

Validation of Self-reported Colorectal Cancer (CRC) Screening in a Study of Ethnically Diverse First-Degree Relatives of CRC Cases

Roshan Bastani, Beth A. Glenn, Annette E. Maxwell, Patricia A. Ganz, Cynthia M. Mojica, and L. Cindy Chang

School of Public Health and Jonsson Comprehensive Cancer Center, University of California at Los Angeles, Los Angeles, California

Abstract

Background: Evidence about the accuracy of self-reports of colorectal cancer (CRC) screening is lacking. We conducted a validation protocol in a randomized trial to increase CRC screening among high-risk individuals.

Methods: First-degree relatives ($n = 1,280$) of CRC cases who were due for CRC screening were included in the parent trial. All subjects who completed the follow-up interview ($n = 948$) were asked to participate in validation activities. Self-reports of receipt of CRC screening during the 12-month study period were verified via physicians.

Results: Although 60% ($n = 567$) verbally agreed, only 171 subjects (18% of original sample) returned the signed validation form with the physician name and contact information and a medical information release statement. The signed forms were mailed to physicians with a \$10 incentive and the request to list the dates of

recent CRC screening tests. One hundred twenty-three physicians (72% of physicians contacted, 13% of original sample) returned completed validation forms. Rates of agreement were low across all three screening types with physicians verifying self-reported screening for 29% of fecal occult blood testing, 56% of sigmoidoscopy, 55% of colonoscopy, and 57% of any screening test.

Conclusion: Validation of self-report using the type of protocol we used for subjects receiving medical care in many community settings may be unfeasible and cost inefficient. Given the overall low participation rate in validation activities and considerable challenges in collecting high quality data, conclusions about the accuracy of self-reported CRC screening are difficult to make based on the results of this study. (Cancer Epidemiol Biomarkers Prev 2008;17(4):791-8)

Introduction

The most commonly used outcome in cancer screening intervention studies is self-report. However, prior research has found that self-report of cancer screening may be inaccurate and prone to biases (1, 2). Therefore, validation of self-reports through medical chart audit has long been viewed as the "gold standard" for assessing the veracity of self-reports. Most validation studies have focused on breast and cervical cancer screening, and therefore, relatively little information is available about the accuracy of self-report for colorectal cancer (CRC) screening (1). Results of the handful of validation studies focused on CRC screening have yielded inconsistent results. Several studies have found that self-reports of endoscopy may be fairly accurate, with high concordance between self-report and medical records (2-5). The accuracy of self-report may be higher for colonoscopy than sigmoidoscopy (4, 6), and concordance

between self-report and medical records has been found to be substantially lower for fecal occult blood testing (FOBT; refs. 2-4). Although overreporting is suspected across CRC screening tests, one study found that subjects underreported versus overreported FOBT receipt (7).

The majority of studies validating self-report of cancer screening have been conducted in health maintenance organizations, within a single health care system, or with patients enrolled in large cohort studies (2-5). In these settings, researchers can often directly access medical charts, including billing information, procedure documentation, and physician notes. However, many cancer screening intervention studies are conducted outside of these settings. Cancer screening interventions are often delivered remotely to individuals recruited through cancer registries, surname lists, and random digit dialing. In these settings, subjects are likely to receive cancer screening at numerous diverse settings dispersed over broad geographic areas. In addition, many cancer screening intervention studies are conducted in community settings. This approach is necessary to reach populations that may not already be connected with the health care system, including low-income, uninsured, and ethnic minority populations. Nonclinical settings for cancer screening intervention research have included churches, community-based organizations, social clubs,

Received 9/27/07; revised 1/23/08; accepted 1/30/08.

Grant support: National Cancer Institute (NIH) grant 1R01 CA75367.

Requests for reprints: Roshan Bastani, University of California at Los Angeles School of Public Health, 650 Charles Young Drive South, Room A2-125 CHS, Los Angeles, CA 90095-6900. Phone: 310-206-9266; Fax: 310-206-3566. E-mail: bastani@ucla.edu

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-2625

barber shops, and work sites. The majority of these studies do not arrange for subjects to receive screening from any particular site. Consequently, subjects may be screened in various settings within the community, including individual physician offices, community clinics, health fairs, and county health department facilities. Researchers in these settings face many challenges to validating self-reports of screening, including being dependent on health care providers to provide the necessary information and having to navigate many different health care settings to receive the requested information.

