

Accuracy of Self-Reported Cancer-Screening Histories: A Meta-analysis

Garth H. Rauscher,¹ Timothy P. Johnson,² Young Ik Cho,² and Jennifer A. Walk⁴

¹Division of Epidemiology and Biostatistics, ²Survey Research Laboratory, and ³Institute for Health Research and Policy, University of Illinois at Chicago, and ⁴Department of Medicine, The University of Chicago, Chicago, Illinois

Abstract

Background: Survey data used to study trends in cancer screening may overestimate screening utilization while potentially underestimating existing disparities in use. **Methods:** We did a literature review and meta-analysis of validation studies examining the accuracy of self-reported cancer-screening histories. We calculated summary random-effects estimates for sensitivity and specificity, separately for mammography, clinical breast exam (CBE), Pap smear, prostate-specific antigen testing (PSA), digital rectal exam, fecal occult blood testing, and colorectal endoscopy.

Results: Sensitivity was highest for mammogram, CBE, and Pap smear (0.95, 0.94, and 0.93, respectively) and lowest for PSA and digital rectal exam histories (0.71 and 0.75). Specificity was highest for endoscopy, fecal occult blood testing, and PSA (0.90, 0.78, and 0.73,

respectively) and lowest for CBE, Pap smear, and mammogram histories (0.26, 0.48, and 0.61, respectively). Sensitivity and specificity summary estimates tended to be lower in predominantly Black and Hispanic samples compared with predominantly White samples. When estimates of self-report accuracy from this meta-analysis were applied to cancer-screening prevalence estimates from the National Health Interview Survey, results suggested that prevalence estimates are artificially increased and disparities in prevalence are artificially decreased by inaccurate self-reports.

Conclusions: National survey data are overestimating cancer-screening utilization for several common procedures and may be masking disparities in screening due to racial/ethnic differences in reporting accuracy. (Cancer Epidemiol Biomarkers Prev 2008;17(4):748–57)

Introduction

Screening for breast, cervical, and colorectal cancer are generally recognized as effective in reducing morbidity and mortality from these cancers, and both the U.S. Preventive Services Task Force (USPSTF) and the American Cancer Society (ACS) provide recommendations regarding screening frequencies for each. ACS currently recommends annual Pap smears to detect cervical cancer, beginning at age 21 years, and annual mammograms to detect breast cancer, beginning at age 40 years (1), whereas USPSTF recommends Pap smears every 3 years and mammograms every 1 to 2 years (2, 3). For detection of colorectal cancer, each recommends either fecal occult blood test (FOBT) annually (ACS) or biannually (USPSTF); sigmoidoscopy every 5 years; or colonoscopy every 10 years, beginning at age 50 years (1, 4). For prostate cancer, ACS (but not USPSTF) recommends annual digital rectal exam (DRE) and prostate-specific antigen (PSA) testing at age 50 years

for men at average risk and at age 45 years for African-American men (1). USPSTF recommends neither for nor against PSA screening.

Currently, epidemiologic surveys are the most common method used for monitoring trends in population compliance with these screening recommendations. According to data from the 2000 National Health Interview Survey (NHIS), ~82% of women over age 25 years obtained a Pap smear within the past 3 years, 70% obtained a mammogram in the past 2 years, and 41% of adults obtained colorectal cancer screening in accordance with guidelines (5). The Healthy People goals call for rates of adherence to national cancer-screening guidelines of 97% for Pap smear, 70% for biannual mammography, and 50% for colorectal cancer screening by 2010 (6). Another set of goals for Healthy People 2010 relates to reducing racial/ethnic and socioeconomic disparities in cancer screening. These goals implicitly assume that the data that are relied upon to estimate national rates of cancer screening are accurate.

Numerous validation studies have documented the levels of agreement/disagreement between survey reports of cancer-screening practices and actual behavior as documented in patient charts, laboratory records, and other auxiliary sources (7–61). We conducted a literature review and meta-analysis of studies validating cancer-screening self-reports. Our aims were to establish estimates of accuracy that could be used to inform the interpretation of national survey data on cancer-screening utilization, and that could be used by clinicians when

Received 9/27/07; revised 1/21/08; accepted 1/29/08.

Grant support: National Cancer Institute grant 5 P50 CA 106743.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Garth H. Rauscher, School of Public Health, Division of Epidemiology and Biostatistics, University of Illinois at Chicago, M/C 923, Chicago, IL 60612. Phone: 312-413-4317; Fax: 312-996-0064. E-mail: garthr@uic.edu

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-2629

deciding whether to refer a patient for screening based on self-reported prior screening. We also sought to determine whether survey methodology, documentation strategy, sample race/ethnicity, and other sample socio-demographic characteristics were associated with self-report accuracy.

Materials and Methods

Search Strategy. We searched for articles published in Medline between January 1, 1966, and July 2005. We searched for article titles containing any of the words "accuracy," "validity," "specificity," "sensitivity," "reliability," or "reproducibility" along with titles containing either "self-report," "recall," or "patient reports" and one of the following: "cancer," "cervical," "pap," "colon," "colorectal," "fobt," "fecal," "prostate," "psa," "breast," "mammo," "skin," or "melanoma." We searched the references of identified articles to locate additional articles of interest that might have been missed in the original search. We identified 55 articles presenting data on self-reported and medically documented screening histories (7-61). We excluded 10 studies that were not conducted in the United States (7, 9, 10, 11, 18, 44, 45, 53, 56, 57), including one study each on clinical skin cancer exam and stomach fluoroscopy (7, 53). We excluded three additional studies that compared self-reported screening rates in one sample of individuals with medically documented screening rates in another sample because accuracy measures could not be calculated (37, 38, 60). In addition, one study, conducted within the United States, validated the date of the most recent screening exam (61) without calculating accuracy measures for specific time frames and was excluded.

These 14 exclusions left 41 studies of the accuracy of self-reported breast, cervical, prostate, and colorectal cancer screening. From these studies, we limited our analysis to the 29 studies that collected data on all four contingencies (true positives, false positives, false negatives, and true negatives) and could therefore estimate sensitivity and specificity (and positive predictive value; refs. 8, 12-15, 17, 20, 22, 25-27, 30-32, 34-36, 39-41, 43, 47, 49, 50, 52, 54, 55, 58, 59).

