Journal of Cellular Biochemistry Supplement 34:7–12 (2000)

Breast Cancer Chemoprevention Trials Using the Fine-Needle Aspiration Model

Bruce F. Kimler,[1]* Carol J. Fabian,[2] and Dennis D. Wallace[3]

[1]Department of Radiation Oncology, University of Kansas Medical Center, Kansas City, Kansas 66160
[2]Division of Clinical Oncology, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, Kansas 66160
[3]Department of Preventive Medicine, University of Kansas Medical Center, Kansas City, Kansas 66160

Abstract Selection of surrogate endpoint biomarkers (SEBs) and appropriate study design are two of the main challenges in evaluating potential chemopreventive agents. In a prospective random fine-needle aspiration (FNA) study of women at high risk of development of breast cancer, we previously demonstrated that cytologic evidence of epithelial hyperplasia with or without atypia, as well as abnormalities of several cellular biomarkers (DNA ploidy; immunocytochemical expression of p53, EGFR, ER, and/or Her-2/*neu*), were more prevalent in high-risk women than in low-risk controls. We also demonstrated that the subsequent development of breast cancer was best predicted by an initial presentation of hyperplasia with atypia, as well as by multiple biomarker abnormalities. These findings indicate that FNA cytology and biomarkers can be used to identify women who are appropriate subjects for chemoprevention trials, and can then be used as surrogate endpoint biomarkers to monitor efficacy of potential agents. An example of this use in an ongoing single-agent phase II trial is provided. Several options for study design of possible multi-agent breast cancer chemoprevention trials are discussed, depending upon the existing preclinical and clinical data, the questions being asked, and the number of eligible subjects available. J. Cell. Biochem. Suppl. 34:7–12, 2000.

Key words: biomarker; breast; breast cancer development; chemoprevention; clinical trials; cytology; ER; EGFR; fine-needle aspiration (FNA); Her-2/*neu*; high risk; p53; ploidy; risk assessment; study design; surrogate endpoint biomarker

For phase II trials of potential agents that might prevent breast cancer, one must identify appropriate cohorts of subjects, select optimum tissue biomarkers for testing, standardize tissue sampling methods, and otherwise develop a reliable model for chemoprevention trials [Dhingra, 1995; Dhingra et al., 1993; Fabian and Kimler, 1997; Fabian et al., 1998; Kelloff et al., 1993; O'Shaughnessy, 1996]. The most appropriate subjects for phase II trials will be those who consider themselves at short-term high risk of developing breast cancer and who possess breast tissue biomarkers that are predictive for breast cancer development and are theoretically reversible.

*Correspondence to: Bruce F. Kimler, Department of Radiation Oncology, University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, KS 66160-7321. E-mail: bfkimler@kumc.edu

Received 3 September 1998; Accepted 9 March 1999

BIOMARKERS FROM FINE-NEEDLE BREAST ASPIRATION

Since 1989 at the University of Kansas Medical Center, we have performed breast fine-needle aspiration (FNA) and characterized the aspirated ductal epithelial cells of more than 500 women at high risk of the development of breast cancer on the basis of family history or prior cancerous or precancerous biopsy. The procedures for aspiration, tissue processing, and biomarker analysis have been previously detailed [Fabian et al., 1993, 1994, 1995, 1996, 1997c; Zalles et al., 1995]. Markers studied included cytologic morphology characterization, DNA ploidy, and immunocytochemical expression of p53 [Kamel et al., 1998], epidermal growth factor receptor (EGFR), estrogen receptor (ER), and Her-2/*neu*.

We have reported that women at high risk of cancer development are more likely to exhibit cytological abnormalities (approximately 60% with epithelial hyperplasia or hyperplasia with