We describe here the results of validation activities that were conducted in the context of a study that evaluated the effect of a print and telephone counseling intervention on CRC screening rates in a sample of first-degree relatives recruited through a population-based cancer registry. Only one prior study in the literature has validated self-reported CRC screening among individuals with a family history of CRC. That study was conducted in Canada (4) and found that colonoscopy was reported very accurately, but that, in comparison, self-reports of sigmoidoscopy and FOBT were less reliable. In addition, no prior studies have assessed the validity of self-reported CRC screening in the context of a screening intervention trial.

Materials and Methods

Brief Description of the Parent Intervention Trial.

The parent study was a two-group randomized intervention trial to increase CRC screening within an ethnically diverse sample of first-degree relatives of CRC cases. First-degree relatives were recruited by initially contacting a random sample of CRC cases identified through the state-wide California Cancer Registry. African-American, Asian, and Latino cases were oversampled to achieve ethnic diversity in the study population. Cases were asked to provide information about their first-degree relatives. A total of 1,645 cancer cases identified 5,073 first-degree relatives. Baseline telephone interviews were conducted with eligible first-degree relatives referred to the study. Eligibility included the following: 40 to 80 years of age; living in the United States or Canada; English or Spanish speaking; sister, brother, or child of case; no personal history of CRC; and no history of very high-risk syndromes such as inflammatory bowel disease, hereditary nonpolyposis CRC, and familial adenomatous polyposis. Of the 3,667 eligible relatives, 2,595 (71%) completed the baseline survey. Of these, 49% ($n = 1,253$) had obtained CRC screening recently, defined as a FOBT in the past year, and/or sigmoidoscopy in the past 5 years, and/or colonoscopy in the past 10 years. Another 2% ($n = 62$) were excluded from this analysis because ethnicity was missing or they were not African-American, Asian, Latino, or White. The remaining 1,280 first-degree relatives (siblings or children) who did not report recent CRC screening constituted the analytic sample for the trial. A two-group nested design was used. All subjects within the same family were assigned to the same study arm to decrease potential contamination. The intervention group received a tailored print intervention within

2 weeks of the baseline interview. Both groups were contacted 6 months after baseline to assess screening receipt, and intervention subjects not adherent to screening at 6 months received a telephone counseling intervention appended to the end of the telephone interview. Twelve-month follow-up interviews were conducted with both groups. Of the 1,280 nonadherent subjects randomized to the intervention, 80% ($n = 1,030$) completed the 6-month interview and 74% ($n = 948$) completed the 12-month interview. The control group received no intervention until after the trial was completed. Study data were collected from September 1999 to April 2004. The study protocol was approved by the University of California at Los Angeles Institutional Review Board. Additional details about the study results are forthcoming.

Validation Procedure. At the end of the 12-month follow-up telephone interview, all subjects were asked to allow the study team to contact their health care provider to verify receipt of screening during the study period. Subjects were informed that we would like to contact their provider regardless of whether they received a recent screening test and that we would like to obtain contact information (i.e., name, address, and telephone number) for the provider (physician or clinic) they had seen for CRC screening or their usual provider (if they were not screened during the study period). All subjects who verbally agreed to the validation activities (567 of 948 subjects at follow-up) were sent a validation form to sign and return to our office. The top portion of the form was pre-populated with participant information such as name and address and included a place for the participant's signature indicating agreement to participate in the validation protocol. The bottom portion of the form was for the medical provider to indicate participant receipt of CRC screening. Health care providers were sent the signed validation form, a cover letter briefly explaining the study and the need for validation, a study abstract, and \$10 to offset the cost of completing and returning the form (e.g., staff time). Health care providers who did not return the validation form were contacted by telephone. Multiple attempts were made to contact each provider by telephone. Providers were allowed to submit the information via mail and fax or provide information over the telephone.

Measures

Self-reported CRC Screening. At the 6- and 12-month follow-up interviews, subjects were asked if they had received a FOBT, sigmoidoscopy, or colonoscopy since their last telephone interview (i.e., baseline or 6-month interview). A brief definition of each test was provided before assessing screening receipt. Subjects reporting any of the tests at the 6- or 12-month follow-up were asked if it was for screening or diagnostic purposes. Only subjects who had had a test for screening purposes (routine checkup, no symptoms, and no prior abnormal findings) were considered "screened" for both the intervention and validation studies.