Self-report time frames varied from study to study. Two studies asked subjects whether they had received a test on the same day as the interview (17, 43), whereas other studies asked subjects to report if they had ever had the procedure (47, 54, 58). The most frequent time frame was for the past 2 years (14, 20, 25, 26, 27, 31, 35, 45), followed by the past 1, 3, and 5 years, respectively. For studies that calculated measures for multiple time frames, we included only one time frame in our analysis to reduce dependencies in our data and to ensure that no single study contributed more information to the meta-analysis simply because it obtained multiple time frames on the same individuals. When forced to make this choice, we chose a reporting interval that corresponded closely to national screening guidelines for a given procedure; (a) 3 years for Pap smear reports; (b) 5 years for colorectal endoscopy; and (c) 2 years for other screening procedures. Several of the 29 studies included in our analysis reported accuracy estimates for more than one sample subgroup; in all, the 29 included

studies contained 39 sample subgroups in our final analysis data set.

Study. From each study, we abstracted the following: first author, year of publication, mid-point year of interviews, and whether the validation study seemed to be part of a larger intervention study or a validation substudy from a larger sample of respondents.

Study Subgroup. For each study subgroup, we abstracted the following: interview mode (face-to-face, telephone, self-administered report), interview location (home, clinic), type of facility from which medical documentation was obtained, and the number of "don't know" responses and how they were handled in the analysis. We also abstracted sample distribution of age, sex, education, and race/ethnicity. We categorized samples as dominant (>70%) White, Hispanic, Black, Native American, Asian American, mixed non-White dominant, and race/ethnicity not specified. We abstracted types of documentation sought from these facilities, including "charts"; "medical records"; "outpatient medical records"; "medical record audits"; "facility records"; and pathology, radiology, and "computerized databases," "HMO records," and many other brief and relatively uninformative descriptions of the documentation process.

Accuracy Measures. For each self-reported versus documented screening history, we identified the number of positive reports that were documented as positive (true positives), the number of positive reports that were documented as negative (false positives), the number of negative reports that were documented as positive (false negatives), and the number of negative reports that were documented as negative (true negatives). From these numbers, we calculated values for sensitivity, specificity, and positive predictive value. Sensitivity was defined as the probability that a history documented as positive was reported as positive (number of true positives divided by the number of true positives and false negatives). Specificity was defined as the probability that a history documented as negative was reported as negative (number of true negatives divided by the number of true negatives plus false positives). Positive predictive value was defined as the probability that a positive self-report was documented as positive (number of true positives divided by the number of true positives plus false positives). We included positive predictive value in our analyses because it may be of particular interest to clinicians. Reports-to-records ratio was defined as the percentage of self-reported positive screening histories divided by the percentage of documented positive histories. A report-to-records ratio greater than 1 was consistent with net overreporting of positive histories by respondents, and a report-to-records ratio less than 1 was consistent with net underreporting. In a few instances, one of the two cells involved in a given measure had a value of zero, prohibiting the simple calculation of a SE. In these rare instances, we added one observation to each cell and recalculated the accuracy estimate.

Statistical Analysis. Sensitivity, specificity, and positive predictive value were each nonnormally distributed. After taking a logit-transformation of the individual accuracy estimates, distributions were approximately

Accuracy of Self-Reported Cancer-Screening Histories

normal. We estimated the SE of each logit-transformed estimate, and plotted point estimates and 95% confidence intervals on the probability scale in order of increasing SE.

Summary Estimates Screening by Procedure. We used the meta command in Stata version 9 (StataCorp) to generate summary estimates of sensitivity, specificity, and positive predictive value for each screening procedure by taking weighted averages of individual study/subgroup estimates. We estimated both "fixed-effects" and "random-effects" summary estimates. Fixed-effects estimates were generated by taking an inverse-variance weighted average of the study-specific, logit-transformed estimates of accuracy. Fixed-effects estimates give more weight to larger studies and assume that each is an estimate of a common value. Random-effects estimates (also known as DerSimonian and Laird estimates) assume that there are true differences in estimates between studies (62). This between-study variance is estimated and then added to the denominator of the precision weights. The result is that larger and smaller studies are more evenly weighted and the confidence interval for the summary effect is usually wider than for the corresponding fixed-effect estimate. Random-effects estimation also provides a test for homogeneity in

study-/subgroup-specific estimates. *P* values for the vast majority of tests of homogeneity were highly statistically significant ($P < 0.0005$), indicating a great deal of heterogeneity among study estimates; therefore, we relied on random-effects summary estimates when interpreting results.

Publication Bias. Individual validation studies of cancer-screening histories, especially substudies undertaken to support results from a larger study, may not be published if accuracy is determined to be poor. The resulting absence of these studies in the literature would tend to bias summary estimates upward in the meta-analysis. To test for the possibility of publication bias, we constructed funnel plots and calculated Egger's statistics as implemented in Stata's metabias and metafunnel commands (63). When the *P* value for the bias statistic was < 0.10 , we implemented Stata's metatrim command, which uses a nonparametric "trim and fill" method to reestimate the summary estimate, after accounting for possible publication bias (64).

Corrected Estimates from NHIS 2005. We applied standard formulae for sensitivity analysis of misclassification (65) to adjust cancer-screening prevalence estimates

Table 1. List of studies ($n = 29$), study subgroups ($n = 39$), and time frame included in these analyses

Study	First author	Publication year	No. strata	Subgroup	Procedures	Time frame (y)
1	Warnecke	1976	1		Pap	Lifetime
2	Sawyer	1989	1		Pap	3
3	Boyce	1990	1		Pap	5
4	King	1990	1		Mam	1
5	Michielutte	1991	1		Pap	0
6	Brown	1992	1		Mam, DRE, FOBT	0.21
7	Degnan	1992	1		Mam	2
8	Mckenna	1992	1		Pap	3
9	Gordon	1993	1		Mam, CBE, Pap, DRE, FOBT, Sig	2
10	Sudman	1994	1		Mam, CBE, Pap, FOBT	5
11	Hiatt	1995	4	Non-Hispanic men Hispanic men Non-Hispanic women Hispanic women	Mam, CBE, Pap, DRE, FOBT Sig	2
12	Lipkus	1996	1		DRE	2
13	Paskett	1996	1		Mam, Pap	Lifetime
14	Fowles	1997	1		Pap	3
15	McGovern	1998	1		Mam, Pap	1
16	Jordan	1999	1		PSA, DRE	2
17	Lawrence	1999	2	Military health system County health system	Mam	1.167
18	Mandelson	1999	1		FOBT	5
19	Thompson	1999	1		Mam	7
20	Baier	2000	1		FOBT, Sig, Col	4
21	Pizarro	2002	1		Pap	1
22	Volk	2002	1		PSA	1
23	Lipkus	2003	1		FOBT	1
24	Caplan	2003a	1		Mam	2
25	Caplan	2003b	1		Mam, CBE, Pap	2
26	Chan	2004	2	Clinic 1 Clinic 2	PSA	Same day
27	Hall	2004	3	Black men White men Women	PSA, DRE, FOBT, Sig, Col	2
28	Tumiel-Berhalter	2004	3	Puerto-Rican Black White	Mam, Pap	1.5
29	Johnson	2005	2	Telephone Self-interview (ACASI)	Mam Pap CBE	3

