A Randomized Clinical Trial of 4-Hydroxyphenylretinamide for High-Grade Squamous Intraepithelial Lesions of the Cervix[1]

Michele Follen,[2] E. Neely Atkinson, David Schottenfeld, Anais Malpica, Loyd West, Scott Lippman, Changping Zou, Walter N. Hittelman, Reuben Lotan, and Waun Ki Hong

Departments of Gynecologic Oncology [M. F., L. W., C. Z.], Biomathematics [E. N. A.], Pathology [A. M.], Experimental Therapeutics [W. N. H.], Cell Biology [R. L.], and Head and Neck Thoracic Oncology [S. L., W. K. H.], The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030; Department of Obstetrics, Gynecology, and Reproductive Biology, The University of Texas Health Science Center; Houston, Texas [M. F., L. W., C. Z.]; and Department of Epidemiology, The University of Michigan, School of Public Health I, Ann Arbor, Michigan 48104 [D. S.]

ABSTRACT

Purpose: **Previous trials of topical *trans*-retinoic acid treatment of cervical intraepithelial neoplasia (CIN) grades 2 and 3 led to a statistically significant regression of CIN 2, but not CIN 3. We tested *N*-(4-hydroxyphenyl)retinamide (4-HPR), a promising oral retinoid that has been shown to induce apoptosis through nonretinoic receptor acid-mediated pathways, for its toxicity and efficacy against CIN 2/3.**

Experimental Design: **In a blinded randomized trial, 4-HPR at 200 mg/day for 6 months (with a 3-day/month drug holiday) was compared with placebo in patients with biopsy-proven CIN-2/3 [high-grade squamous intraepithelial lesions (HGSILs)]. Patients were treated with placebo or 4-HPR for 6 months, biopsied, and then followed for an additional 6 months. At the 12-month end point, they underwent either loop excision if a histological lesion was present or a biopsy from the original area of the lesion if no lesion was present.**

Results: **An interim analysis of blinded data showed a significantly worse prognosis at 12 months for one group. When the code was broken because of the poorer outcomes, we discovered that the 4-HPR treatment arm was performing more poorly than was the placebo at 6 and 12 months (25 *versus* 44% response rates at 6 months; 14 *versus* 50% at 12 months). Toxicity was not significant in either arm.**

Conclusions: **4-HPR at 200 mg/day with a 3-day/month drug holiday is not active compared with placebo in the treatment of HGSIL. Because 4-HPR is active in the laboratory, the lack of effect in our trial may indicate that higher doses are needed in patients to achieve comparable results.**

INTRODUCTION

4-HPR[3] is one of the most promising retinoid compounds in both laboratory and animal models (1). In 1979, Moon *et al.* demonstrated that 4-HPR was a potent inhibitor of mammary carcinogenesis in the rat (2). Its toxicity has been extensively studied and found to be less than that of other retinoids (3, 4). The Chemoprevention Branch of the NCI has active 4-HPR trials in several organ sites, including the prostate, lung, oral cavity, breast, bladder, and cervix (5).

In part, the wide enthusiasm for 4-HPR comes from its ability to decrease cell growth and drive cells into apoptosis in many cell culture systems (6–9). The mechanism of effect of 4-HPR has been consistently demonstrated to be independent of retinoic acid receptors (10–12) and to involve the production of reactive oxygen species, at least in some cells, including cervical carcinoma (12).

The five Phase I-II trials of 4-HPR for various cancers have been completed and were summarized by Veronesi *et al.* (1). Significant ocular toxicity, manifest as nyctalopia or night blindness, was noted at doses above 200 mg/day and was shown to be related to low serum retinol levels. Serum retinol levels returned to the normal range if 4-HPR was stopped for 3 days each month; therefore, Phase II studies funded by the NCI used an oral dose of 200 mg/day monthly with a 3-day/month drug holiday.

This is the first trial of 4-HPR in patients with cervical lesions. The objectives of this clinical trial were to determine the efficacy of 4-HPR *versus* placebo for the treatment of high-grade CIN 2 and 3, to document the qualitative and quantitative toxicity of 4-HPR compared with placebo, and to explore the modulation of the several biomarkers of cervical carcinogenesis. Herein we report the results of an interim analysis of trial participation, the efficacy judged by histological regression, and both reported and measured toxicities. Results of the biomarker studies will be published later.

Received 1/16/01; revised 8/16/01; accepted 9/13/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This study was funded by the National Cancer Institute Chemoprevention Branch Contract NO1-CN-25433A. A preliminary version of this report was presented at the Society of Gynecologic Investigation in Toronto, Canada, on March 16, 2001.

[2] To whom requests for reprints should be addressed, at M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Box 193, Houston, Texas 77030. Phone: (713) 745-2564; Fax: (713) 792-4856; E-mail: mfollen@mdanderson.org.

[3] The abbreviations used are: 4-HPR, *N*-(4-hydroxyphenyl)retinamide; NCI, National Cancer Institute; CIN, cervical intraepithelial neoplasia; LBJ, Lyndon Baines Johnson; UTHSC, University of Texas Health Science Center; STD¸ sexually transmitted disease; HPV, human papillomavirus; CIS, carcinoma *in situ*; 4-MPR, *N*-(4-medroxyphenyl)retinamide; OCP, oral contraceptive.

PATIENTS AND METHODS

Subjects and Recruitment. The recruitment of patients began in July 1995, and the study remained open until October 2000. We screened for eligibility women 18 years and older with no prior malignancy who agreed to use contraception for the duration of the study. All of the patients underwent Pap smear, and those who had abnormal Pap smears showing CIN2/3 were referred for colposcopy at the M. D. Anderson Cancer Center, the UTHSC at Houston, or the LBJ Hospital, Houston, Texas. Exclusion criteria included age less than 18 years old, a positive pregnancy test, desire to become pregnant, and HIV positivity.

Colposcopy-directed biopsies were performed, and only patients with high-grade lesions in the cervical biopsy and a negative endocervical curettage (indicating no involvement of the endocervical canal) were eligible for the study. The pathology slides were reviewed at M. D. Anderson Cancer Center, and only after consensus by the study pathologist (A. M.) and the clinical pathologist were patients considered eligible. For a patient to be eligible, the cervical lesion had to be four to five times larger than a 2- by 2-mm biopsy and involve at least one-third of the surface area of the cervix. Thus, a histological diagnosis was made and the lesion size determined by colposcopy in each patient. Any patient who developed any sign of invasive cervical cancer (on a Pap smear, colposcopy, or cervical biopsy) was to be removed from the study and treated.