Physician Report of CRC Screening. The bottom half of the validation form included space for the health care provider to indicate the two most recent dates of CRC screening (FOBT, sigmoidoscopy, and colonoscopy). Providers were also asked to indicate the reason for the

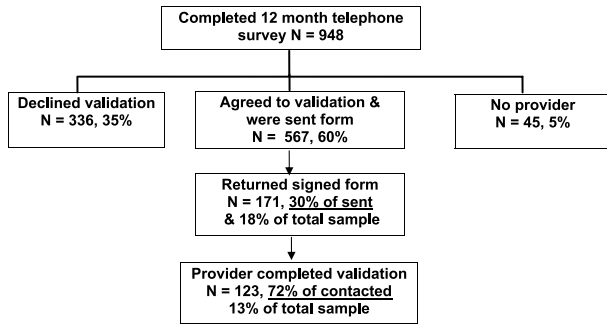


Figure 1. Validation process.

test: the screening of an asymptomatic patient versus an indicated procedure following symptoms or abnormal finding on test. They were also asked to indicate whether the test result was positive or negative.

Analyses. We compared subgroups of subjects created by inclusion versus exclusion from the various steps involved in the validation process (depicted in Fig. 1) using χ^2 tests to assess bivariate relationships and logistic regression for multivariate comparisons. All variables listed in Table 1 were included in the multivariate analyses, except for income (due to substantial missing data) and country of birth (applicable only to Latinos and

Asians). Because all subjects within the same family were randomized to the same study arm (the average family size was 1.4), we used the GENMOD procedure, with a logit link, in Statistical Analysis System (Windows version 9.1) for our multivariate analyses. This procedure fits models to correlated responses using generalized estimating equations equivalent to logistic regression with SEs adjusted for correlated data. Concordance was calculated by assessing the percent total agreement (% of instances when self-reports matched provider reports) between self-report and medical records for receipt of FOBT, sigmoidoscopy, and colonoscopy during the 12-month study period. We also calculated sensitivity (% true positives), specificity (% true negatives), positive predictive value (PPV; % of self-reports of screening verified by provider report), and negative predictive value (NPV; % of self-reports of "no screening" verified by provider report). The exact procedure in Stata 9.0 was used to calculate the 95% confidence intervals (95% CI) for the various measures of agreement (8). Fisher's exact test was used to compare the differences in measures of agreement between the intervention and control group on FOBT, colonoscopy, and any screening.

Results

Response Rates. Figure 1 depicts the response rates for the validation activities. Of the 948 individuals who

Table 1. Subject characteristics by participation in validation activities

Characteristic	Follow-up sample (n = 948), n (%)	Verbally agreed (n = 567), n (%)	Verbally declined* (n = 381), n (%)	Subject returned medical release (n = 171), n (%)	Subject did not return medical release (n = 396), n (%)	Provider did not return info (n = 48), n (%)	Final validation sample (n = 123), n (%)	Not validated (n = 825), n (%)
	A	B	C	D	E	F	G	H
Ethnicity								
White	273 (29)	165 (29)	108 (28)	58 (34)	107 (27)	17 (35)	41 (33)	232 (28)
African-American	204 (22)	117 (21)	87 (23)	26 (15)	91 (23)	9 (19)	17 (14)	187 (23)
Latino	279 (29)	162 (29)	117 (31)	41 (24)	121 (31)	11 (23)	30 (24)	249 (30)
Asian	192 (20)	123 (22)	69 (18)	46 (27)	77 (19)	11 (23)	35 (28)	157 (19)
Age (y)								
40-49	487 (51)	285 (50)	202 (53)	95 (56)	190 (48)	29 (60)	66 (54)	421 (51)
50-64	342 (36)	217 (38)	125 (33)	63 (37)	154 (39)	14 (29)	49 (40)	293 (36)
≥65	119 (13)	65 (11)	54 (14)	13 (8)	52 (13)	5 (10)	8 (7)	111 (13)
Sex								
Male	405 (43)	250 (44)	155 (41)	73 (43)	177 (45)	25 (52)	48 (39)	357 (43)
Marital status								
Married	644 (68)	395 (70)	249 (65)	123 (72)	272 (69)	31 (65)	92 (75)	552 (67)
Education								
Some college	590 (63)	361 (64)	229 (60)	129 (75)	232 (59)	34 (71)	95 (77)	495 (60)
Income								
≥\$50,000	445 (50)	271 (50)	174 (49)	97 (58)	174 (46)	22 (47)	75 (62)	370 (48)
Insurance								
Yes	851 (90)	536 (95)	315 (83)	165 (97)	371 (94)	46 (96)	119 (98)	732 (89)
Born in the United States†								
Yes	299 (63)	187 (66)	112 (60)	68 (78)	119 (60)	17 (77)	51 (78)	248 (61)
Group status								
Intervention	489 (52)	305 (54)	184 (48)	94 (55)	211 (53)	27 (56)	67 (54)	422 (51)
Self-reported screening								
Yes	286 (30)	241 (43)	45 (12)	102 (60)	139 (35)	34 (71)	68 (55)	218 (26)

NOTE: χ^2 tests were completed to compare the following columns: (a) B versus C, (b) D versus E, (c) F versus G, and (d) G versus H. Adjacent areas are shaded to indicate statistically significant differences.