Abbreviations: Mam, mammogram; Pap, Pap smear; Sig, sigmoidoscopy; Col, colonoscopy; ACASI, audio computer-assisted self-interview.

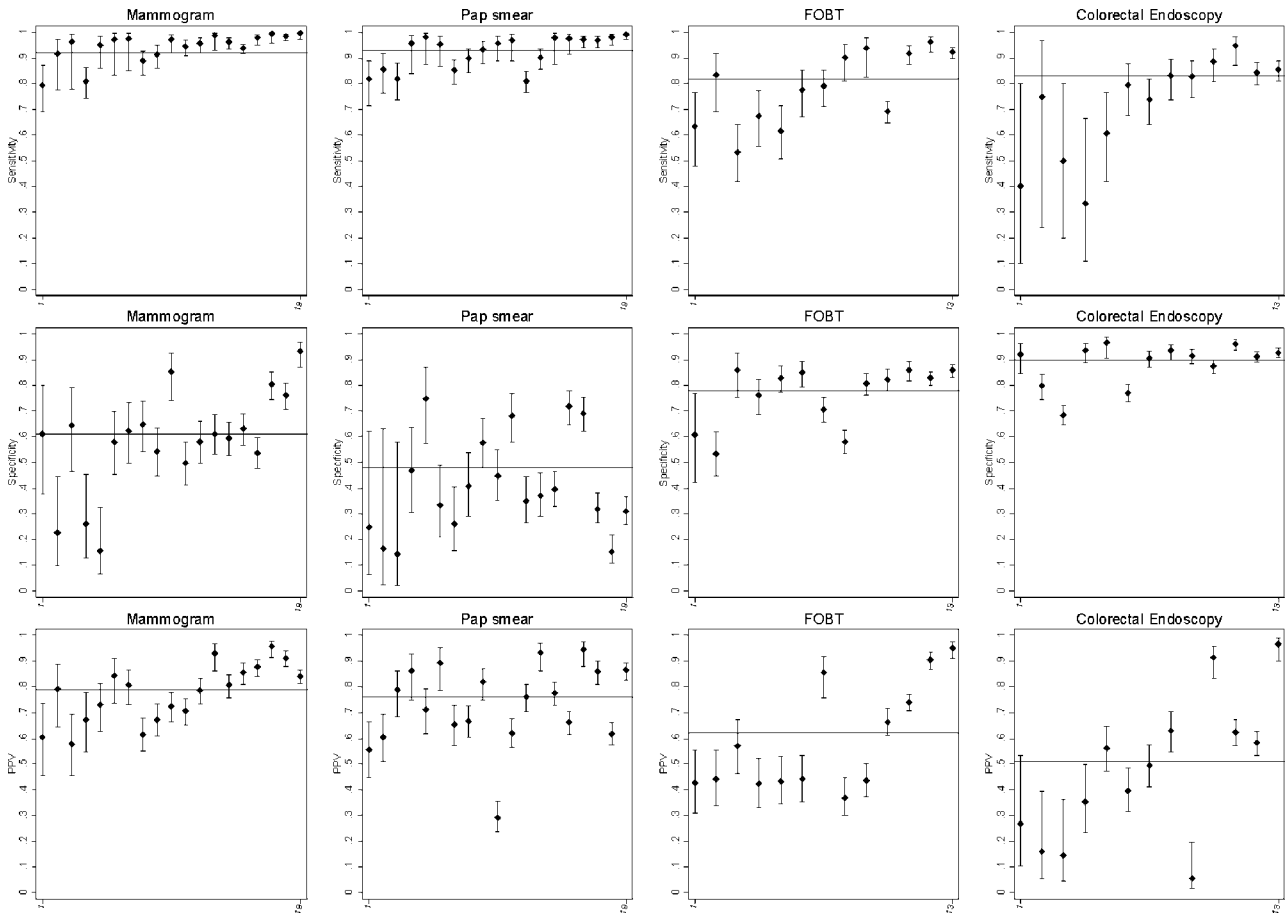


Figure 1. Plot of random-effects estimates and 95% confidence intervals for sensitivity, specificity, and positive predictive value for mammogram, Pap smear, FOBT, and colorectal endoscopy self-reports. Within each panel, estimates are arranged in order of increasing SE.

from the 1998 and 2000 NHIS for potential underreporting and overreporting as estimated in our meta-analysis (5, 66).

Results

Table 1 lists the 29 studies and 39 study strata that contributed a total of 87 observations to the final analysis data set. Individual point estimates and 95% confidence intervals for sensitivity, specificity, and positive predictive value are presented in Fig. 1 for self-report histories of mammogram, Pap smear, FOBT, and colorectal endoscopy. There was a great deal of variation in estimated accuracy across studies for a given screening procedure. Sensitivity was positively correlated with the prevalence of self-reported use and with positive predictive value, but inversely correlated with specificity ($\rho = 0.76, 0.60,$ and $-0.49,$ respectively). In addition, specificity was inversely correlated with both prevalence of self-reported use and with positive predictive value ($\rho = -0.86$ and $-0.24,$ respectively; results not shown). The mean (median) values for reports-to-records ratio were 1.2 (1.1) for clinician breast exam, 1.2 (1.2) for mammography, 1.4 (1.2) for Pap smear, 0.9 (0.9) for PSA

testing, 1.3 (1.2) for DRE, 1.4 (1.4) for FOBT, and 2.2 (1.7) for colorectal endoscopy (results not shown).

