

Why Most Randomized Phase II Cervical Cancer Chemoprevention Trials are Uninformative: Lessons for the Future

Michele Follen, Frank L. Meyskens, Jr., E. Neely Atkinson, David Schottenfeld

According to the worldwide cancer incidence database maintained by the World Health Organization, cervical cancer is the third most common malignancy in women worldwide, exceeded in incidence only by breast and colorectal cancers (1). Cervical cancer is an important cause of mortality in women worldwide, and the cervix is a well-established clinical and histopathologic model of carcinogenesis. Human papillomavirus (HPV) is the major etiologic agent in cervical cancer. The cervix is easily accessible for examination, and colposcopy provides a visual model of angiogenesis and tumor development, making the cervix a good model for preventive interventions (2).

The precursor lesions to cervical cancer are squamous intraepithelial lesions (SILs) or cervical intraepithelial neoplasia (CIN). SILs are used in the cytologic classification, and CINs are used in the histopathologic classification. Several chemoprevention trials have been conducted in women with SIL/CIN by use of retinoids, micronutrients, α -difluoromethylornithine, and indole-3-carbinol. Although the results of phase I/IIa studies have appeared to be promising, the phase IIb/III studies have usually been negative (3–13) (Table 1).¹ The failure of the phase IIb/III studies to demonstrate an effect may be due to several factors not being adequately considered in the design of these trials.

Advances in the design and implementation of clinical trials have been made over the last decade. Factors that need to be considered in the design of cervical chemoprevention trials are as follows: the natural history of the disease in the absence of any intervention, the optimal range of anticipated clinical response to the randomly assigned intervention, and the validity and predictiveness of the primary (histologic regression) and secondary (surrogate endpoint biomarker modulation) outcome measures (14–16). The likelihood that phase II cervical chemoprevention trials will be uninformative can be minimized with careful attention to critical features of study design: enrollment of a sufficient number of patients to permit the study to reveal differences in response rates; careful selection of the type and dose of chemopreventive agent, based on results from preclinical studies and phase I and IIa clinical trials; accurate classification of patients' disease status, both at enrollment and at study end; and selection of appropriate primary and secondary outcome measures.

To ensure that cervical cancer chemoprevention trials have appropriate statistical power, the natural history of the disease must be taken into account. Patton wrote an extensive review of the natural history of SIL/CIN in the 1950s [reviewed in (17,18)], and two recent reviews of the natural history of SIL/CIN have also been reported (17,18). Estimated rates of regression of CIN vary widely, both within each CIN grade and among studies. The regression rate decreases as the CIN grade increases, but

estimates of spontaneous regression based on randomized trials vary from 27% to 66%, depending on the study and the entry criteria (Table 1). Apparent spontaneous regression of CIN has been observed from the natural history studies, from 57% regression of CIN1 and 43% regression of CIN2 to 32% regression of CIN3 (18). Thus, with the exception of the studies by Meyskens et al. (3), de Vet et al. (8), and Childers et al. (13), many of the published phase IIb trials did not have sufficient power to detect a clinically significant difference in response rate.

The dose and duration of medications used in chemoprevention trials should be selected on the basis of results from phase I/IIa trials, which are, unfortunately, rarely performed before phase IIb/III trials in the cervix. Phase I/IIa trials preceded the trials of Meyskens et al. (3), Romney et al. (4,5), and Keefe et al. (6,7) but not de Vet et al. (8), Fairley et al. (9), Mackerras et al. (10), Butterworth et al. (11,12), or Childers et al. (13). As such, the dose of medication used may have been inactive. A phase I/IIa trial is an opportunity to carefully examine toxicity, tissue drug levels, and modulation of surrogate endpoint biomarkers. It is important that selection of the dose and duration of medication for a cervical chemoprevention trial be based on a trial in the cervix, not taken from another organ where tissue drug levels and biomarker modulation may not be the same.