atypia) than are women who do not possess any recognized risk factors [Zalles et al., 1995]. Similarly, we have observed [Fabian et al., 1993, 1994, 1995, 1997a,c] that the five biomarkers are abnormally expressed (i.e., DNA aneuploidy and immunocytochemical overexpression) in the high-risk women, whereas they are uncommon in the low-risk group. Abnormal biomarker expression, in turn, is associated with cytologic abnormality, being most common in FNAs that demonstrate hyperplasia with atypia. Likewise, the incidence of multiple abnormal biomarker expression increases with increasing cytologic abnormality. Finally, we have demonstrated that these abnormalities are associated with increased short-term risk of developing breast cancer. Cytologic evidence of hyperplasia with atypia at initial aspiration was the best predictor of subsequent development of breast cancer, with a median follow-up of approximately 3 years [Fabian et al., 1996, 1997a,b,c]. Thus, we have identified a suitable population, as well as a set of cytologic and molecular markers that may be considered for use as surrogate endpoint biomarkers (SEBs) in chemoprevention trials. Because of the inherent limitation of low cell yield from FNA, one cannot expect to be able to reliably acquire data on cytology plus all five molecular biomarkers from each aspirate from each woman. For chemoprevention trials, it would be advantageous to restrict biomarkers to those that are not only associated with subsequent cancer development, but that are also most frequently abnormal. A high frequency of expression provides the greatest opportunity for normalization as a result of the intervention.

CURRENT CHEMOPREVENTION TRIAL USING THE FNA MODEL

In an ongoing phase II trial of α-difluoromethylornithine (DFMO) using the FNA model, we have focused on cytology and three biomarkers to define eligibility for study participation (Table I). Cytologic characteristics of hyperplasia with and without atypia were selected as they are associated in univariate (hyperplasia) and multivariate (hyperplasia with atypia) analyses with subsequent cancer development. EGFR and p53 were selected as they are likewise associated with cancer development in univariate analyses, as well as associated with hyperplasia with atypia in multivariate analysis. EGFR and p53 were also overexpressed in 40% and 32%, respectively, of our high-risk group. DNA ploidy (aneuploidy was detected in 28% of aspirates) was selected since it can be performed with nuclear mophometry which is being explored as a potential method of quantifying cytologic morphology. ER and Her-2/*neu* were not selected as ER overexpression in cytospin preparations occurs infrequently in the high-risk population; and Her-2/*neu* overexpression by our methodology was not predictive of cancer development even in univariate analysis.

TABLE I. Use of Biomarkers in Phase II Chemoprevention Trial of DFMO in Women at High Risk of Breast Cancer

Identification/selection of cohort—eligibility criteria
Cytological evidence of epithelial hyperplasia with or without atypia
EGFR overexpression[a]
p53 overexpression[a]
DNA aneuploidy[a]
Potential surrogate endpoint biomarkers
Cytology category
Cytological grading score (as per Masood's criteria)
p53 expression
EGFR expression
PCNA expression (labeling index)
DNA ploidy
Nuclear morphometry (individual factors or combined z-score)
Insulin-like growth factor-1 levels (serum)
Insulin-like growth factor binding protein-3 levels (serum)
Mammographic breast density
Drug effect/biochemical activity markers
Urinary polyamine (spermine, spermidine, putrescine) levels

DFMO, α-difluoromethylornithine; EGFR, epidermal growth factor; PCNA, proliferating cell nuclear antigen.
[a]At least one required if no atypia present.

To be eligible for participation in the trial, women must have cytologic evidence of hyperplasia, plus either atypia or at least one biomarker abnormality (from the set of DNA aneuploidy, or overexpression of EGFR or p53). Women with these characteristics have a higher likelihood of subsequent cancer development than do those without [Fabian et al., 1998]. Some 47% of our high-risk population satisfy these criteria and are potentially eligible for study participation. Approximately 18% of the women in our high-risk cohort exhibit epithe-

lial hyperplasia with atypia, and are thus eligible (Figure 1). For those with hyperplasia without atypia, the single most productive biomarker test (in terms of demonstrating an abnormality) is EGFR, with another 18% of the entire cohort identified as potentially eligible. Another 8% are identified by the addition of a test for p53 overexpression, and another 3% by ploidy analysis. As mentioned above, the inclusion of DNA aneuploidy is justified by the potential to acquire considerable nuclear morphometric data from a Fuelgen-stained slide prepared for simply image analysis of DNA content. On the other hand, inclusion of ER and her-2/*neu*, which would add at most another 5% of eligible subjects, is not justified, as it raises the risk of being unable to perform any of the five tests because the available cells were diluted between too many test slides.