Summary estimates of fixed and random effects are presented in Table 2. In almost every case, there was highly statistically significant between-estimate variance; therefore, we relied on the random-effects estimates when interpreting results. Sensitivity was highest for mammogram, clinical breast exam (CBE), and Pap smear (0.95, 0.94, and 0.93, respectively) and lowest for PSA and colorectal screening histories. Specificity was highest for endoscopy, FOBT, and PSA (0.90, 0.78, and 0.73, respectively) and lowest for CBE, Pap smear, and mammogram histories (0.26, 0.48, and 0.61, respectively). For the 21 random-effects summary estimates across the seven screening procedures and three measures, we found eight instances of possible publication bias, six of which seemed to inflate the accuracy estimate by 2 to 8 percentage points, and two that seemed to lower the estimate by 4 and 8 points, respectively (Table 2).

Socioeconomic and Study Design Factors. Information on socioeconomic status was limited; the distribution of age, income, insurance status, and other factors was usually not available. We were able to categorize a

Accuracy of Self-Reported Cancer-Screening Histories

Table 2. Summary random effects estimates by screening procedure

Procedure	<i>n</i> *		PPV (95% CI)	Se (95% CI)	Sp (95% CI)
Mammogram	19	Fixed [†]	79 (77-80)	93 (92-94)	62 (60-64)
		Random [‡]	79 (74-84)	95 (93-97)	61 (54-68)
		Filled [§]		92 (88-95)	
CBE	7	Fixed	78 (76-80)	93 (92-95)	29 (25-33)
		Random	82 (73-88)	94 (91-96)	26 (18-36)
		Filled	76 (65-84)	92 (88-95)	
Pap smear	26	Fixed	70 (69-72)	91 (90-92)	47 (45-49)
		Random	76 (69-81)	93 (91-95)	48 (40-57)
		Filled	68 (61-74)	88 (83-91)	
PSA test	6	Fixed	76 (73-79)	71 (68-74)	73 (70-76)
		Random	79 (71-85)	71 (68-74)	73 (62-81)
		Filled			
DRE	10	Fixed	63 (61-65)	74 (71-76)	60 (58-62)
		Random	59 (52-66)	75 (66-82)	60 (55-66)
		Filled	67 (59-74)		
FOBT	13	Fixed	62 (60-64)	78 (76-80)	77 (76-79)
		Random	62 (50-73)	82 (73-88)	78 (71-83)
		Filled			
Screening endoscopy	13	Fixed	56 (54-59)	82 (80-84)	85 (84-86)
		Random	51 (41-62)	79 (73-84)	90 (85-93)
		Filled		83 (77-88)	83 (77-88)

Abbreviations: PPV, positive predictive value; Se, sensitivity; Sp, specificity; CI, confidence interval.

*Number of study subgroups.

† Fixed effects summary estimate.

‡ Random effects summary estimate.

§ Summary random effects estimate adjusted for publication bias.

little more than half of the samples on the percentage with a high school degree; this somewhat crude measure was not associated with reporting accuracy (results not shown). Sex was also not associated with self-report accuracy (results not shown). Neither the type of documentation facility nor the types of documentation

relied on was consistently associated with self-report accuracy (results not shown). Interviews conducted in the clinic and face-to-face interviews, in general, tended to be associated with reduced self-report accuracy compared with telephone-administered or self-administered interviews (results not shown).

Table 3. Random-effects summary estimates of sensitivity and specificity, stratified by dominant race/ethnicity, among studies that examined more than one racial/ethnic group

Author	Sensitivity			Specificity		
	White	Black	Hispanic	White	Black	Hispanic
	Se (95% CI)	Se (95% CI)	Se (95% CI)	Sp (95% CI)	Sp (95% CI)	Sp (95% CI)
Mammogram						
Hiatt (1995)	0.92 (0.86-0.95)		0.81 (0.74-0.86)	0.50 (0.41-0.58)		0.63 (0.57-0.69)
Tumiel-Berhalter (2004)	0.96 (0.78-0.99)	0.98 (0.85-1.00)	0.97 (0.84-1.00)	0.23 (0.10-0.44)	0.26 (0.13-0.45)	0.16 (0.07-0.32)
CBE						
Hiatt (1995)		0.92 (0.88-0.95)	0.87 (0.82-0.91)	0.24 (0.16-0.33)		0.34 (0.27-0.42)
Pap smear						
Hiatt (1995)	0.90 (0.84-0.94)		0.85 (0.80-0.89)	0.37 (0.29-0.46)		0.39 (0.33-0.47)
Tumiel-Berhalter (2004)	0.98 (0.88-1.00)	0.96 (0.89-0.98)	0.98 (0.91-0.99)	0.14 (0.02-0.58)	0.17 (0.02-0.63)	0.25 (0.06-0.62)
PSA test						
Hall (2004)	0.70 (0.65-0.74)	0.75 (0.68-0.80)		0.77 (0.73-0.81)	0.68 (0.60-0.75)	
DRE exam						
Hall (2004), men	0.83 (0.80-0.87)	0.85 (0.79-0.89)		0.57 (0.52-0.62)	0.48 (0.40-0.56)	
Hiatt (1995), men	0.69 (0.55-0.81)		0.63 (0.51-0.72)	0.66 (0.58-0.74)		0.72 (0.66-0.76)
Hiatt (1995), women	0.66 (0.58-0.73)		0.48 (0.41-0.55)	0.69 (0.61-0.77)		0.66 (0.59-0.73)
FOBT						
Hall (2004), men	0.78 (0.67-0.85)	0.83 (0.69-0.92)		0.86 (0.83-0.88)	0.86 (0.82-0.89)	
Hiatt (1995), men	0.63 (0.48-0.77)		0.62 (0.51-0.72)	0.76 (0.69-0.82)		0.81 (0.76-0.85)
Hiatt (1995), women	0.68 (0.56-0.77)		0.53 (0.42-0.64)	0.83 (0.77-0.87)		0.82 (0.78-0.86)
Colonoscopy						
Hall (2004), men	0.74 (0.64-0.82)	0.61 (0.42-0.77)		0.93 (0.91-0.94)	0.90 (0.87-0.93)	
Sigmoidoscopy						
Hall (2004), men	0.84 (0.79-0.88)	0.83 (0.74-0.98)		0.77 (0.74-0.80)	0.80 (0.74-0.84)	
Hiatt (1995), men	0.50 (0.20-0.80)		0.40 (0.10-0.80)	0.94 (0.89-0.96)		0.92 (0.88-0.94)
Hiatt (1995), women	0.33 (0.11-0.67)		0.75 (0.24-0.97)	0.94 (0.90-0.96)		0.96 (0.93-0.97)

