**Agreement between self-reported and registered colorectal cancer screening
A meta-analysis**

Dodou, D; de Winter, JCF

DOI

[10.1111/ecc.12204](https://doi.org/10.1111/ecc.12204)

Publication date

2015

Document Version

Final published version

Published in

European Journal of Cancer Care

Citation (APA)

Dodou, D., & de Winter, JCF. (2015). Agreement between self-reported and registered colorectal cancer screening: A meta-analysis. *European Journal of Cancer Care*, 24(3), 286-298.
<https://doi.org/10.1111/ecc.12204>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Agreement between self-reported and registered colorectal cancer screening: a meta-analysis

D. DODOU, ^{PHD}, *Department of BioMechanical Engineering, Delft University of Technology, Delft*, & J.C.F. DE WINTER, ^{PHD}, *Department of BioMechanical Engineering, Delft University of Technology, Delft, the Netherlands*

DODOU D. & DE WINTER J.C.F. (2015) *European Journal of Cancer Care* **24**, 286–298

Agreement between self-reported and registered colorectal cancer screening: a meta-analysis

This random-effects meta-analysis investigates the accuracy of self-reported colorectal cancer screening history as a function of screening mode (colonoscopy, flexible sigmoidoscopy, faecal occult blood testing – FOBT, double-contrast barium enema – DCBE) and survey mode (written, telephone, face-to-face). Summary estimates of sensitivity, specificity, positive predictive value (PPV) and area under the receiver operating characteristic curve (AUC) were calculated. Medical record data were used as reference. We included 23 studies comprising 11 592 subjects. Colonoscopy yielded higher AUC [0.948, 95% confidence interval (CI) = 0.918, 0.968] than flexible sigmoidoscopy (0.883, 95% CI = 0.849, 0.911) and FOBT (0.869, 95% CI = 0.833, 0.898). Colonoscopy showed the highest sensitivity (0.888, 95% CI = 0.835, 0.931), whereas specificity was comparable between screening modes (ranging from 0.802 for FOBT to 0.904 for DCBE). AUC was not significantly different between survey modes. Prevalence of screening history correlated positively with sensitivity and negatively with specificity, possibly because of errors in the medical records. In conclusion, the accuracy of self-reported cancer screening is generally moderate, and higher for colonoscopy than for sigmoidoscopy and FOBT.

Keywords: cancer, colonoscopy, mass screening, medical audit, self-report.

INTRODUCTION

Colorectal cancer (CRC) screening reduces mortality by facilitating detection and removal of precancerous lesions and detection and treatment of early-stage adenocarcinomas (Levin *et al.* 2008). Four screening tests are recognised as effective for reducing mortality from CRC: faecal occult blood testing (FOBT), flexible sigmoidoscopy, colonoscopy and double-contrast barium enema (DCBE), with DCBE becoming uncommon (Levin *et al.* 2008; Smith *et al.* 2012). According to 2010 data

from the Behavioral Risk Factor Surveillance System based on 223 980 individuals aged between 50 and 75 years, who participated in a random-digit dialled telephone survey in the USA, 60.3% of the surveyed had undergone colonoscopy within the past 10 years, 11.7% a FOBT within the previous year, and 5.7% a flexible sigmoidoscopy within the past 5 years (Joseph *et al.* 2012). The Survey of Health, Aging and Retirement in 11 European countries revealed that 15.2% and 20.2% of 18 139 adults older than 50 years reported having undergone a colonoscopy or FOBT, respectively, within the last 10 years (Stock & Brenner 2010).

Surveys are a broadly used tool for identifying population trends in screening compliance (cf. the National Health Interview Survey in the United States and the European Health Interview survey). Reporting autobiographic information is a cognitively demanding task that requires interpretation of the question, retrieval of

Correspondence address: Dimitra Dodou, Department of BioMechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, the Netherlands (e-mail: d.dodou@tudelft.nl).

Conflict of interest: None

Accepted 3 April 2014

DOI: 10.1111/ecc.12204

European Journal of Cancer Care, 2015, **24**, 286–298

information from memory, judgement formation and decision making about information disclosure (Tourangeau 1984, 1987; Strack & Martin 1987; Prohaska *et al.* 1998; Sirken *et al.* 1999). Common errors when responding to CRC screening surveys include confusing screening tests with each other (e.g. flexible sigmoidoscopy with colonoscopy) (Baier *et al.* 2000; Vernon *et al.* 2004) or with other medical investigations (e.g. DCBE with upper gastrointestinal roentgenography) (Jones *et al.* 2008; Fisher *et al.* 2009), and underestimating the time since the last screening (so-called telescoping error; Prohaska *et al.* 1998; Tisnado *et al.* 2006).

In a meta-analysis examining the accuracy of self-reported history of seven types of cancer screening (colorectal endoscopy, faecal occult blood testing, digital rectal exam, mammography, clinical breast exam, Pap smear and prostate-specific antigen testing), Rauscher *et al.* (2008) estimated accuracies of self-reported CRC screening based on eight studies of FOBT and four studies of colorectal endoscopy (i.e. colonoscopy and/or flexible sigmoidoscopy) published up to 2000. Their summary estimates of sensitivity, specificity and positive predictive value, with medical records as reference, were 82%, 78% and 62% for FOBT and 79%, 90% and 56% for endoscopy.

Screening history can be reported through a written questionnaire, telephone interview or face-to-face interview. It has been suggested that telephone interviews are experienced as lengthier than face-to-face interviews of equal duration, and that telephone and face-to-face respondents are more likely to respond in a socially desirable manner (Holbrook *et al.* 2003; Bowling 2005) at cost of specificity (Johnson *et al.* 2005), and more reluctant to disclose information regarding sensitive topics (Holbrook *et al.* 2003) at cost of sensitivity than written-questionnaire respondents. A study evaluating self-reported mammography in low-income women (Armstrong *et al.* 2004) showed that specificity was lower for telephone interviews than for written questionnaires, possibly due to socially desirable responding in verbal communication. Sensitivity was higher for telephone interviews than for written questionnaires, possibly because in the latter case the surveyed had difficulty understanding the questions and no opportunity for clarifications. Others reported no significant effects of survey mode (see Zapka *et al.* 1996 for mammography and Beebe *et al.* 2008; Vernon *et al.* 2008 for CRC). In their meta-analysis, Rauscher *et al.* (2008) showed that, compared with written and face-to-face interviews, self-reports from telephone interviews tended to be less accurate in terms of sensitivity, specificity and positive predictive value, but, as noted by the authors, the number of available studies was small.