The principal investigator (M. F.) was the only physician involved in the care of the study patients at the three sites. She was assisted by two certified nurse practitioners who are expert colposcopists with >16 years of experience each. Two research nurses worked on the study during its course, one nurse from 1995 to 1998 and a second from 1998 to the present. NCI auditors reviewed patient data at 6-month intervals. All of the eligible patients were asked to participate and signed an informed consent. The research nurses spoke English and Spanish.

After informed consent, patients were randomly assigned by computer to 4-HPR and placebo arms by the "research pharmacy" at M. D. Anderson Cancer Center. There was no attempt to stratify the patients by any criteria. The principal investigator, nurse practitioners, patient, and research nurses were blinded to whether the patient was receiving drug or placebo. After randomization, all of the patients underwent colposcopy and colpophotography by the nurse practitioners to establish baseline measurement. All of the colpophotographs (for lesion size), pathology slides (baseline, 6- and 12-month responses), bone density scans (baseline, and increases or decreases), eye tests (baseline and changes), and laboratory studies were evaluated three times. These studies were first reviewed by the clinician assigned to the task, a second time by the participating study specialist (pathologist, nuclear medicine radiologist, opthalmological researcher), and a third time by the study specialist for consensus. All of these were blinded at each review to the treatment arm and the study outcome.

Study Plan. The study evaluation plan is summarized in Fig. 1. Pretreatment laboratory work included blood counts and measurement of bilirubin and creatinine levels. A baseline bone density study and electroretinogram and/or nyctalopia testing were also performed. HPV testing was performed at baseline using *in situ* hybridization and the ViraPap test (Digene, Gaithersburg, MD). The demographic history included race, birth date, history of STDs, number of sexual partners, smoking history (past and current), and OCP use (past and current). The baseline biopsies were read at M. D. Anderson. Lesion size was classified as >1/3, 1/3–2/3, and >2/3 of the surface area of the cervix by the first colposcopic exam and verified with colpophotography. The positive results for *in situ* hybridization or the Virapap test were recorded by the Pathology Department at M. D. Anderson. Patients were asked to participate in follow-up for 1 year at 3-month intervals.

Patients were randomized to 6 months of placebo or oral 4-HPR, 200 mg per day, with a 3-day drug holiday monthly. Colposcopy with colpophotography and Pap smears were repeated at 3, 6, 9, and 12 months, as were blood counts and serum chemistry assays. Colposcopy-directed biopsies were repeated at 6 and 12 months, as were bone density evaluations. Electroretinograms and/or nyctalopia testing were repeated at 6 and 12 months.

Response was evaluated histologically at the 6- and 12-month visits. Although Pap smears were obtained, the findings were of interest only if the cells showed a higher-grade lesion than those of the biopsy. If progressive disease was present histologically at the 6-month visit, code was broken by the research pharmacy. If the patient with progressive disease was on 4-HPR, she was treated with loop excision. If the patient was on placebo, she was crossed over to 4-HPR, treated for 6 months, evaluated again, and then monitored for another year on the study. The colpophotographs were used to judge clinical response, but only retrospectively.

Histological response was classified as: (*a*) no change; (*b*) partial response; (*c*) complete response; or (*d*) progressive disease. Lesions were considered to have partially regressed if the histological diagnosis decreased from CIS or CIN3 to CIN2, CIN 1, or HPV infection, and to have completely regressed if tissue samples were normal or negative by histological examination. Patients' lesions were considered to have progressed if CIN 2/3 progressed to CIS, microinvasion, or invasive cancer or if CIS progressed to microinvasive or invasive cancer. For some analyses, the patients with no change or with progressive disease would be considered nonresponders, and those with partial responses or with complete responses would be considered responders.

All of the biopsies were read three times, once by a gynecological pathologist assigned to patient care on the day the biopsy was obtained and twice by the study pathologist (A. M.). Discrepancies between the first and second reading were resolved by the study pathologist at the third reading, and a final diagnosis was assigned. The primary end point of interest was the response to therapy at 6 months. The 12-month end point indicated whether the response was durable.

Compliance was evaluated by counting the unused pills in the bottle at the end of each 3-month course. Serial assays of serum 4-HPR and retinol were performed. All of the patients were asked to fill out the NCI toxicity criteria sheets daily. These records were collected at the conclusion of 3 and 6 months of treatment. Patients were also called weekly to check the toxicities and their compliance with taking the medication.

Levels of 4-HPR, its metabolite 4-MPR, and retinol were

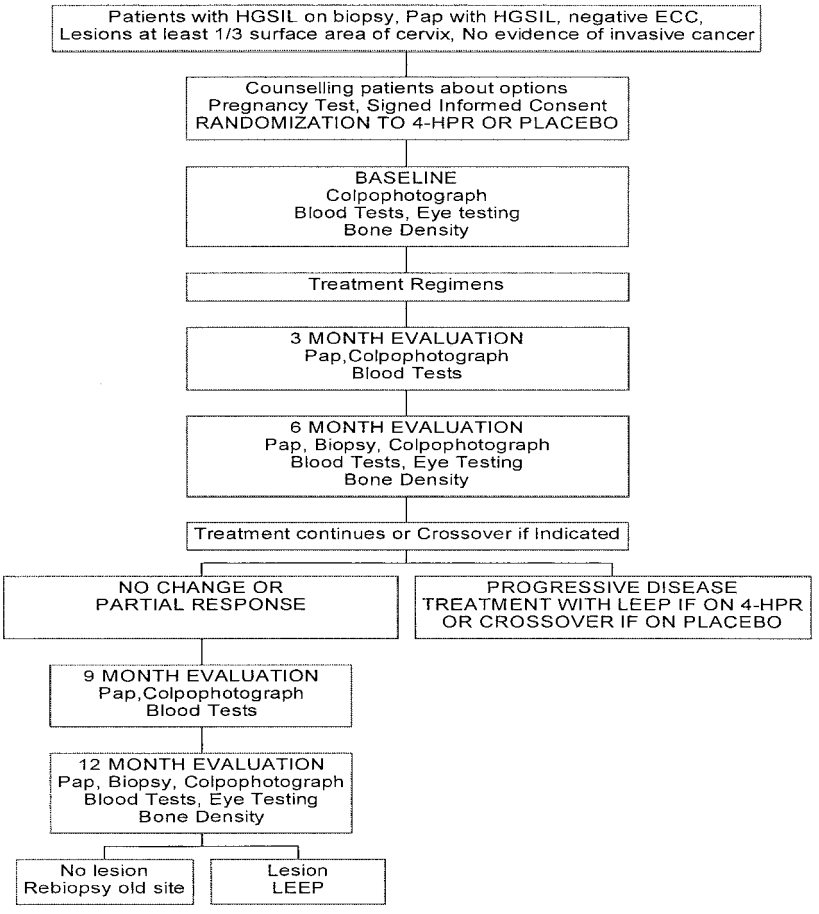

Fig. 1 Schema of study and evaluation for Phase II randomized clinical trial of 4-HPR and placebo in CIN 2/3. *HGSIL*, high-grade squamous intraepithelial lesion; ECC, endocervical curettage; LEEP, loop excision.