The primary endpoint that will be evaluated for response is the change in cytological category between the prestudy and post-treatment aspirations (Table I). In an effort to define a more objective endpoint, changes in a semiquantitative cytological index score [Masood, 1995; Masood et al., 1990] will be studied. We will compare immunocytochemical expression of p53 and EGFR; plus proliferating cell nuclear antigen (PCNA) will be used as an index of proliferation. In addition to DNA ploidy, we will explore automated nuclear morphometry with its ability to provide objective, quantitative information on a number of morphologic variables [Boone et al., 1997]. These variables may be considered individually or summed into a z-score that represents a global deviation from normalcy. We will also explore biomarkers that do not require sampling of breast tissue. Serum levels of insulin-like growth factor-1 (IGF-1) and one of its binding proteins (IGFBP-3) will be assessed, as well as mammographic breast density. Finally, urinary polyamine levels will be assayed as biochemical activity markers to document that DFMO, at the dose used, inhibits ornithine decarboxylase and lowers polyamine levels.

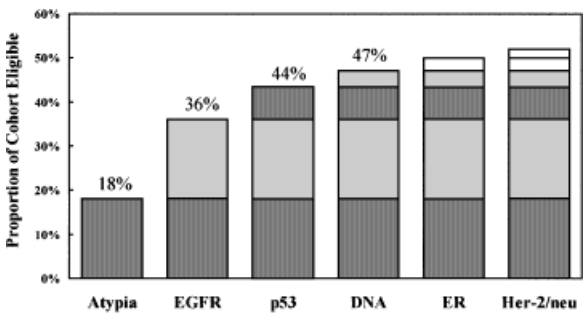

Fig. 1. Cumulative frequency (expressed as a percentage of all individuals in the high-risk cohort) of subjects potentially eligible for trial. This demonstrates the effect of stepwise addition to a baseline requirement of epithelial hyperplasia, additional test considerations by which the secondary criterion of an additional single abnormality can be fulfilled.

For the primary endpoint, we define a response as an improvement in cytological characterization (e.g., from hyperplasia with atypia to hyperplasia without atypia; or from hyperplasia to nonproliferative cytology) and/or a reduction in the cytology index of 3 points (scale from 6 to 18 for normal to hyperplasia with atypia). Moreover, we considered that such changes should be observed in 30% of subjects in order to be of practical importance and to warrant further study of DFMO as a chemopreventive agent. Using this 30% response rate and conventional power considerations for determination of sample sizes (see below), the trial was designed as a simple two-arm (DFMO vs placebo) study with a planned entry of 120 subjects. The goal is to obtain 100 evaluable subjects, 50 in each arm. The trial was initiated in June of 1997 and has accrued 61 subjects in less than 1 year. Thus, the accrual target is well within reach.

DESIGN CONSIDERATIONS FOR MULTI-AGENT TRIALS

When one considers trials of multiple agents, several options are available for study design. The specific type of multi-agent (e.g., drug A and drug B) trial design will depend on the preclinical and clinical data available, as well as the questions that one wishes to address. For simplicity, we will assume that multiple dose issues need not be addressed (i.e., at least a tolerable dose, and in some cases a minimum effective dose, for each agent has already been determined); otherwise, the number of possible drug-dose combinations is overwhelming.

Two-Arm Study: Placebo vs A+B

In a situation in which neither drug A or drug B is an effective clinical chemopreventive in minimally toxic doses, but preclinical data suggest that the combination may be effective, one could perform a standard randomized, double-blind, two-arm study of placebo vs the combination of A + B. Such a trial would only address

the question of whether the combination is active; it would not address the possible efficacy of either agent alone. For purposes of sample size calculations, let us assume that we are interested in a true change in the primary endpoint in 30% of the treated cohort. Further, let us assume (using a conservative estimate) a 20% change in the placebo group, such that we are effectively looking for a difference in SEB response rate of 20% vs 50%. In this situation, we would be permitted to use a one-tailed approach. If we power the study with type I error (α) values of 0.05 and power (1-β) of 0.90 (see below), and take into account a reasonable drop-out and/or inevaluability rate (10%), we are left with an overall requirement to enter approximately 60 subjects per arm, for a total of 120 eligible subjects. In essence, this is the same study design used for our current two-arm phase II trial of DFMO.