The previous meta-analysis (Rauscher *et al.* 2008) needs updating as a large number of studies investigating the accuracy of self-reported CRC screening history have been published since 2000, offering the opportunity for a more powerful synthesis. This paper presents a meta-analysis of the accuracy of self-reported CRC screening as a function of screening mode and survey mode.

METHOD

We conducted a literature search (last search for updates: 8 February 2014) in PubMed (all-fields) using the Boolean operation: ('self-report' OR 'self-reporting' OR 'self-reported') AND colorectal AND cancer AND screening AND (colonoscopy OR sigmoidoscopy OR endoscopy OR FOBT OR FOB OR barium OR 'immunochemical testing') and in Google Scholar using the search query: ('self-report' colorectal cancer screening colonoscopy OR sigmoidoscopy OR endoscopy OR FOBT OR FOB OR barium OR 'immunochemical testing'). No year restriction was set in either search engine. The Google Scholar search was conducted with the Publish or Perish software (Harzing 2011). Because this software limits the results of one query to 1000 hits, we repeated the Google Scholar search multiple times, each time for a different time period, so that each query yielded fewer than 1000 hits. PubMed was chosen as it allows for Boolean operations and Medical Subject Headings (MeSH) terms. Google Scholar was chosen because it uses full-text search and provides access to a large number of articles, reports, theses and conference papers, and is a recommended tool for systematic reviews (Gehanno *et al.* 2013; Shariff *et al.* 2013; De Winter *et al.* 2014).

Our inclusion criterion was whether CRC screening histories reported by patients were provided and compared with medical records from a health foundation, health maintenance organisation, hospital, laboratory, primary care practice or medical office. Studies in which physician records/reports, administration data from a medical office or laboratory receipts were used as reference were also included. Studies in which insurance claims were used as reference were not included, as claims exhibit lower accuracy than medical records (Fowles *et al.* 1995; Fiscella *et al.* 2006).

The abstracts and/or full texts of the retrieved articles were read to identify whether the inclusion criterion was fulfilled. The reference lists of Rauscher *et al.* (2008) and of each study fulfilling the inclusion criterion were reviewed with the aim to retrieve additional studies. Study extraction from PubMed and Google Scholar was conducted by the first author. Checking of reference lists

and additional, opportunistic Google Scholar searches were conducted by both authors. Study selection was done by the first author, with uncertainties about inclusion eliminated through discussion with the second author.

The following data were collected per study (when available): publication year, sample size, time frame in which last screening was conducted, screening mode, survey mode (colonoscopy, flexible sigmoidoscopy, endoscopy, FOBT, DCBE; where endoscopy denotes studies that provided data for colonoscopy and flexible sigmoidoscopy combined), response rates, whether the surveyed were involved in an intervention aiming to increase awareness of CRC or not, country, recruitment method (i.e. whether subjects were recruited from a health organisation, or via random digit dialling), gender (% men), ethnicity/race (% of the surveyed being White, Black, Hispanic or Asian), and level of completed education (% of surveyed having completed high school and beyond, or college and beyond). For each study and screening mode, the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were retrieved, with medical record data as reference. If TP, TN, FP and FN values were not reported, these were calculated by solving a 4×4 system of equations of sensitivity $[TP/(TP + FN)]$, specificity $[TN/(FP + TN)]$, positive predictive value [PPV; $TP/(TP + FP)$]; also called precision rate; it estimates the probability that a positive result is truly positive], and concordance $[(TP + TN)/(TP + FP + TN + FN)]$ – or other functions of TP, TN, FP and FN available, such as report-to-record ratio, negative predictive value, etc. To calculate test accuracy measures for Vernon *et al.* (2008), in which subjects were randomised in three groups, each group subjected to a different survey mode, we merged subjects by summing TP, TN, FP and FN across survey modes.

All data were extracted twice by the first author, in separate occasions. Uncertainties that were raised during the calculation of TP, TN, FP and FN were eliminated through discussion with the second author. Large studies and studies with outlying results were checked by both authors separately. To reduce the risk of data extraction errors, sample size as well as sensitivity, specificity and/or other diagnostic accuracy estimates available in each study were re-calculated using the reported/estimated TP, TN, FP and FN, and compared with the values provided in that study. In addition, for the studies meta-analysed by Rauscher *et al.* (2008), sensitivity, specificity and PPV were calculated and compared with the values in Rauscher *et al.* (comparison values extracted from plots with graphics software). Disagreements between our values and the values in the original studies or in Rauscher *et al.* were resolved through discussion between the two authors.

Summary estimates of sensitivity, specificity, PPV and prevalence of screening history $[(TP + FN)/(TP + FN + FP + TN)]$ were calculated per screening mode, both per survey mode and for all survey modes combined, by applying the DerSimonian-Laird random-effects method. Summary estimates for endoscopy were calculated only when the corresponding TP, FP, TN and FN were provided in/were calculable from the original studies. We did not merge TP, FP, TN and FN of colonoscopy and flexible sigmoidoscopy ourselves, as a FP of colonoscopy could be TP for flexible sigmoidoscopy and so on. Diagnostic odds ratios $[DOR = (TP/FN)/(FP/TN)]$ were estimated per sample, and an overall DOR was estimated by applying the DerSimonian-Laird method. Overall DOR was not used as a metric itself but to determine a symmetric summary receiver operating characteristic (ROC) curve and to calculate the area under this curve (AUC) (Zamora *et al.* 2006a,b), as a measure of overall diagnostic accuracy. AUC can be seen as an average sensitivity across all specificities (Glas *et al.* 2003).

A random-effects method was chosen, as homogeneity is typically low in meta-analyses of test accuracy studies (Macaskill *et al.* 2010). Homogeneity among studies was assessed by means of the I^2 statistic, defined as the percentage of the observed variances that is due to real differences in effect sizes (Higgins *et al.* 2003; Higgins & Green 2011). Pearson correlations weighted by sample size between sensitivity and specificity were also calculated. A strong negative correlation would indicate that a univariate meta-analysis (i.e. pooling sensitivity and specificity separately) is not appropriate and that a bivariate method (i.e. pooling pairs of sensitivity and specificity) is required instead (Reitsma *et al.* 2005). We calculated sample-size weighted Spearman correlation coefficients to assess the relationship between prevalence of screening history versus AUC, sensitivity and specificity. We also calculated sample-size weighted Spearman correlation coefficients between the time frame since the last screening, ethnicity (% Whites), educational level (% of surveyed having completed high school and beyond, or college and beyond) and gender (% men), on the one hand, and prevalence, AUC, sensitivity and specificity, on the other. All analyses were conducted in MATLAB (Version R2012b, The MathWorks, Inc., Natick, MA, USA). The codes used are provided as Supporting Information (including the TP, TN, FP and FN per study).

