Cumulative Incidence of False-Positive Results in Repeated, Multimodal Cancer Screening

Jennifer Miller Croswell, MD*

Barnett S. Kramer, MD, MPH*

Aimee R. Kreimer, PhD

Phil C. Prorok, PhD

Jian-Lun Xu, PhD

Stuart G. Baker, ScD

Richard Fagerstrom, PhD

Thomas L. Riley

Jonathan D. Clapp

Christine D. Berg, MD

John K. Gohagan, PhD

Gerald L. Andriole, MD

David Chia, PhD

Timothy R. Church, PhD

E. David Crawford, MD

Mona N. Fouad, MD, MPH

Edward P. Gelmann, MD

Lois Lamerato, PhD, MS

Douglas J. Reding, MD, MPH

Robert E. Schoen, MD, MPH

CORRESPONDING AUTHOR

Jennifer M. Croswell, MD
Office of the Director
Office of Disease Prevention
National Institutes of Health
6100 Executive Blvd, Suite 2B-03
Bethesda, MD 20892
croswellj@od.nih.gov

**Drs Croswell and Kramer contributed equally as
lead authors of this article.*

ABSTRACT

PURPOSE Multiple cancer screening tests have been advocated for the general population; however, clinicians and patients are not always well-informed of screening burdens. We sought to determine the cumulative risk of a false-positive screening result and the resulting risk of a diagnostic procedure for an individual participating in a multimodal cancer screening program.

METHODS Data were analyzed from the intervention arm of the ongoing Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, a randomized controlled trial to determine the effects of prostate, lung, colorectal, and ovarian cancer screening on disease-specific mortality. The 68,436 participants, aged 55 to 74 years, were randomized to screening or usual care. Women received serial serum tests to detect cancer antigen 125 (CA-125), transvaginal sonograms, posteroanterior-view chest radiographs, and flexible sigmoidoscopies. Men received serial chest radiographs, flexible sigmoidoscopies, digital rectal examinations, and serum prostate-specific antigen tests. Fourteen screening examinations for each sex were possible during the 3-year screening period.

RESULTS After 14 tests, the cumulative risk of having at least 1 false-positive screening test is 60.4% (95% CI, 59.8%-61.0%) for men, and 48.8% (95% CI, 48.1%-49.4%) for women. The cumulative risk after 14 tests of undergoing an invasive diagnostic procedure prompted by a false-positive test is 28.5% (CI, 27.8%-29.3%) for men and 22.1% (95% CI, 21.4%-22.7%) for women.

CONCLUSIONS For an individual in a multimodal cancer screening trial, the risk of a false-positive finding is about 50% or greater by the 14th test. Physicians should educate patients about the likelihood of false positives and resulting diagnostic interventions when counseling about cancer screening.

INTRODUCTION

Numerous cancer screening tests are promoted to the healthy public.[1-8] The motivating factor behind regular cancer screening is the theory that the earlier one detects a malignancy or premalignancy, the more likely treatment is to be effective in increasing lifespan while minimizing harms caused by the therapy.[9] Although this model has intuitive appeal, it is often used without actual proof in hand and without full consideration of potential adverse consequences. The most common potential adverse consequence is a false-positive result, which often brings with it physical, psychological, and economic burdens of further diagnostic testing.[10-13]

The false-positive rate of a single screening test has been studied, but the cumulative false-positive rate of repeating the test at regular intervals is infrequently reported,[14-17] and the cumulative false-positive rate of multiple tests has not, to our knowledge, been reported at all. The ongoing Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial is designed to assess the benefits and harms of screening for 4 major causes of cancer mortality. As such, it represents an ideal opportunity to assess the cumula-

tive false-positive rate and resulting diagnostic procedure rates of a combined-modality screening program.

METHODS

Methods for the PLCO Cancer Screening Trial

The PLCO trial is a randomized, 2-armed trial to determine the effects of prostate, lung, colorectal, and ovarian cancer screening on disease-specific mortality.[18-23] Ten screening centers are participating. Randomization began in 1993 and concluded in 2001. The total number of enrolled participants was 154,935. Individuals were eligible to enroll if they were aged between 55 and 74 years with no history of prostate, lung, colorectal, or ovarian cancer. Further exclusion criteria included current treatment for any cancer except basal or squamous skin cancer; prior removal of the entire prostate, 1 lung, or the entire colon; concurrent participation in another cancer screening or primary prevention study; use of finasteride within 6 months of enrollment; and, after 1995, men reporting more than 1 prostate-specific antigen test in the 3 years before enrollment, or any colonoscopy, sigmoidoscopy, or barium enema in the 3 years before enrollment.

Eligible participants were randomly assigned to the control arm (normal health care routines), or to the intervention arm. Within the control arm, patients were informed about screening tests for PLCO cancers and could have potentially received testing outside the confines of the trial through private physicians. Randomization to the intervention arm meant that women were offered annual tests for the cancer antigen 125 (CA-125) and transvaginal ultrasonography for ovarian cancer for 4 years, posteroanterior chest radiographs for lung cancer at baseline and yearly for 2 years (nonsmokers) or 3 years (smokers), and baseline and 3- or 5-year flexible sigmoidoscopy for colorectal cancer. Men were offered the same tests for lung and colorectal cancers; additionally, they received annual digital rectal examinations (DREs) and prostate-specific antigen (PSA) blood tests for 4 years.

The order in which the tests were offered to a given patient during a round of screening varied, with the exception that PSA had to be performed before a DRE, and serum to test for CA-125 was drawn before a transvaginal ultrasonography was offered. Flexible sigmoidoscopy was usually scheduled to be the last procedure during a round of testing, as it was believed that delaying the more unpleasant examination might improve overall compliance. Positive screening results were specifically defined in the study protocol,[18] but follow-up was left to the discretion of the participants' personal physicians, who were notified of all abnormal results. According to PLCO protocol, colonic lesions

found during flexible sigmoidoscopy were not immediately biopsied and removed; instead, they prompted referral to the participant's primary physician for follow-up. Criteria for a positive screening examination are listed in Table 1.[24]

Methods for the PLCO False-Positive Screening Study

Study Population and Eligibility

Enrollment exclusions used by the PLCO trial as a whole applied. We began with the intervention arm of the PLCO trial (n = 77,464). Additional exclusion criteria for this study were (1) death before the first screening test (n = 39); (2) missing all screening tests (n = 5,028); and (3) inadequate follow-up time (less than 3 years after the last screening test taken) (n = 3,961) (Figure 1). We observed the remaining 68,436 participants up to a maximum possible 14 tests (although not all participants consented to or received all 14 examinations). At the time of the initial screening and 3 years later, participants were offered all 4 tests (chest radiograph, flexible sigmoidoscopy, CA-125, and transvaginal ultrasonography for women; and chest radiograph, flexible sigmoidoscopy, PSA, and DRE for men); 1 and 2 years after the initial screening, participants were offered all tests except flexible sigmoidoscopy.