* "Verbally declined" category includes 45 participants who reported having no physician.

† For Latinos and Asians only.

CRC Screening Validation among First-Degree Relatives

Table 2. Multivariate analyses predicting participation in validation activities

Characteristic	Verbally agreed vs not (<i>n</i> = 940)		Returned medical release vs not (<i>n</i> = 564)		In final validation sample vs not (<i>n</i> = 940)	
	OR	95% CI	OR	95% CI	OR	95% CI
Ethnicity						
White (reference)	—		—		—	
Asian	1.17	0.78-1.78	1.06	0.62-1.79	1.15	0.67-2.00
African-American	0.90	0.59-1.36	0.54*	0.30-0.98	0.55	0.28-1.09
Latino	1.00	0.69-1.47	0.74	0.45-1.23	0.83	0.48-1.44
Age (continuous)	0.91	0.77-1.06	0.85	0.69-1.06	0.85	0.69-1.05
Men (vs women)	1.18	0.88-1.56	0.87	0.59-1.27	0.76	0.50-1.14
Married (vs. not)	0.98	0.72-1.33	0.96*	0.62-1.48	1.20	0.75-1.91
Some college (vs less)	0.99	0.72-1.36	1.96	1.26-3.07	1.88*	1.15-3.07
Insured (vs not)	2.75*	1.72-4.41	1.75	0.56-5.48	2.71	0.87-8.40
Intervention (vs control)	1.14	0.85-1.53	0.89	0.60-1.32	1.00	0.66-1.51
Self-reported screening (yes vs no)	5.14*	3.54-7.48	3.07*	2.05-4.58	3.46*	2.31-5.18

*Statistically significant at $P \leq 0.05$.

completed the 12-month survey, 567 (60%) verbally agreed to validation and were sent a validation form for signature and return to our office. Of these, 171 (30%) subjects returned the signed form. We were able to validate screening status for 123 (72%) of this group based on information provided by 121 different physicians. Due to the dropout that occurred at the various steps in the process, we were able to complete validation for only 13% of the total sample or 22% of those who verbally agreed to participate in validation.

Characteristics of Validation Sample and Comparison of Participants and Nonparticipants. Characteristics of the total follow-up sample ($n = 948$) appear in the first column of Table 1. Of the 948 subjects in our 12-month follow-up sample, 771 (81.5%) resided in California, and the remaining 177 were distributed among 37 other states across the country (data not shown). Among the subjects residing in California, 231 resided in Los Angeles County. Thus, 30% of the California residents and 24% of the total sample resided in Los Angeles.

To examine the representativeness of our validation sample, we compared the characteristics of the subgroups created by the various steps in the validation process both bivariate (Table 1) and multivariate (Table 2). First, we compared those who verbally agreed to participate in the validation protocol to those who did not (Table 1, columns B versus C). The 45 subjects who indicated that they had no provider were placed in the "did not verbally agree" category because bivariate analyses indicated that these two groups did not differ on demographic characteristics. Subjects who verbally agreed to validation were more likely, than subjects who did not agree, to have health insurance and to self-report CRC screening during the study period. Multi-

variate logistic regression analyses (Table 2) found that insurance status [odds ratio (OR), 2.75] and self-reported screening (OR, 5.14) remained independent predictors of verbally agreeing to validation. Within the subgroup of 567 subjects who agreed to the validation, we compared the 171 subjects who returned the signed validation form to the 396 subjects who did not (Table 1, columns D versus E). Bivariate, subjects who returned the validation form and provided medical release were more likely to be White or Asian compared with Black or Latino, have higher levels of education, have been born in the United States (only calculated for Latinos and Asians), and self-report CRC screening within the study period. Ethnicity (OR, 0.54 for African-American versus White), education (OR, 1.96), and self-reported screening (OR, 3.07) were retained in the multivariate model (Table 2). There were no statistically significant differences between the subgroup for whom physicians returned completed validation forms ($n = 123$) and those subjects for whom physicians did not respond ($n = 48$; Table 1, columns F versus G). Finally, we made a comparison between the 123 subjects for whom we were able to complete the validation process and the remainder of subjects ($n = 825$) in our 12-month follow-up sample (Table 1, columns G versus H). Bivariate analyses indicated that members of the final validation sample ($n = 123$) were more likely than the remainder of subjects ($n = 825$) to be White and Asian, to have higher levels of education and health insurance, to self-report CRC screening during the study period, and to have been born in the United States (among Latinos and Asians only). Education (OR, 1.88) and screening status (OR, 3.46) were the only variables retained in the multivariate analysis.