RESULTS

PubMed yielded 143 titles and Google Scholar yielded 3404 titles. After removing 92 duplicates between the two

search engines, 3455 titles were reviewed and 2829 of these were excluded as not relevant or duplicates within a search engine. The abstracts of the remaining 626 studies were reviewed and 369 studies were excluded either because they were reviews or not relevant, or because of not providing data on screening history or because of investigating reports by physicians about the screening history of their patients instead of reports by patients themselves. The full texts of the remaining 257 publications were reviewed. Of these, 21 studies fulfilled the inclusion criteria and 236 were excluded because of not providing medical record data, not providing self-reported data, not comparing self-reported histories to medical records, not providing sufficient data to calculate TP, TN, FP and FN, or because of reporting the same or overlapping data with one of the 21 studies. Of the 21 studies, 17 were retrieved by both PubMed and Google Scholar and four by Google Scholar only. By reviewing the reference list of Rauscher *et al.* (2008) and the reference lists of the 21 studies fulfilling the inclusion criteria, two more suitable studies were retrieved, summing to 23 studies included in the analysis. Seven of these 23 studies were also included in the meta-analysis by Rauscher *et al.* (2008). Two of the 23 studies used the same sample, each reporting survey responses to different screening modes (Schenck *et al.* 2007, 2008). Summarising, our meta-analysis included 23 studies, and 22 unique samples consisting of 11 592 unique subjects. The number of unique subjects (TP + TN + FP + FN) per screening mode was 6222 for colonoscopy, 7736 for flexible sigmoidoscopy, 4940 for endoscopy, 10 769 for FOBT and 1753 for DCBE. In 16 of the 23 studies the subjects were surveyed about two or more screening modes. All but three of the included studies (Madlensky *et al.* 2003; Hoffmeister *et al.* 2007; Khoja *et al.* 2007) were conducted in the USA. Vernon *et al.* (2008) split their sample into three groups, each group subjected to a different survey mode (written, telephone and face-to-face). All other 21 samples were surveyed using only one survey mode (six written, 10 telephone and five face-to-face). Figure 1 shows the flow diagram of study selection and Table 1 lists the studies included in the meta-analysis.

I^2 of sensitive and specificity was generally above 90% for all screening modes (sensitivity: colonoscopy: 91.0%; flexible sigmoidoscopy: 85.9%; endoscopy: 91.0%; FOBT: 95.3%; DCBE: 35.1%; specificity: endoscopy: 92.7%; colonoscopy: 95.8%; flexible sigmoidoscopy: 96.7%; FOBT: 94.9%; DCBE: 93.7%). The Pearson correlation coefficients between sensitivity and specificity were overall moderate (endoscopy: $r = -0.282$, $n = 10$; colonoscopy: $r = 0.293$, $n = 12$; flexible sigmoidoscopy:

$r = 0.315$, $n = 13$, FOBT: $r = 0.496$, $n = 20$; DCBE: $r = -0.349$, $n = 5$), justifying a univariate meta-analysis.

The response rate of the studies included in the meta-analysis, defined here as the number of subjects who (fully) completed the survey over the number of eligible subjects, ranged from 13.4% to 88.8%, with an average weighted with sample size of 61.9% across survey modes, 66.5% for written surveys, 61.6% for telephone interviews and 60.4% for face-to-face interviews – rates that are typical for these survey modes (Hox & De Leeuw 1994). Response rates per study are provided in the Supporting Information.

Figure 2 provides summary estimates of AUC, sensitivity, specificity and PPV per screening mode and survey mode. The corresponding 95% confidence intervals are reported in the Supporting Information. For all survey modes combined, colonoscopy and endoscopy tended to yield higher AUC than flexible sigmoidoscopy, DCBE and FOBT, with the difference being statistically significant for colonoscopy versus flexible sigmoidoscopy and FOBT, and for endoscopy versus FOBT (significance judged by whether the confidence intervals were overlapping). Sensitivity was highest for endoscopy, followed by colonoscopy, FOBT, flexible sigmoidoscopy and DCBE (with the differences being statistically significant for: endoscopy versus flexible sigmoidoscopy, FOBT and DCBE; colonoscopy versus flexible sigmoidoscopy, FOBT and DCBE; and FOBT versus DCBE). Specificity was comparable among screening modes (no difference was statistically significant). PPV was highest for endoscopy and colonoscopy, followed by FOBT, flexible sigmoidoscopy and DCBE.

Inspection of the AUC estimates in Figure 2 suggests that written surveys performed better than the other two survey modes, but the difference was not statistically significant for any pairwise combination of the screening modes. Face-to-face interviews tended to yield higher sensitivity than telephone interviews for endoscopy and colonoscopy (difference being significant for colonoscopy), and written surveys yielded higher sensitivity than face-to-face interviews for FOBT. There were no effects of survey mode on specificity or PPV.

The prevalence of flexible sigmoidoscopy had a strong correlation with sensitivity (positive) and specificity (negative). Prevalence of FOBT correlated moderately and positively with sensitivity and negatively with specificity. Prevalence of endoscopy and colonoscopy did not reveal statistically significant correlations with any of the three accuracy estimates (Table 2; see Fig. 3 for scatter plots).

Prevalence correlated positively with time frame for all screening modes except DCBE, the correlation being

Figure 1. Flow diagram of study selection.

statistically significant for colonoscopy (sample-size weighted Spearman $\rho = 0.724$, $P = 0.008$). The correlation coefficients between time frame and sensitivity were positive for all screening modes, and statistically significant for FOBT ($\rho = 0.485$, $P = 0.030$). The correlation coefficients between the accuracy measures and % Whites, % of surveyed having completed high school and beyond, or college and beyond, and % men in the samples, were

generally inconsistent or difficult to interpret (numerical results provided in the Supporting Information).

