

Validity of Four Self-reported Colorectal Cancer Screening Modalities in a General Population: Differences over Time and by Intervention Assignment

Resa M. Jones,^{1,2} Steven J. Mongin,³ DeAnn Lazovich,⁴ Timothy R. Church,³ and Mark W. Yeazel⁵

¹Department of Epidemiology and Community Health, School of Medicine and ²Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia and ³Division of Environmental Health Sciences, School of Public Health, ⁴Division of Epidemiology and Community Health, School of Public Health, and ⁵Department of Family Medicine and Community Health, School of Medicine, University of Minnesota, Minneapolis, Minnesota

Abstract

Little is known about the validity of self-reported colorectal cancer screening. To date, few published studies have validated all four screening modalities per recommended guidelines or included a general population-based sample, and none has assessed validity over time and by intervention condition. To estimate the validity of self-reported screening, a random sample of 200 adults, ages ≥ 50 years, was selected from those completing annual surveys on screening behavior as part of an intervention study. Approximately 60% of the validation sample authorized medical record review. Sensitivity, specificity, and positive and negative predictive values were calculated for baseline and year 1 follow-up reports for each test and for overall screening adherence. Sensitivity at baseline ranged from 86.9% (flexible sigmoidoscopy) to 100% (colonoscopy). Sensitivity at follow-up was slightly lower.

Adjusting for validity measures, the sample over-reported screening prevalence at baseline for each of the four modalities. At follow-up, overreporting was greatest for fecal occult blood test (13.0%). Overreporting across intervention conditions was highest for fecal occult blood test (10.8% for control; 24.8% for the most intense intervention) and overall screening adherence (10.9% for control; 14.3% for the most intense intervention). Sensitivity and specificity of self-reported colorectal cancer screening compared with medical records were high; however, adjusting self-reported screening rates based on relative error rates reduced screening prevalence estimates. Those exposed to more intense interventions to modify screening behavior seemed more likely to overestimate their screening rates compared with those who were not exposed. (*Cancer Epidemiol Biomarkers Prev* 2008;17(4):777–84)

Introduction

Compared with the information on mammography and Pap smear, less is known about the validity of self-reported colorectal cancer screening. Further, of the 10 previously published studies (1–10) examining the accuracy of colorectal cancer screening self-report, only 1 has included a general population-based sample (1), and none has assessed validity of self-report over time. Further, until now (11, 12), none has attempted to validate all modalities of colorectal cancer screening currently recommended for average-risk adults by the American Gastroenterological Association and American Cancer Society [i.e., annual fecal occult blood test (FOBT), flexible sigmoidoscopy within 5 years, colonoscopy

within 10 years, or barium enema within 5 years; refs. 13, 14]. Consequently, with the exception of the study by Partin et al. (11) in this issue of *Cancer Epidemiology, Biomarkers & Prevention*, the validity of a composite measure of overall screening adherence according to recommended frequency has not been estimated.

The evaluation of the validity of self-report is critical as most epidemiologic and behavioral intervention studies rely on self-reported data to estimate the prevalence of screening and to assess the effectiveness of interventions. In addition, self-report is the means by which surveys such as the National Health Interview Survey and the Behavioral Risk Factor Surveillance System monitor screening prevalence in the United States. The reliance on self-report has become even more essential since the adoption of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and additional restrictions associated with the Act, which took effect in 2003, making access to medical records more difficult.

Validation studies of self-reported screening that have been published show that men and women do not accurately recall whether they have had the tests (2–6, 8). In addition, forward telescoping—the act of recalling that

Received 5/15/07; revised 1/17/08; accepted 2/19/08.

Grant support: Allina Health Systems Foundation, Minnesota Medical Foundation, and National Cancer Institute training grant PHS T32 CA09607.

Requests for reprints: Resa M. Jones, Department of Epidemiology and Community Health, School of Medicine, Virginia Commonwealth University, 1000 East Clay Street, P. O. Box 980212, Richmond, VA 23298. Phone: 804-628-2519; Fax: 804-828-9773. E-mail: rmjones2@vcu.edu

Copyright © 2008 American Association for Cancer Research.
doi:10.1158/1055-9965.EPI-07-0441

Validity of Self-reported Colorectal Cancer Screening

screening was obtained more recently than it actually was—has been documented (8, 10). If screening behavior is overestimated by self-reports due to telescoping or other factors, inaccurate conclusions may be drawn about screening prevalence or intervention effects.

Although most studies find overreporting, we do not know much about the potential causes and effects of overreporting. Participation in an intervention study could influence participants' recall and result in differentially biased reporting. Two studies with conflicting information have been published examining differential self-report by intervention assignment (7, 15). The findings of Paskett et al. (15) suggest that there may be differential overreporting such that intervention group participants are more likely to overreport mammography screening compared with controls.