Table 3. Prevalence of CRC screening by source of report and study sample

Type of screening	Validation sample (<i>n</i> = 123)		Not validated (<i>n</i> = 825)
	Provider report, <i>n</i> (%)	Self-report, <i>n</i> (%)	Self-report, <i>n</i> (%)
FOBT	15 (12)	38 (31)	134 (16)
Sigmoidoscopy	6 (5)	9 (7)	36 (4)
Colonoscopy	25 (20)	38 (31)	81 (10)
Endoscopy	31 (25)	47 (38)	114 (14)
Any screening	42 (34)	68 (55)	218 (26)

Table 4. Sensitivity, specificity, PPV, NPV, and concordance between self-report and provider report by CRC screening test

Type of test	Test receipt by information source				Measures of concordance				
	Self and provider (A) <i>n</i>	Self only (B) <i>n</i>	Provider only (C) <i>n</i>	Neither (D) <i>n</i>	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Concordance (95% CI)
FOBT	11	27	4	81	73 (45-92)	75 (66-83)	29 (15-46)	95 (88-99)	75 (66-82)
Intervention	7	17	1	42	88 (47-100)	71 (58-82)	29 (13-51)	98 (88-100)	73 (61-83)
Control	4	10	3	39	57 (18-90)	80 (66-90)	29 (8-58)	93 (81-99)	77 (64-87)
Sigmoidoscopy	5	4	1	113	83 (36-100)	97 (91-99)	56 (21-86)	99 (95-100)	96 (91-99)
Colonoscopy	21	17	4	81	84 (64-95)	83 (74-90)	55 (38-71)	95 (88-99)	83 (75-89)
Intervention	12	13	3	39	80 (52-96)	75 (61-86)	48 (28-69)	93 (81-99)	76 (64-86)*
Control	9	4	1	42	90 (55-100)	91 (79-98)	69 (39-91)	98 (88-100)	91 (80-97)*
Endoscopy	29	18	2	74	94 (79-99)	80 (71-88)	62 (46-75)	97 (91-100)	84 (76-90)
Any screening	39	29	3	52	93 (81-99)	64 (53-75)	57 (45-69)	95 (85-99)	74 (65-81)
Intervention	24	21	1	21	96 (80-100)	50 (34-66)*	53 (38-68)	95 (77-100)	67 (55-78)
Control	15	8	2	31	88 (64-99)	79 (64-91)*	65 (43-84)	94 (80-99)	82 (70-91)

NOTE: Concordance measures were assessed using the following formulas: sensitivity = $[A / (A + C)] \times 100$; specificity = $[D / (B + D)] \times 100$; PPV = $[A / (A + B)] \times 100$; NPV = $[D / (C + D)] \times 100$; concordance = $[(A + D) / (A + B + C + D)] \times 100$.

*Statistically significant difference between intervention and control groups at $P < 0.05$ (two sided), Fisher's exact test.

Stratified bivariate logistic regression analyses conducted separately among Latinos and Asians indicated that country of birth was statistically significantly associated with participation in validation activities only among Asians (data not shown): compared with foreign-born Asians, those born in the United States were more likely to agree verbally to be part of the validation study (OR, 2.14; 95% CI, 1.17-3.91; $P < 0.05$), to return the signed medical release form (OR, 4.23; 95% CI, 1.79-9.98; $P < 0.05$), and to be in the final validation sample (OR, 4.74; 95% CI, 2.00-11.27; $P < 0.05$). Country of birth was not a statistically significant predictor of participation among Latinos.

CRC Screening Rates. Table 3 displays CRC screening rates obtained in the study for the validation sample ($n = 123$) by provider and self-reports as well as the self-reported screening rates for the 825 subjects for whom validation was not conducted. For every type of screening test, except sigmoidoscopy, the self-reported rate of screening for the validation sample was substantially higher than the provider rate of screening. For example, 31% of subjects self-reported FOBT screening, whereas the providers reported 12% screening. Furthermore, self-reported rates of screening for the validation sample were higher than for the nonvalidated sample for all tests, except sigmoidoscopy.