measured by high-performance liquid chromatography in thawed serum samples taken from 0, 3, 6, and 12 months, frozen at −70°C and protected from exposure to light. High-performance liquid chromatography was performed using 4-ethoxyphenylretinamide as an internal standard. The chromatographic separations was performed on a Vydac 20 ITP column (0.46 × 25 cm). The isocratic mobile phase was 55% acetonitrile, 10% *n*-butyl alcohol, 35% water, and 0.01 M ammonium acetate. The detector was programmed at 364 nm for the last 8 min to correspond to the elution times of 4-HPR, 4-MPR, and 4-ethoxyphenylretinamide.

Sample Size and Power Calculations. The sample size calculation and planned interim analyses for the study were based on the primary end point of histological regression. The sample size was generated using a logistic regression as described by Self *et al.* (13) using the formula:

$$\mathrm{Log}\left(\frac{\theta}{1-\theta}\right) = \beta_0 + \beta_1 I_{\mathrm{trt}} + \beta_2 I_{\mathrm{rec}} + \beta_3 I_{\mathrm{CIN\,2}}$$

where $I_{\mathrm{trt}} = 0$ if the patient was on placebo and $I_{\mathrm{trt}} = 1$ if the patient was on 4-HPR, where $I_{\mathrm{rec}} = 0$ if the patient had a new lesion and $I_{\mathrm{rec}} = 1$ if the patient had a recurrent lesion, and where $I_{\mathrm{CIN\,2}} = 0$ if the patient had CIN 3 and $I_{\mathrm{CIN\,2}} = 1$ if the patient had CIN 2.

We assumed that patients with CIN 2 and CIN 3 would be equally represented in the trial, as would patients with new and recurrent lesions. Initially, we assumed that recurrence had no effect on the probability of response, that is, that $\beta_2 = 0$. We took as our null hypothesis $\beta_0 = -2.19$, $\beta_1 = 0$, and $\beta_3 = 1.35$. We took as our alternate hypothesis that $\beta_1 = 1.79$, with β_0 and β_3 unchanged. These values produced the following probabilities of response at 6 months under the null hypothesis: for placebo, a regression rate of 0.10 for CIN 3 and 0.30 for CIN 2, and for 4-HPR, a response rate of 0.40 for CIN 3 and 0.72 for CIN2. Under these assumptions, testing with a power of 80% required a total sample size of 57 patients, and a power of 90% required 75 patients. To allow for the possibility that recurrent lesions could be more resistant than new lesions, we also considered the case in which $\beta_2 = -0.747$; this yielded the following probabilities of response under the alternate hypothesis: a regression rate with placebo of 0.05 for CIN 3 and 0.17 for CIN2, and a response rate of 0.24 with 4-HPR for CIN3 and 0.55 for CIN 2. Under these assumptions, a power of 80% required a total sample size of 64 patients, and a power of 90% required 84 patients. Permission was requested to treat up to 120 patients because of all of the assumptions made. An interim analysis was planned after 40, 60, and 90 patients or if requested by the NCI, the funding agency. The NCI requested an interim

Table 1 Randomized clinical trial participants and distribution by treatment arm and histological end point

Treatment group	Patients randomized [crossover at 6 mo]	Patients not seen after baseline]	Total patients with 6-mo histological end point [total considering crossover]	Total patients with 12-mo histological end point [total considering crossover]
Total	39 [2]	3	36 [36]	30 [30]
4-HPR		2	20 [22]	14 [16]
Placebo		1	16 [14]	16 [14]

Table 2 Histological response to 4-HPR and placebo at 6 and 12 months[a]

Histological end point	Progressive disease	No change	Partial response	Complete response	No return after baseline	Total patients in study	Total patients evaluable
6-mo	4	20	6	6	3	39	36
4-HPR	2	13	4	1	2	22	20
Placebo	2	7	2	5	1	17	16
12-mo	3	17	4	6	3	33	30
4-HPR	2	10	1	1	2	16	14
Placebo	1	7	3	5	1	17	16

[a] This Table does not reflect the crossover of two patients, which is reflected in Fig 2.

Table 3 Serum 4-HPR, 4-MPR, and retinol levels for trial participants at 6-month visit

Treatment arm	4-HPR (ng/ml) Mean (95% CI)[a]	4-MPR (ng/ml) Mean (95% CI)	Retinol (ng/ml) Mean (95% CI)
4-HPR	34.8 (22.4–47.3)	53.1 (37.3–68.9)	453.3 (406.9–499.7)
Placebo	0	0	615.4 (576.6–654.1)

[a] CI, confidence interval.

analysis after 39 patients were accrued to assess efficacy and toxicity.

Statistical Analysis. The sample size for the study was based on the histological end point of regression at 6 months. All of the recorded events were included in the analysis, regardless of treatment duration and compliance levels, according to an intention-to-treat principle. We also conducted analyses that considered the effects of crossover. The differences in response proportions and their 95% confidence intervals were calculated for both arms of the trial.

The differences in response proportions and their 95% confidence intervals were calculated for three end points: the primary end points, which were actual 6-month and 12-month histological responses, and the secondary end points, which were 6-month histological response and 12-month histological responses if 4-HPR proved effective in the two patients randomized to 4-HPR and lost to follow-up, and 6-month and 12-month histological responses if 4-HPR proved ineffective in the two patients randomized and lost to follow-up.

χ^2 tests were applied to the categories "no change," "progressive disease," "partial response," and "complete response." Patients were subsequently reclassified as "responders" if there was a partial or complete response or as "nonresponders" if there was progressive disease or no change.

Patients were further stratified by age, CIN grade (grade 3 *versus* 2), HPV status by *in situ* hybridization or the ViraPap test (positive *versus* negative), lesion size ($>2/3$ *versus* $<2/3$), OCP use (user *versus* nonuser), number of sexual partners (>1 partner *versus* 1 partner), smoking (smoker *versus* nonsmoker), and history of prior treatment for a STD (positive history *versus* negative history). χ^2 tests were used to compare the distribution of these confounders between treated and placebo arms. The *t* test was used to compare the continuous variable age. All of the tests were two-sided at the 5% level of significance. Computations were made using StatView,[4] StatXact,[5] Mathematica,[6] and SAS[7] software programs. An analysis of odds ratios for confounders was carried out using StatView. A logistic regression was carried out using SAS. The variable STD was rejected by SAS because of the small number of observations. LogXact[8] was then used to carry out the logistic regression analysis of this single variable.