Eligibility for cervical chemoprevention trials should be ascertained on the basis of colposcopically directed biopsy (15). The standard of clinical care is to perform colposcopically directed biopsies in patients in whom Pap smears are abnormal. Since the "gold standard" (or criterion standard) used for treatment is the colposcopically directed biopsy, this same standard should hold for chemoprevention trials. There are several metrics used to compare the performance of clinical tests, such as sensitivity, specificity, positive and negative predictive values, receiver operating characteristic (ROC) curve analysis, and the area under the ROC curve. Tests may perform differently in screening populations, in which the disease prevalence is low, and in diagnostic or follow-up populations, in which the disease prevalence is high. Fahey et al. (19) conducted a meta-analysis

Affiliations of authors: M. Follen (Department of Gynecologic Oncology), E. N. Atkinson (Department of Biomathematics), The University of Texas M. D. Anderson Cancer Center, Houston; F. L. Meyskens, Jr., Chao Family Comprehensive Cancer Center, University of California-Irvine; D. Schottenfeld, Department of Epidemiology, The University of Michigan, School of Public Health I, Ann Arbor.

Correspondence to: Michele Follen, M.D., Ph.D., Department of Gynecologic Oncology, Biomedical Engineering Center, Box 193, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030 (e-mail: mfollen@mdanderson.org).

See "Notes" following "References."

© Oxford University Press

Table 1. Phase IIb/III trials of cervical chemopreventives*

Investigator(s) (reference Nos.)	Retinoid studies		Micronutrient studies			
	Meyskens et al. (3)		Romney et al. (4,5)	Keefe et al. (6,7)	de Vet et al. (8)	
Study design	Phase IIb		Phase II	Phase II	Phase III	
No. randomly assigned (No. evaluable for efficacy)	301 (232)		98 (69)	124 (103)	369 (278)	
Medication, dose and timing	β -trans-RA (topical); 0.372% for 4 days at baseline and 2 days at 3 mo and 2 days at 6 mo	β -trans-RA (topical); 0.372% for 4 days at baseline and 2 days at 3 mo and 2 days at 6 mo	β -Carotene, 30 mg Placebo: lactose, 9 mo of oral medication	β -Carotene, 30 mg Placebo, 6 mo of oral medication	β -Carotene, 10 mg Placebo, 3 mo of oral medication	
Disease	CIN2	CIN3	CIN1-3	CIN2-3	CIN1-3	
Results/regression rates in patients, evaluable for efficacy						
CR	RA, 32/75 = 43%* Placebo, 18/66 = 27%	RA, 10/40 = 25% Placebo, 16/51 = 31%	β -Carotene, 9/39 = 23% Placebo, 14/30 = 47%*	β -Carotene, 47% Placebo, 47%	β -Carotene, 22/137 = 16% Placebo, 15/141 = 11%	
P †	.04	.33	.03	.69	.36	
Partial and complete responses	NA	NA	β -Carotene, 18/39 = 46% Placebo, 15/30 = 50%	NA	β -Carotene, 44/137 = 32% Placebo, 45/141 = 32%	
P †	NA	NA	NS	NA	.68	
Adequate power (0.80)?‡	Yes	Yes	No	No	Yes	
Regression defined as	CIN3 to CIN1, CIN2 to normal	CIN3 to CIN1, CIN2 to normal	Pap and biopsy negative for CR, one grade for PR	CIN3 to CIN1, CIN2 to normal	CIN3 to CIN1, CIN2 to normal	
Study entry test	Colposcopically directed biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	
Study exit test	Colposcopically directed biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	Pap or colposcopy or if progression, biopsy	