Agreement between Self-report and Provider Report. Table 4 displays the percent agreement between self-report and provider report for the three types of CRC screening tests separately and for any endoscopy (sigmoidoscopy or colonoscopy). Overall concordance between provider and self-report ranged from "fair" (74%) for any screening to "excellent" (96%) for sigmoidoscopy based on criteria suggested by Tisnado et al. (9). Observed NPVs (proportion of "not screened" self-reports confirmed by physicians as not screened) for all screening tests were excellent. In contrast, the PPVs (proportion of screened self-reports confirmed by physician as screened) for all screening tests were poor. The lowest PPV was observed for FOBT. Provider report corroborated self-report of a FOBT in only 29% of cases.

PPVs for sigmoidoscopy and colonoscopy were substantially higher (56% and 55%, respectively) but still poor. Because our work and previous research (3) has found that patients are unsure of the type of endoscopic procedure they received, we examined the rate of agreement for receipt of any endoscopy within the study period (i.e., colonoscopy or sigmoidoscopy). The PPV for any endoscopy was slightly higher (62%) than that for either procedure separately, suggesting that a few subjects may have inaccurately reported which type of endoscopy they had undergone. The PPV for "any screening" (FOBT and/or sigmoidoscopy and/or colonoscopy) was also poor. We also assessed whether provider corroboration of self-reported screening was influenced by including dates of screening reported by physicians that were outside of the study period. Previous research has found that patients often inaccurately remember the timing of medical tests (10). We asked providers to indicate the two most recent dates on which subjects received any of the three cancer screening tests but did not specify the study period. The PPV increased only slightly (<3% on average; data not shown) when comparing self-reports with provider reports of screening that occurred outside the study period. This suggests that inaccurate recall of the timing of tests did not contribute statistically to the poor PPVs observed.

Agreement by Intervention and Control Condition. Because this validation study was conducted in the context of a screening intervention trial, we computed the various measures of agreement, by intervention versus control, for FOBT, colonoscopy, and any screening, although small cell sizes warrant caution in interpretation of findings. We did not compute similar information for sigmoidoscopy due to the extremely low prevalence of this procedure in our sample (e.g., one provider report and four self-reports). For FOBT, overall concordance was very similar by intervention versus control (73% and 77%) as was the PPV. Observed sensitivity and specificity seemed to differ by intervention group status; however, substantial overlap in the confidence intervals indicated lack of statistical significance. For colonoscopy, most measures of agreement

were lower in the intervention versus the control group, although the differences were not statistically significant, except for the overall concordance rate (76% intervention versus 91% control group). Similar findings were observed for any screening test, with the only statistically significant difference being for specificity (50% intervention versus 79% control).

Discussion

Our most substantial challenge was obtaining participant written consent to inclusion in the validation substudy. Consequently, we were able to complete the validation with only 13% of our follow-up sample. The parent trial was conducted entirely via mail or telephone contact. Therefore, to obtain written informed consent, as required by our Human Subjects Protection Committee, we obtained verbal consent at the end of our last follow-up telephone interview (12 months), mailed a validation form to those agreeing, and had to rely on subjects mailing back the signed form. The largest dropout occurred at this last step. Although 60% of subjects verbally agreed to the validation, only 30% of those who agreed or 18% of the total sample actually returned the completed signed form. This low response proportion is somewhat surprising given that the vast majority of the sample had already participated in three study interviews and about half had received educational materials and a telephone counseling session as part of the intervention. No telephone reminders were included for this portion of the trial due to resource limitations, which may explain the low response. In addition, all subjects were aware that they were contacted for this study because they were at increased risk for CRC due to family history and that screening was important for early detection. Therefore, subjects who were not screened during the study period (70% according to self-report) may have been hesitant to allow study staff to check with their providers to verify screening, resulting in lower proportions of verbal agreement to participate in validation as well as a lower proportions of return of the medical release forms. This hypothesis is consistent with our findings that subjects who self-reported screening were statistically significantly more likely to participate in validation activities. Subjects may also have been hesitant to allow us to contact their providers due to concerns that it might prompt providers to encourage subjects to be screened, something subjects may or may not be interested in pursuing. In contrast to the low response rate among subjects, the response rate among physicians was fairly high at 72%. This is probably due to the aggressive telephone follow-up that was implemented for this phase of the validation study. We were able to cover the cost of the physician follow-up due to the small numbers involved. This was a labor-intensive process because almost every subject identified a unique physician. Had the subject response rate been higher, it is unlikely that we could have implemented this intensive physician reminder protocol.