DISCUSSION

In this study, the accuracy of self-reported CRC screening history was assessed per screening mode (colonoscopy, flexible sigmoidoscopy, endoscopy, FOBT, DCBE) and

Table 1. Studies included in the analysis

	Author [year]	Sample size	Time frame	Screening mode	Survey mode	Response rate	Intervention	Country	Recruitment method	% men	Ethnicity/race	Education
1	Baier <i>et al.</i> (2000)	229	4 years 4 years 4 years 4 years	FOBT COL SIG ENDO	Telephone		No	USA	Kaiser Permanente		85% White, 3% Black, 7% Hispanic	
2	Bastani <i>et al.</i> (2008)	123	1 year 1 year 1 year 1 year	FOBT COL SIG ENDO	Telephone	13.4%	Yes	USA	Relatives of cancer cases from California Cancer Registry	39%	33% White, 14% Black, 24% Hispanic, 28% Asian	
3	Brown and Adams (1992)	380	1 year 1–3 months	FOBT ENDO	Telephone	88.4%	No	USA	Harvard Community Health Plan	50%		
4	Daly <i>et al.</i> (2010)	417*	Up-to-date	COL	Written	53.2%	No	USA	Iowa Research Network	50%		50% ≥ High school
5	Ferrante <i>et al.</i> (2008)	723 321	1 year 5 years SIG/DCBE/ 10 years COL	FOBT ENDO	Face-to-face	80.2%	Yes	USA	Supporting Colorectal Cancer Outcomes through Participatory Enhancements (SCOPE)	40%	69% White, 17% Black, 8% Hispanic	88% ≥ High school, 40% ≥ College
6	Fisher <i>et al.</i> (2009)	201	5 years 10 years 10 years 10 years	FOBT COL SIG DCBE	Face-to-face	66.3%	No	USA	Veterans Affairs	74%	78% White, 16% Black	92% ≥ High school, 29% ≥ College
7	Gordon <i>et al.</i> (1993)	601 672	2 years 2 years	FOBT SIG	Written	75.1%	No	USA	Kaiser Permanente	45%	75% White, 12% Black, 3% Hispanic, 10% Asian/Other	65% ≥ High school, 35% ≥ College
8	Hall <i>et al.</i> (2004)	2086 2068 2080 2070	2 years 5 years 5 years 5 years	FOBT COL SIG ENDO	Telephone	64.8%	No	USA	Kaiser Permanente & HealthPartners	57%		92% ≥ High school, 34% ≥ College
9	Hiatt <i>et al.</i> (1995)	1274 1266	2 years 2 years	FOBT SIG	Telephone	78.8%	No	USA	Kaiser Permanente	47%	37% White, 63% Hispanic	41% ≥ High school
10	Hoffmeister <i>et al.</i> (2007)	252†	5 years	ENDO	Face-to-face	99.8%	No	Germany	DACHS study	54%		41% ≥ High school
11	Jones <i>et al.</i> (2008)	119‡ 120 118 118	1 year 5 years 5 years 5 years	FOBT COL SIG DCBE	Written	80.9%	Yes	USA	Wright County Colorectal Cancer Screening Project	50%		17% ≥ College
12	Khoja <i>et al.</i> (2007)	146	2 years 5 years	FOBT ENDO	Telephone	37.0%	No	Canada	Random-digit dial telephone survey in Calgary Health Region	38%		92% ≥ High school
13	Liptkus <i>et al.</i> (2003)	658	1 year	FOBT	Telephone	76.5%	Yes	USA	New Jersey Carpenters Fund	99%	95% White	52% ≥ High school

Table 1. Continued

	Author [year]	Sample size	Time frame	Screening mode	Survey mode	Response rate	Intervention	Country	Recruitment method	% men	Ethnicity/race	Education
14	Madlensky <i>et al.</i> (2003)	306	Ever	FOBT	Written	55.1%	No	Canada	Ontario Familial Colon Cancer Registry			
		332	Ever	COL								
		323	Ever	SIG								
15	Mandelson <i>et al.</i> (1999)	1021	5 years	FOBT	Telephone	80.3%	No	USA	Group Health Cooperative of Puget Sound	0%	89% White, 5% Black, 4% Asian	91% ≥ High school, 34% ≥ College
16	Mathews <i>et al.</i> (2005)	24	1 year	FOBT	Face-to-face	76.9%	No	USA	General Internal Medicine Clinic	33%	38% White, 58% Black, 2% Hispanic	38% ≥ High school, 15% ≥ College
		52	5/10 years	ENDO								
17	Maxwell <i>et al.</i> (2011)	142	1 year	FOBT	Written	68.3%	Yes	USA	Community-based organisations and churches	34%	100% Asian	
									Veterans Affairs		American	
18	Partin <i>et al.</i> (2008)	325	1 year	FOBT	Written	77.1%	No	USA		63%	70% White, 25% Black, 1% Hispanic	40% ≥ High school, 27% ≥ College
		323	10 years	COL								
		345	5 years	SIG								
		328	5/10 years	ENDO								
		304	5 years	DCBE								
19	Reiter <i>et al.</i> (2013)	721	1 year	FOBT	Telephone	65.2%	No	USA	Randomly selected residents of Appalachian Ohio counties	41%	97% White	59% ≥ High school
			10 years	COL								
			5 years	SIG								
20/21	Schenck <i>et al.</i> (2007, 2008)	561	5 years	COL	Telephone		No	USA	The Carolinas Center for Medical Excellence, the Quality Improvement Organization for Medicare	39%	76% White, 24% Black	82% ≥ High school
			4 years	SIG								
			4/5 years	ENDO								
			1 year	FOBT								
22	Shokar <i>et al.</i> (2011)	271	1 year	FOBT	Face-to-face	53.0%	No	USA	University-based family medicine clinic	43%	33% White, 33% Black, 34% Hispanic	89% ≥ High school
			10 years	COL								
			5 years	SIG								
23	Vernon <i>et al.</i> (2008)	857	5 years	DCBE								
			5 years	FOBT	Written, telephone and face-to-face	39.5%	No	USA	Kelsey-Seybold Clinic	34%	54% White, 29% Black, 11% Hispanic	98% ≥ High school, 55% ≥ College
			5 years	COL								
			5 years	SIG								
			5 years	ENDO								
			5 years	DCBE								

*Only the group of non-impaired subjects was included in the analysis.

†Only the control group was included in the analysis.

‡Only the baseline data were included in the analysis.

Abbreviations: COL, colonoscopy; DCBE, double-contrast barium enema; ENDO, endoscopy; FOBT, faecal occult blood testing; SIG, flexible sigmoidoscopy.

Figure 2. Random-effects summary estimates for area under the receiver operating characteristic curve (AUC), sensitivity, specificity and positive predictive value (PPV) of self-reported history of colorectal cancer screening per screening mode and survey mode. COL, colonoscopy; DCBE, double-contrast barium enema; ENDO, endoscopy; FOBT, faecal occult blood testing; SIG, flexible sigmoidoscopy.

