Variability in Flexible Sigmoidoscopy Performance Among Examiners in a Screening Trial

PAUL F. PINSKY,* ROBERT E. SCHOEN,† JOEL L. WEISSFELD,† BARNETT KRAMER,§ RICHARD B. HAYES,‖ and LANCE YOKOCHI,¶ FOR THE PLCO PROJECT TEAM

*Division of Cancer Prevention, ‖Division of Cancer Epidemiology and Genetics, National Cancer Institute, and §Office of Prevention, National Institutes of Health, Bethesda, Maryland; †University of Pittsburgh, Pittsburgh, Pennsylvania; ¶Pacific Health Research Institute, Honolulu, Hawaii

Background & Aims: **The efficacy of flexible sigmoidoscopy (FSG) in reducing colorectal cancer mortality is being evaluated in randomized trials. In 2 European trials, wide variability across examiners in FSG performance was noted. We report on the performance of examiners in the US randomized trial: the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.** ***Methods:*** **Screening was performed at 10 geographically dispersed clinical centers. Patients with screens positive for a lesion or mass were referred to their private health care providers for endoscopic follow-up evaluation; lesions were not removed and a biopsy examination was not performed at screening. FSG performance among 64 examiners at these centers, each performing 100 or more baseline FSG examinations, with an aggregate of almost 50,000 examinations, was analyzed.** ***Results:*** **Screen-positivity results among examiners ranged from 9%–58%, with a coefficient of variation (CV) of 36%. CVs were 29% for distal polyp detection and 21% for distal adenoma detection. Inadequate rates ranged from 1%–27% (CV, 52%). Examiners with higher screen-positivity rates had higher false-positive rates, defined as a positive screen with no distal lesion found on endoscopic follow-up evaluation.** ***Conclusions:*** **Considerable variability exists in the rates of positive screens and in polyp and adenoma detection rates among FSG examiners performing the procedures using a common protocol.**

Colorectal carcinoma is the second leading cause of cancer-related mortality in the United States.[1] Screening for colorectal cancer using fecal occult blood testing with follow-up colonoscopy has been shown to reduce colorectal cancer mortality and also to reduce colorectal cancer incidence by removing adenomas before they have a chance to progress to cancer.[2,3] Flexible sigmoidoscopy (FSG) detects both adenomas and early colorectal cancers in the distal colon and rectum and thus also potentially is capable of reducing colorectal cancer incidence and mortality. The efficacy of FSG in reducing colorectal cancer mortality currently is being evaluated in several large screening trials.[4–6] In addition, several case-control studies have shown that FSG is associated with reductions in colorectal cancer incidence and mortality.[7,8]

The performance characteristics of FSG are highly dependent on the examiner. Recently, the UK Flexible Sigmoidoscopy Screening Trial reported on variability among the examiners performing FSG examinations for that study.[9] They found large variations in polyp detection and adenoma detection rates across examiners. In this study, we analyzed the variability in FSG performance among examiners in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, a multicenter, ongoing, randomized trial evaluating, for the colorectal component, the effect of FSG screening on colorectal cancer mortality. We analyze here variability among examiners in the rates of positive and inadequate screens, as well as the rates of polyp and adenoma detection.

Methods

Randomization of men and women aged 55–74 years to the screened or usual care arm of the PLCO trial began in November of 1993 and was completed in July of 2001. The 10 PLCO centers were located in the following cities: Washington, DC, Pittsburgh, PA, Birmingham, AL, Detroit, MI, Marshfield, WI, Minneapolis, MN, St. Louis, MO, Denver, CO, Salt Lake City, UT, and Honolulu, HI, and enrolled a total of almost 155,000 patients. Patients in the screened arm received FSG at year 0 and year 5 (patients randomized before the middle of 1995 received FSG at year 3 instead of year 5). Men in the screened arm also received annual prostate-specific antigen tests, digital rectal examinations, and chest radiographs whereas women received annual cancer antigen 125

Abbreviations used in this paper: CI, confidence interval; CV, coefficient of variation; FSG, flexible sigmoidoscopy; GE, gastroenterologist; NP, nurse practitioner; OR, odds ratio; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

1542-3565/05/$30.00
PII: 10.1053/S1542-3565(05)00286-7

(CA-125) tests, transvaginal ultrasound examinations, and chest radiographs. A baseline questionnaire was administered around the time of randomization. The details of the design, conduct, and recruitment of the trial have been reported previously.[6] The study was approved by the institutional review boards of each study center. Eligibility criteria for the trial included no current treatment for cancer (except for basal or squamous cell skin cancer); no known prior cancer of the colorectum, prostate, lung, or ovaries; and, for patients randomized after April, 1995, no colonoscopy, sigmoidoscopy, or barium enema in the past 3 years.