Using the prior probabilities of response and nonresponse, the probability of being CIN 2 or 3, and the probability of being randomized to treatment or placebo, a futility analysis was conducted using Mathematica.[6] Assuming uniform prior distributions, posterior β distributions for the probability of response for patients on placebo and 4-HPR in patients with CIN 2 and CIN 3 were derived from the results of the study. The posterior β distribution for the probability that a patient would be CIN 3 was also derived from study results. These posteriors were then used as the priors for 1000 simulations of the remaining portion of the trial. For each simulated outcome, we computed the probability of the difference-in-response of proportions between treated and control patients. We computed the difference was at least 0 and the probability that the difference was at least 0.30. In addition to the above Bayesian analysis, we also analyzed the

[4] StatView, version 5.0, 1998; SAS, Cary, NC.
[5] StatXact, version 3.0, 1996; Cytel Software Corp, Cambridge, MA.
[6] Mathematica, version 4.0, 1999; Wolfram Research, Champaign, IL.
[7] SAS, version 6.12, 1997; SAS, Cary, NC.
[8] LogXact, version 1.0, 1992; Cytel Software Corp., Cambridge, MA.

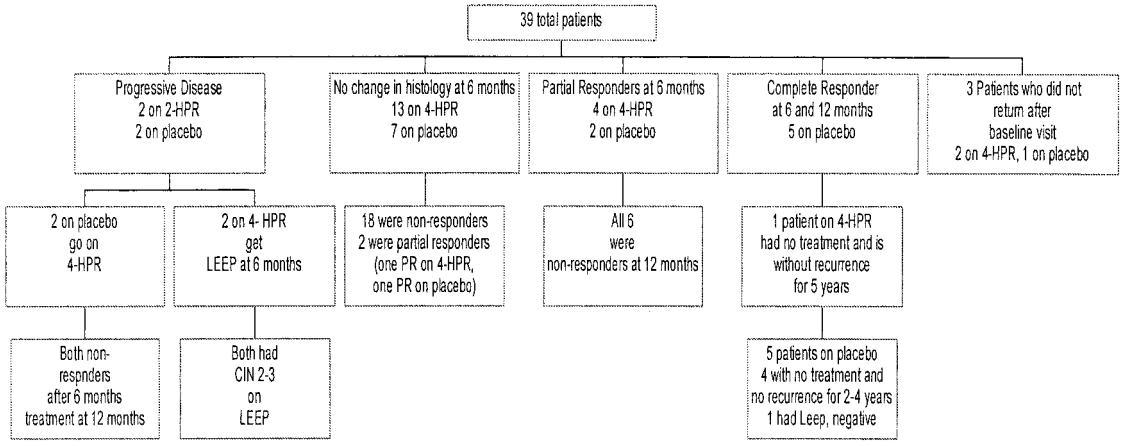

Fig. 2 Schema of patient outcomes for all of the patients in the randomized clinical trial of 4-HPR and placebo; *PR*, partial responder.

Table 4 Response rates at 6 and 12 months with patients classified as responders and nonresponders[a] ($n = 36$)

End point	Treatment group	
	Placebo	4-HPR
6-mo responders	7	5
6-mo nonresponders	9	15
Response rate	44%	25%
6-mo proportionate difference (95% confidence interval)		0.19 (−0.13–0.49) $P = 0.25$
12-mo responders	8	2
12-mo nonresponders	8	12
Response rate	50%	14%
12-mo proportionate difference (95% confidence interval)		0.36 (0.2–0.65) $P = 0.04$

[a] Not considering crossover.

Table 5 Intent-to-treat analysis of histological response rates at 6 and 12 months, assuming that the outcome in dropouts favors 4-HPR[a] ($n = 39$)

End point	Treatment group	
	Placebo	4-HPR
6-mo responders	7	7
6-mo nonresponders	10	15
Response rate	41%	32%
6-mo proportionate difference (95% confidence interval)		0.09 (−0.21–0.39) $P = 0.56$
12-mo responders	8	4
12-mo nonresponders	9	12
Response rate	47%	25%
12-mo proportionate difference (95% confidence interval)		0.22 (−0.11–0.53) $P = 0.18$

[a] Two of the three unknown patients were randomized to 4-HPR.

results of the simulations using standard frequentist methods. The results of each simulation were tested using Fisher's exact test whenever at least one expected value was <5 and using the standard χ^2 analysis with Yates' continuity correction otherwise. The differences in response rates were considered significant if $P \leq 0.05$.

RESULTS

Demographic Data and Intention to Participate. All of the eligible patients were interviewed and invited to participate. A total of 327 patients from all of the three study sites were considered eligible based on biopsy. Further review excluded 100 patients for reasons such as a positive endocervical curettage, inadequate lesion size, past history of multiple missed appointments with two "return request receipt" letters, possible invasive cancer according to Pap smear, history of a previous cancer, current pregnancy, or pathological review not consistent with CIN 2 or 3. Of the 227 remaining patients, 39 (17%) had agreed to participate by the time of this interim analysis, leaving 188 patients not participating.

All of the patients who were eligible were asked to state in their own words why they chose or chose not to participate. Those who agreed to participate reported that they did so because of several motivations: a family history of cancer, desire to help future generations, fear of eventual surgery and desire to avoid surgery if possible, hope that the medication would work and further therapy would not be needed, appeal of a less invasive treatment, and hope that 4-HPR would "help cells in other parts of the body." Of the 188 patients who chose not to participate in the trial, 8 (4%) gave four reasons for nonparticipation, 30 (16%) gave only three reasons, 93 (49%) gave only two reasons, and 57 (30%) gave one reason. The reasons for nonparticipation varied, but the most frequently mentioned were: desire for definitive treatment as soon as possible; inability to schedule multiple visits; family opposition to patient participating in research; fear of developing cervical cancer;

Table 6 Intent-to-treat analysis of histological response rates at 6 and 12 months, assuming that the outcome in dropouts does not favor 4-HPR[a] ($n = 39$)

End point	Treatment group	
	Placebo	4-HPR
6-mo responders	8	5
6-mo nonresponders	9	17
Response rate	47%	23%
6-mo proportionate difference (95% confidence interval)		0.24 (−0.06–0.53) $P = 0.12$
12-mo responders	9	2
12-mo nonresponders	8	14
Response rate	53%	13%
12-mo proportionate difference (95% confidence interval)		0.40 (0.09–0.68) $P = 0.01$

[a] Two of the three unknown patients were randomized to 4-HPR.