Flexible sigmoidoscopy frequency was decreased from every 3 to every 5 years during the trial to better match changing clinical practices. Given our 3-year timeframe, flexible sigmoidoscopies offered at year 5 were not included in this study. Approximately 25% of participants had the opportunity to undergo 2 flexible sigmoidoscopies; the rest were scheduled for only 1 (at baseline). As a posteroanterior chest radiograph was offered only twice after the baseline examination for nonsmokers, not all participants underwent the fourth chest radiograph at year 3.

Definitions

False-Positive Result

By consensus of the investigators, a positive screen was considered a false positive if the participant had at least 3 years of follow-up after the positive result and the target cancer was not diagnosed by that time. For flexible sigmoidoscopy, advanced adenomas (including those with villous histopathologic findings, or severe cellular dysplasia, or that were ≥1 cm in diameter) were considered true positives for the base-case analysis. There is ongoing debate concerning the target lesion in colorectal cancer screening. Although clinicians generally agree that colorectal lesions 1 cm and larger warrant work-up, discordant recommendations are found for smaller polyps, with some clinicians advocating for the removal of any observable lesion.[25] The

Table 1. Criteria for a Positive Screening Examination and Classification of Diagnostic Follow-Up Procedures

Screening Modality	Definition of a Positive Screening Result	Diagnostic Follow-Up Procedures		
		Minimally Invasive Procedures	Moderately Invasive Procedures	Major Surgical Procedures
PSA	>4 ng/mL	Endoscopy	Biopsy Cystourethroscopy Cystoscopy Cystopanendoscopy Lymphadenectomy Ureterogram Transurethral resection of the prostate (TURP)	Prostatectomy

DRE	One or more of the following findings:
	  Nodularity
	  Induration
	  Asymmetry
	  Loss of anatomic landmarks
		As for PSA	As for PSA	As for PSA

CA-125	>35 U/mL	Colonoscopy	Biopsy Culdocentesis Hysteroscopy Intra-abdominal washings Laparoscopy Paracentesis	Hysterectomy Laparotomy Omentectomy Oophorectomy Ovarian lymphadenectomy (with laparotomy)

TVU	One or more of the following findings:
	  Ovary or cyst >10 cc[a]
	  Solid area or papillary projection extending into the cavity of a cystic ovarian tumor of any size
	  Mixed (solid/cystic) component within a cystic ovarian tumor
		As for CA-125	As for CA-125	As for CA-125

CXR	One or more of the following findings:
	  Nodule
	  Mass
	  Hilar or mediastinal lymph node enlargement
	  Major atelectasis or lobar collapse
	  Infiltrate, consolidation, or alveolar opacity
		Bronchoscopy	Biopsy
			Lymphadenectomy
			Mediastinoscopy
			Thoracoscopy
			Transbronchial aspiration
			Transthoracic aspiration
				Lung resection
				Thoracotomy

FSG	One or more of the following findings[b]:
	  Rectal nodule(s)
	  Rectal and/or colon mass(es)
	  Colon polyp(s)
		Colonoscopy (without biopsy)
		Endoscopy
		Rigid sigmoidoscopy
			Biopsy
			Cystoscopy
			Lymphadenectomy
				Colon resection
				Hemicolectomy
				Laparotomy

CA-125 = cancer antigen 125; CXR = chest radiograph; DRE = digital rectal examination; FSG = flexible sigmoidoscopy; PSA = prostate-specific antigen; TVU = transvaginal ultrasonography.

[a] Assuming a perfect sphere and using the prolate ellipsoid formula, this volume correlates to an approximate diameter of 2.67 cm. A 10-cc cutoff was chosen, as it corresponds to 2 standard deviations above mean ovarian volume for postmenopausal women.[24]

[b] We define a true-positive FSG as both overt cancer and advanced adenomas (lesions with villous histologic findings, severe cellular dysplasia, and/or ≥1 cm in diameter).
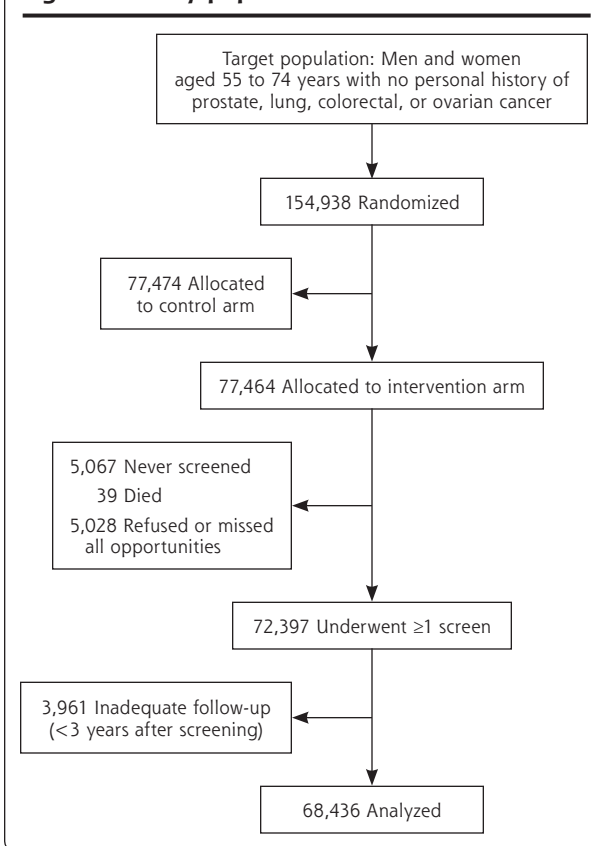
natural history of all of these lesions (how often they progress to cancer and how long it takes), including advanced adenomas, however, remains poorly understood.[26] We therefore chose the midpoint along this spectrum of opinions for our primary analysis.

Diagnostic Follow-Up
Diagnostic follow-up for positive results was defined by consensus of investigators and divided into the following categories: (1) minimally invasive procedure (ie, simple endoscopy with conscious sedation); (2) moderately invasive procedure (ie, tissue removal, more

involved instrumentation, or general anesthesia, including laparoscopy); and (3) major thoracic, abdominal, or pelvic surgery (ie, prostatectomy or colectomy). (The adjective "invasive" is not intended as a value-laden term but is being used according to the medical definition: "Involving puncture or incision of the skin or insertion of an instrument or foreign material into the body" [http://www.dorlands.com/].) Repeated screening examinations, physical examinations, chart reviews, and basic imaging examinations were not considered invasive procedures and did not contribute to these categories.