Table 4. Random-effects summary estimates stratified by dominant race/ethnicity

	<i>n</i>	Se (95% CI)	Sp (95% CI)	PFV (95% CI)
Mammogram				
White (reference)	9	0.97 (0.95-0.98)	0.62 (0.64-0.61)	0.83 (0.76-0.88)
African American	3	0.97 (0.91-0.99)	0.49 (0.57-0.42)	0.82 (0.51-0.95)
Hispanic	3	0.83 (0.73-0.90)*	0.46 (0.49-0.43)	0.64 (0.55-0.72)*
Pap smear				
White (reference)	6	0.95 (0.89-0.97)	0.48 (0.82-0.16)	0.78 (0.70-0.84)
African American	7	0.95 (0.88-0.98)	0.31 (0.52-0.16)*	0.70 (0.50-0.85)
Hispanic	2	0.93 (0.66-0.99)	0.39 (0.22-0.60)	0.82 (0.36-0.97)
DRE				
White (reference)	4	0.76 (0.67-0.83)	0.61 (0.61-0.60)	0.64 (0.50-0.75)
African American	2	0.82 (0.74-0.88)	0.50 (0.17-0.83)	0.63 (0.48-0.76)
Hispanic	2	0.45 (0.40-0.68)*	0.69 (0.14-0.97)	0.47 (0.25-0.71)
FOBT				
White (reference)	8	0.83 (0.72-0.90)	0.79 (0.36-0.96)	0.68 (0.47-0.84)
African American	2	0.89 (0.73-0.96)	0.73 (0.13-0.98)	0.44 (0.37-0.51)
Hispanic	2	0.58 (0.51-0.66)	0.82 (0.06-1.00)	0.43 (0.37-0.50)
Endoscopy				
White (reference)	9	0.81 (0.75-0.87)	0.90 (0.20-1.00)	0.60 (0.35-0.80)
African American	2	0.74 (0.52-0.90)	0.86 (0.03-1.00)	0.50 (0.25-0.75)†
Hispanic	2	0.54 (0.22-0.83)	0.94 (0.94-0.94)	0.10 (0.03-0.26)†

**P* ≤ 0.10.†*P* ≤ 0.05.

Race/Ethnic Comparisons. Of the 39 sample subgroups analyzed, 15, 9, and 4 subgroups were predominantly White, Black, and Hispanic, respectively. The remainder were either mixed minority dominant (*n* = 2), did not have a dominant race/ethnicity (*n* = 7), or were missing data on race/ethnicity (*n* = 2). Three studies directly compared predominantly White subgroups to predominantly Black and/or Hispanic subgroups with respect to mammography (27, 54), CBE (27), Pap smear (27, 54), PSA (26), DRE (27, 26), FOBT (26, 27), and colorectal endoscopy (26, 27). These comparisons are presented in Table 3. These within-study comparisons suggested that there was a reduced sensitivity of self-reporting for Hispanics compared with Whites for both mammography and DRE (27), but no differences were suggested for specificity (Table 3).

When meta-regressions were done on all studies together, sensitivity tended to be lowest for predominantly Hispanic samples, whereas specificity tended to be lower for both predominantly Hispanic and predominantly Black samples compared with predominantly White samples (Table 4). Compared with White samples, Hispanic ethnicity was associated with reduced sensitivity of mammogram and DRE self-reporting. Hispanic ethnicity was also associated with reduced specificity of Pap smear self-reporting and reduced

positive predictive value for mammogram and endoscopy self-reporting. Predominantly Black samples were associated with reduced specificity of Pap smear reporting (Table 4).

Adjusted Prevalence Estimates from the NHIS. We adjusted prevalence estimates from the 1998 and 2000 NHIS using the procedure-specific random-effects summary estimates of sensitivity and specificity that were generated by the meta-analysis (Table 2). The estimated prevalence of mammography in the past 2 years was lowered with adjustment from 0.70 to 0.56, and the prevalence of Pap smear in the past 3 years was lowered from 0.82 to 0.74 (Table 5). Prevalence estimates for PSA and colorectal screening were also lowered after adjustment for self-report accuracy in this manner. We then adjusted race/ethnicity-specific mammogram and Pap smear prevalence estimates from NHIS, using the procedure-specific random-effects summary estimates of sensitivity and specificity that were generated separately for non-Hispanic White, non-Hispanic Black, and Hispanic individuals (Table 4). After adjusting prevalence estimates in this manner, Black-White and Hispanic-White disparities in mammogram and Pap smear prevalence estimates seemed to be considerably larger than those based on the observed estimates alone (Table 6).

Table 5. Applying sensitivity and specificity estimates from the present meta-analysis to results from the 1998 and 2000 NHIS

	Time frame (y)	Ages	Observed prevalence	Se	Sp	Adjusted prevalence
Mammogram	2	40+	70	0.95	0.61	56
Pap smear	3	25+	82	0.93	0.48	74
PSA test	1	50+	41	0.71	0.73	32
FOBT (women)	2	50+	26	0.82	0.78	<10
FOBT (men)	2	50+	29	0.82	0.78	11
Endoscopy (women)	3	50+	10	0.79	0.90	<10
Endoscopy (men)	3	50+	19	0.79	0.90	13

NOTE: Mammogram, Pap smear, and PSA prevalence estimates are from the 2000 NHIS; FOBT and endoscopy prevalence estimates are from the 1998 NHIS.

Discussion

The main goal of this analysis was to generate estimates of self-report accuracy for specific cancer-screening procedures that could then be applied to national survey data to improve the accuracy of national prevalence estimates for colorectal cancer screening and screening for other cancers, estimates that are largely based on self-reports. We examined sensitivity as a measure of underreporting and examined both specificity and positive predictive value as measures of overreporting.