Table 7 Possible confounders and their distribution by treatment arm

Confounder	Placebo	4-HPR	Fisher's exact test result, *P*
Age			
≤35 yr	15/17	21/22	0.57
≥35 yr	2/17	1/22	
Histology			
CIN 2	9/17	4/22	0.04
CIN 3	8/17	18/22	
Lesion size			
=1/3	6/13	3/17	0.12
>1/3	7/13	14/17	
HPV			
Negative	4/13	8/15	0.27
Positive	9/13	7/15	
History of STD			
No	11/17	15/22	>0.99
Yes	6/17	7/22	
Number of partners			
1	3/17	4/19	>0.99
>1	14/17	15/19	
Smoke			
No	14/17	10/22	0.02
Yes	3/17	12/22	
OCP use			
No	8/17	17/22	0.09
Yes	9/17	5/22	

need for child care during clinic visits; and a "schedule too busy for this type of study."

Thirty-nine patients were randomized before the interim analysis (Table 1). The interim analysis showed no statistically significant differences at 6 months, but at 12 months there were statistically significant differences in the two groups; therefore, at the request of NCI, the code was broken, and the study was unblinded. Of the 39 patients randomized, 22 patients received 4-HPR and 17 received placebo. During the study course, 35 of 39 patients were referred to M. D. Anderson Cancer Center for the study, including patients from the UTHSC Hospital and LBJ Hospital. During the final year of the study, four patients were seen and evaluated at the LBJ Hospital.

Table 8 Odds ratios for confounders for response using intention-to-treat

Confounder	Odds ratio for response	95% confidence interval
Age, >35 yr	0.89	(0.07–10.7)
CIN 3	0.85	(0.21–3.36)
Size, >1/3 of cervix	0.25	(0.04–1.3)
HPV positive	0.63	(0.13–3.03)
History of STD	1.18	(0.29–4.7)
Number of partners, >1	0.70	(1.13–3.8)
Smoking	0.51	(0.13–2.07)
OCP use	0.35	(0.08–1.56)

The 39 patients ranged in age from 18–42 years, with a mean of 26 years. All of the patients were premenopausal. There were 19 Caucasian, 15 Latin-American, 4 African-American, and 1 Native-American participant. Thirteen patients had CIN 2, and 26 had CIN 3. Only 2 of 39 patients had received prior therapy to the cervix, specifically cryotherapy in one case and loop excision in the other; both had identifiable lesions.

At the 6-month point, 36 patients were considered evaluable; the 3 patients considered inevaluable failed to return after the baseline exam (Tables 1 and 2). Thirty of 36 patients completed the 12-month study as intended. Six patients took the medication for less than 6 full months but underwent loop excision and had histopathological results available and, therefore, were evaluable. Two of these 6 did not finish the full 6 months because they became pregnant, despite monthly counseling by the research nurse. Of the two pregnant patients, one was on placebo and had stopped the placebo one month before conception. The other patient was on 4-HPR and had stopped taking the drug 6 weeks before conception. The pregnant patients continued to be seen by our team and outcomes were known. Both infants were normal and healthy and both women were treated with loop electrosurgical excision of the cervix after normal vaginal deliveries of the infants.

Compliance and 4-HPR Levels. Compliance was measured by pill count and verified with 4-HPR levels. Most patients (28 of 30) took more than 80% of the pills prescribed. One patient who was crossed over took 58 and 81% of the two 3-month courses of placebo, and, after being crossed over for progressive disease, she took 73 and 83% of the 3-month courses of 4-HPR. A second patient on placebo took 61% of the first 3-month course of placebo and 80% of the second 3-month course of placebo. The 4-HPR, 4-MPR, and retinol levels are presented in Table 3. These levels confirm that patients on the 4-HPR arm had measurable levels of 4-HPR and its metabolite, 4-MPR. Retinol levels in patients treated with 4-HPR decreased as expected, compared with patients on the placebo arm.

Histological Response. The histological responses of the patients by treatment arm at and 6 and 12 months are summarized in Table 2. No patient had progression to microinvasive or invasive cancer. Fig. 2 is a schema of end-of-study evaluation, which permits a view of responders and nonresponders.

Responders included both complete and partial responders. Of the 6 patients who were complete responders at 6 months,

Table 9 Logistic regression using arm (4-HPR *versus* placebo) as the dependent variable

Dependent variable	Variable	Additional variable	Degree of freedom	Parameter estimate for arm	SE	Wald χ^2	*P*	Odds ratio
6 months, intention to treat								
	4-HPR arm	None	1	−0.85	0.72	1.38	0.24	0.43
	4-HPR arm	Grade	1	−0.91	0.78	1.35	0.245	0.40
	4-HPR arm	HPV	1	−0.35	0.82	0.18	0.671	0.70
	4-HPR arm	Lesion size	1	0.00	0.86	0.00	1.00	1.00
	4-HPR arm	OCP	1	−1.17	0.79	2.20	0.138	0.31
	4-HPR arm	Smoke	1	−0.76	0.79	0.93	0.334	0.47
	4-HPR arm	No. of partners	1	−0.82	0.75	1.19	0.276	0.44
	4-HPR arm	STD	1	−0.84	0.72	1.36	0.244	0.43
6 months, considering crossover								
	4-HPR arm	None	1	−0.41	0.78	0.27	0.603	0.67
	4-HPR arm	Grade	1	−0.55	0.81	0.46	0.499	0.58
	4-HPR arm	HPV	1	−0.35	0.82	0.18	0.671	0.70
	4-HPR arm	Lesion size	1	0.00	0.86	0.00	1.00	1.00
	4-HPR arm	OCP	1	−0.61	0.84	0.52	0.469	0.55
	4-HPR arm	Smoke	1	−0.29	0.86	0.11	0.737	0.75
	4-HPR arm	No. of partners	1	−0.30	0.82	0.14	0.713	0.74
	4-HPR arm	STD	1	−0.35	0.79	0.19	0.660	0.71
12 months, intention to treat								
	4-HPR arm	None	1	−2.48	0.94	6.93	0.009	0.08
	4-HPR arm	Grade	1	−3.00	1.08	7.70	0.006	0.05
	4-HPR arm	HPV	1	−2.24	0.99	5.13	0.024	0.11
	4-HPR arm	Lesion size	1	−2.50	0.99	6.39	0.012	0.08
	4-HPR arm	OCP	1	−3.47	1.28	7.31	0.007	0.03
	4-HPR arm	Smoke	1	−3.03	1.23	6.11	0.013	0.05
	4-HPR arm	No. of partners	1	−2.36	0.96	6.00	0.014	0.10
	4-HPR arm	STD	1	−2.72	[a]	[a]	0.008	0.07
12 months, considering crossover								
	4-HPR arm	None	1	−2.71	0.97	7.79	0.005	0.07
	4-HPR arm	Grade	1	−3.05	1.07	8.04	0.005	0.05
	4-HPR arm	HPV	1	−2.39	1.00	5.74	0.017	0.09
	4-HPR arm	Lesion size	1	−2.64	0.99	7.12	0.008	0.07
	4-HPR arm	OCP	1	−3.51	1.28	7.59	0.006	0.03
	4-HPR arm	Smoke	1	−3.15	1.22	6.66	0.010	0.04
	4-HPR arm	No. of partners	1	−2.62	1.00	6.86	0.009	0.07
	4-HPR arm	STD	1	−2.78	[a]	[a]	0.007	0.06