The causal relationship between a diagnostic pro-

Figure 1. Study population selection.

Target population: Men and women aged 55 to 74 years with no personal history of prostate, lung, colorectal, or ovarian cancer

154,938 Randomized

77,474 Allocated to control arm

77,464 Allocated to intervention arm

5,067 Never screened
39 Died
5,028 Refused or missed all opportunities

72,397 Underwent ≥1 screen

3,961 Inadequate follow-up (<3 years after screening)

68,436 Analyzed

cedure and a positive findings on a screening examination was determined by trained chart abstractors at each study center. An organ-specific standardized diagnostic evaluation form was completed for every positive screening result that captured follow-up care through 1 year after the screening date.

Statistical Methods

We sought to answer the question, What is the probability that an individual entering a multimodal screening program will obtain at least 1 false-positive test in a given number of tests? The approach applied a model for a single screening test to multiple tests.[27] The model is a type of survival analysis based upon number of tests taken rather than time elapsed. The tests are ordered by appearance and analyzed by tests-to-event rather than by time-to-event; this method allows for appropriate handling of missing tests while enhancing use of available data. For example, if a participant took the first 7 tests, skipped 3 tests, and then returned for the last 4 tests, his or her results contribute to the first 11 points along the cumulative incidence curve. A participant no longer contributes to the curve either after the first false-positive result (event reached) or after the last test taken (censored).

Because participants could receive different screening tests at each testing point, the model treats each test in order of appearance as a composite of all modalities. It makes a single informative censoring assumption, applied only to individuals who did not receive a false-positive result during the study period and who took some, but not all, 14 tests: the unobserved probability of receiving a first false-positive result among the censored tests of various modalities follows a geometric distribution with a constant hazard.

This method was also used to determine the cumulative risk of any single invasive diagnostic procedure resulting from a false-positive test result. Sensitivity analysis was performed by using a recently published modified Kaplan-Meier approach.[28]

RESULTS

The demographic characteristics of the PLCO trial have been described elsewhere.[23,29] There were 68,436 participants who underwent at least 1 screening test and had adequate follow-up (Figure 1). Most participants (65%) ranged in age from 55 to 64 years, were white, non-Hispanic (88.3%), and had a college education or higher (57.5%). Participant sex was equally distributed (50% men, 50% women).

False-Positive and Follow-up Diagnostic Procedure Rates

The maximum number of false-positive results observed in a female was 8, and in a male, 10, from a potential 14 (Table 2). Of the 68,436 participants, 43.1% (n = 29,517) had 1 or more false-positive findings: 35.3% of women and 50.9% of men. In men flexible sigmoidoscopy accounted for the greatest

Table 2. Proportion of Participants With 1 or More False-Positive Results

False-Positive Result	Participants No. (%)
1	18,394 (26.9)
2	6,043 (8.8)
3	2,531 (3.7)
4	1,535 (2.2)
5	642 (0.9)
6	228 (0.3)
7	78 (0.1)
8	53 (0.1)
9	11 (0.0)
10	2 (0.0)
11	0 (0.0)
12	0 (0.0)
13	0 (0.0)
14	0 (0.0)

percentage of false-positive results (26.8%), then chest radiograph (18.6%), DRE (15.0%), and PSA (10.4%). In women flexible sigmoidoscopy accounted for the greatest percentage of false-positive results (17.2%), then chest radiograph (16.3%); transvaginal ultrasonography (7.8%), and CA-125 (3.0%) (Table 3).

Of the participants only 3.0% had minimally invasive diagnostic procedures, and 15.8% underwent moderately invasive procedures; 1.6% of participants incurred major surgery. When examining diagnostic responses by screening modality, flexible sigmoidoscopy resulted in the highest rates of minimally and moderately invasive diagnostic follow-up (for men, 3.4% and 15.8%, respectively; for women, 3.0% and 10.1%, respectively). Transvaginal ultrasonography accounted for the preponderance of major surgeries

resulting from false-positive findings. Of the screened women 3% underwent a major surgical procedure for false-positive findings on a transvaginal sonogram.

Cumulative Risk of a False-Positive Result

For men the risk of having at least 1 false-positive finding is 36.7% (95% CI, 36.2%-37.3%) by the 4th screening test (the end of the first day of screening), and 60.4% (95% CI, 59.8%-61.0%) by the 14th examination. Women have a 26.2% risk (95% CI, 25.7%-26.8%) by the 4th screening test, and 48.8% risk (95% CI, 48.1%-49.4%) by the 14th examination. Sensitivity analysis by modified Kaplan-Meier approach provides similar results at 14 tests: 57.2% for men (95% CI, 56.3%-58.2%) and 44.1% for women (95% CI, 43.4%-44.7%). The cumulative risk of receiving at least

Table 3. Initial Entry Round (Year 1) Through Last Entry Round (Year 4) of Screening: False-Positive and Resulting Diagnostic Follow-Up Rates

Subgroup	Ever False Positive No. (%)	Physician Response									
		Repeated Screening		Imaging		Minimally Invasive Procedure[a]		Moderately Invasive Procedure[b]		Major Surgical Procedure[c]	
		No. (%)[d]	% of FPs[e]	No. (%)[d]	% of FPs[e]	No. (%)[d]	% of FPs[e]	No. (%)[d]	% of FPs[e]	No. (%)[d]	% of FPs[e]
Overall rates	29,517 (43.1)	13,576 (19.8)	46.0	7,306 (10.7)	24.8	2,049 (3.0)	6.9	10,822 (15.8)	36.7	1,062 (1.6)	3.6
Men	17,432 (50.9)	9,251 (27.0)	53.1	4,056 (11.9)	23.3	1,088 (3.2)	6.2	6,972 (20.4)	40.0	47 (0.1)	0.3
Women	12,085 (35.3)	4,325 (12.6)	35.8	3,250 (9.5)	26.9	961 (2.8)	8.0	3,850 (11.3)	31.9	1,015 (3.0)	8.4
By screening modality: men[f]											
PSA	3,388 (10.4)	2,884 (8.9)	85.1	1,494 (4.6)	44.1	1 (0.0)	0.0	1,491 (4.6)	44.0	6 (0.0)	0.2
DRE	4,882 (15.0)	3,881 (12.0)	79.5	1,326 (4.1)	27.2	0 (0.0)	0.0	1,259 (3.9)	25.8	1 (0.0)	0.0
CXR	6,320 (18.6)	3,216 (9.5)	50.9	1,466 (4.3)	23.2	52 (0.2)	0.8	77 (0.2)	1.2	35 (0.1)	0.6
FSG	8,186 (26.8)	645 (2.1)	7.9	157 (0.5)	1.9	1,036 (3.4)	12.7	4,821 (15.8)	58.9	5 (0.0)	0.1
By screening modality: women[f]											
CA-125	888 (3.0)	567 (1.9)	63.9	349 (1.2)	39.3	0 (0.0)	0.0	103 (0.4)	11.6	125 (0.4)	14.1
TVU	2,310 (7.8)	745 (2.5)	32.3	1,394 (4.7)	60.4	1 (0.0)	0.0	677 (2.3)	29.3	874 (3.0)	37.8
CXR	5,531 (16.3)	2,907 (8.6)	52.6	1,498 (4.4)	27.1	56 (0.2)	1.0	93 (0.3)	1.7	40 (0.1)	0.7
FSG	5,239 (17.2)	321 (1.1)	6.1	151 (0.5)	2.9	906 (3.0)	17.3	3,072 (10.1)	58.6	7 (0.0)	0.1