Given the low participant response, it was important to examine how the validation sample may have differed from the overall sample on demographic factors and receipt of screening. Overall, at each step in the process,

subjects who participated in the validation protocol were statistically significantly more socioeconomically advantaged and less likely to be African-American or Latino than those who declined participation. Among Asian subjects, participants in validation activities were more likely to be born in the United States compared with nonsubjects. Furthermore, subjects who self-reported having received a CRC screening test during the study period were more likely to participate. Given these differences, we cannot generalize the results of our validation to the total sample.

Among the 123 subjects for whom we were able to obtain provider verification of screening status, we observed a wide variation in level of agreement between provider and self-report for the three CRC screening procedures assessed. Agreement also varied by specific screening modality. In general, NPVs were uniformly high for all tests and combination of tests, indicating that provider reports corroborated self-reports when subjects reported no screening. However, PPVs (self-report of screening receipt confirmed by provider report) were low across the board but varied by type of procedure: the PPV was 62% for any endoscopic test and 29% for FOBT. The PPV observed for FOBT in this study was almost identical to that observed by Madlensky et al. (4) who also recruited a sample of first-degree relatives in Canada. This finding of inaccurate self-report of FOBT is reflected in other measures of agreement that we calculated in this study (sensitivity, specificity, and concordance), supports the results of other studies in the literature (2-4), and has been attributed to less thorough documentation of FOBT in medical records and the fact that the test is completed at home and returned to the provider's office or laboratory, increasing the possibility of misfiling or misplacing results. It is also possible that subjects could not accurately recall the physician to whom they returned the FOBT samples. For subjects who had an endoscopy, multiple physicians may have been involved in the process (e.g., primary care physician made the referral, whereas gastroenterologist conducted the procedure); however, we sought validation from only one physician. The overreporting of all three tests that was observed in our study is generally supported by the literature (1, 2, 4-6, 11).

It is important to keep in mind that the findings of this study may be specific to the characteristics of the population we recruited, the time interval over which we assessed screening, and the fact that the data were collected from subjects participating in a screening intervention trial. Most previous studies have been conducted with average-risk participants. It is possible that first-degree relatives, aware of their increased need for screening, may overreport screening more frequently than average-risk populations, although no data are available to directly answer this question. However, the levels of agreement we observed in our study for sigmoidoscopy, colonoscopy, and endoscopy are higher, in almost every instance, than those in a recent study by Schenck et al. (6) in an average-risk sample. Although multiple factors probably contributed to these differences, one contributing factor may be that our study attempted to verify receipt of screening over a much smaller time interval (1 year versus 5 years).

This is one of the first studies to examine the accuracy of self-reports in the context of an intervention trial. This is of particular importance in community intervention trials because the majority of such trials rely on self-reports to assess the efficacy or effectiveness of the interventions being tested. There is little available information about how screening promotion interventions may influence the accuracy of self-report. Although we attempted to examine whether the accuracy of self-report differed by intervention versus control group assignment, we hesitate to draw conclusions given our overall small sample size coupled with very small and unbalanced cell sizes. Our results suggest that, if a bias according to group assignment exists, it is likely to be more prominent for self-report of endoscopy versus FOBT. Overall, intervention and control groups were not statistically significantly different in all but two instances. There was a statistically nonsignificant trend toward higher accuracy in the control group compared with the intervention group, suggesting that intervention trials may overestimate effect sizes if they rely on self-reports to assess screening outcomes. Further research is needed to fully examine this troubling possibility.

The norm in cancer screening validation research is to regard the provider report as the gold standard against which the accuracy of self-report can be evaluated. However, the majority of studies that have evaluated the accuracy of self-report have been conducted in settings very different from the present one. Most studies have taken place within the context of health maintenance organizations (3, 11, 12), within a singular health care system (7), or with subjects in a limited geographic area attending a limited number of health centers (5, 13). In these settings, researchers can often directly access medical charts, including billing information, procedure documentation, and physician notes. In contrast, our subjects were drawn from 38 different states and there was practically no overlap among the identified physicians. We did not have access to subjects' medical records and had to rely on subjects to provide the correct information about their health care provider and to return the signed medical release. We then had to rely on the health care provider to gather information from any pertinent sources that might document screening (i.e., procedural notes, billing documentation, and clinical notes). The accuracy of provider report in this situation is not clear. Furthermore, validation in nonclinical settings in which subjects receive care from numerous providers has low feasibility. The cost involved in conducting a validation study in such a situation may not be sustainable within the budget limitations of a typical community intervention trial. For the particular protocol we used, further research may be warranted to determine whether investing in more intensive follow-up with subjects who did not return the signed consent forms could result in a more representative study sample.