Table 2. Spearman correlations weighted by sample size between the prevalence of screening history versus area under the receiver operating characteristic curve (AUC), sensitivity and specificity, per screening mode

	<i>n</i>	Summary prevalence % (95% CI)	Correlations with prevalence (<i>P</i> -value)		
			AUC	Sensitivity	Specificity
COL	12	35.2 (26.9, 44.0)	0.137 (0.671)	0.425 (0.169)	−0.311 (0.325)
SIG	13	14.5 (7.2, 23.8)	0.513 (0.073)	0.849 (<0.001)	−0.816 (<0.001)
ENDO	10	45.2 (37.8, 52.7)	0.443 (0.200)	0.503 (0.138)	0.188 (0.603)
FOBT	20	27.5 (18.3, 37.7)	0.119 (0.619)	0.464 (0.039)	−0.556 (0.011)
DCBE	5	6.5 (3.0, 11.2)	0.164 (0.792)	0.164 (0.792)	0.735 (0.157)

Abbreviations: CI, confidence interval; COL, colonoscopy; DCBE, double-contrast barium enema; ENDO, endoscopy; FOBT, faecal occult blood testing; *n*, number of samples; SIG, flexible sigmoidoscopy.

survey mode (written, telephone, face-to-face), with medical record data as reference. Twenty-three studies with 11 592 unique subjects were included in the analysis.

Results showed that the accuracy of self-reports varied with screening mode. Colonoscopy yielded higher overall diagnostic accuracy (measured by AUC) than flexible sigmoidoscopy and FOBT, and endoscopy yielded higher overall diagnostic accuracy than FOBT. Among tests, endoscopy and colonoscopy yielded the highest sensitivity. PPV estimates showed that in the best case (i.e. colonoscopy and endoscopy), about 77% to 80% of the self-reported positive screening histories were confirmed

by the medical record. For FOBT, flexible sigmoidoscopy and DCBE, 46%, 56% and 34% of the self-reported data, respectively, were confirmed by the medical record. Our sensitivity (78%), specificity (80%) and PPV (56%) estimates for FOBT are in line with Rauscher *et al.* (2008; 82%, 78% and 62% respectively), whereas for endoscopy, we found higher sensitivity (91% vs. 79%) and PPV (81% vs. 56%) and lower specificity (82% vs. 90%) than the estimates in Rauscher *et al.* (2008).

A possible error when reporting on past flexible sigmoidoscopy is that the procedure is easily confused with colonoscopy (Baier *et al.* 2000; Vernon *et al.* 2004). A

Figure 3. Scatter plots of sensitivity and specificity versus prevalence of screening history. The area of the circles corresponds to the sample size. The circles in the legends correspond to a sample size of 300. DCBE, double-contrast barium enema; FOBT, faecal occult blood testing.

possible reason for the low sensitivity of FOBT could be that it is a simple procedure often conducted at home and may therefore be more easily forgotten than, for example, colonoscopy, which is an invasive procedure that requires planning (e.g. taking time off work), is conducted in the outpatient department of a hospital, in a clinic or in a physician's office, and involves discomfort (cf. Cahill &

McGaugh 1998, describing cognitive and neural mechanisms that modulate memory formation, storage and recall of emotionally arousing events). Another possible source of error in reporting FOBT is that respondents may confuse in-office with home-based FOBT (Jones *et al.* 2010). In 15 of the 19 studies with FOBT data, respondents were explicitly asked about home-based FOBT, whereas the other four studies do not mention what type of FOBT was analysed and whether the type of FOBT was clarified/defined in the survey. A possible cause of the low accuracy of self-reporting for DCBE is that, despite the test descriptions provided in the surveys, DCBE can be confused with upper gastrointestinal roentgenography (Jones *et al.* 2008; Fisher *et al.* 2009).

Our meta-analysis found that endoscopy yielded higher sensitivity, higher PPV and lower specificity than colonoscopy and flexible sigmoidoscopy. This result can be explained by the fact that in nine of the 10 studies reporting results on endoscopy, the endoscopy data were a retrospective combination of the responses to separate questions on colonoscopy and flexible sigmoidoscopy, rather than responses to a question mentioning the word 'endoscopy'. This retrospective combination by definition increases the number of TP and reduces the number of TN, with respect to the corresponding values of the two screening modes separately.

In agreement with Rauscher *et al.* (2008), written questionnaires tended to yield higher overall diagnostic accuracies than telephone and face-to-face interviews, but the number of studies was small and none of the comparisons reached statistical significance. Survey mode seemed to have a differential effect on sensitivity, with face-to-face interviews yielding higher sensitivity than telephone interviews for colonoscopy, and written surveys being related to higher sensitivity than face-to-face interviews for FOBT.

Study quality and heterogeneity

In this study, medical records were used as reference and every deviance of self-reported data from the medical record was considered as error in the self-report. Medical records, however, are also susceptible to error. Human errors in transcription may occur, and screening tests may be missing (Madlensky *et al.* 2003), delayed or unrecorded (e.g. non-billable, Ferrante *et al.* 2008). Lack of standardised audit processes makes the quality of medical records strongly dependent on the policy adopted by the physician's office (Fowles *et al.* 1995; Diamond *et al.* 2001). Jones *et al.* (2008) found that although some progress notes indicated that a CRC test had taken place during a

consultation, no test results were found in the medical record. For home-based FOBT, results could possibly be misplaced or returned to another physician than the respondent's primary medical provider (Bastani *et al.* 2008; Schenck *et al.* 2008). Moreover, of all 20 studies with FOBT data, only Schenck *et al.* (2008) mentioned the types of FOBT (guaiac, immunochemical or not specified) registered in the medical records. Older studies most possibly used guaiac FOBT. Guaiac FOBT is less regularly reimbursed than FOBT and may therefore be more often missing from the medical record (Schenck *et al.* 2008). The accuracy of the medical records in the studies included in our meta-analysis may also depend on whether or not researchers looked for data in multiple clinics and physician offices, whether they carried out counterchecks for the abstracted data, whether the abstractor was a trained researcher/nurse or a 'busy physician' (Brown & Adams 1992), and whether medical record data were collected prospectively or retrospectively. It is also possible that some subjects conducted the screening test outside their primary health provider and that this information was missing from their medical record. More information about the data collection quality per study is provided in the Supporting Information.

Errors in the medical records may explain the positive correlation that we found between prevalence and sensitivity and negative correlation between prevalence and specificity. According to their definitions, sensitivity and specificity are independent of prevalence. As early as 1966, however, Buck and Gart (1966) recognised that sensitivity and specificity may correlate with prevalence, a topic since addressed in a number of articles (Hui & Walter 1980; Thibodeau 1981; Vacek 1985; Brenner & Gefeller 1997; Rutjes 2005; Leeftang *et al.* 2009). Boyko *et al.* (1988) showed that errors in the reference (i.e. sensitivity and specificity of the reference itself being lower than 100%) can lead to an underestimation of sensitivity, particularly for low prevalence, and an underestimation of specificity, particularly for high prevalence. Considering that for 47 of the 56 samples included in our meta-analysis, the reported prevalence was lower than 0.5 (with 34 samples having a prevalence lower than 0.3), sensitivity estimates may have been underestimated. This phenomenon may (partially) account for the comparatively low summary estimates of sensitivity for DCBE, FOBT and flexible sigmoidoscopy.