CA-125 = cancer antigen 125; CXR = chest radiograph; DRE = digital rectal examination; FP = false positive; FSG = flexible sigmoidoscopy; PSA = prostate-specific antigen; TVU = transvaginal ultrasonography.

Note: These percentages reflect only the subset of diagnostic procedures directly caused by a false-positive result; they do not include procedures associated with a true-positive or negative.

[a] Example: simple endoscopy with conscious sedation (see Table 1 for the complete list of procedure classifications).

[b] Example: tissue removal, more involved instrumentation, or general anesthesia, including laparoscopy.

[c] Example: prostatectomy, or laparotomy with colectomy or oophorectomy.

[d] Percentage among participants in the intervention arm that took at least 1 screening examination.

[e] Percentage among participants in the intervention arm that received at least 1 false-positive examination (change in denominator). This percentage (and the accompanying crude numbers) does not include procedures that resulted in a true positive (diagnosis of cancer).

[f] In each case, the denominator is the total number of participants taking examinations for a given screening modality; this number varied depending on differences in compliance with different tests.

1 false-positive result in the 14 tests is displayed in Figure 2, panel A. The exact estimates and 95% confidence intervals for each test number (t) can be found in Supplemental Table 1, available online at http://www. annfammed.org/cgi/content/full/7/3/212/DC1.

When comparing individual screening modalities, flexible sigmoidoscopy accounted for the highest cumulative false-positive risk: 41.8% (95% CI, 40.9%-42.7%) for men and 29.2% (95% CI, 28.3%-30.0%) for women after 2 examinations. Chest radiograph accounted for the next highest false-positive risk, with a 22.3% (95% CI, 21.4%-23.2%) probability after 4 tests for men and a 21.5% (95% CI, 20.6%-22.4%) probability for women. The cumulative risks of receiving at least 1 false-positive result by screening modality is displayed in Figure 2, panel B. The exact estimates and 95% confidence intervals for each test number (t) can be found in Supplemental Table 2, available online at http://www.annfammed. org/cgi/content/full/7/3/212/DC1.

Cumulative Risk of an Invasive Diagnostic Procedure Resulting from a False-Positive Result

For a woman the cumulative risk of undergoing a false-positive–prompted invasive diagnostic procedure* is 12.3% (95% CI, 11.8-12.8%) after 4 tests, and 22.1% (95% CI, 21.4%-22.7%) after 14 examinations. For men, the risk is slightly higher: 17.2% (95% CI, 16.7%-17.6%) after 4 tests, and 28.5% (95% CI, 27.8%-29.3%) after 14 examinations. Sensitivity analysis confirmed similar percentages: for women, the risk after 14 tests is 20.6% (95% CI, 20.4%-20.9%); for men, 27.5% (95% CI, 27.2%-27.9%). The cumulative probability of receiving at least 1 invasive diagnostic procedure is displayed in Figure 3, panel A. The exact estimates and 95% confidence intervals for each test number (t) are shown in Supplemental Table 3, available online at http://www.annfammed. org/cgi/content/full/7/3/212/DC1.

We also compared cumulative risks of false-positive–prompted invasive diagnostic interventions by individual modalities. Flexible sigmoidoscopy accounts for the greatest probability of a false-positive–prompted invasive diagnostic procedure: a 30.1% (95% CI, 29.2%-31.1%) probability after 2 tests for men, and a 22.1% (95% CI, 21.2%-23.0%) probability for women. A false-positive transvaginal sonogram is associated with the next highest risk, with a 6.7% (95% CI, 5.6%-7.7%) probability after 4 tests. The probability of receiving at least 1 invasive diagnos-

tic procedure by screening modality is displayed in Figure 3, panel B. The exact estimates and 95% confidence intervals for each test number (t) are shown shown for all modalities in Supplemental Table 4, available online at http://www.annfammed. org/cgi/content/full/7/3/212/DC1.

DISCUSSION

The cumulative risk of an individual obtaining a false-positive result in a multimodal screening program increases with number of screening tests; by the 4th screening test—which in the PLCO trial would mean the end of day 1—the risk is about 37% for men and 26% for women. By the 14th test, the risk is approximately 60% and 49% for men and women. The risk of undergoing any false-positive–prompted invasive diagnostic procedure is about 17% for men and 12% for women after 4 tests, and 29% for men and 22% for women after 14 screening tests.

Historically, recommendations for individual screening modalities have varied and will likely continue to vary across health care systems. The overarching principle shown by this study, however—that of rapidly increasing cumulative false-positive rates associated with multiphasic screening—remains important and underappreciated when recommending screening. To our knowledge, this study is the first to use randomized trial data to estimate the general magnitude of false positives that might be expected with a multimodal screening program; the fundamental approach and techniques used in this work should be applicable across a wide range of screening regimens.

The PLCO trial does not include an evaluation of mammography; however, as this modality is in widespread use, the associated false-positive rates are of clear importance to women and clinicians. Elmore et al estimated the cumulative incidence of false positives associated with mammography to be 49.1% after 10 examinations.[14] To put this estimate in context with our own findings, the cumulative incidence of a false-positive mammogram was estimated at approximately 18% after 3 examinations (the maximum number that would have been observed during our study's timeframe). Although this percentage is not directly additive to our cumulative incidence curve, it is a mathematical certainty that adding another testing modality would increase the final risk of obtaining at least 1 false positive for women.

Our study was a test-to-event rather than time-to-event analysis: we included any participant who took at least 1 screening examination sometime during the course of the 3 rounds (or 14 possible tests), and we analyzed the test results by total number rather than

---

*Invasive categories were previously defined as minimally invasive procedure, moderately invasive procedure, or major surgical procedure. These categories exclude repeated screening examinations, chart reviews, physical examinations, and basic radiographic imaging.