In this study, we compared self-reports for four different colorectal cancer screening modalities at two points in time with information obtained from medical records in a subsample of subjects who participated in a colorectal cancer screening intervention study. We addressed the following questions: (a) Are self-reports of the four commonly recommended colorectal cancer screening modalities accurately recalled compared with medical records? (b) Is a composite measure of having any type of screening according to guidelines (i.e., overall adherence) a valid measure of actual screening behavior? (c) What are the estimated screening rates for each modality across time and by intervention condition after adjusting for misclassification of self-reported screening relative to medical records?

Materials and Methods

Wright County Colorectal Cancer Screening Project.

The Wright County Colorectal Cancer Screening Project (i.e., parent study) was a community-based intervention study to increase colorectal cancer screening rates among average-risk adults ages ≥ 50 years conducted in five greater Minnesota counties beginning in March 2000 (16). The 1999 Minnesota State Driver's License and Identification Card data tape was used to randomly select 2,600 adults, ages ≥ 50 years, living in the five rural study counties. A single county received the intervention as the community health foundation of that county agreed to implement the county-wide education and screening promotion campaign that blanketed the county. Participants within the intervention county were randomly assigned to one of three conditions: (a) county-wide education and promotion, (b) county-wide education and promotion with direct mailing of FOBT kits without reminders, and (c) county-wide education and promotion with direct mailing of FOBT kits and reminder mailings of FOBT kits. Four other counties were selected as control counties because they, too, were largely rural and had a relatively similar proportion of residents ages ≥ 50 years. To limit contamination, control counties were not contiguous with the intervention county. The main goal of the intervention was to increase overall colorectal cancer screening rates among those who were ages ≥ 50 years, regardless of their access to or use of medical services.

All participants in the parent study completed a mailed, self-administered questionnaire at baseline and 12 months

later, with telephone follow-up if the self-administered survey was not received (16, 17). Using these two methods, 80.0% of the eligible parent study cohort completed the baseline survey (1,698 of 2,099 estimated eligible), and 1,558 responded to the follow-up survey. More specifically, $\sim 63.0\%$ of the eligible parent study cohort completed self-administered surveys at baseline and follow-up, whereas $\sim 20.0\%$ completed telephone-administered questionnaires at baseline and follow-up. The follow-up intervention results are published elsewhere (17). Briefly, the 1-year absolute percentage changes (i.e., increases from baseline to 1-year follow-up) for self-reported adherence to FOBT use were 16.9% for the direct-mail-FOBT-with-no-reminders group and 23.2% for the direct-mail-FOBT-with-reminders group.

Validation Sample. The original sample of 1,698 parent study respondents was stratified by whether they reported adherence to any colorectal cancer screening modality at baseline. From each of these two strata, 100 were randomly selected for the validation study in proportion to the representation of the counties in the parent study, yielding 200 subjects overall. This study was approved by the University of Minnesota Institutional Review Board.

Data Collection

Self-report Questionnaire. The questionnaire, developed to assess the effectiveness of the parent study intervention, included questions about the four commonly recommended colorectal cancer screening tests and procedures (questionnaire available online).⁶ To help participants differentiate among tests, screening questions were prefaced with a succinct, but distinctive, description of the test. Individuals were asked to report whether they had (a) ever had a stool blood test (home based), (b) ever had a flexible sigmoidoscopy, (c) ever had a colonoscopy, and (d) ever had a barium enema. Each question had seven response categories: (a) no, have never had one; (b) yes, within the last 12 months; (c) yes, more than 1 but less than 5 years ago; (d) yes, more than 5 but less than 10 years ago; (e) yes, 10 or more years ago; (f) yes, but not sure when; or (g) don't know if I have. Although our screening questions were created before the 2004 publication of the National Cancer Institute–recommended core colorectal cancer screening measures (18), they were similar in content and format.

Obtaining Authorization for Release of Medical Records. Participants in the validation study were contacted by telephone to obtain the names and addresses of all health care facilities they had visited and the length of time they had received services from each facility. A HIPAA authorization form was then mailed to respondents who agreed to participate and reminder calls were made if the authorization form was not received within 2 weeks. Of the 200 selected participants, 120 (60.0%) returned a signed form. Responders and nonresponders were similar according to age, sex, marital status, education, income, and intervention condition assignment.

⁶ <http://jncicancerspectrum.oupjournals.org/cgi/content/full/jnci.96/10/770/DC1>

However, self-reported colorectal cancer screening according to guidelines was more common among responders than nonresponders at baseline (59.2% versus 36.3%; $P = 0.001$) and at follow-up (66.7% versus 52.5%; $P = 0.044$).

Medical Record Abstraction. Each Minnesota health care facility identified by consenting participants was visited by research staff to abstract medical records. The entire medical record, both paper and electronic, was made available for the abstractor. If a facility refused on-site abstraction or was located in another state, copies of medical records were obtained by mail or fax. The medical record abstractor and all clinic staff were unaware of each participant's reported screening history.