Note also that the observed prevalences are causally related to the time frame used in the survey. That is, the wider the time frame, the higher the probability that one has been screened in this time frame. Our correlational

analysis concurs that time frame tended to correlate positively with prevalence as well as with sensitivity.

Respondent bias is at play as respondents are more likely to have an updated screening test than non-respondents (Jones *et al.* 2008). The majority of studies used enrollees in a managed care health plan and/or convenience samples (e.g. consecutive or non-consecutive patients in the waiting room of a clinic). In three studies (Mandelson *et al.* 1999; Madlensky *et al.* 2003; Bastani *et al.* 2008), the samples consisted of first-degree relatives of CRC patients, who are known to have an increased risk for CRC (Johns & Houlston 2001) and increased perception of this risk (Stark *et al.* 2006; Rees *et al.* 2008; Pieper *et al.* 2012). In four studies (Lipkus *et al.* 2003; Bastani *et al.* 2008; Ferrante *et al.* 2008; Jones *et al.* 2008), the surveyed were involved in an intervention aimed to increase awareness of CRC (e.g. mailing of FOBT kits), which may have decreased specificity due to social desirability (Jones *et al.* 2008). Research has shown that racial differences exist in the accuracy of self-reported CRC screening (Zapka *et al.* 1996; McPhee *et al.* 2002; Fiscella *et al.* 2006; Katz *et al.* 2011), but we found no such systematic or clearly interpretable differences in the correlational analysis.

Twenty out of the 23 studies were conducted in the USA. Whether persons actually are invited for screening depends upon the healthcare provider, with some organisations (e.g. Veterans Affairs or Kaiser Permanente North California; Swan *et al.* 2012) providing programmatic screening and other organisations providing only advisory guidelines or only opportunistic screening. Awareness, willingness to respond and adherence to past testing is generally higher in subjects recruited from health organisations with organised screening (Baier *et al.* 2000).

Differences between surveys are another source of heterogeneity. Questionnaires use diverse terminology (e.g. FOBT is named variously 'stool blood', 'hemoccult', or 'stool guaiac'), test descriptions, time frames (e.g. testing within the last year, within the last 5 years, or ever; Vernon *et al.* 2008), and question wording is known to influence self-reporting behaviour (Schwarz 1999). Baier *et al.* (2000) found that adding a sentence stating that sedation is used in colonoscopy but not in flexible sigmoidoscopy improved the accuracy of self-reports for both tests. Based on cognitive interviewing (i.e. think-aloud and verbal probing) focused on issues related to comprehending and interpreting the questions, Vernon *et al.* (2008) revised the National Cancer Institute Colorectal Cancer Screening questionnaire (NCI CRCS) by replacing 'screening' with 'checkup' and 'procedure' with 'test', and by using multiple names for the same test

(e.g. both 'stool blood test' and 'FOBT') in the description of the test. Three of the included studies in our meta-analysis used the NCI CRCS questionnaire, while the rest used other, purpose-made, questionnaires. None of the included studies used web-based surveys, a survey type that may deserve further investigation, as the Internet is increasingly used by the aged population. Web-based surveys may prevent transcription errors (Van Gelder *et al.* 2010), and include validation checks that warn when a respondent provides implausible answers or tries to submit an incomplete questionnaire. Finally, the accuracy of self-report for emerging screening methods

(e.g. stool DNA, virtual colonoscopy) requires investigation as well.

Summarising, agreement between self-reports and medical records was generally moderate, with sensitivity and specificity values of about 0.8. Colonoscopy yielded higher accuracies than FOBT, possibly because the intensive preparation and invasiveness of endoscopic screening are better remembered than FOBT sampling. The positive correlation between prevalence and sensitivity points towards errors in the medical records, which, in combination with the low prevalence, may have led to an underestimation of sensitivity.

REFERENCES

- Armstrong K., Long J.A. & Shea J.A. (2004) Measuring adherence to mammography screening recommendations among low-income women. *Preventive Medicine* **38**, 754–760.
- Baier M., Calonge N., Cutter G., McClatchey M., Schoentgen S., Hines S., Marcus A. & Ahnen D. (2000) Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiology, Biomarkers and Prevention* **9**, 229–232.
- Bastani R., Glenn B.A., Maxwell A.E., Ganz P.A., Mojica C.M. & Chang L.C. (2008) Validation of self-reported colorectal cancer (CRC) screening in a study of ethnically diverse first-degree relatives of CRC cases. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 791–798.
- Beebe T.J., Jenkins S.M., Anderson K.J., Davern M.E. & Rockwood T.H. (2008) The effects of survey mode and asking about future intentions on self-reports of colorectal cancer screening. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 785–790.
- Bowling A. (2005) Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health* **27**, 281–291.
- Boyko E.J., Alderman B.W. & Baron A.E. (1988) Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease. *Journal of General Internal Medicine* **3**, 476–481.
- Brenner H. & Gefeller O. (1997) Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in Medicine* **16**, 981–991.
- Brown J.B. & Adams M.E. (1992) Patients as reliable reporters of medical care process: recall of ambulatory encounter events. *Medical Care* **30**, 400–411.
- Buck A.A. & Gart J.J. (1966) Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence. *American Journal of Epidemiology* **83**, 586–592.
- Cahill L. & McGaugh J.L. (1998) Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences* **21**, 294–299.
- Daly J.M., Levy B.T., Joshi M., Xu Y. & Jogerst G.J. (2010) Patient clock drawing and accuracy of self-report compared with chart review for colorectal cancer (CRC) screening. *Archives of Gerontology and Geriatrics* **50**, 341–344.
- De Winter J.C.F., Zadpoor A.A. & Dodou D. (2014) The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics* **98**, 1547–1565.
- Diamond C.C., Rask K.J. & Kohler S.A. (2001) Use of paper medical records versus administrative data for measuring and improving health care quality: are we still searching for a gold standard? *Disease Management* **4**, 121–130.
- Ferrante J.M., Ohman-Strickland P., Hahn K.A., Hudson S.V., Shaw E.K., Crosson J.C. & Crabtree B.F. (2008) Self-report versus medical records for assessing cancer-preventive services delivery. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 2987–2994.
- Fiscella K., Holt K., Meldrum S. & Franks P. (2006) Disparities in preventive procedures: comparisons of self-report and Medicare claims data. *BMC Health Services Research* **6**, 122.
- Fisher D.A., Voils C.I., Coffman C.J., Grubber J.M., Dudley T.K., Vernon S.W., Bond J.H. & Provenzale D. (2009) Validation of a questionnaire to assess self-reported colorectal cancer screening status using face-to-face administration. *Digestive Diseases and Sciences* **54**, 1297–1306.
- Fowles J.B., Lawthers A.G., Weiner J.P., Garnick D.W., Petrie D.S. & Palmer R.H. (1995) Agreement between physicians' office records and Medicare Part B claims data. *Health Care Financing Review* **16**, 189–199.
- Gehanno J.-F., Rollin L. & Darmoni S. (2013) Is the coverage of google scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making* **13**, 7.
- Glas A.S., Lijmer J.G., Prins M.H., Bonsel G.J. & Bossuyt P.M. (2003) The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology* **56**, 1129–1135.
- Gordon N.P., Hiatt R.A. & Lampert D.I. (1993) Concordance of self-reported data and medical record audit for six cancer screening procedures. *Journal of the National Cancer Institute* **85**, 566–570.
- Hall H.I., Van den Eeden S.K., Tolsma D.D., Rardin K., Thompson T., Sinclair A.H., Madlon-Kay D.J. & Nadel M. (2004) Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Preventive Medicine* **39**, 27–35.
- Harzing A.W. (2011) *The Publish or Perish Book: A Guide to the Software*. Tarma Software Research, Melbourne, Australia.
- Hiatt R.A., Perezstable E.J., Quesenberry C., Sabogal F., Oterosabogal R. & McPhee S.J. (1995) Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Preventive Medicine* **24**, 278–285.
- Higgins J.P., Thompson S.G., Deeks J.J. & Altman D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.)* **327**, 557–560.
- Higgins J.P.T. & Green S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration. Available at: <http://www.cochrane-handbook.org>
- Hoffmeister M., Chang-Claude J. & Brenner H. (2007) Validity of self-reported endoscopies of the large bowel and implications for estimates of colorectal cancer