[a] These data were derived from LogXact, which does not produce a SE or the Wald χ^2.

four received placebo and no excisional treatment and remained free of disease for 4–5 years. A fifth patient, who received placebo, had a slight regression in a colposcopy-detected lesion. She underwent loop excision, but the specimen was negative on histological examination of multiple, serial sections. The sixth patient, who received 4-HPR and serial follow-up using colposcopy and colposcopy-directed biopsy, was disease free for 5 years and remained so at data analysis. There were six partial responders at 6 months, all of whom underwent loop excision at the 12-month follow-up and were classified as nonresponders.

Nonresponders included those patients with no change or progressive disease in biopsy. All 20 patients in whom there was no change, according to histological examination of the biopsy taken at 6 months, underwent loop excision at 12 months; 18 proved to be nonresponders, and 2 proved to be partial responders (1 patient on 4-HPR and 1 on placebo). For 4 patients with progressive disease at 6 months, the code was broken at 6 months. Two who were on the 4-HPR arm underwent loop excision and proved to be nonresponders; the two on the placebo arm were crossed over to the 4-HPR arm, eventually underwent loop excision, and were also classified as nonresponders.

The response rates, their proportionate differences, confidence intervals, and *P*s are summarized in Table 4. There were two end points, one before crossover and one after, adding two additional patients to the 4-HPR arm. At the 6-month end point, there were no statistically significant differences in responders *versus* nonresponders, whether crossover was considered or not (*P*, not significant), although the placebo arm fared better than the 4-HPR arm. At trial end (12 months), there was a significantly greater response rate in the placebo group, whether crossover was considered or not.

Tables 5 and 6 show an "Intent to Treat" analysis that includes the three patients who dropped out of the study. As in Fig. 2, two of the three patients who dropped out had been randomized to 4-HPR. Table 5 assumes a favorable response to 4-HPR in those two of three and an unfavorable response in the patient randomized to placebo. Table 6 assumes a negative response to 4-HPR in the two patients and a positive response in the patient in the placebo arm. These results indicate that had these patients stayed enrolled, the overall outcomes of the study at 6 and 12 months would not have been different (for 6 months, *P*s = 0.56, assuming favorable response to 4-HPR, and $P = 0.12$ for unfavorable response to 4-HPR; for 12 months, $P = 0.18$ assuming a favorable response to 4-HPR, and $P = 0.01$ for

Table 10 Reported toxicities using the NCI toxicity scale[a]

	4-HPR grade		Placebo grade	
Toxicity	1	2	1	2
Alopecia	1	0	2	0
Anisocoria	1	0	0	0
Anorexia	0	0	2	0
Cheilitis	3	0	0	0
Conjunctivitis	1	0	0	0
Diarrhea	1	0	1	1
Dry eyes	1	0	0	0
Dry nasal mucosa	1	1	0	0
Dry skin	1	0	1	0
Fatigue	1	0	1	0
Headache	0	0	1	0
Insomnia	0	0	1	0
Mood	1	0	0	0
Nausea alone	1	0	2	0
Nyctalopia	1	0	0	0
Peeling finger tip	1	1	0	0
Photosensitivity	1	2	1	0
Skin rash	3	0	0	2
Skin reaction	1	0	1	0
Stomatitis	0	0	2	1
Totals	20	4	15	4

[a] There were no instances of grade 3–5 toxicity in either arm.

Table 11 Reported toxicity summary

Agent	Grade 1 toxicities	Grade 2 toxicities
4-HPR	20/22 (90.9%)	4/22 (18.2%)
Placebo	15/17 (88.2%)	4/17 (23.5%)

an unfavorable response). Although the *P* at 12 months for the assumption of a favorable response lost significance, the response rates still overwhelmingly favored the placebo group (47 *versus* 25%).

Of the potential principal confounders of histological response, only two showed significant differences, histology, and smoking history (Table 7). More patients in the 4-HPR arm had CIN 3 (82%) than in the placebo arm (48%). The importance of this variable was examined in logistic regression and found not to be significant. Additionally, 3 of 17 placebo patients were smokers, compared with 12 of 22 4-HPR patients. The importance of this variable was also examined in logistic regression and found not to be significant.

The odds ratios for these variables and their confidence intervals are presented in Table 8. Neither CIN 3 histology nor smoking were statistically significant predictors of response. These results were further confirmed by a logistic regression in which response at 6 and 12 months, considering both intention to treat and crossover, was fit as a function of treatment arm (Table 9). The results show that for both 6- and 12-month response rates, whether analyzed by intention to treat or crossover, the parameter estimates were negative. Although at 6 months, none were statistically significant, at 12 months, randomization to the treatment arm predicted no-response with statistical significance. Thus, treatment with 4-HPR was associated with a significantly lower probability of response at 12 months. If some other predictor had accounted for the lower response rate, then when fitting the logistic regression, the other predictor would have resulted in an insignificant estimate for the treatment arm. Because all of the estimates remained significant at 12 months, the negative association between the treatment arm and the response cannot be explained by any of the other variables in the Table.

Futility Analysis. The results of the Bayesian futility analysis show that, based on 1000 simulations, there was a less than 1/1000 chance of seeing a statistically significant difference in the response rate given the probability of response noted at 6 or 12 months. The results of the frequentist futility analysis were similar, showing that there was a 0/1000 chance of 4-HPR being more effective than placebo, 974/1000 chance that placebo was more effective than 4-HPR, and 26/1000 chance that there would be no difference between them.