Figure 2. Cumulative probability of a false-positive result for a multimodal cancer screening regimen.

Panel A shows the cumulative probability of receiving at least 1 false-positive screening test, for men and women, all modalities combined, over the course of 14 tests.[a]
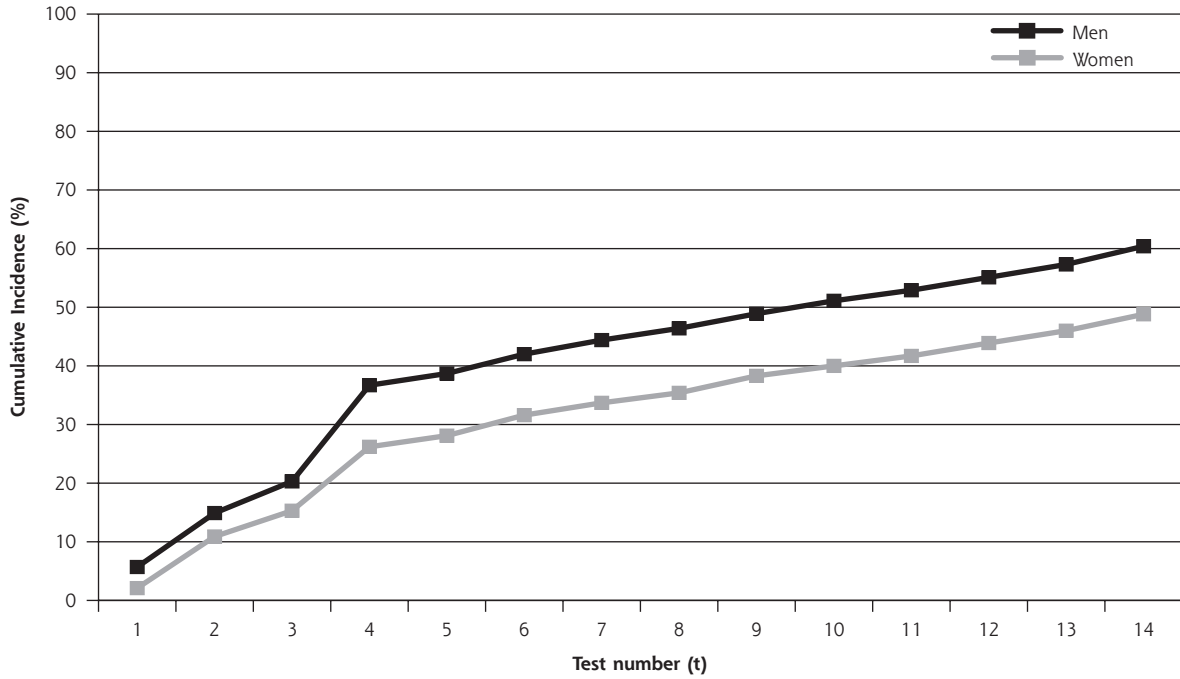
Panel B shows the cumulative probability of receiving at least 1 false-positive screening test, for men and women, by individual screening modality (maximum number of tests per screening modality was 4).[a,b]
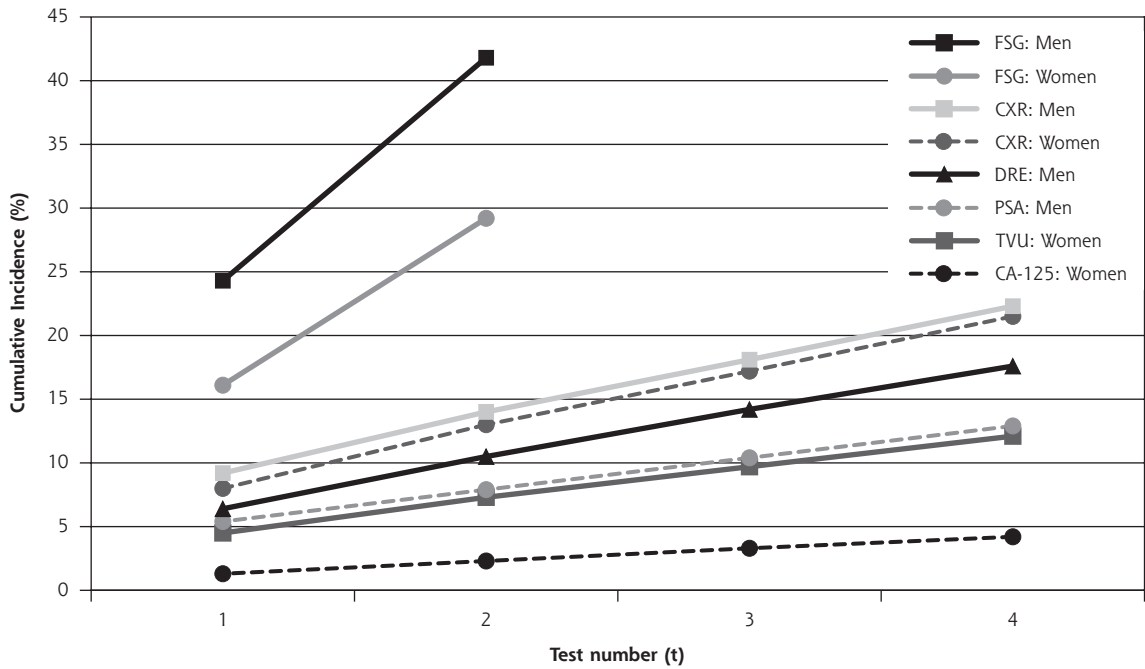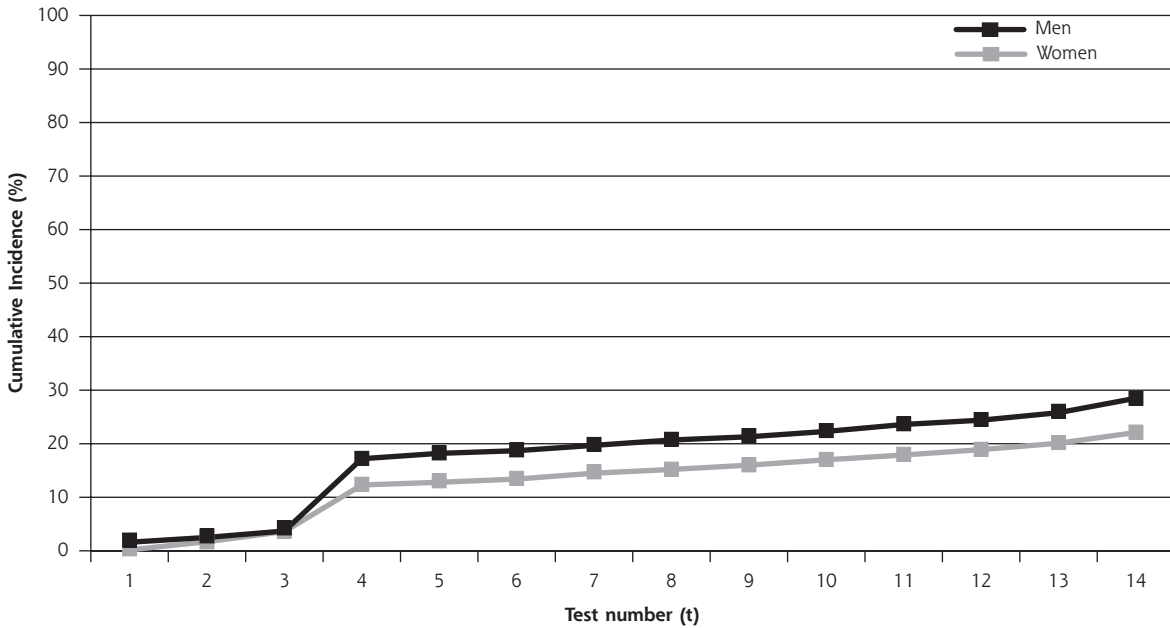