Investigator(s) (reference Nos.)	Micronutrient studies				
	Fairley et al. (9)	Mackerras et al. (10)	Butterworth et al. (11)	Butterworth et al. (12)	Childers et al. (13)
Study design	Phase II	Phase II	Phase II	Phase II	Phase III
No. randomly assigned (No. evaluable for efficacy)	117 (111)	147 (141)	78 (47)	235 (177)	331 (246)
Medication, dose and timing	β -Carotene, 30 mg Placebo (400 mg lecithin) 12 mo of oral medication	β -Carotene, 30 mg Vitamin C, 500 mg β -Carotene, 30 mg + vitamin C, 500 mg Placebo, 6 mo of oral therapy	10 mg of folate Placebo, 10 mg of vitamin C 3 mo of oral therapy	10 mg of folate Placebo, 10 mg of vitamin C 6 mo of oral therapy	5 mg of folic acid/day Placebo, 6 mo of oral therapy
Disease	Atypia-CIN2	Atypia-CIN1	CIN1-2	CIN1-2	HPV, CIN1, CIN2
Results/regression rates in patients, evaluable for efficacy					
CR	NA	β -Carotene, 16/36 = 46% Vitamin C, 9/35 = 26% Both, 8/35 = 23% Placebo, 10/35 = 29%	Folate, 3/22 = 4% Placebo, 1/25 = 4%	Folate, 58/91 = 64% Placebo, 45/86 = 66%	Folate, 9/129 = 7% Placebo, 7/117 = 6%
P †	NA	.14	NS	.23	.75
Partial and complete response	β -Carotene, 37/59 = 63% Placebo, 31/52 = 60%	NA	Folate, 8/22 = 36%* Placebo, 4/25 = 16%	NA	NA
P †	.84	NA	<.05	NA	NA
Adequate power (0.80)?‡	No	No	No	No	Yes
Regression defined as	CIN2 to HPV, CIN1 to normal	Normal or HPV on Pap and colposcopy	Normal biopsy for CR, one grade lower for PR	Normal biopsy for CR, one grade lower for PR	Pap and colposcopy, both improved
Study entry test	Pap smear	Colposcopically directed biopsy	Pap and colposcopy	Pap and colposcopy	Colposcopically directed biopsy
Study exit test	Pap smear	Pap or colposcopy of progression biopsy	Colposcopically directed biopsy	Colposcopically directed biopsy	Pap and colposcopy

*RA = retinoic acid, CIN = cervical intraepithelial neoplasia, NA = not available, NS = not significant, HPV = human papillomavirus, CR = complete response, PR = partial response.

† $P \leq .05$.

‡Adequate power, given entry diagnosis and observed regression in the placebo arm.

examining the metrics of the Pap smear by use of histology as a gold standard. Fahey et al. showed that the sensitivity of the Pap smear was 58% if used for screening and 66% if used for follow-up and that the specificity was 69% if used for screening and 66% if used for follow-up. Mitchell et al. (20,21) used the same studies and calculated an area under the ROC curve for the performance of the Pap smear of 0.70 for screening and 0.76 for follow-up or diagnosis. They performed meta-analyses of the performance of colposcopy in both the screening and diagnostic settings. Colposcopy used in the screening setting, verifying positive colposcopies only with histology and by use of the Pap smear for specificity, has a sensitivity of 86% and a specificity of 83%, with an area under the ROC curve of 0.89 (20). Colposcopy used in the diagnostic setting has a mean sensitivity of 96% and a specificity of 48%, with an area under the ROC curve of 0.84 (21). Although the performance of colposcopy used for screening is slightly higher than that for diagnosis, the differences are not statistically significant. Since in screening studies, only areas that are positive colposcopically are given a biopsy, this may account for the higher specificity of colposcopy used for screening; it may be due to the low sensitivity of the Pap smear. The highest sensitivity is achieved with a colposcopically directed biopsy. Since sensitivity and specificity are a trade-off, it is best to have high sensitivity and identify lesions accurately. Lower specificity is not detrimental to chemoprevention trials, since lesions that are false positive (positive colposcopically but not histologically) will be appropriately excluded from the trial. In summary, using the Pap smear or colposcopy without biopsy could lead to the misclassification of 15%–40% of patients at study entry (9,11,12). While biopsy may induce a slightly higher rate of regression than that observed with Pap smears (17), the increased accuracy in study entry and response ascertainment obtained with biopsy probably outweighs the negative effects of the biopsy approach. Similarly and equally important, the response needs to be ascertained by use of the same test at the study entry as at the end of the study. Using a different test at the entry and end of a study, a common practice in phase II studies reported to date (8,10–13) introduces the error of the test differences.

Choosing a relevant anticipated response is critical for the design of the trial. According to the natural history literature, CIN/SIL lesions, grades 1–3, have a regression rate of 32%–57% if followed with biopsy (18). Randomized clinical trials of treatments for SIL/CIN, grades 1–3, such as cryotherapy, laser therapy, and loop excision, demonstrated 2-year complete response rates of more than 80% (22,23). While investigators may choose any level of anticipated benefit in their studies, conventional therapy yields good cure rates, with minimal complications. A chemopreventive agent should achieve an incremental benefit over conventional therapy; a response rate of 40%–50% greater than that expected in the placebo arm is anticipated. Only five studies (3,6,7,10,12) set an anticipated level of response. The sample size for the chemoprevention trial must take into account the natural history of the lesion grade, the use of biopsy to follow patients and the anticipated regression rate associated with it, and the sensitivity of the diagnostic method for detecting the anticipated response in the treatment arm.