Colpophotographs and Pap Smears. In the 36 evaluable patients, colpophotography demonstrated lesion regression in 11 (6 in the 4-HPR group and 5 in the placebo group). These assessments did not correspond to actual lesion response as assessed histologically, and, thus, colpophotographs were not useful in the study for monitoring response. Biopsies taken at the site of the original-study biopsy proved positive despite the colpophotography-demonstrated reduction in lesion size.

The Pap smear was a grade or two lower than the histological diagnosis in 36% of the patients, a grade higher in 17% of the patients, and consistent in the remaining 47% of the patients. The Pap smears were, thus, not helpful in assessing response.

Toxicities. All of the patients were compliant with the laboratory, bone density, and eye testing requirements. The serum chemistries and blood counts were run at M. D. Anderson for 37 of the 39 patients and at the LBJ Hospital for 2 patients. All of the bone density testing was performed at M. D. Anderson by the Department of Nuclear Medicine. All of the eye testing was performed by one collaborator at the UTHSC Department of Ophthalmology. The initially planned electroretinagrams were tolerated by few patients; all of the patients tolerated the nyctalopia testing, which was completed in all of the patients.

Toxicity reported during the trial was minimal. Results are presented in Table 10 and a comparison of the grade 1 and 2 toxicities in the two arms is presented in Table 11. No patients stopped taking the drug because of toxicity. There were no statistically significant differences between grade 1 and 2 toxicities among the placebo and 4-HPR groups.

The laboratory, bone density, and nyctalopia testing detected many unsuspected abnormalities (Table 12). Baseline abnormalities included anemia, hypercholesteremia, hypertriglyceridemia, increased liver enzymes, and decreased bone density. These were unexpected in a population of young women, and their discovery led to intervention after the trial.

Table 12 Number of patients with abnormalities in results of laboratory, bone density, and eye testing

	4-HPR baseline	Placebo baseline	4-HPR 6 mo	Placebo 6 mo	4-HPR 12 mo	4-HPR 12 mo
Blood chemistries						
Elevated liver enzymes	0	1	2	0	0	0
Elevated lipid profiles	2	3	0	1	1	0
Anemia	1	1	3	3	0	0
Bone densities, osteopenia	0	2	2	0	0	1
Eye testing, abnormal eye testing	0	0	0	0	0	0

Table 13 Concentration of 4-HPR from cell culture experiments recalculated to compare with plasma or tissue levels of 4-HPR

Concentration used in cell culture in the laboratory (μmol)	Comparable plasma or tissue level (ng/ml)
1.0	400
2.0	800
3.0	1200
5.0	2000
10.0	4000

DISCUSSION

Despite the promising laboratory and animal studies demonstrating that 4-HPR is a potent inhibitor of carcinogenesis, this is the third negative clinical trial of oral 4-HPR at the dose level of 200 mg/day with a 3-day/month drug holiday (14, 15).

In one breast cancer prevention study (14), 2972 Italian women, 30–70 years old, with stage I breast cancer or ductal CIS, were randomized to receive 4-HPR or follow-up for 5 years. The dose of 4-HPR was 200 mg/day with a 3-day drug holiday monthly. The primary end points were reduction in cancer incidence in the contralateral breast and the ipsilateral breast 7 years after randomization. At a median observation time of 97 months, there were no statistically significant differences between the groups in the primary outcomes of interest. *Post hoc* analyses showed that premenopausal women may benefit from taking 4-HPR, with decreases in contralateral but not ipsilateral breast cancers. The analyses of postmenopausal women did not show a benefit.

Kurie *et al.* (15) reported no histological regression in 82 smokers enrolled in a randomized placebo-controlled trial of 4-HPR at 200 mg/day with a 3-day/month drug holiday. Patients were treated for 6 months. 4-HPR had no measurable effect on histopathology despite reasonable serum 4-HPR levels. Several biomarkers were also not found to be modulated by 4-HPR in this trial, including retinoic acid receptor β and loss of heterozygosity of 3p, 9p, and 17p. Kurie *et al.* concluded that 4-HPR was not effective in reversing squamous metaplasia, dysplasia, or genetic and phenotypic abnormalities in bronchial epithelium.

In our study, 4-HPR at 200 mg/day for 6 months, did not appear to be more active than placebo in CIN-2 or -3 lesions. Although there were no significant differences in histological regression at 6 months between the 4-HPR and placebo arms, the proportionate difference favored placebo. The 12-month histological response rates significantly favored the placebo group. Although sample sizes were small, it seemed best to stop the trial, because the medication was not efficacious. The analysis of several confounders showed that none affected response. The futility analysis showed that there was very little likelihood that continuation of the trial would show histological regression in patients treated with 4-HPR.

In retrospect, perhaps the oral dose of 200 mg/day is too low. A comparison of concentrations of 4-HPR used in the laboratory to suppress cell growth (in micromoles) with plasma and breast levels of 4-HPR (in ng/ml) is of interest for dosage evaluation. The molecular weight of 4-HPR is ~400 g/mol. The formula for the conversion of concentration used in cell culture to comparable levels in the plasma is concentration in μmol/l = drug concentration in μg/ml × 1000 divided by the molecular weight of the drug in g/mol (16). Table 13 shows the expected plasma or tissue concentrations that correspond to the concentrations of 1, 3, 5, and 10 micromolar 4-HPR that were used in the laboratory. Because the 5- and 10-micromolar concentrations were the most effective in suppressing cell growth and driving cells into apoptosis and because the levels in the breast measured by Formelli *et al.* (17) approached the 5-micromolar concentration, the dosage of 200 mg/day for breast chemoprevention is probably reasonable. Because of ocular toxicity, the NCI set the dose for all of the Phase II studies in the cervix, oral cavity, lung, prostate, and bladder at 200 mg/day (1, 5). Currently, no investigators have published tissue levels from these organs other than breast. Because vitamin A is a fat-soluble vitamin, one might expect higher levels in fatty tissue like the breast than in nonfatty tissues like those mentioned. Formelli *et al.* (17) reported a mean serum level of 360 ng/ml on a 200-mg/day dose with a 3-day drug holiday; this level corresponds to the 1-μmol laboratory dose. Khuri is conducting a Phase 1 trial of 4-HPR in lung cancers and has shown that that doses greater than 1 gram can be tolerated.[9]

Our trial yielded important information about participation, histological regression, and toxicity in an age group younger than those women who participated in the trial reported by Veronesi *et al.* (14) or in the trial reported by Kurie *et al.* (15). Although many patients were eligible for the study at the three sites, only 17% agreed to participate. This participation was less than expected at trial outset, and our group is exploring the reasons for the low rate of participation. Fear of waiting for treatment emerged as an important reason to decline participation. Equally important were the multiple visits required by this trial. Chemoprevention trials and other clinical trials of HGSIL

[9] F. Khuri, personal communication.

may need to be designed to accommodate the patient's need for fewer visits, transportation, and child care.