We found that estimates of self-report accuracy varied substantially from study to study, making it difficult to arrive at stable estimates of self-report accuracy for different procedures and accuracy measures. Study heterogeneity could be the result of many different factors related to sample socioeconomic and cultural characteristics, the context in which self-reports were elicited (e.g., survey mode and location, wording of question), and the completeness and accuracy of documentation of true screening histories, which are functions of both the strategy for documenting procedures and, equally importantly, the thoroughness with which that strategy is implemented. Details on the strategy and implementation of documenting cancer-screening procedures were scant, and determining the intensity of documentation was generally not possible. The quality of the sources and strategies used to document survey answers might influence conclusions about self-report accuracy in this meta-analysis (67-69). In particular, we thought that studies with a more rigorous documentation strategy would yield higher (and more accurate) specificity and positive predictive value estimates; however, there was little evidence to suggest that this was true. It is likely that some of the heterogeneity in results was due to unmeasured differences in the quality of documentation.

Some validation studies might be conducted to support the results of a larger study rather than to produce generalizable information about self-report accuracy. For such studies, publication bias is likely because there would be less incentive to publish low-accuracy estimates that reflected poorly on the implementation of the larger study. With this in mind, we excluded, from our analysis, validation studies that confirmed only the accuracy of positive self-report histories (from which only positive predictive values could be estimated). Publication bias was stronger for the 15 excluded study subgroups with Pap smear results than was observed in our main analysis, but there was no

evidence of publication bias for the 14 excluded study subgroups with mammography results (results not shown).

Relative to underreporting, overreporting seemed to be a greater concern for most of the screening procedures examined. Overreporting for both mammography and Pap smear has been shown to be related to forward telescoping of dates, when an event is remembered as being more recent than it actually was (42, 61). In addition, Pap smears may be overreported if women mistake a routine gynecologic exam without a Pap test as including a Pap test. Incomplete documentation of a test in the medical record would cause a true positive screening history to be classified as a false positive, thereby biasing specificity estimates downward. Specificity for both CBE and DRE are likely to be biased downward by this mechanism, because they do not generate a separate report and are not reimbursed well, further reducing the incentive to document. For this reason, determining the true extent of overreporting of these two tests may not be possible using current documentation strategies. The reason for overreporting of colorectal endoscopies is not clear but may be a reflection of a general tendency for respondents to overreport socially desirable behaviors.

Underreporting seems to be infrequent for breast and cervical cancer-screening self-reports, but more of a concern for prostate and colorectal screening. The relatively poor sensitivity for PSA testing (compared with that for other procedures) may be due to the lack of saliency for the patient. PSA testing does not require a separate appointment; the only cue to the patient is a blood draw indistinguishable from any other blood draw, and men may not always be told that one purpose of the blood draw is for PSA testing. In contrast, colorectal endoscopy is a highly salient event because of the preparation required on the part of the patient and the invasiveness of the procedure. Nonetheless, underreporting of colorectal endoscopy seemed to be similar in magnitude to the other two colorectal exams (FOBT and DRE). Unlike breast and cervical screening, prostate cancer screening is unique to men, whereas colorectal screening is obtained by both men and women. It is possible that the greater underreporting for these tests is due to a greater tendency of men to underreport cancer-screening histories, although we found no evidence of this in either stratified analyses or meta-regressions (results not shown). Underreporting for PSA and DRE could also be a function of embarrassment, although this

Table 6. Applying race/ethnicity-specific sensitivity and specificity estimates from the present meta-analysis to results from the 2000 NHIS

	Observed		Self-report accuracy		Adjusted	
	Prevalence	Disparity	Se	Sp	Prevalence	Disparity
Pap smear in the past 3 y (ages 25+)						
White	83	Reference	0.95	0.48	73	Reference
Black	84	-1	0.95	0.31	58	15
Hispanic	77	6	0.93	0.39	51	22
Mammogram in the past 2 y (ages 40+)						
White	72	Reference	0.97	0.62	58	Reference
Black	68	4	0.97	0.49	37	21
Hispanic	61	11	0.83	0.48	28	30

has not been studied to our knowledge. Because sensitivity is calculated only for positively documented histories, subjects with incomplete documentation are excluded from the calculation of sensitivity. If the true sensitivity was differentially higher among individuals with incomplete documentation of their screening history, this would artificially lower the observed sensitivity, but this pattern of misclassification seems unlikely. For this reason, incomplete documentation in the medical record is not a likely explanation for low sensitivity of prostate and colorectal screening self-reports.

African American and Hispanic women seemed to have a greater tendency to overreport mammogram and Pap smear histories compared with Whites, although the number of studies compared was small and race/ethnicity-stratified estimates were unstable for all of the procedures we examined. The apparently greater overreporting observed in Black and Hispanic versus White subgroups could be partly due to differences in documentation quality or survey context. For example, being part of an intervention program, interviewed face-to-face, and interviewed in the clinic, all of which might be associated with overreporting, were all somewhat more likely to be a study design characteristic associated with predominantly Black as opposed to White subgroups. Although not evident from our meta-analysis, another possibility is that documentation quality is lower at facilities where minority patients are more likely to receive health care. Differences in self-report accuracy may also be partly a consequence of variability in the comprehension of survey questions (70).

Two studies of racial/ethnic differences in self-report accuracy that were not a part of this meta-analysis deserve mention. First, a validation study by McPhee et al. (42), which was excluded from this meta-analysis because it was limited to confirming positive self-reports, found that African American, Hispanic, Chinese, and Filipino women were all more likely than White women to overreport both mammogram and Pap smear histories and that forward telescoping of dates was more common among ethnic minority women. In addition, these differences in self-report accuracy remained after controlling for type of facility. Second, in a study of mammography use (estimated using both claims and survey data) among older Medicare enrollees, the authors found that the survey data failed to identify racial/ethnic disparities in screening that were apparent from analyses using claims data (71). These two studies, in combination with our meta-analysis, suggest that racial and ethnic minority women are more likely to overreport mammography (and perhaps Pap smear histories) when compared with White women. The data are less clear for other screening procedures.

In addition to the apparent differences in overreporting cancer-screening histories, Hispanic individuals may have a greater tendency than Whites to underreport mammogram and colorectal screening histories, according to our meta-analysis. These are two highly invasive and salient procedures. Perhaps cultural differences may cause Hispanics to underreport these procedures out of embarrassment or privacy concerns. Alternatively, it may be that underreporting by Hispanics is a more general phenomenon that was easier to detect in these two

generally well-documented procedures; differences in sensitivity could be masked by less complete documentation of other procedures.