The primary endpoint for phase IIb/III trials should be histologic response; the secondary endpoint should be modulation of surrogate endpoint biomarkers, such as quantitative histopathologic and cytologic markers, proliferation markers, regulation

markers, differentiation markers, general genomic instability markers, and tissue maintenance markers. Histologic response should be determined by use of a sample obtained by colposcopically directed biopsy. Since the kappa for intraobserver and interobserver agreement among pathologists for the reading of cervical biopsies is in the 0.40–0.60 range (17), representing moderate to good agreement, consideration should be given to a consensus review and to quantitative histomorphometric assessment (16). Surrogate endpoint biomarkers may be useful in determining biologic responses. Before being used in a phase IIb/III trial, biomarkers should have been validated in a phase I trial. Although biomarkers will be chosen on the basis of the medication under study, some of these markers, including viral load and HPV oncoprotein expression, may be of interest in all cervical cancer chemoprevention trials (24). Preclinical laboratory work, including suppression of HPV oncoprotein expression in cell lines or prevention of HPV-induced tumors in mice, would strengthen the biologic rationale (25). A review of surrogate endpoint biomarkers and their modulation in chemoprevention trials of cervical neoplasia shows that quantitative measures of histology have been validated (26).

Cervical cancer chemoprevention studies require familiarity with clinical oncologic, molecular biologic, and epidemiologic principles. Much is currently known about the natural history and pathobiology of cervical cancer that may be incorporated into future study designs that evaluate the efficacy of chemoprevention agents.

REFERENCES

- (1) Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer* 1999;80:827–41.
- (2) Follen M. A clinical trial using 4-hydroxyphenylretinamide for the treatment of high-grade squamous intraepithelial lesions of the cervix: an interim analysis. Ann Arbor (MI): University of Michigan; 2000.
- (3) Meyskens FL Jr, Surwit E, Moon TE, Childers JM, Davis JR, Dorr RT, et al. Enhancement of regression of cervical intraepithelial neoplasia II (moderate dysplasia) with topically applied all-*trans*-retinoic acid: a randomized trial. *J Natl Cancer Inst* 1994;86:539–43.
- (4) Current Clinical Trials Oncology. National Cancer Institute PDQ. Vol 1. No. 2. March–April 1994 (protocol 09315).
- (5) Romney SL, Ho GY, Palan PR, Basu J, Kadish AS, Klein S, et al. Effects of β -carotene and other factors on outcome of cervical dysplasia and human papillomavirus infection. *Gynecol Oncol* 1997;65:483–92.
- (6) Current Clinical Trials Oncology. National Cancer Institute PDQ. Vol 1. No. 2. March–April 1994 (protocol 09088).
- (7) Keefe KA, Brewer C, Schell MJ, Chapman JA, Wilczynski SP, Meyskens FL, et al. Effects of β -carotene on cervical intraepithelial neoplasia: six-month interim results of a two-year phase II/III trial. *Proc Ann Meet Soc Gynecol Oncol*; 1998.
- (8) de Vet HC, Knipschild PG, Willebrand D, Schouten HJ, Sturmans F. The effect of beta-carotene on the regression and progression of cervical dysplasia: a clinical experiment. *J Clin Epidemiol* 1991;44:273–83.
- (9) Fairley CK, Tabrizi SN, Chen S, Baghurst P, Young H, Quinn M, et al. A randomized clinical trial of beta carotene vs placebo for the treatment of cervical HPV infection. *Int J Gynecol Cancer* 1996;6:225–30.
- (10) Mackerras D, Irwig L, Simpson JM, Weisberg E, Cardona M, Webster F, et al. Randomized double-blind trial of beta-carotene and vitamin C in women with minor cervical abnormalities. *Br J Cancer* 1999;79:1448–53.
- (11) Butterworth CE Jr, Hatch KD, Gore H, Mueller H, Krumdieck CL. Improvement in cervical dysplasia associated with folic acid therapy in users of oral contraceptives. *Am J Clin Nutr* 1982;35:73–82.
- (12) Butterworth CE Jr, Hatch KD, Soong SF, Cole P, Tamura T, Sauberlich HE, et al. Oral folic acid supplementation for cervical dysplasia: a clinical intervention trial. *Am J Obstet Gynecol* 1992;166:803–9.
- (13) Childers JM, Chu J, Voigt LF, Feigl P, Tamimi HK, Franklin EW, et al. Chemoprevention of cervical cancer with folic acid: a phase II Southwest