There were no serious toxic reactions in these patients. In fact, many patients benefited from participating in the trial. Those patients with anemia and elevated lipid profiles received interventions that would benefit their long-term health. Similarly, those patients with decreased bone densities were referred after the trial for intervention. We did not anticipate finding bone density abnormalities in women this young. Finally, we saw no eye toxicity from 6 months of 4-HPR in these young women.

4-HPR is a very promising medication in the laboratory and deserves additional testing at higher dose levels in clinical trials. Nevertheless, this interim analysis suggests that the final results may show no significant treatment effects. Laboratory and animal data can be used to guide the trial, but tissue levels in each organ site need to be examined. Ultimately, toxicity plays a role in the use of any medication but particularly in chemoprevention trials. The biomarker analyses that are under way may prove helpful in sorting out the negative effects at the 12-month end point and ensuring that the 12- and 6-month effects are consistent. The concepts of Phase I study design need to be incorporated into the trial design for each organ site under study. The field of chemoprevention and the use of surrogate end point biomarkers in trials are evolving and may provide women with new choices of therapy in the future.

REFERENCES

1. Veronesi, U., De Palo, G., Costa, A., Formelli, F., and Decensi, A. Chemoprevention of breast cancer with fenretinide. IARC Sci. Publ., *136:* 87–94, 1996.

2. Moon, R. C., Thompson, H. J., Becci, P. J., Grubbs, C. J., Gander, R. J., Newton, D. L., Smith, J. M., Phillips, S. L., Henderson, W. R., Muller, L. T., Brown, C. C., and Sporn, M. B. *N*-(4-hydroxyphenyl) retinamide, a new retinoid for prevention of breast cancer in the rat. Cancer Res., *39:* 1339–1346, 1979.

3. Paulson, J. D., Oldham, J. W., Preston, R. F., and Newman, D. Lack of genotoxicity of the cancer chemopreventive agent *N*-(4-hydroxyphenyl) retinamide. Fundam. Appl. Toxicol., *5:* 144–150, 1995.

4. Formelli, F., Barua, A. B., and Olson, J. A. Bioactivities of *N*-(4-hydroxyphenyl) retinamide and retinoyl β-glucuronide. FASEB J., *10:* 1014–1024, 1996.

5. Kelloff, G. J., Boone, C. W., Crowell, J. A., Steele, V. E., Lubet, R. A., Doody, L. A., Malone, W. F., Hawk E. T., and Sigman, C. C. New agents for cancer chemoprevention. J. Cell. Biochem., *26S:* 1–28, 1996.

6. Hong, W. K., and Itri, L. M. Retinoids and human cancer. *In:* M. B. Sporn, A. B. Roberts, D. S. Goodman (eds.), The Retinoids, Ed. 2, p. 597. New York: Raven Press, 1994.

7. Lippman, S. M., Batsakis, J. G., Toth, B. B., Weber, R. S., Lee, J. J., Martin, J. W., Hays G. L., Goepfert, H., and Hong, W. K. Comparison of low-dose isotretinoin with β carotene to prevent oral carcinogenesis. N. Engl. J. Med., *328:* 15–20, 1993.

8. Lippman, S. M., Kessler, J. F., and Meyskens, F. L., Jr. Retinoids as preventive and therapeutic anticancer agents. Cancer Treat. Rep, *71:* 391–405, 493–515, 1987.

9. Hong, W. K., Lippman, S. M., Itri, L. M., Karp, D. D., Lee, J. S., Byers, R. M., Schantz, S. P., Kramer, A. M., Lotan, R., and Peters, L. J. Prevention of second primary tumors with isotretinoin in squamous-cell carcinoma of the head and neck. N. Engl. J. Med., *323:* 795–801, 1990.

10. Oridate, N., Suzuki, S., Higuchi, M., Mitchell, M. F., Hong, W. K., and Lotan, R. Involvement of reactive oxygen species in *N*-(4-hydroxyphenyl)retinamide-induced apoptosis in cervical carcinoma cells. J. Natl. Cancer Inst. (Bethesda), *89:* 1191–1198, 1997.

11. Dmitrovsky, E. *N*-(4-hydroxyphenyl) retinamide activation of a distinct pathway signaling apoptosis. J. Natl. Cancer Inst. (Bethesda), *89:* 1179–1181, 1997.

12. Sun, S. Y., Li, W., Yue, P., Lippman, S. M., Hong, W. K., and Lotan, R. Mediation of *N*-(4-hydoxyphenyl)retinamide-induced apoptosis in human cancer cells by different mechanisms. Cancer Res., *59:* 2493–2498, 1999.

13. Self, S. G., Mauritsen, R. H., and Ohara, J. Power calculations for likelihood ratio tests in generalized linear models. Biometrics, *48:* 31–39, 1992.

14. Veronesi, U., De Palo, G., Marubini, E., Costa, A., Formelli, F., Mariani, L., Decensi, A., Camerini, T., Del Turco, M. R., Di Mauro, M. G., Muraca, M. G., Del Vecchio, M., Pinto, C., D'Aiuto, G., Boni, C., Campa, T., Magni, A., Micheli, R., Perloff, M., Malone, W. F., and Sporn, M. B. Randomized trial of fenretinide to prevent second breast malignancy in women with early breast cancer. J. Natl. Cancer Inst. (Bethesda), *91:* 1847–1856, 1999.

15. Kurie, J. M., Lee, J. S., Khuri, F. R., Mao, L., Morice, J. C., Lee, J. J., Walsh, G. L., Broxson, A., Lippman, S. M., Ro, J. Y., Kemp, B. L., Liu, D., Fritsche, H. A., Xu, X., Lotan, R., and Hong, W. K. *N*-(hydroxyphenyl)retinamide in the chemoprevention of squamous metaplasia and dysplasia of bronchial epithelium. Clin. Cancer Res., *6:* 2973–2979, 2000.

16. Anderson, P. O., Knoben, J. E., and Troutman, W. G. Conversion factors. *In:* Handbook of Clinical Drug Data, Ed. 9, p. 1009. Stanford, CT: Appleton & Lange, 1999.

17. Formelli, F., Clerici, M., Campa, T., Di Mauro, M. G., Magni, A., Mascotti, G., Moglia, D., De Palo, G., Costa, A., and Veronesi, U. Five-year administration of fenretinide: pharmacokinetics and effects on plasma retinol concentrations. J. Clin. Oncol., *11:* 2036–2042, 1993.