cancer in a group of women who have major risk factors for the disease. Women with hyperplasia with atypia in their random fine-needle aspirates and a projected 10-year Gail risk above our cohort’s median of 4% have an observed incidence of breast cancer development of 15% at 3 years.

The prevalence of cytologically defined hyperplasia with atypia observed in our high-risk cohort may seem high, in that histologically defined atypical hyperplasia is present only in 4% of benign breast biopsy specimens (24,56) and in 3%–13% of autopsied women (57,58). However, the median age and/or risk
level for women in the above biopsy and autopsy series is likely to be different from that for our high-risk cohort. Furthermore, our prevalence rate for cytologic evidence of hyperplasia with atypia is based on pooling of two aspirations done 6–12 months apart to form the initial dataset. The prevalence rate of cytologically defined hyperplasia with atypia from a single aspiration is 12%. This percentage is in agreement with the findings of Ward et al. (59) and Khan et al. (39), who reported the prevalence of hyperplasia with atypia in random fine-needle aspirates in high-risk cohorts to be 8% and 16%, respectively. Considering that the three studies differed in fine-needle aspirate technique, cell processing, cytologic criteria, and type of high-risk population, these percentages are remarkably consistent.

Because hyperplasia with atypia is such a strong predictor of the development of breast cancer and is theoretically reversible, efforts are currently being expended to evaluate periareolar fine-needle aspirate cytology as a surrogate end point biomarker in intermediate-length phase II trials (31). To extend the use of fine-needle aspirate cytology as a surrogate end point biomarker, morphologic changes will need to be quantitated or, at the minimum, semiquantitated. Masood and co-workers (60,61) have proposed a semiquantitative method for evaluating fine-needle aspirate cytology, in which six different morphologic features are assigned 1–4 points. Using diagnostic smears of fine-needle aspirates taken from palpable masses, Masood et al. (60,61) defined specimens that scored between 6 and 10 as nonproliferative breast disease, those scoring between 11 and 14 as proliferative breast disease without atypia, those scoring between 15 and 18 as proliferative breast disease with atypia, and those that scored 19 or higher as carcinoma. Using filtered preparations as opposed to smears, we found that nonproliferative specimens generally scored 10 and below, hyperplasia without atypia scored between 11 and 14, and hyperplasia with atypia scored between 15 and 18. Efforts are being made to fully quantify morphology through image morphometry (62,63). Nuclear and nucleolar size and shape and chromatin pattern can be characterized individually or jointly as a z score (deviation from a normal nonproliferative specimen) (21,64).

Molecular assessments in breast samples are more readily quantified than is morphologic characterization and, therefore, should be less susceptible to interpretive variance. Immunocytochemical detection of protein expression is currently the method most widely used in small tissue samples. Our findings are consistent with those of other investigators (65–83), in that many molecular markers previously thought to be expressed only in breast cancer are also expressed to some degree in benign breast tissue. Staining intensity or pattern may be different between benign and malignant lesions, however (22,52,53,84–89).

The degree to which antibody localization (membrane versus cytoplasm versus nucleus), pattern (heterogeneous versus diffuse), or even staining intensity has prognostic or predictive importance has not been defined clearly for many markers. Despite the theoretical advantages of molecular markers as risk or surrogate end points, in our study, only the expression of multiple markers from the three-biomarker set (EGFR, ER, and p53) and expression of ER alone were predictive of later cancer development. This result was observed only in univariate analysis when neither cytology nor the Gail risk was considered. In a previous analysis of 224 subjects (six cancers, median follow-up of 32 months) (43) in which tests of all five biomarker tests were attempted or performed, expression of multiple markers from the three-biomarker set (EGFR, ER, and p53) was predictive of subsequent cancer development in multivariable and univariate analyses. There are several possibilities for why the molecular biomarkers were not predictive for cancer in the current multivariable analysis: 1) Our marker immunostaining or scoring criteria may have lacked sensitivity or specificity; 2) all subjects in the present analysis did not have all tests for the biomarkers attempted, and even for those tests attempted a portion of the samples did not have sufficient cells to adequately analyze markers; and 3) biomarker expression may be heterogeneous in precancerous breast tissue, similar to the situation in invasive cancer, and thus morphology as a final common pathway will have the advantage over molecular markers in multivariable analysis. All of these potential explanations are likely to be operative in regard to our study. In view of the predictive ability for some markers (EGFR, ER, and p53) in a subcohort in which tests for all markers were attempted or performed (43), the latter two considerations are likely to be the most important. This underscores the technical limitations of trying to immunocytochemically assay a large number of biomarkers on a small amount of tissue. Would serial core breast biopsies be better in this regard? Although the intuitive answer would appear to be “yes,” the study by Mansoor et al. (32) would indicate otherwise, unless biopsies could be directed to a nonfatty portion of the breast tissue.

The 60% prevalence rate of EGFR associated with hyperplasia with atypia in fine-needle aspirates is similar to what other investigators (68,69,82,83) have noted in histologically defined proliferative breast disease. Although our rate in samples with atypia is higher than what is observed in invasive cancer, it must be remembered that EGFR protein expression is lost in the majority of cases of invasive cancer, even while adjacent benign tissue retains EGFR immunoreactivity (83). Moreover, those samples retaining strong EGFR expression are generally ER negative (83). The 39% prevalence of ER expression in fine-needle aspirate samples with hyperplasia with atypia is somewhat lower than what others (79,81) have observed and may be related to loss of immunoreactivity during cytospin preparation (54).