Data collection forms for the medical record abstraction included preprinted information with each participant's baseline and follow-up survey completion dates. Dates of service for each of the four screening modalities conducted before the survey completion dates were recorded by the abstractor in relation to the baseline and year 1 follow-up self-report survey completion dates. For example, for an individual who completed the baseline and follow-up questionnaires on March 1, 2000 and April 1, 2001, respectively, the abstractor would record the most recent date of each modality before March 1, 2000 and before April 1, 2001. The abstractor reviewed the entire chart to find any documentation of the four procedures.

The mean number of unique medical health systems/clinics from which charts were reviewed per participant was 1.5 (range, 1-4). A total of 93.2% of medical records were directly abstracted on site by study staff, whereas the remainder were obtained from copies provided by health care facilities. The majority (70.8%) of participants had 10 or more years of records available for review. On average, 19.1 years (range, 4-56 years) of appointments and services were available for review per participant.

Statistical Analysis. To assess whether self-reports of the four commonly recommended colorectal cancer screening modalities were accurately recalled compared with medical records as well as whether a composite measure of having any type of screening according to guidelines (i.e., overall adherence) is a valid measure of actual screening behavior, we calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The PPV was determined by assessing, for example, the proportion of all FOBTs reported to have occurred within the last 12 months (per recommendations) that were subsequently confirmed by medical records. The NPV for FOBT, for example, was the proportion of all reports of no tests within the last 12 months that were confirmed by medical records. Because the parent study sample was stratified by whether participants reported any screening adherence at baseline or not before being randomly selected for the validation study, sensitivity and specificity were calculated using the parent study screening positivity rate (p) and the validation sample PPV and NPV. Appendix A provides an explanation of the use of the parent study screening positivity rate to calculate sensitivity and specificity. PPV, NPV, sensitivity, and specificity, for flexible sigmoidoscopy, colonoscopy, and barium enema, were calculated similarly, relative to the time frame of each recommendation (i.e., flexible sigmoidoscopy within the

last 5 years, colonoscopy within the last 10 years, and barium enema within the last 5 years). Dates of screening procedures obtained from medical records were coded so they mirrored the survey response categories for time since the test was completed. Thus, the definition and ultimate coding of screening adherence were identical for the self-reported measures and those obtained from medical records. Specifically, an individual was considered adherent to guidelines if (a) FOBT was completed within the last year, (b) flexible sigmoidoscopy was done within the last 5 years, (c) colonoscopy was done within the last 10 years, or (d) barium enema was done within the last 5 years. Overall adherence was defined as screening by any of the four modalities according to guidelines. Ambiguous survey screening responses for screening done per guidelines (i.e., "yes, but not sure when" and "don't know if I have") were coded as not screened. Overall, ~5% had an ambiguous response depending on the screening end point.

The frequency of telescoping, the act of reporting that screening was obtained in a more recent time interval than documented in the medical records, was also calculated. We compared the self-reported time because each test was completed to the medical record dates of screening, which were coded so they mirrored the survey response categories (i.e., never had test, had test within the last 12 months, had test more than 1 but less than 5 years ago, had test more than 5 but less than 10 years ago, and had test 10 or more years ago). Note that only telescoping that would change the time intervals as reported on the questionnaire can be observed; telescoping within an interval would not be detected.

The adjustment for misclassification was done for each modality and overall adherence across time and by intervention condition across time. We used the self-reported screening rates in the parent study, which were adjusted for nonresponse and eligibility [methods described in detail elsewhere (17, 19, 20)], and applied the PPV and NPV of each test to further adjust the parent study self-reported screening rates to estimate the rates expected from medical record reviews [corrected screening rate = self-report rate \times PPV + (1 - self-report rate) \times (1 - NPV); refs. 21, 22]. This approach (i.e., adjusting for validity measures) quantifies how the validity changes the screening rates in question and provides estimated screening rates for each modality and overall adherence after adjusting for misclassification of self-reported screening relative to medical records. Note, however, that the PPV and NPV are not adjusted for nonresponse. We also calculated the report-to-records ratio using the estimated parent study self-report rates divided by the "corrected" rates (to medical records), which is analogous to that presented by Warnecke et al. (23) where values >1.0 indicate overreporting [report-to-records ratio = (true positives + false positives) / (true positives + false negatives)]. We did not calculate the report-to-records ratio directly from the 2×2 tables of the validation study counts [as did Warnecke et al. (23)] because our validation sample was selected from the parent sample, which had been stratified according to survey-reported screening adherence status at baseline. Thus, our report-to-records ratios incorporate adjustment. All statistical analyses were done using Statistical Analysis System 8.2.

Validity of Self-reported Colorectal Cancer Screening

Results

At baseline, 50.0% of the validation study participants were female and 39.2% were ages ≥ 65 years (mean age was 63.9 years). The majority of the sample (77.0%) was married. Roughly 17.0% had a college degree or higher education, 17.4% had annual household incomes $< \$15,000$, and 2.3% reported not having any type of health care coverage. Approximately 74.2% of the validation sample was in the parent study intervention condition. Overall, at baseline, 29.2% of the sample was adherent to only one test and 20.8% of the sample was adherent to more than one test.