CA-125 = cancer antigen 125; CXR = chest radiograph; DRE = digital rectal examination; FSG = flexible sigmoidoscopy; PSA = prostate-specific antigen; TVU = trans-vaginal ultrasonography.

[a] Lines correspond to labels as ordered top to bottom in legend.

[b] Note change in y-axis scale.

Figure 3. Cumulative probability of receiving an invasive diagnostic procedure[a] as a direct result of a false-positive test in a multimodal cancer screening regimen.

Panel A shows the cumulative probability of receiving at least 1 invasive diagnostic procedure as a direct result of a false positive, for men and women, all modalities combined, over the course of 14 tests.[b]

Panel B shows the cumulative probability of receiving at least 1 invasive diagnostic procedure as a result of a false positive, for men and women, by individual screening modality (maximum number of tests per screening modality was 4).[b,c]
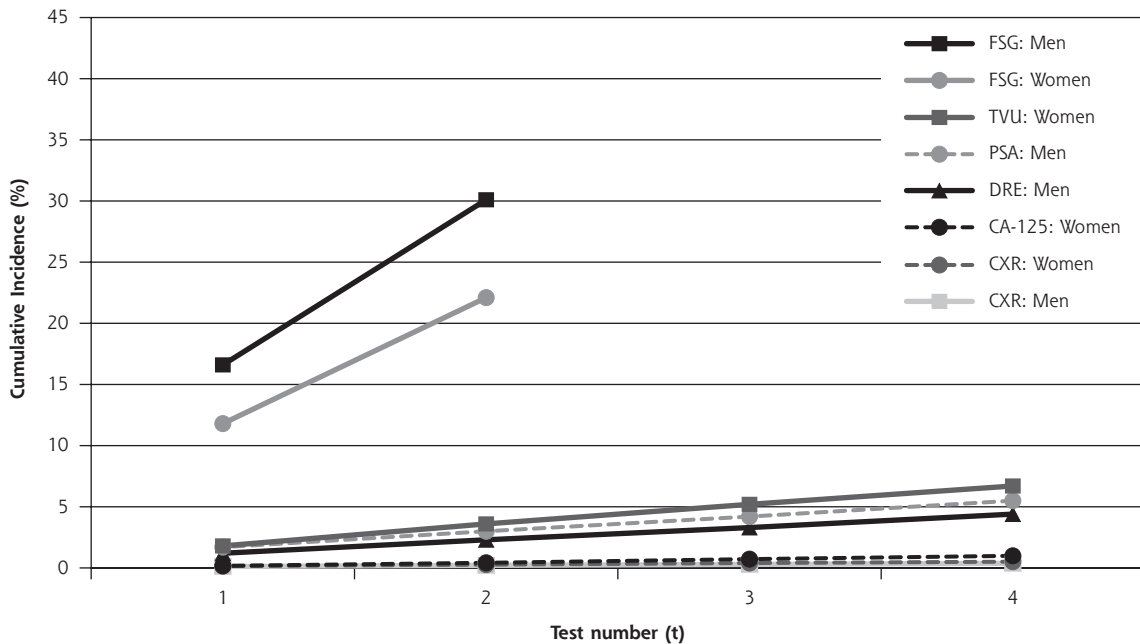
CA-125 = cancer antigen 125; CXR = chest radiograph; DRE = digital rectal examination; FSG = flexible sigmoidoscopy; PSA = prostate-specific antigen; TVU = transvaginal ultrasonography.

[a] Invasive diagnostic procedures are defined by 3 categories: minimally invasive (example: simple endoscopy with conscious sedation); moderately invasive (example: involving tissue removal, more invasive instrumentation, or general anesthesia (including laparoscopy); and major surgical procedures (example: prostatectomy, or laparotomy with colectomy or oophorectomy). See Table 1 for the complete list of procedure classifications.

[b] Lines correspond to labels as ordered top to bottom in legend.

[c] Note change in y-axis scale.

by time elapsed to allow for appropriate handling of missing tests while maximizing the use of available data. Individuals in the intervention arm who received no screening examinations were not included in the results: we chose this approach because we believe it to be most accurate and informative for the real-world scenario of a patient visiting a clinic with the intention of being screened and who wants to understand the risks involved if he or she undergoes at least some proportion of a multiphasic program of screening.

Additionally, not all participants received all 14 tests; as a result, the calculated false-positive cumulative incidence curves may underestimate true false-positive rates for a multiphasic screening program with 100% compliance. We believe, however, that the population included in this analysis—with good but not perfect compliance—and the cumulative risks obtained from this population best approximate what could be expected in real-world practice.

Because a given individual could receive a different screening modality at each point along the cumulative risk curves, the curves show the general accumulated burden better than they serve as precise indicators of risk at a given time. That is, the curves provide rule-of-thumb estimates at the point of each screening examination taken (t), but most accurately predict exact percentages at the end of the first day of screening and once all of the tests have been undergone (at t = 14). The curves provide rough approximations of time-to-event: if all tests are taken according to protocol, estimated rates at t = 4 are equivalent to the end of day 1 of testing, rates at t = 7 are equal to year 1 of testing, rates at t = 10 are equal to year 2 of testing, and rates at t = 14 are equal to year 3 of testing. Our methods, however, maximize the use of available data from variably compliant individuals and analyze tests by total number rather than elapsed time. For that reason, it is more accurate to interpret results by final number of tests taken than by years within the screening program.