- risk. *American Journal of Epidemiology* **166**, 130–136.
- Holbrook A.L., Green M.C. & Krosnick J.A. (2003) Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly* **67**, 79–125.
- Hox J.J. & De Leeuw E.D. (1994) A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity* **28**, 329–344.
- Hui S.L. & Walter S.D. (1980) Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–171.
- Johns L.E. & Houlston R.S. (2001) A systematic review and meta-analysis of familial colorectal cancer risk. *The American Journal of Gastroenterology* **96**, 2992–3003.
- Johnson T.P., O'Rourke D.P., Burris J.E. & Warnecke R.B. (2005) An investigation of the effects of social desirability on the validity of self-reports of cancer screening behaviors. *Medical Care* **43**, 565–573.
- Jones R.M., Mongin S.J., Lazovich D., Church T.R. & Yeazel M.W. (2008) Validity of four self-reported colorectal cancer screening modalities in a general population: differences over time and by intervention assignment. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 777–784.
- Jones R.M., Devers K.J., Kuzel A.J. & Woolf S.H. (2010) Patient-reported barriers to colorectal cancer screening: a mixed-methods analysis. *American Journal of Preventive Medicine* **38**, 508–516.
- Joseph D.A., King J.B., Miller J.W. & Richardson L.C. (2012) Prevalence of colorectal cancer screening among adults – behavioral risk factor surveillance system, United States, 2010. *MMWR. Morbidity and Mortality Weekly Report* **61**, 51–56.
- Katz M.L., Ferketich A.K., Paskett E.D., Harley A., Reiter P.L., Lemeshow S., Westman J.A., Clinton S.K. & Bloomfield C.D. (2011) Cancer screening practices among Amish and non-Amish adults living in Ohio Appalachia. *The Journal of Rural Health* **27**, 302–309.
- Khoja S., McGregor S.E. & Hilsden R.J. (2007) Validation of self-reported history of colorectal cancer screening. *Canadian Family Physician* **53**, 1192–1197.
- Leeftang M.M., Bossuyt P.M. & Irwig L. (2009) Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *Journal of Clinical Epidemiology* **62**, 5–12.
- Levin B., Lieberman D.A., McFarland B., Smith R.A., Brooks D., Andrews K.S., Dash C., Giardiello F.M. & Glick S. (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians* **58**, 130–160.
- Lipkus I.M., Samsa G.P., Dement J., Skinner C.S., Green L.S.G., Pompeii L. & Ransohoff D.F. (2003) Accuracy of self-reports of fecal occult blood tests and test results among individuals in the carpentry trade. *Preventive Medicine* **37**, 513–519.
- Macaskill P., Gatsonis C., Deeks J., Harbord R.M. & Takwoingi Y. (2010) Analysing and presenting results. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0* (eds Deeks J.J., Bossuyt P.M. & Gatsonis C.). The Cochrane Collaboration. Available at: <http://srdta.cochrane.org/>
- Madlensky L., McLaughlin J. & Goel V. (2003) A comparison of self-reported colorectal cancer screening with medical records. *Cancer Epidemiology, Biomarkers and Prevention* **12**, 656–659.
- Mandelson M.T., LaCroix A.Z., Anderson L.A., Nadel M.R. & Lee N.C. (1999) Comparison of self-reported fecal occult blood testing with automated laboratory records among older women in a health maintenance organization. *American Journal of Epidemiology* **150**, 617–621.
- Matthews B.A., Nattinger A.B. & Anderson R.C. (2005) Accuracy and certainty of self-report for colorectal cancer screening among ambulatory patients. *Psychology, Health and Medicine* **10**, 1–16.
- Maxwell A.E., Crespi C.M., Danao L.L., Antonio C., Garcia G.M. & Bastani R. (2011) Alternative approaches to assessing intervention effectiveness in randomized trials: application in a colorectal cancer screening study. *Cancer Causes and Control* **22**, 1233–1241.
- McPhee S.J., Nguyen T.T., Shema S.J., Nguyen B., Somkin C., Vo P. & Pasick R. (2002) Validation of recall of breast and cervical cancer screening by women in an ethnically diverse population. *Preventive Medicine* **35**, 463–473.
- Partin M.R., Grill J., Noorbaloochi S., Powell A.A., Burgess D.J., Vernon S.W., Halek K., Griffin J.M., Van Ryn M. & Fisher D.A. (2008) Validation of self-reported colorectal cancer screening behavior from a mixed-mode survey of veterans. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 768–776.
- Pieper C., Kolankowska I. & Jöckel K.H. (2012) Does a screening questionnaire for familial and hereditary colorectal cancer risk work in a health insurance population? *European Journal of Cancer Care* **21**, 758–765.
- Prohaska V., Brown N.R. & Belli R.F. (1998) Forward telescoping: the question matters. *Memory (Hove, England)* **6**, 455–465.
- Rauscher G.H., Johnson T.P., Cho Y.I. & Walk J.A. (2008) Accuracy of self-reported cancer-screening histories: a meta-analysis. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 748–757.
- Rees G., Martin P.R. & Macrae F.A. (2008) Screening participation in individuals with a family history of colorectal cancer: a review. *European Journal of Cancer Care* **17**, 221–232.
- Reiter P.L., Katz M.L., Oliveri J.M., Young G.S., Llanos A.A. & Paskett E.D. (2013) Validation of self-reported colorectal cancer screening behaviors among Appalachian residents. *Public Health Nursing* **30**, 312–322.
- Reitsma J.B., Glas A.S., Rutjes A.W., Scholten R.J., Bossuyt P.M. & Zwinderman A.H. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* **58**, 982–990.
- Rutjes A.W.S. (2005) *Sources of Bias and Variation in Diagnostic Accuracy Studies* (dissertation). Amsterdam, the Netherlands: University of Amsterdam (UvA).
- Schenck A.P., Klabunde C.N., Warren J.L., Peacock S., Davis W.W., Hawley S.T., Pignone M. & Ransohoff D.F. (2007) Data sources for measuring colorectal endoscopy use among Medicare enrollees. *Cancer Epidemiology, Biomarkers and Prevention* **16**, 2118–2127.
- Schenck A.P., Klabunde C.N., Warren J.L., Peacock S., Davis W.W., Hawley S.T., Pignone M. & Ransohoff D.F. (2008) Evaluation of claims, medical records, and self-report for measuring fecal occult blood testing among Medicare enrollees in fee for service. *Cancer Epidemiology, Biomarkers and Prevention* **17**, 799–804.
- Schwarz N. (1999) Self-reports: how the questions shape the answers. *American Psychologist* **54**, 93–105.
- Shariff S.Z., Bejaimal S.A., Sontrop J.M., Iansavichus A.V., Haynes R.B., Weir M.A. & Garg A.X. (2013) Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *Journal of Medical Internet Research* **15**, e164.
- Shokar N.K., Vernon S.W. & Carlson C.A. (2011) Validity of self-reported colorectal cancer test use in different racial/ethnic groups. *Family Practice* **28**, 683–688.
- Sirken M.G., Herrmann D.J., Schechter S., Schwarz N., Tanur J.M. & Tourangeau R. (1999) *Cognition and Survey Research*. John Wiley & Sons, New York, USA.