Self-reported Screening Compared with Medical Records. At baseline and follow-up, sensitivity was lowest for barium enema and highest for colonoscopy, whereas specificity was lowest for overall adherence and highest for barium enema (Table 1). PPV was highest for overall adherence at baseline and follow-up, whereas NPV was highest for colonoscopy (Table 1).

At baseline, 17.5% of all validation sample participants reported having FOBT more recently than confirmed by medical records, whereas 22.5% had evidence of this telescoping at follow-up. Roughly 12.5% reported more recent baseline flexible sigmoidoscopy and 10.8% had telescoping at follow-up. Similarly, for colonoscopy, there was evidence of telescoping at baseline (6.7%) and follow-up (7.5%). At both baseline and follow-up, 10.8% reported having barium enema more recently than according to medical records.

Estimated Screening Rates across Time and by Intervention Condition after Adjustment for Misclassification. For all screening procedures, the direction in reporting errors resulted in an overestimate of actual screening done as determined from medical records (Table 2). At baseline, the highest percent change (i.e., percent of overreporting) after adjustment was 10.9% for colonoscopy, and at follow-up, the highest percent change was 13.0% for FOBT. The report-to-records ratio was > 1.0 for all screening modalities and overall adherence to screening per recommendations.

Table 3 provides an example of estimating the differences in overreporting by treatment condition assign-

ment in the overall Wright County Colorectal Cancer Screening Project sample by applying the overall estimated PPV and NPV. When these estimates are taken at face value, at baseline, overreporting across treatment conditions seems to be similar within each modality, but at follow-up, overreporting across treatment conditions is highest for FOBT and overall screening adherence. Overreporting of FOBT and overall adherence seem to be greater at follow-up than at baseline. Similarly, a dose-response relationship seems to exist for overreporting FOBT at follow-up by intensity of intervention assignment within the parent study.

Discussion

This is one of the first studies to assess simultaneously the accuracy of self-report relative to medical records for all four modalities of colorectal cancer screening and to consider screening in relation to recommended guidelines for the time frame within which each test should be done. Further, this is the first study to use validity measures to examine the relative accuracy of colorectal cancer screening self-report across time and by the treatment condition assignment.

Self-reported Screening Compared with Medical Records. Direct comparisons of validity measures are difficult because accuracy in most previous studies was not assessed according to the recommended time frame for each modality as it is in the current study or others in this issue of *Cancer Epidemiology, Biomarkers & Prevention* (11, 12). Given that caveat, the sensitivity of self-reported data in our study is higher (particularly at baseline) than that reported in the meta-analysis by Rauscher et al. (24) in this issue of *Cancer Epidemiology, Biomarkers & Prevention* as well as those reported by Partin et al. (11) and Vernon et al. (12). Similarly, our specificity of self-report is generally higher (particularly at baseline) than found by others (1-3, 5, 9, 24), with the exceptions of Baier et al. (6), Madlensky et al. (8), and Vernon et al. (12) who reported higher specificity for colonoscopy. Partin et al. (11) and Vernon et al. (12) also reported higher specificity for barium enema. In addition, our PPV for

Table 1. Validity of self-reported screening of each type according to recommendations: sensitivity, specificity, PPV, and NPV at baseline and follow-up

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Baseline (%)				
FOBT ($n = 119$)	93.0 (75.1-99.6)	89.7 (85.7-93.6)	60.0 (42.6-75.9)	98.9 (94.8-100.0)
Flexible sigmoidoscopy ($n = 118$)	86.9 (74.4-95.4)	87.1 (81.4-92.4)	72.5 (57.8-84.5)	94.9 (88.4-98.4)
Colonoscopy ($n = 120$)	100.0 (85.6-100.0)	86.6 (82.4-90.9)	60.5 (45.8-74.0)	100.0 (96.2-100.0)
Barium enema ($n = 120$)	73.5 (50.3-90.8)	92.7 (89.5-96.0)	55.0 (34.3-74.6)	97.0 (92.3-99.3)
Overall adherence ($n = 120$)*	95.8 (88.7-99.4)	80.6 (72.8-87.9)	81.7 (71.6-89.5)	95.9 (87.7-99.4)
Year 1 follow-up (%)				
FOBT ($n = 119$)	77.3 (56.8-92.7)	81.4 (77.8-85.6)	40.5 (26.7-55.3)	96.1 (90.1-99.1)
Flexible sigmoidoscopy ($n = 120$)	79.8 (67.2-90.0)	81.1 (75.0-87.3)	64.3 (49.5-77.5)	91.0 (83.4-96.1)
Colonoscopy ($n = 118$)	91.8 (78.8-98.5)	86.0 (80.8-91.2)	66.7 (51.4-79.9)	97.5 (92.2-100.0)
Barium enema ($n = 117$)	48.8 (27.6-71.4)	90.3 (87.3-93.8)	38.9 (19.6-61.0)	93.9 (88.1-97.6)
Overall adherence ($n = 120$)*	93.9 (86.4-98.4)	70.2 (62.0-78.5)	76.3 (66.3-84.6)	92.5 (81.7-98.1)

NOTE: Sensitivity (S_e) and specificity (S_p) can then be computed by the following formulae: $S_e = (PPV \times p) / [PPV \times p + (1 - NPV)(1 - p)]$; $S_p = [NPV \times (1 - p)] / [(1 - PPV) \times p + NPV \times (1 - p)]$.