An FSG examination was considered positive if the examiner noted a polypoid lesion or mass. The location, shape, and size (largest diameter) of each of the 4 largest lesions were recorded by the examiner. Lesions were not removed and did not undergo a biopsy examination. An FSG examination was considered inadequate if no lesion was found and the depth of insertion was less than 50 cm or less than 90% of the mucosa could be visualized. Patients were referred to their personal physicians for evaluation of screen-detected abnormalities. Information on diagnostic follow-up evaluation was collected using trained medical record abstractors who recorded the pathology, size, and location of each lesion found on endoscopy. We define a false-positive examination as a positive screening examination in which no distal polyp was found on endoscopic follow-up evaluation; in this article *distal* refers to the descending colon, sigmoid colon, and rectum. An adenoma that was 10 mm or greater in diameter, villous or tubulovillous, or had severe dysplasia was classified as advanced.

The PLCO protocol required that all FSG examiners be either physicians, registered nurses, nurse practitioners (NPs), or physician assistants. All examiners, except board-certified gastroenterologists or physicians with hospital privileges to perform FSG or colonoscopy, underwent training and certification by PLCO staff. Training and certification involved watching a videotape, observing 10 procedures and performing 10 practice procedures (ie, where one learns how to operate manual controls and withdraw the scope), and then performing as many successful training procedures under the guidance of a training gastroenterologist (a minimum of 25) as deemed necessary to show competence. Because the majority of PLCO examinations were performed by either gastroenterologists (GEs) or NPs, we limited the analysis to these groups. In addition, because there were a number of examiners who performed only a small number of examinations, we included in the analysis only results from examiners who performed at least 100 examinations. Only examinations from the baseline screening round were analyzed.

Statistical Methods

We were interested in estimating the true variability across FSG examiners in rates of screen positivity (and other outcomes). The observed variability in examiner positivity rates may not reflect true underlying variability because of several factors. First, because some examiners performed relatively few examinations there is some random error (noise) associated with the observed examiner rates that tends to make the observed variability an overestimate of the true underlying variability. Second, examiners in this study had a different mix of patients with respect to sex, age, and other factors that correlate with screen positivity; thus, some apparent variability in positivity rates could be caused by different subpopulations of patients for different examiners.

To deal with the earlier-described issues, we used a statistical tool known as mixed models.[10] Mixed models postulate that there are fixed effects that account for the effect of fixed patient or examiner covariates (such as age, sex, or smoking status) on the outcome of interest (eg, screen positivity) as well as random effects that account for the fact that one is sampling (in theory randomly) from a population of examiners and that each examiner in the population may have a different underlying rate of the outcome of interest. For each outcome of interest (positivity rate, inadequate rate), we first ran a full mixed model that incorporated random effects as well as fixed effects for patient age (4 age groups), sex, and smoking status (never, former, current), and examiner credential (NP vs GE). Note that inclusion of other patient covariates had little additional effect on estimates of variability. A backward stepwise procedure then was used to generate the final model. In addition to modeling positivity rates, we also modeled the rate of finding lesions of reported size (by the screening examiner) at least 10 mm in diameter, at least 5 mm in diameter, and at most 4 mm in diameter. The model details are given in Appendix 1. The coefficient of variation (CV) is defined here as the SD of the examiners' underlying rates divided by the mean of the examiners' underlying rates; it is multiplied by 100 and expressed as a percentage. In addition to calculating an overall CV across all examiners, CVs also were calculated separately for GEs and for NPs.

About 25% of positive-screen patients did not have endoscopic follow-up evaluation. We calculated the false-positive rate (false-positive screens over all positive screens) based only on positive screens with endoscopic follow-up evaluation. In addition, we calculated the distal polyp and distal adenoma detection rates by multiplying the proportion of screens that were positive by the proportion of positive screens with follow-up evaluation that had distal polyps or adenomas identified. A similar mixed model as described earlier was used to estimate variability in polyp and adenoma detection rates, and to estimate the correlation between screen-positive rates and false-positive rates (see Appendix 1).

Results

A total of 57,124 T0 FSG examinations were performed through June of 2000 by 158 different examiners. Table 1 shows the examinations performed by the 31 GEs and 33 NPs who performed at least 100 examinations. These 64 examiners at 10 clinical centers performed 49,955 examinations (87% of the total). The GEs came from 6 different screening centers and the NPs came from 8 different centers. Twelve NPs and 2 GEs