Three-Arm Study: Placebo vs A vs A+B

Next, we might consider another scenario wherein clinical trial evidence (e.g., from two phase II trials) indicates that one agent (A) is active but the other (B) is either inactive or minimally active. However, preclinical data from animal carcinogenesis models suggest that the combination of A + B may be more effective than A alone. In this case, we could perform a three-arm study of placebo vs A vs A + B. Such a trial design would be capable of addressing two questions: Is A effective? and Does B modify A (either to increase or decrease activity)? In this case, because of the potential to either gain or lose from the addition of B to A, a two-tailed approach seems warranted. For sample size considerations, let us assume that the extent of the effect that addition of B adds to A is 20%, such that we are projecting SEB response rates of 20% (placebo) and 50% (agent A alone) as above, and 70% for the combination of A + B. Because of the existing information and the questions being asked, we maintain the size of the placebo arm at 60 subjects, but the two treatment arms (A and A + B) must be increased to 120 subjects each. Thus, a total requirement for 300 subjects is predicted. Note that this provides a minimum power of 0.90 to test A vs A+B and a power in excess of 0.97 to test A vs placebo and A+B vs placebo.

Four-Arm Study: Placebo vs A vs B vs A+B

Finally, let us consider the scenario wherein there may or may not be clinical trial evidence that either A or B is active. However, preclinical data indicate that each agent may be effective and that the combination may be even more so. Here, we would perform a four-arm study of placebo vs A vs B vs A + B. Such a trial would be designed to answer the following questions: Is A active? Is B active? and Does the combination provide increased activity relative to the better single agent? For purposes of sample size calculations, let us assume the same response rates as above, with the addition that agent B is projected to have a 20% true response rate when used alone (i.e., not quite as good as agent A alone). Thus, the response rates we are projecting are 20% (placebo), 50% (A), 40% (B), and 70% (A + B). We again may maintain the placebo arm at 60 subjects, but will require 120 subjects in each of the three treatment arms. Thus, a total of 420 subjects for this four-arm study would be required. This sample size produces a power of 0.90 to test A vs A+B, a power of >0.97 to test A vs placebo and A+B vs placebo, but only a power of 0.77 to test B vs placebo. It is worth noting that such a four-arm study, although not specifically powered for it, may provide additional information as to whether the combination of A + B is superior to B alone, and whether the increased activity is more than additive.

Choice of Study Design

Overall, we observe that depending upon what is known and what is desired to be tested, one may choose a two-arm, a three-arm, or a four-arm trial design. With increasing number of arms, an increase in the total number of subjects will be required (Table II). However, with the increased requirement for subjects, one does gain the potential to address more questions. In fact, given the right set of circumstances and a particular database of preclinical

TABLE II. Study Design Considerations for Multi-agent Chemoprevention Trials

Study design	Subjects required	Questions addressed
Two-arm	120	1
Three-arm	300	2
Four-arm	420	3 (possibly 5)

and clinical data, an initial four-arm study may be more efficient than sequential (two or three) two-arm studies, followed by a three-arm or four-arm study. Pragmatically, it may be much easier to convince a potential subject worried about her risk of developing breast cancer to enroll in a trial in which she has only a 1:7 chance of being assigned to a placebo group than to a trial in which her chances of not receiving a potentially active agent are 1:2. Another pragmatic consideration is that while it is possible to conduct a trial requiring 120 subjects at a single institution, this is not the case when 300–420 subjects are required. For this, a consortium of several institutions must be developed in order to reach accrual goals in the desired time-frame of 1–2 years. Conversely, if the potential number of subjects is not sufficient, then three- and four-arm studies should not be considered.

Regardless of the specific trial design chosen for a multi-agent trial, one must have a solid preclinical and/or clinical basis and rationale to predict that multi-agent chemoprevention will provide substantial improvement over single-agent chemoprevention.

Changing the Assumptions

As an addendum to the above discussion, we might consider the assumptions used for power calculations: Type I error (α) values of 0.05 and a power of 0.90. Although it is conventional in most chemoprevention phase I/II trials to use a power of only 0.80, we would argue that the power of 0.90 is appropriate for such trials in order to reduce the possibility of "discarding" from future consideration an agent that is actually active, but for which activity is not detected in a single trial. If a power of 0.80 were to be used, the sample size requirements for the studies described above would be reduced by approximately 30%.

SUMMARY

Random fine-needle breast aspiration of women at high epidemiologic risk of development of breast cancer provides a minimally noninvasive approach to tissue sampling in phase II chemoprevention trials. Comparison of cytologic and molecular biomarkers over the treatment interval can be used to document response to a possible chemopreventive agent. Several trial designs are available for multi-agent trials, but all except the simplest two-arm study require fairly large number of subjects in order to detect a substantial change in SEB expression.