Implications for Clinicians. According to this analysis, nearly half of self-reported positive colorectal screening histories and approximately one fifth of self-reported positive breast and cervical cancer-screening histories are likely to be negative. During a doctor's visit, individuals may be even more likely to overreport their cancer-screening histories because of social desirability influences arising from face-to-face contact in a clinic setting. There was some evidence for this in our analyses, although the number of clinic-based validation studies was small. Positive predictive value for mammography, which was 0.79 across all studies in our analysis, was only 0.66 for clinic-based interviews. Given the apparent extent of overreporting of cancer-screening histories, clinicians should not rely on patient reports when making recommendations for future screening. Reliance on patient self-reports could create a barrier to implementation of timely follow-up screening.

Implications for Tracking National Utilization. The Healthy People 2010 goals call for increasing the percentage of women adhering to national cancer-screening guidelines. Healthy People 2010 calls for an increase, by 2010, in Pap smear utilization in the preceding 3 years from 92% to 97%, mammography in the preceding 2 years from 67% to 70%, annual FOBT from 35% to 50%, and colorectal endoscopy in the preceding 5 years from 37% to 50% (6). Results from this meta-analysis indicate that we are probably further from these goals than survey data suggest. Another broad goal of Healthy People 2010 is the reduction of disparities in health and health care utilization. Again, according to this meta-analysis, disparities in cancer screening by race/ethnicity are likely to be larger than they seem to be in national survey data. These inaccuracies need to be taken into account when interpreting progress toward the Healthy People 2010 goals of increasing utilization and reducing disparities. Because the NHIS is the major source of data on cancer screening used for tracking prevalence in the U.S. population, validation studies should be undertaken for a sample of respondents within the NHIS, and designed with enough power to detect meaningful differences in sensitivity and specificity for different racial/ethnic and socioeconomic groups.

References

1. American Cancer Society. Available from: http://www.cancer.org/docroot/PED/content/PED_2_3X_ACS_Cancer_Detection_Guidelines_36.asp, accessed 2007 Aug 14.
2. National Cancer Institute cervical cancer screening guidelines. Available from: <http://www.cancer.gov/newscenter/pressreleases/cervicalscreen>, accessed 2005 Nov 11.
3. NCI statement on mammography screening (Update 2002 Feb 21). Available from: <http://www.cancer.gov/newscenter/mammstatement31Jan02>, accessed 2005 Nov 11.
4. NCI website: National Cancer Institute trial yields new data on colon cancer screening test. Available from: <http://www.cancer.gov/newscenter/pressreleases/PLCO>, accessed 2005 Nov 11.
5. Swan J, Breen N, Coates RJ, Rimer BK, Lee NC. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. *Cancer* 2003;97:1528-40.
6. Healthy People 2010 website. Available from: <http://www.healthypeople.gov/>, accessed 2007 Jan.