Abbreviation: 95% CI, 95% confidence interval.

*Overall screening adherence via any one of the four tests per recommendations.

Table 2. Estimated self-reported screening rates per recommendations: estimated parent study self-reported rates and corrected rates using the PPV and NPV

	Self-report rate*	Corrected rate (to medical record) [†]	Δ [‡]	Report-to-records ratio [§]
Baseline (%)				
FOBT (<i>n</i> = 1,697)	21.8	13.9	7.8	1.57
Flexible sigmoidoscopy (<i>n</i> = 1,695)	33.2	27.5	5.7	1.21
Colonoscopy (<i>n</i> = 1,690)	27.7	16.7	10.9	1.66
Barium enema (<i>n</i> = 1,691)	14.4	10.5	3.9	1.37
Overall adherence (<i>n</i> = 1,698)	54.8	46.6	8.2	1.18
Year 1 follow-up (%)				
FOBT (<i>n</i> = 1,558)	26.6	13.6	13.0	1.96
Flexible sigmoidoscopy (<i>n</i> = 1,556)	36.6	29.2	7.4	1.25
Colonoscopy (<i>n</i> = 1,544)	31.6	22.8	8.8	1.39
Barium enema (<i>n</i> = 1,544)	14.1	10.7	3.4	1.32
Overall adherence (<i>n</i> = 1,553)	61.3	49.7	11.7	1.23

*Estimated self-reported screening rates adjusted for parent study eligibility and nonresponse.

[†] Screening rates "corrected" (adjusted) for PPV and NPV of self-reported screening relative to medical records; predictive values were unadjusted for validity study nonresponse due to small sample size.

[‡] Reduction in screening rate after "correction" (adjustment) = Δ = (estimated self-reported screening rate – medical record screening rate).

[§] Because our validation study sample was stratified by survey-reported screening adherence status, the report-to-records ratio was not derived directly from the 2 × 2 table of validation study counts [e.g., as in Warnecke et al. (23)]. Instead, it is the estimated self-report rate divided by the corrected rate (to medical record), which incorporates adjustment.

^{||} Overall screening adherence via any one of the four tests per recommendations.

FOBT is similar to the results in the meta-analysis by Rauscher et al. (24); however, our PPV is higher for flexible sigmoidoscopy and colonoscopy compared with their PPV for endoscopy.

In our study, as in others, sensitivity was highest for colonoscopy (100.0 and 91.8, respectively, at baseline and follow-up). The invasive nature of colonoscopy may ensure that those who had the procedure would better remember having the exam. Further, because a colonoscopy within the last 10 years is considered done according to guidelines, people have a larger window of time within which to report screening and hence need to be less accurate in the timing. Conversely, barium enema had the lowest sensitivity at baseline and year 1 follow-up (73.5 and 48.8). Although care was taken to provide a succinct description of all tests, this finding could be a result of people confusing a barium enema

with barium X-rays for upper gastrointestinal diagnostic purposes. Anecdotally, in the review of medical records, we did see reports for upper gastrointestinal procedures; however, we did not collect such data for this study.

On average, all validity measures were lower at follow-up than at baseline. As discussed in detail later, a possible explanation could be that there was a different pattern of overreporting by intervention group compared with control.

Overall Adherence Compared with Medical Records.

Although the sensitivity for overall adherence per recommendations is high (i.e., >93.9), overall adherence specificity in our study is rather low. Our overall adherence sensitivity (with ambiguous responses coded as not screened) is similar to that of Partin et al. (where ambiguous answers were coded as noncompliant).

Table 3. Example of using PPV and NPV to compute percent overreporting by applying them to the intervention study screening rates per recommendations by treatment group

	Baseline				Year 1 follow-up			
	Percent overreporting				Percent overreporting			
	Rx0* (<i>n</i> = 436)	Rx1 [†] (<i>n</i> = 423)	Rx2 [‡] (<i>n</i> = 436)	Rx3 [§] (<i>n</i> = 403)	Rx0* (<i>n</i> = 409)	Rx1 [†] (<i>n</i> = 404)	Rx2 [‡] (<i>n</i> = 389)	Rx3 [§] (<i>n</i> = 351)
FOBT	7.9	7.6	7.1	8.2	10.8	10.6	19.5	24.8
Flexible sigmoidoscopy	5.7	6.1	5.2	6.0	7.5	7.1	6.9	7.4
Colonoscopy	11.0	11.5	9.7	11.3	8.8	8.7	7.2	10.6
Barium enema	3.9	5.4	3.1	3.3	3.5	3.8	2.7	2.4
Overall adherence	8.1	8.6	7.8	8.7	10.9	12.7	13.3	14.3

NOTE: Predictive values used in computations are averaged over all treatment groups rather than specific to treatment and are not themselves adjusted for nonresponse.