Our 46% prevalence rate of p53 associated with hyperplasia with atypia in fine-needle aspirates is consistent with the rate of 25% in a variety of benign breast disease samples reported by Millikan et al. (74). Schmitt (75) has also noted p53 immunopositivity in intraductal hyperplasia without atypia. Rohan et al. (80) reported that p53 immunodetection in benign breast tissue was associated with a weak but a statistically significantly increased incidence of later breast cancer development. Strong, diffuse nuclear immunostaining is often a reflection of protein stabilization as a result of missense mutation and is generally accompanied by loss of p53 transcriptional activity (90–93). p53 staining in our study and in several reports of benign breast tissue tends to be heterogeneous and to be less intense than that generally observed in high-grade DCIS or invasive cancer, in which there is an underlying p53 missense mutation (90–99). It is unlikely that p53 mutation was responsible for protein stabilization in the majority of our samples (100). Although mutations have been observed in mouse mammary hyperplasia (101), p53 mutations in human breast neoplasia appear infrequent before in situ cancer (98,102). p53 protein stabilization and immunodetection have also been reported in response to hypoxic stress, hyperproliferative signals, and altered protein–protein interactions.
interactions, such as loss of adhesion and DNA damage (103–107). It is tempting to speculate that the high rate of p53 expression found in our samples of hyperplasia with atypia is associated with hyperproliferation and loss of adhesion.

Cytoplasmic and low-level HER2/NEU membrane staining has been detected in normal and proliferative breast disease specimens when frozen sections and non-formalin-based fixatives are used; these methods are similar to methods used in this study (88,89,108–110). Strong membrane staining in formalin-fixed tissue is often associated with HER2/NEU gene amplification, high proliferation rates, and poor prognosis in invasive cancer (111–114). Cytoplasmic staining for HER2/NEU may be associated with activated receptor, internalization, low proliferation rates, and cellular differentiation and has been observed in both cancer and benign tissues (113,115–117). The prognostic implications of cytoplasmic staining (113,118–120) or detection of low levels of HER2/NEU in invasive cancer by immunocytochemistry or other methodology is controversial (87,121–123). Whether we nor Rohan et al. (80) found HER2/NEU staining in benign breast samples to be predictive of later cancer, although, in our study, it was associated with concomitant atypia. The staining pattern for HER2/NEU in our study, as for p53, tended to be focal and less intense than that observed for cancer. Recent studies [reviewed in (88,89)] indicate that HER2/NEU may function largely as a coreceptor within the EGFR/HER2 family. Gene amplification (reflected by intense, diffuse, membrane stain uptake) often results in HER2/NEU homodimerization and strong proliferative potential. However, the physiologic consequences of lower levels of HER2/NEU expression in the absence of gene amplification are varied and seem to be dependent on concomitant levels of other receptors, growth factors, and hormones.

The role of molecular markers in cells from random fine-needle aspirates assayed by immunocytochemistry is at present unclear. Molecular marker assessment may not substantially contribute to risk prediction if morphology is also being assessed. Because some molecular markers (most notably expression of multiple markers from the three-biomarker set—EGFR, ER, and p53) are associated with hyperplasia with atypia, which is a strong predictor of cancer, they may still be useful in chemoprevention trials as surrogate end point biomarkers, depending on the mechanism(s) of action proposed for the agent being tested.

Finally, caution needs to be applied in regard to extending the random fine-needle aspirate technique beyond a young, high-risk population. The majority of subjects in our study were premenopausal or, if postmenopausal, were on HRT. Although not specifically measured, the majority of women had mammographically dense breasts. Aspiration of predominantly fatty/involutorial breast tissue is associated with poor cell yield in our experience and in the experience of others (124).

In conclusion, cytologic evidence of hyperplasia with atypia in random periareolar fine-needle aspirates and Gail model risk estimates were independently predictive of subsequent cancer development/detection in a high-risk cohort. Use of the two factors together provided the best prediction of short-term cancer risk and can be used to select very high risk subjects for prevention interventions. Cytomorphology from random fine-needle aspirates is being explored as a surrogate end point biomarker in breast cancer chemoprevention trials.

REFERENCES

(18) Jernstrom HC, Olsson H, Borg A. Reduced testosterone, 17 beta-oestradiol and sexual hormone binding globulin, and increased insulin-like growth factor-I concentrations, in healthy nulligravid women aged 19–25 years who were first and/or second degree relatives to breast cancer patients. Eur J Cancer Prev 1997;6:330–40.
(21) Boone CW, Kelloff GJ. Intraepithelial neoplasia, surrogate endpoint bio-


Journal of the National Cancer Institute, Vol. 92, No. 15, August 2, 2000 ARTICLES 1225


NOTES

This project, conducted over many years, was made possible by the generous support of the Kansas Cancer Institute, the Kansas Masonic Foundation, Eastern Star, the Ladies Auxiliary of the Veterans of Foreign Wars, the Oppenheimer Foundation, and numerous private donations.

We acknowledge with gratitude the devotion and effort of many individuals too numerous to identify.

Manuscript received November 22, 1999; revised May 16, 2000; accepted May 23, 2000.