- Smith R.A., Cokkinides V. & Brawley O.W. (2012) Cancer screening in the United States, 2012. A review of current American Society guidelines and current issues in cancer screening. *CA: A Cancer Journal for Clinicians* **62**, 129–142.
- Stark J.R., Bertone-Johnson E.R., Costanza M.E. & Stoddard A.M. (2006) Factors associated with colorectal cancer risk perception: the role of polyps and family history. *Health Education Research* **21**, 740–749.
- Stock C. & Brenner H. (2010) Utilization of lower gastrointestinal endoscopy and fecal occult blood test in 11 European countries: evidence from the Survey of Health, Aging and Retirement in Europe (SHARE). *Endoscopy* **42**, 546–556.
- Strack F. & Martin L. (1987) Thinking, judging, and communicating: a process account of context effects in attitude surveys. In: *Social Information Processing and Survey Methodology* (eds Hippler H.J., Schwarz N. & Sudman S.), pp. 123–148. Springer-Verlag, New York, USA.
- Swan H., Siddiqui A.A. & Myers R.E. (2012) International colorectal cancer screening programs: population contact strategies, testing methods and screening rates. *Practical Gastroenterology* **21**, 20–29.
- Thibodeau L.A. (1981) Evaluating diagnostic tests. *Biometrics* **37**, 801–804.
- Tisnado D.M., Adams J.L., Liu H., Damberg C.L., Chen W.P., Hu F.A., Carlisle D.M., Mangione C.M. & Kahn K.L. (2006) What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Medical Care* **44**, 132–140.
- Tourangeau R. (1984) Cognitive science and survey methods. In: *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines* (eds Jabine T.B., Straf M.L., Tanur J.M. & Tourangeau R.), pp. 73–100. National Academies Press, Washington, DC, USA.
- Tourangeau R. (1987) Attitude measurement: a cognitive perspective. In: *Social Information Processing and Survey Methodology* (eds Hippler H.J., Schwarz N. & Sudman S.), pp. 149–162. Springer-Verlag, New York, USA.
- Vacek P.M. (1985) The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **41**, 959–968.
- Van Gelder M.M., Bretveld R.W. & Roeleveld N. (2010) Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology* **172**, 1292–1298.
- Vernon S.W., Meissner H., Klabunde C., Rimer B.K., Ahnen D.J., Bastani R., Mandelson M.T., Nadel M.R., Sheinfeld-Gorin S. & Zapka J. (2004) Measures for ascertaining use of colorectal cancer screening in behavioral, health services, and epidemiologic research. *Cancer Epidemiology, Biomarkers and Prevention* **13**, 898–905.
- Vernon S.W., Tiro J.A., Vojvodic R.W., Coan S., Diamond P.M., Greisinger A. & Fernandez M.E. (2008) Reliability and validity of a questionnaire to measure colorectal cancer screening behaviors: does mode of survey administration matter? *Cancer Epidemiology, Biomarkers and Prevention* **17**, 758–767.
- Zamora J., Abaira V., Muriel A., Khan K. & Coomarasamy A. (2006a) Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Medical Research Methodology* **6**, 31.
- Zamora J., Muriel A. & Abaira V. (2006b) Statistical methods, Clinical BioStatistics Unit – Hospital Ramón y Cajal.
- Zapka J.G., Bigelow C., Hurley T., Ford L.D., Egelhofer J., Cloud W.M. & Sachsse E. (1996) Mammography use among sociodemographically diverse women: the accuracy of self-report. *American Journal of Public Health* **86**, 1016–1021.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Summary estimates of area under the receiver operating characteristic curve (AUC), sensitivity (Se), specificity (Sp) and positive predictive value (PPV) per screening mode and survey mode.

Table S2. Spearman correlations weighted by sample size between the time frame, % Whites, % of surveyed having completed high school and beyond, % of surveyed having completed college and beyond, and % men, on one hand, and prevalence of screening history, area under the receiver operating characteristic curve (AUC), sensitivity and specificity, per screening mode, on the other.

Table S3. Data collection quality.

Table S4. Dropout rates.

MATLAB code for the simulations presented in the paper (including the TP, TN, FP and FN per study).