*Rx0 = control counties (in Wright County Colorectal Cancer Screening Project sample).

[†]Rx1 = intervention county, community intervention only (in Wright County Colorectal Cancer Screening Project sample).

[‡]Rx2 = intervention county, direct mailing of FOBT kits with no reminders plus community intervention (in Wright County Colorectal Cancer Screening Project sample).

[§]Rx3 = intervention county, direct mailing of FOBT kits with reminders plus community intervention (in Wright County Colorectal Cancer Screening Project sample).

^{||} Overall screening adherence via any one of the four tests per recommendations.

Validity of Self-reported Colorectal Cancer Screening

However, our baseline and year 1 follow-up specificity are higher than Partin et al.'s (80.6 and 70.2, respectively, versus 65; ref. 11).

When participants stated they had not had any test done according to recommendations, evidence to the contrary was confirmed by medical records. Thus, overall adherence is perhaps more accurate for people who state they have had some type of screening according to recommendations than for those who report not screening according to guidelines for any exam.

Estimated Screening Rates across Time and by Intervention Condition after Adjustment for Misclassification. The report-to-records ratio showed over-reporting bias (i.e., net values >1.0) for all of the screening modalities and overall adherence at baseline and follow-up. In general, our baseline report-to-records ratio calculations for FOBT and flexible sigmoidoscopy are similar to those reported by others in this issue of *Cancer Epidemiology, Biomarkers & Prevention* (11, 12). Conversely, whereas Vernon et al. (12) reported under-reporting of barium enema, we found overreporting but to a much lesser extent than reported by Partin et al. (11).

Rothman and Greenland (25) suggest that the actual misclassification probabilities should be assessed by their effect on the outcomes of interest. Thus, a more informative approach that quantifies how the validity actually changes screening rates based on self-report was included here as a more useful examination of the effect of reporting errors on screening prevalence.

After "correcting" the self-reported screening rates given the validity measures (PPV and NPV), we found that self-reported FOBT screening rates exceeded those estimated by the medical records by 7.8% to 13.0%. Similarly, Rauscher et al. (24) found that self-reported FOBT prevalence exceeded adjusted prevalence by 16% for women and 18% for men. Some caution must be used in interpreting these rates, as FOBT may be more likely to be omitted from the medical record than the other procedures.

The fact that overreporting increased from baseline to follow-up for three (FOBT, flexible sigmoidoscopy, and overall adherence) of the five screening measures and given that the majority of the validation sample was assigned to an intervention group within the parent study, it is important to look at overreporting by treatment condition to evaluate whether differential misclassification exists. We found greater discrepancy with more intense intervention, intended to increase overall screening adherence via FOBT promotion. Specifically, at baseline, FOBT self-report exceeded the rate in the medical records similarly for all conditions in the Wright County Colorectal Cancer Screening Project; however, at year 1 follow-up after the implementation of intervention activities, overreporting increased by increasing intensity of intervention activity. Paskett et al. (15) also found differential overreporting in self-report of mammography among the screening intervention group compared with the control group. However, mammography self-report data were collected by the interventionists immediately after conducting the intervention; thus, there is the likelihood that social desirability influenced respondents' self-reports.

Our findings could be due to differential social desirability or differential recall because the parent study

intervention promoted FOBT. Between baseline and follow-up, the majority of participants in this parent study were exposed to a community intervention to increase colorectal cancer screening rates. Those in the intervention county who were randomized to the community-only intervention were exposed to messages about the importance of colorectal cancer screening; however, in addition to the community intervention, those in the direct mail group without reminders received a FOBT kit and promotional materials and those in the direct mail group with reminders received multiple mailings of FOBT kits and promotional materials. People, who heard messages about the importance of colorectal cancer screening for adults ages ≥ 50 years and received FOBT kits in the mail, could have falsely reported screening at follow-up because they perceived it to be the "correct" or acceptable response.

In adjusting for parent study nonresponse, especially in year 1 follow-up for FOBT, the up-weighting for nonresponders yields a lower estimate of the corrected self-report rate than we find without adjusting for parent study nonresponse. This suggests that if we could have convinced parent study nonresponders to respond, they would have tended to report that they were not adherent to screening recommendations or were less adherent than the parent study responders. It is possible that, rather than admit this fact, they chose to not respond.

Some caution must be used in interpreting these results as our study has some limitations. First, although medical records are generally considered the "gold standard," they are not a perfect tool for assessing screening behavior. If the medical record does not contain all pertinent information, a respondent could indeed report accurately, but the medical record would not be able to confirm screening. During the record review, we found a few progress notes suggesting the completion of an exam; however, no confirmation or test result was found. In addition, due to the lack of proximity of certain health care facilities to the University of Minnesota, some medical records were copied by clinic staff and sent to the research staff. Clinic staff were asked to review the entire chart and send all records pertaining to colorectal cancer screening. Given various time demands, medical records staff could potentially miss pertinent records; however, very few charts (6.8%) were abstracted in this way. Further, we were not able to obtain records covering at least 10 years of service for 29.1% of participants, but we were able to confirm overall screening per recommendations for 87.5% of all participants and had at least 5 years of records for 91.7%. Collecting clinic information and abstracting records for any facility where the participant stated they ever had screening and any other facility they had visited in the last 15 years allowed more complete ascertainment of any procedure done.