The major limitation of our study was that we were able to conduct validation activities on only a very small proportion (13%) of our intervention trial sample, thus calling into question the representativeness of our findings. Future studies will need to invest more resources into achieving a higher subject response rate.

Repeated mail and/or telephone reminders accompanied by monetary or other incentives for returning medical release forms may be required. The fact that we used aggressive follow-up accompanied by a \$10 incentive for physicians may have been responsible for the higher provider response rate (72%) achieved. Providers may also have been motivated to provide the information because the validation form included their patient's signature. If resource constraints are an issue, it may be more efficient to focus intense validation efforts on a subsample of the intervention trial sample rather than trying to obtain validation data on all subjects. For example, validation studies could select all subjects self-reporting screening and a random subsample of other subjects. Validation studies may also be more feasible if associated with community trials that take place in geographic areas where there are fewer providers and significant proportions of subjects share the same providers or health care facilities. This would increase the familiarity of the study team with the protocol for obtaining medical record information from various health care settings and may increase the responsiveness on the part of both subjects and provider offices. In addition, although validation of self-report of screening through medical record reviews may in theory be a goal for every screening study, it may be more practical to gather information about the accuracy of self-report from studies that primarily focus on this issue and thus have the resources available to implement a rigorous protocol (14). Funding agencies will need to be more willing to provide additional resources, over and above those available for implementing the intervention trial, to include validation protocols in community intervention studies.

The results of our study seriously call into question the feasibility of attempting to conduct validation of self-reports in community settings in the United States in which subjects are dispersed over a very broad geographic area and receive care from numerous nonoverlapping providers and there are no centralized records. In this context, direct access to patient medical records is often not available and instead researchers must rely on the voluntary participation of physicians who may not be familiar with the investigators or their home institutions. Further research is needed to develop feasible and cost-effective validation protocols in such situations or at the very least to establish consensus guidelines about the appropriate ways to assess screening.

References

1. Warnecke RB, Sudman S, Johnson TP, O'Rourke D, Davis AM, Jobe JB. Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears, clinical breast examinations, and mammograms. *Am J Epidemiol* 1997;146:982-92.
2. Vernon SW, Briss PA, Tiro JA, Warnecke RB. Some methodologic lessons learned from cancer screening research. *Cancer* 2004;101:1131-45.
3. Hall HI, Van Den Eden SK, Tolsma DD, et al. Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Prev Med* 2004;39:27-35.
4. Madlensky L, McLaughlin J, Vivek G. Comparison of self-reported colorectal cancer screening with medical records. *Cancer Epidemiol Biomarkers Prev* 2003;656:659-62.

CRC Screening Validation among First-Degree Relatives

5. Hoffmeister M, Chang-Claude J, Brenner H. Validity of self-reported endoscopies of the large bowel and implications for estimates of colorectal cancer risk. *Am J Epidemiol* 2007;130:136–66.
6. Schenck AP, Klabunde CN, Warren JS, et al. Data sources for measuring colorectal endoscopy use among Medicare enrollees. *Cancer Epidemiol Biomarkers Prev* 2007;16:2118–27.
7. Lipkus IM, Samsa GP, Dement J, et al. Accuracy of self-reports of fecal occult blood tests and test results among individuals in the carpentry trade. *Prev Med* 2003;513:519–37.
8. Clopper C, Pearson S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404–13.
9. Tisnado DM, Adams JL, Liu H, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care* 2006;44:132–40.
10. McPhee SJ, Nguyen TT, Shema SJ, et al. Validation of recall of breast and cervical cancer screening by women in an ethnically diverse population. *Prev Med* 2002;463:473–35.
11. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. *J Natl Cancer Inst* 1993;85:566–70.
12. Caplan LS, McQueen DV, Qualters RQ, Leff M, Garrett C, Calonge N. Validity of women's self-reports of cancer screening test utilization in a managed care population. *Cancer Epidemiol Biomarkers Prev* 2003;1182:1187–12.
13. Paskett ED, Tatum CM, Mack DW, Hoen H, Case LD, Velez R. Validation of self-reported breast and cervical cancer screening tests among low-income minority women. *Cancer Epidemiol Biomarkers Prev* 1996;721:726–5.
14. Vernon S, Tiro J, Vojvodic R, et al. Reliability and validity of a questionnaire to measure colorectal cancer screening behaviors: does mode of survey administration matter? *Cancer Epidemiol Biomarkers Prev* 2008.