A steep rise in the cumulative risk for false-positive examination results can be seen during the initial screening session (tests 1 to 4, Figure 2, panel A) There is likely to be a subset of healthy individuals in the general population who do not have cancer but chronically express a benign abnormality detectable by blood test, radiographic test, or endoscopy. Upon initial screening, these individuals would likely have been immediately evaluated further, and once a false positive was identified, these individuals did not contribute further to the overall multimodal risk curve.

The classification of follow-up procedures into 3 categories is subjective in that the diagnostic tests performed were widely disparate; different clinicians may have conflicting opinions about what relegates a

procedure to a minimally or moderately invasive category. For this reason, we have presented the full list of procedures and how they fit in our own classification system for the sake of transparency.

Clinicians may have varying thresholds for what they consider a true- vs false-positive result. Some researchers currently estimate a 5- to 15-year lead-time for PSA and prostate cancer.[30] Our study used a standardized 3-year negative follow-up period across all modalities to define a false positive; we did not feel it appropriate, absent definitive knowledge, to vary the time-interval criteria for a false positive by individual modality.

In the case of colorectal cancer, as previously noted, some clinicians argue that any adenoma, regardless of size, should be considered a true-positive result. We performed a sensitivity analysis to examine the effect that including any adenoma would have on false-positive rates in the PLCO trial. The percentage of men receiving false-positive flexible sigmoidoscopy results using this broader definition changes from 26.8% to 20.2%, and for women, from 17.2% to 13.5%. (In this situation, the remaining false positives encompassed such findings as inflammation, hyperplastic polyps, or a lack of demonstrable lesion on diagnostic follow-up.) The percentage of participants receiving at least 1 false positive of any modality shifts from 43.1% to 39.5% (for men, 50.9% to 46.6%; for women, 35.3% to 32.3%). If, conversely, the definition of a true-positive is restricted to overt cancer only, as some PLCO investigators favored, the percentage of men receiving at least 1 false-positive flexible sigmoidoscopy result increases to 30.3%; in women, 19.3%. For all modalities combined, the absolute percentage of participants receiving at least 1 false positive rises to 45.2% (men, 53.5%; women, 36.8%). Establishing a threshold for referral to colonoscopy after abnormal flexible sigmoidoscopy is complex. Randomized trials in Europe used biopsy at the time of flexible sigmoidoscopy to assist in triaging patients to reduce the number sent for colonoscopy,[31,32] an approach not used in the PLCO trial.[33] Ultimately, understanding the benefit of a higher compared with a lower rate of referral for colonoscopy after flexible sigmoidoscopy must await the outcome of the randomized trials.

Since the PLCO trial began, the use of flexible sigmoidoscopy in the United States has become less common, while the use of screening colonoscopy has increased substantially with the introduction of Medicare reimbursement coverage for this procedure.[34] The Centers for Disease Control and Prevention estimates that approximately 2.8 million flexible sigmoidoscopy examinations were performed in 2002; flexible sigmoidoscopy also remains a recommended screening modality by many professional organizations, including

the US Preventive Services Task Force and, jointly, the American Cancer Society, the US Multi-society Task Force on Colorectal Cancer, and the American College of Radiology.[35-37] Because colonoscopy allows evaluation of more of the total area of the colon than does flexible sigmoidoscopy (and thus, has the potential to discover more abnormalities of uncertain significance), it is likely that substituting this modality would have increased the cumulative incidence of false positives. Our observed flexible sigmoidoscopy false-positive rate could thus be considered a conservative estimate of the false-positive risks for colonoscopy.

The rate of prostatic biopsy follow-up (ie, the percentage of moderately invasive procedures) for abnormal PSA results may appear low to some in the urologic community (44%). This percentage does not include men whose prostatic biopsies resulted in a diagnosis of cancer (ie, true positives), and thus it represents only a fraction of the total number of biopsies performed in the PLCO trial. Our cutoff point for an abnormal PSA was 4 ng/mL; however, because there is no absolute value below which a man can be assured he does not have prostate cancer,[38] some researchers and clinicians have used lower cut-points for screening (between 2.5 to 4 ng/mL). Attempting to increase sensitivity in this manner would likely increase the cumulative false-positive rates we report here.

Our study has several limitations. The PLCO trial examines a series of cancer screening tests, but not all of these modalities are used by a majority of the public. Nevertheless, all have been publicly advocated at one time by professional organizations or advocacy groups. The trial participants may not be entirely representative of the general population in that by agreeing to join a long-term cancer screening trial, they may be more involved in health promotion activities and may have higher screening compliance rates than nonparticipants.[29]

The false-positive and diagnostic follow-up rates found in this study have important practical implications, especially given that the study was limited to 4 rounds of screening examinations. Most formal screening recommendations are open-ended; none advise stopping after 4 rounds. Given the rate at which false-positive results can accumulate in multimodal screening programs and the potential iatrogenic burden these regimens can generate in healthy individuals, we propose that future guidelines begin determining risk-benefit equations for entire screening regimens rather than continuing to evaluate individual tests separately.

This study has developed a clearer picture in a previously unexplored aspect of burden and risks associated with scheduled multimodal cancer screening programs.

The benefit of regular combined-modality cancer screenings in reducing mortality is not yet known; we will not have that information until the conclusion of the PLCO trial. Physicians and patients must therefore examine the balance of known risks vs potential benefits to determine the most appropriate course of action for each individual.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/7/3/212.

Authors' affiliations: National Institutes of Health, Office of the Director, Office of Disease Prevention, Bethesda, Maryland (Croswell, Kramer); National Institutes of Health, National Cancer Institute, Division of Cancer Prevention, Bethesda, Maryland (Kreimer, Xu, Baker, Fagerstrom, Berg, Gohagan); IMS, Incorporated, Washington, DC (Riley, Clapp); Washington University School of Medicine, St. Louis, Missouri (Andriole); UCLA School of Medicine, Los Angeles, California (Chia); University of Minnesota School of Public Health, Minneapolis, Minnesota (Church); University of Colorado Health Sciences Center, Aurora, Colorado (Crawford); University of Alabama at Birmingham, Birmingham, Alabama (Fouad); Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York (Gelmann); Henry Ford Health System, Detroit, Michigan (Lamerato); Marshfield Clinic Research Foundation, Marshfield, Wisconsin (Reding); University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania (Schoen).

References

1. Lung Cancer Alliance. Lung cancer screening and early detection. http://www.lungcanceralliance.org/facing/screening.html. Accessed Apr 24, 2008.