Second, although $\sim 90.0\%$ of the validation study participants completed the phone interview and provided health care facility information, only 60.0% provided HIPAA consent, resulting in a sample size of 120. Although the validation study response rate was not high, it is on par with response rates for other population-based studies and could be considered quite good given the request for HIPAA authorization (26). Further, based on *a priori* calculations, we determined

that a total sample of 100 to 120 participants would be an adequate sample size for SD of 0.14 to 0.16 based on a hypothesized 80% observed agreement given a binomial distribution with 50% expected agreement.

Third, selection or nonresponse bias for the validity sample is an issue. For example, validation study participants were more likely, than nonresponders, to report screening per guidelines at baseline (59.2% versus 36.3%; $P = 0.001$); however, validity study responders were not statistically significantly different from the overall parent study responders. If those who were most likely to erroneously report screening behavior did not agree to participate in the validation study because they felt they were being "tested" or that their original self-report was under suspicion and did not want the investigators to learn their true screening status, the validity of self-report could be overestimated. On the other hand, if those who were most susceptible to social desirability gave HIPAA consent, validity could be underestimated. Given the smaller sample size and the fact that we could not find specific evidence to support the notion that validation study responders would or would not be more accurate than nonresponders, we did not adjust for validity study nonresponse. Thus, although we corrected the self-reported screening rates by adjusting for validity measures as well as parent study eligibility and nonresponse, these conclusions should be interpreted with caution in the absence of all sources of variability of the estimates and the potential effect of validation study nonresponse bias, neither of which is a topic of this article.

Fourth, the percent of ambiguous responses overall was ~5%, depending on the screening end point. Some bias was probably introduced by coding these as not screened; however, the effect of this assumption on discrepancies, as measured by Δ in Table 2, is probably to attenuate them. Because NPVs are generally high, the right-hand side of the equation [corrected screening rate = self-report rate \times PPV + (1 - self-report rate) \times (1 - NPV)] is dominated by the first term, self-report rate \times PPV, which is an estimate of the true positive rate. If some proportion of ambiguous cases should instead be coded as screened, we expect the estimated true positive rate to increase by the same proportion as the self-report rate because ambiguous cases were included in the validation sample. Proportionate increases in both of these rates can only increase the absolute difference, Δ . We are assuming that the actual state of such ambiguous cases is not greatly different between the validation set and its complement. Thus, our assumption of underlying negativity tends to reduce the estimated discrepancy producing a more conservative estimate of overreporting.

Fifth, although this study sample may be more representative of the general population than participants in other studies assessing the validity of self-report, which were done in more narrowly defined populations (e.g., health maintenance organization samples or a specific occupational group; refs. 2-12), it is not likely to be representative of all adults ages ≥ 50 years, which may influence the generalizability. These study participants are unique because they responded to a request to participate in an annual colorectal cancer screening questionnaire and subsequently agreed to release their medical records.

Last, although all measures were self-reported and most were self-administered, ~25% of completed surveys were telephone administered at both baseline and follow-up. Although different modes of survey administration may increase the possibility for a variant of differential misclassification bias, Vernon et al. (12) found that reliability and validity were similar for mail, telephone, and face-to-face modes of survey administration.

Conclusions

Despite limitations, our findings suggest that, whereas sensitivity and specificity measures of self-reported colorectal cancer screening, compared with medical records, can be quite high and deemed within an acceptable range, adjusting self-reported screening rates based on relative error rates can reduce the estimates of screening prevalence. Further, those exposed to more intense interventions to modify behavior seem to be more likely, than those who are not exposed to intense interventions, to overestimate their FOBT screening rates. Researchers who use self-reported data to determine screening prevalence or the effectiveness of an intervention must consider the implications of differential bias and should conduct validation studies to obtain estimates of measurement error, which can be used to assess misclassification and provide more precise estimates of actual population screening rates.

Appendix A

Certain parameters, such as sensitivity, are of interest as they apply to the parent study cohort because this cohort is more representative of the target population. For brevity, let "truth" denote the medical record and "test" refer to the self-reported survey response. Due to the validation study sampling design, survey sensitivity in this larger parent study cohort cannot be estimated solely from the 2×2 "truth versus test" contingency table of counts from the validation sample because the numbers of positive survey responses ("test positive"; i.e., reporting adherence with any screening) and negative survey responses ("test negative"; i.e., not reporting adherence) were fixed by the study design at 100 of each response. Fixing these counts along the test margin of the 2×2 table affects what the subcounts will be, conditional on each level of truth.