- Oncology Group Intergroup study. *Cancer Epidemiol Biomarkers Prev* 1995;4:155–9.
- (14) Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–9.
- (15) Mitchell MF, Hittelman WN, Lotan R, Nishioka K, Tortolero-Luna G, Richards-Kortum R, et al. Chemoprevention trials in the cervix: design, feasibility, and recruitment. *J Cell Biochem Suppl* 1995;23:104–12.
- (16) Boone CW, Kelloff GJ, Steele VE. Natural history of intraepithelial neoplasia in humans with implications for cancer chemopreventive strategy. *Cancer Res* 1992;52:1651–9.
- (17) Mitchell MF, Hittelman WN, Hong WK, Lotan R, Schottenfeld D. The natural history of cervical intraepithelial neoplasia: an argument for intermediate endpoint biomarkers. *Cancer Epidemiol Biomarkers Prev* 1994;3:619–26.
- (18) Ostor AG. Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol* 1993;12:186–92.
- (19) Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol* 1995;141:680–9.
- (20) Mitchell MF, Cantor SB, Brookner C, Utzinger U, Schottenfeld D, Richards-Kortum R. Screening for squamous intraepithelial lesions with fluorescence spectroscopy. *Obstet Gynecol* 1999;94:889–96.
- (21) Mitchell MF, Cantor SB, Ramanujam N, Tortolero-Luna G, Richards-Kortum R. Fluorescence spectroscopy for diagnosis of squamous intraepithelial lesions of the cervix. *Obstet Gynecol* 1999;93:462–70.
- (22) Mitchell MF, Tortolero-Luna G, Cook E, Whittaker L, Rhodes-Morris H, Silva E. A randomized clinical trial of cryotherapy, laser vaporization, and loop electrosurgical excision for the treatment of squamous intraepithelial lesions of the cervix. *Obstet Gynecol* 1998;92:737–44.
- (23) Follen M. Preinvasive diseases of the female lower genital tract. In: Gershenson DM, De Cherney AH, Curry SL, editors. *Operative gynecology*. Philadelphia (PA): Saunders; 2001. p. 273–99.
- (24) Ho GY, Burk RD, Klein S, Kadish AS, Chang CJ, Palan P, et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *J Natl Cancer Inst* 1995;87:1365–71.
- (25) Jin L, Qi M, Chen DZ, Anderson A, Yang GY, Arbeit JM, et al. Indole-3-carbinol prevents cervical cancer in human papilloma virus type 16 (HPV 16) transgenic mice. *Cancer Res* 1999;59:3991–7.
- (25) Follen M, Schottenfeld D. Surrogate endpoint biomarkers and their modulation in cervical chemoprevention trials. *Cancer* 2001;91:1758–76.
- (25) Goodman GE. The clinical evaluation of cancer chemoprevention agents: defining and contrasting phase I, II, and III objectives. *Cancer Res* 1992;52(9 Suppl):2752s–2757s.

NOTES

¹Definition of chemoprevention trial phases (27). Phase I trials measure the incidence and spectrum of side effects associated with a chemopreventive agent in the first 12 months of treatments, utilizing a single arm consisting of, usually, a relatively small number of subjects. Phase II trials further evaluate side effects in a randomized, placebo-controlled, double-blind study design. They may look at multiple doses, and last from 1 to 5 years. They are larger than phase I trials. Phase III trials, which are also randomized, double-blind, and placebo controlled, have more power than phase I and II trials, and can examine more uncommon and longer-term side effects than phase I and II trials. Chemoprevention trials can be divided into parts A and B. Part A trials encompass all clinical endpoints; part B trials focus on intermediate endpoint biomarkers.

Supported by Public Health Service grant CN25433 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

M. Follen is published formerly under Michele Follen Mitchell.

Manuscript received January 31, 2001; revised June 5, 2001; accepted June 8, 2001.