ACKNOWLEDGMENTS

The secretarial assistance of Rebecca McNeely is acknowledged with gratitude, as are the efforts of Jennifer Klemp in coordinating the entire High Risk Breast Cancer program.

REFERENCES

Boone CW, Bacus JW, Bacus JV, Steele VE, Kelloff GJ. 1997. Properties of intraepithelial neoplasia relevant to cancer chemoprevention and to the development of surrogate end points for clinical trials. Proc Soc Exp Biol Med 216:151–165.

Dhingra K. 1995. A phase II chemoprevention trial design to identify surrogate endpoint biomarkers in breast cancer. J Cell Biochem 23(suppl):19–24.

Dhingra K, Vogel V, Sneige N, Sahin A, Aldaz M, Hortobagyi GN, Hittelman W. 1993. Strategies for the application of biomarkers for risk assessment and efficacy in breast cancer chemoprevention trials. J Cell Biochem 17(suppl G):37–43.

Fabian CJ, Kimler BF. 1997. Chemoprevention of breast cancer. Breast J 5:220–226.

Fabian C, Zalles C, Kamel S, McKittrick R, Moore W, Zeiger S, Simon C, Kimler B, Cramer A, Garcia F, Jewell W. 1993. Biomarker and cytologic abnormalities in women at high and low risk for breast cancer. J Cell Biochem 17(suppl G):153–160.

Fabian CJ, Zalles C, Kamel S, Kimler BF, McKittrick R, Tranin AS, Zeiger S, Moore WP, Hassanein RS, Simon C, Johnson N, Vergara G, Jewell WR, Lin F, Bhatia P, Chin T. 1994. Prevalence of aneuploidy, overexpressed ER and EGFR in random breast aspirates of women at high and low risk for breast cancer. Breast Cancer Res Treat 30: 263–274.

Fabian CJ, Kamel S, Kimler BF, McKittrick R. 1995. Potential use of biomarkers in breast cancer risk assessment and chemoprevention trials. Breast J 1:236–242.

Fabian CJ, Kamel S, Zalles C, Kimler BF. 1996. Identification of a chemoprevention cohort from a population of women at high risk for breast cancer. J Cell Biochem 25S:112–122.

Fabian CJ, Kamel S, Zalles C, Kimler BF. 1997a. Biomarkers predictive of short interval cancer development in high risk women. Breast J 3(suppl):42–48.

Fabian CJ, Zalles C, Kamel S, Kimler BF. 1997b. Short term development of breast cancer is predicted by fine needle aspiration cytology and cellular biomarkers. Proc Am Soc Clin Oncol 16:129a.

Fabian CJ, Zalles C, Kamel S, Zeiger S, Simon C, Kimler BF. 1997c. Breast cytology and biomarkers obtained by random fine needle aspiration: use in risk assessment and early chemoprevention trials. J Cell Biochem Suppl 28:101–110.

Fabian CJ, Kimler BF, Elledge RM, Grizzle WE, Beenken SW, Ward JH. 1998. Models for early chemoprevention trials in breast cancer. Hematol Oncol Clin North Am 12:993–1017.

Kamel S, Zeiger S, Zalles C, Tawfik O, Simon C, Kimler BF, Fabian CF. 1998. p53 immunopositivity and gene mutation in a group of women at high risk for breast cancer. Breast J 4:396–404.

Kelloff GJ, Boone CW, Steele VE, Crowell JA, Lubet R, Doody LA, Greenwald P. 1993. Development of breast cancer chemopreventive drugs. J Cell Biochem 17(suppl G):2–13.

Masood S. 1995. Cytomorphology of fibrocystic change, high risk, and premalignant breast lesions. Breast J 1:210–221.

Masood S, Frykberg ER, McLellan GL, Scalapino MC, Mitchum DG, Bullard JB. 1990. Prospective evaluation of radiologically directed fine-needle aspiration biopsy of nonpalpable breast lesions. Cancer 66:1480–1487.

O'Shaugnessy JA. 1996. Chemoprevention of breast cancer. JAMA 275:1349–1353.

Zalles C, Kimler BF, Kamel S, McKittrick R, Fabian CJ. 1995. Cytologic patterns in random aspirates from women at high and low risk for breast cancer. Breast J 1:343–349.