Without such a constraint, assuming that the validation sample of 200 was obtained completely at random, sensitivity could be estimated directly and without bias from the two cells of the 2×2 table where truth is at the positive level. However, with the constraint, estimates of sensitivity will generally be biased. The same issues apply to specificity and the report-to-records ratio. The reason for the constraint (i.e., fixing the test distribution at 100 of each response) was to ensure sufficient quantities of responses for estimating both predictive values (positive and negative), which in turn [via the following equation: corrected screening rate = self-report rate \times PPV + (1 - self-report rate) \times (1 - NPV)] provide optimal estimates of the actual screening rate (per medical records) in the parent study cohort.

Acknowledgments

We thank Carla Cerra, R.N., for assistance with medical record abstraction, staff at Data Collection and Support Services in the Division of Epidemiology and Community Health at the University of Minnesota School of Public Health for assistance in recruiting participants for this validation study, and Sally W. Vernon, Ph.D., for her editorial comments.

References

- Khoja S, McGregor SE, Hilsden RJ. Validation of self-reported history of colorectal cancer screening. *Can Fam Physician* 2007;53:1192–7.
- 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.
- Hiatt RA, Perez-Stable EJ, Quesenberry C, Sabogal F, Otero-Sabogal R, McPhee SJ. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in Northern California. *Prev Med* 1995;24:278–85.
- Lipkus IM, Rimer BK, Lyna PR, Pradhan AA, Conaway M, Woods-Powell CT. Colorectal screening patterns and perceptions of risk among African-American users of a community health center. *J Community Health* 1996;21:409–27.
- Mandelson MT, Curry SJ, Anderson LA, et al. Colorectal cancer screening participation by older women. *Am J Prev Med* 2000;19:149–54.
- Baier M, Calonge N, Cutter G, et al. Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiol Biomarkers Prev* 2000;9:229–32.
- 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;37:513–9.
- Madlensky L, McLaughlin J, Goel VA. A comparison of self-reported colorectal cancer screening with medical records. *Cancer Epidemiol Biomarkers Prev* 2003;12:656–9.
- Hall HI, Van Den Eeden 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.
- Bradbury BD, Brooks DR, Brawarsky P, Mucci LA. Test-retest reliability of colorectal testing questions on the Massachusetts Behavioral Risk Factor Surveillance System (BRFSS). *Prev Med* 2005;41:303–11.
- Partin MR, Grill J, Noorbaloochi S, et al. Validation of self-reported colorectal cancer screening behavior from a mixed mode survey of veterans. *Cancer Epidemiol Biomarkers Prev* 2008;17:768–76.
- Vernon SW, Tiro JA, Vojvodic RW, 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;17:758–67.
- Winawer S, Fletcher R, Rex D, et al. Gastrointestinal Consortium Panel. Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology* 2003;124:544–60.
- Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2003. *CA Cancer J Clin* 2003;53:27–43.
- Paskett E, Tatum C, Rushing J, et al. Randomized trial of an intervention to improve mammography utilization among a triracial rural population of women. *J Natl Cancer Inst* 2006;98:1226–37.
- Yeazel MW, Church TR, Jones RM, et al. Colorectal cancer screening adherence in a general population. *Cancer Epidemiol Biomarkers Prev* 2004;13:654–7.
- Church TR, Yeazel MW, Jones RM, et al. A randomized trial of direct mailing of fecal occult blood tests to increase colorectal cancer screening. *J Natl Cancer Inst* 2004;96:770–80.
- Vernon SW, Meissner H, Klabunde C, et al. Measures for ascertaining use of colorectal cancer screening in behavioral, health services, and epidemiologic research. *Cancer Epidemiol Biomarkers Prev* 2004;13:898–905.
- Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 1952;47:663–85.
- Mongin S. Adjustment for non-response in the Minnesota nurses study. University of Minnesota Health Studies Section technical report [last accessed 2007 Dec 22]. Available from: <http://www1.umn.edu/eoh/NewFiles/researchpdffiles/response.adjustment.1.pdf>.
- Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *J Clin Epidemiol* 1990;43:941–7.
- Tenenbein A. A double sampling scheme for estimating from binomial data with misclassification. *J Am Stat Assoc* 1970;65:1350–61.
- Warnecke RB, Sudman S, Johnson TP, O'Rourke DP, 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.
- Rauscher GH, Johnson TP, Cho YI, Walk JA. Accuracy of self-reported cancer screening histories: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2008;17:748–57.
- Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Pennsylvania: Lippincott-Raven; 1998.
- Partin MR, Burgess DJ, Halek K, et al. Randomized trial showed requesting medical records with a survey produced a more representative sample than requesting separately. *J Clin Epidemiol*. In press.

Validity of Four Self-reported Colorectal Cancer Screening Modalities in a General Population: Differences over Time and by Intervention Assignment

Resa M. Jones, Steven J. Mongin, DeAnn Lazovich, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:777-784.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/17/4/777>

Cited articles This article cites 23 articles, 12 of which you can access for free at:
<http://cebp.aacrjournals.org/content/17/4/777.full.html#ref-list-1>

Citing articles This article has been cited by 7 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/17/4/777.full.html#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.