Accuracy of Self-Reported Cancer-Screening Histories

7. Aitken JF, Youl PH, Janda M, et al. Validity of self-reported skin screening histories. *Am J Epidemiol* 2004;159:1098–105.
8. Baier M, Calonge N, Cutter G, et al. Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiol Biomarkers Prev* 2000;9:229–32.
9. Barratt A, Cockburn J, Smith D, Redman S. Reliability and validity of women's recall of mammographic screening. *Aust N Z J Public Health* 2000;24:79–81.
10. Bowman JA, Redman S, Dickinson JA, Gibberd R, Sanson-Fisher RW. The accuracy of Pap smear utilization self-report: a methodological consideration in cervical screening research. *Health Serv Res* 1991;26:97–107.
11. Bowman JA, Sanson-Fisher R, Redman S. The accuracy of self-reported Pap smear utilization. *Soc Sci Med* 1997;44:969–76.
12. Boyce JG, Fruchter RG, Romanzi L, Sillman FH, Maiman M. The fallacy of the screening interval for cervical smears. *Obstet Gynecol* 1990;76:627–32.
13. Brown JB, Adams ME. Patients as reliable reporters of medical care process. Recall of ambulatory encounter events. *Med Care* 1992;30:400–11.
14. Caplan LS, Mandelson MT, Anderson LA, Health Maintenance Organization. Validity of self-reported mammography: examining recall and covariates among older women in a Health Maintenance Organization. *Am J Epidemiol* 2003;157:267–72.
15. Caplan LS, McQueen DV, Qualters JR, 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;12:1182–7.
16. Champion VL, Menon U, McQuillen DH, Scott C. Validity of self-reported mammography in low-income African-American women. *Am J Prev Med* 1998;14:111–7.
17. Chan EC, Vernon SW, Ahn C, Greisinger A. Do men know that they have had a prostate-specific antigen test? Accuracy of self-reports of testing at 2 sites. *Am J Public Health* 2004;94:1336–8.
18. Coulter A, Baldwin A. Survey of population coverage in cervical cancer screening in the Oxford region. *J R Coll Gen Pract* 1987;37:441–3.
19. Crane LA, Kaplan CP, Bastani R, Scrimshaw SC. Determinants of adherence among health department patients referred for a mammogram. *Women Health* 1996;24:43–64.
20. Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzalez J. Measuring the use of mammography: two methods compared. *Am J Public Health* 1992;82:1386–8.
21. Etzi S, Lane DS, Grimson R. The use of mammography vans by low-income women: the accuracy of self-reports. *Am J Public Health* 1994;84:107–9.
22. Fowles JB, Fowler E, Craft C, McCoy CE. Comparing claims data and self-reported data with the medical record for Pap smear rates. *Evaluation Health Professions* 1997;20:324–42.
23. Fruchter RG, Rones K, Roth T, Webber CA, Camilien L, Boyce JG. Pap smear histories in a medical clinic: accuracy of patients' self-reports. *N Y State J Med* 1992;92:421–4.
24. Fulton-Kehoe D, Burg MA, Lane DS. Are self-reported dates of mammograms accurate? *Public Health Rev* 1992-1993;20:233–40.
25. 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.
26. 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.
27. Hiatt RA, Perez-Stable EJ, Quesenberry C, Jr., 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.
28. Hochstim JR. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 1967;62:976–89.
29. Johnson CS, Archer J, Campos-Outcalt D. Accuracy of Pap smear and mammogram self-reports in a southwestern Native Am tribe. *Am J Prev Med* 1995;11:360–3.
30. Johnson TP, O'Rourke DP, Burriss JE, Warnecke RB. An investigation of the effects of social desirability on the validity of self-reports of cancer screening behaviors. *Med Care* 2005;43:565–73.
31. Jordan TR, Price JH, King KA, Masyk T, Bedell AW. The validity of male patients' self-reports regarding prostate cancer screening. *Prev Med* 1999;28:297–303.
32. King ES, Rimer BK, Trock B, Balshem A, Engstrom P. How valid are mammography self-reports? *Am J Public Health* 1990;80:1386–8.
33. Kottke TE, Trapp MA, Fores MM, et al. Cancer screening behaviors and attitudes of women in southeastern Minnesota. *JAMA* 1995;273:1099–105.
34. Lawrence VA, De Moor C, Glenn ME. Systematic differences in validity of self-reported mammography behavior: a problem for intergroup comparisons? *Prev Med* 1999;29:577–80.
35. Lipkus IM, Rimer BK, Lyna PR, Pradhan AA, Conway M, Woods-Powell CT. Colorectal screening patterns and perceptions of risk among African-Am users of a community health center. *J Community Health* 1996;21:409–27.
36. 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.
37. Loftus EF, Klinger MR, Smith KD, Fielder J. A tale of two questions: benefits of asking more than one question. *Public Opin Q* 1990;54:330–45.
38. Montano DE, Phillips WR. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *Am J Public Health* 1995;85:795–800.
39. Mandelson MT, LaCroix AZ, Anderson LA, Nadel MR, Lee NC. Comparison of self-reported fecal occult blood testing with automated laboratory records among older women in a health maintenance organization. *Am J Epidemiol* 1999;150:617–21.
40. McGovern PG, Lurie N, Margolis KL, Slater JS. Accuracy of self-report of mammography and Pap smear in a low-income urban population. *Am J Prev Med* 1998;14:201–8.
41. McKenna MT, Speers M, Mallin K, Warnecke R. Agreement between patient self-reports and medical records for Pap smear histories. *Am J Prev Med* 1992;8:287–91.
42. 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;35:463–73.
43. Michielutte R, Dignan MB, Wells HB, et al. Errors in reporting cervical screening among public health clinic patients. *J Clin Epidemiol* 1991;44:403–8.
44. Munk C, Kjaer SK, Poll P, Bock JE. Cervical cancer screening: knowledge of own screening status among women aged 20–29 years. *Acta Obstet Gynecol Scand* 1998;77:917–22.
45. Newell S, Giris A, Sanson-Fisher R, Ireland M. Accuracy of patients' recall of Pap and cholesterol screening. *Am J Public Health* 2000;90:1431–5.
46. Norman SA, Localio AR, Zhou L, et al. Validation of self-reported screening mammography histories among women with and without breast cancer. *Am J Epidemiol* 2003;158:264–71.
47. 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;5:721–6.
48. Peters RK, Bear MB, Thomas D. Barriers to screening for cancer of the cervix. *Prev Med* 1989;18:133–46.
49. Pizarro J, Schneider TR, Salovey P. A source of error in self-reports of Pap test utilization. *J Community Health* 2002;27:351–6.
50. Sawyer JA, Earp JA, Fletcher RH, Daye FF, Wynn TM. Accuracy of women's self-report of their last Pap smear. *Am J Public Health* 1989;79:1036–7.
51. Suarez L, Goldman DA, Weiss NS. Validity of Pap smear and mammogram self-reports in a low-income Hispanic population. *Am J Prev Med* 1995;11:94–8.
52. Thompson B, Taylor V, Goldberg H, Mullen M. Mammography status using patient self-reports and computerized radiology database. *Am J Prev Med* 1999;17:203–6.
53. Tsubono Y, Fukao A, Hisamichi S, Hosokawa T, Sugawara N. Accuracy of self-report for stomach cancer screening. *J Clin Epidemiol* 1994;47:977–81.
54. Tumiel-Berhalter LM, Finney MF, Jaen CR. Self-report and primary care medical record documentation of mammography and Pap smear utilization among low-income women. *J Natl Med Assoc* 2004;96:1632–9.
55. Volk RJ, Cass AR. The accuracy of primary care patients' self-reports of prostate-specific antigen testing. *Am J Prev Med* 2002;22:56–8.
56. Wain GV, Farnsworth A, Hacker NF. The Papanicolaou smear histories of 237 patients with cervical cancer. *Med J Aust* 1992;157:14–6.
57. Walter SD, Clarke EA, Hatcher J, Stitt LW. A comparison of physician and patient reports of Pap smear histories. *J Clin Epidemiol* 1988;41:401–10.
58. Warnecke RB, Havlicek P, Graham S. Characteristics of blacks obtaining Papanicolaou smears. *Cancer* 1976;37:2015–25.
59. 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.
60. Whitman S, Lacey L, Ansell D, Chen EH, Dell J, Phillips CW. Do chart reviews and interviews provide the same information about breast and cervical cancer screening? *Int J Epidemiol* 1993;22:393–7.
 61. Zapka JG, Bigelow C, Hurley T, et al. Mammography use among sociodemographically diverse women: the accuracy of self-report. *Am J Public Health* 1996;86:1016–21.
 62. DerSimonian R, Laird N. Meta-analysis in Clin trials. *Control Clin Trials* 1986;7:177–88.
 63. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.
 64. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63.
 65. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia (PA): Lippincott-Raven Publishers; 1998.
 66. Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R. Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *J Natl Cancer Inst* 2001;93:1704–13.
 67. Feigl P, Glaefke G, Ford L, Diehr P, Chu J. Studying patterns of cancer care: how useful is the medical record? *Am J Public Health* 1988;78:526–33.
 68. Koseoff J, Fink A, Brook RH, Chassin MR. The appropriateness of using a medical procedure. Is information in the medical record valid? *Med Care* 1987;25:196–201.
 69. Horwitz RJ, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. *J Chron Dis* 1984;37:825–31.
 70. Johnson TP, Cho YI, Holbrook AL, O'Rourke D, Warnecke RB, Chavez N. Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Ann Epidemiol* 2006;16:661–8.
 71. Holt K, Franks P, Meldrum S, Fiscella K. Mammography self-report and mammography claims: racial, ethnic, and socioeconomic discrepancies among elderly women. *Med Care* 2006;44:513–8.