2. American Urological Association. Prostate cancer screening. http://www.urologyhealth.org/adult/index.cfm?cat=05&topic=250. Accessed Apr 24, 2008.

3. Centers for Disease Control and Prevention. Colorectal (colon) cancer. Screening guidelines. http://www.cdc.gov/cancer/colorectal/basic_info/screening/guidelines.htm. Accessed Apr 24, 2008.

4. Armstrong JT, Brandt K, Read Y. Women's Health Screening in Napa Valley: Uterine, Tube, and Ovarian Screening. http://www.womenshealthscreening.com/pelvic.htm. Accessed Apr 24, 2008.

5. American Cancer Society. American Cancer Society guidelines for the early detection of cancer. http://www.cancer.org/docroot/PED/content/PED_2_3X_ACS_Cancer_Detection_Guidelines_36.asp?sitearea=PED. Accessed Apr 24, 2008.

6. National Cancer Institute. The Pap test: questions and answers. http://www.cancer.gov/cancertopics/factsheet/detection/Pap-test. Accessed Apr 24, 2008.

7. The Skin Cancer Foundation. Early detection. The ugly duckling sign: an early melanoma detection tool. http://www.skincancer.org/early-detection/. Accessed Apr 24, 2008.

8. Bladder Cancer Screening Underway in Communities across the Country Using the NMP22(R) BladderChek(R) Test. *Business Wire.* http://goliath.ecnext.com/coms2/gi_0199-5732062/Bladder-Cancer-Screening-Underway-in.html. Accessed Apr 24, 2008.

9. Meyskens FL. Cancer prevention, screening, and early detection. In: Abeloff MD, Armitage JO, Niederhuber JE, Kastan MB, McKenna WG, editors. *Clinical Oncology.* 3rd ed. Amsterdam: Elsevier Publishing; 2004.

10. Brawley OW, Kramer BS. Cancer screening in theory and in practice. *J Clin Oncol.* 2004;23(2):293-300.

11. Lafata JE, Simpkins J, Lamerato L, et al. The economic impact of false-positive cancer screens. *Cancer Epidemiol Biomarkers Prev.* 2004;13(12):2126-2132.

12. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med.* 2007;146(7):502-510.

13. Gilbert FJ, Cordiner CM, Affleck IR, et al. Breast screening: the psychological sequelae of false-positive recall in women with and without a family history of breast cancer. *Eur J Cancer.* 1998;34(13):2010-2014.

14. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of FP screening mammograms and clinical breast examinations. *N Engl J Med.* 1998;338(16):1089-1096.

15. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian breast cancer screening program. *Cancer.* 2004;101(7):1501-1507.

16. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst.* 2000;92(20):1657-1666.

17. Njor SH, Olsen AH, Schwartz W, Vejborg I, Lynge E. Predicting the risk of a false-positive test for women following a mammography screening programme. *J Med Screen.* 2007;14(2):94-97.

18. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin Trials.* 2000;21(6)(Suppl):273S-309S.

19. Gohagan JK, Prorok PC, Hayes RB, Kramer BS. The prostate, lung, colorectal and ovarian (PLCO) cancer screening trial of the National Cancer Institute: history, organization, and status. *Control Clin Trials.* 2000;21(6)(Suppl):251S-272S.

20. O'Brien B, Nichaman L, Browne JE, et al. Coordination and management of a large multicenter screening trial: the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial. *Control Clin Trials.* 2000;21(6)(Suppl):310S-328S.

21. Hasson MA, Fagerstrom RM, Kahane DC, et al. Design and evolution of the data management systems in the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial. *Control Clin Trials.* 2000;21(6)(Suppl):329S-348S.

22. Weissfeld JL, Fagerstrom RM, O'Brien B. Quality control of cancer screening examination procedures in the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial. *Control Clin Trials.* 2000;21(6)(Suppl):390S-399S.

23. Andriole GL, Reding D, Hayes RB, Prorok PC, Gohagan JK. The prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial: status and promise. *Urol Oncol.* 2004;22(4):358-361.

24. Pavlik EJ, DePriest PD, Gallion HH, et al. Ovarian volume related to age. *Gynecol Oncol.* 2000;77(3):410-412.

25. Rex DK, Ransohoff DF, Achkar E. Colonoscopy is justified for any polyp discovered during computed tomographic colonography. *Am J Gastroenterol.* 2005;100(9):1903-1908.

26. Ransohoff DF. Colon cancer screening in 2005: status and challenges. *Gastroenterology.* 2005;128(6):1685-1695.

27. Xu JL, Fagerstrom RM, Prorok PC, Kramer BS. Estimating the cumulative risk of a false-positive test in a repeated screening program. *Biometrics.* 2004;60(3):651-660.

28. Baker SG, Kramer BS. Estimating the cumulative risk of a false-positive under a regimen involving various types of cancer screening tests. *J Med Screen.* 2008;15(1):18-22.

29. Pinsky PF, Miller A, Kramer BS. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Epidemiol.* 2007;165(8):874-881.

30. Tornblom M, Erikson H, Franzen S, et al. Lead time associated with screening for prostate cancer. *Int J Cancer.* 2004;108(1):122-129.

31. UK Flexible Sigmoidoscopy Screening Trial Investigators. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. *Lancet.* 2002;359(9314):1291-1300.

32. Segnan N, Senore C, Andreoni B, et al. Baseline findings of the Italian multicenter randomized controlled trial of "once-only sigmoidoscopy "—SCORE. *J Natl Cancer Inst.* 2002;94(23):1763-1772.

33. Weissfeld JL, Schoen RE, Pinsky PF, et al. Flexible sigmoidoscopy in the PLCO cancer screening trial: results from the baseline screening examination of a randomized trial. *J Natl Cancer Inst.* 2005;97(13):989-997.

34. Gross CP, Andersen MS, Krumholz HM, et al. Relation between Medicare screening reimbursement and stage at diagnosis for older patients with colon cancer. *JAMA.* 2006;296(23):2815-2822.

35. Seeff LC, Richards TB, Shapiro JA, et al. How many endoscopies are performed for colorectal cancer screening? Results from CDC's survey of endoscopic capacity. *Gastroenterology.* 2004;127(6):1670-1677.

36. US Preventive Services Task Force. Screening for colorectal cancer: recommendations and rationale. Rockville, MD: Agency for Healthcare Research and Quality; 2002. http://www.ahrq.gov/clinic/3rduspstf/colorectal/colorr.htm. Accessed Jul 22, 2008.

37. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin.* 2008;58(3):130-160.

38. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *JAMA.* 2005;294(1):66-70.