

# Variation of diagnostic performance of fecal immunochemical testing for hemoglobin by sex and age: results from a large screening cohort

Hermann Brenner<sup>1-3</sup>

Jing Qian<sup>1</sup>

Simone Werner<sup>1</sup>

<sup>1</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; <sup>2</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany; <sup>3</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

**Objective:** Fecal immunochemical tests (FITs) for hemoglobin in stool are increasingly used for colorectal cancer screening. Reported sensitivities and specificities have strongly varied between studies, but it is unclear to what extent such variation reflects differences between tests or between study population characteristics. We aimed to evaluate the key parameters of FIT performance for detecting advanced neoplasia (AN) according to sex and age.

**Methods:** Sex- and age-specific sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs) for detecting AN of a quantitative FIT (FOB Gold®) were evaluated among 3211 men and women aged 50–79 years who underwent screening colonoscopy in Germany.

**Results:** At the cutoff recommended by the manufacturer (17 µg hemoglobin/g feces), sensitivity was higher (51.2% versus 34.7%,  $p=0.004$ ) and specificity was lower (91.0% versus 94.8%,  $p<0.001$ ) among 65–79 year-old participants compared with 50–64 year-old participants. PPVs and NPVs did not differ significantly between age groups. However, higher NPVs were observed among women compared with men (94.7% versus 92.5%,  $p=0.015$ ). Specificity was also higher among women compared with men (94.7% versus 92.3%,  $p=0.007$ ), while there was only a little variation in sensitivity (40.3% versus 41.8%,  $p=0.789$ ) according to sex. In joint stratification by both factors, sensitivity ranged from 34.1% (95% CI 24.2%–45.2%) in 50–64 year-old women to 51.4% (95% CI 39.3%–63.3%) in 65–79 year-old men ( $p=0.029$ ). The observed age and sex differences were highly consistent across a wide range of alternative cutoffs from 10 to 50 µg hemoglobin/g feces.

**Conclusion:** There are major differences in diagnostic performance parameters according to sex and age, which should receive careful attention in the interpretation and comparison of results of FIT-based colorectal cancer screening studies.

**Keywords:** colorectal cancer, fecal immunochemical tests, hemoglobin, screening

## Plain language summary

Fecal immunochemical tests (FITs) for hemoglobin in stool are increasingly used for colorectal cancer (CRC) screening. FITs detect the majority of CRCs and a relevant proportion of their precursors called advanced adenomas. Reported sensitivity (ie, proportion testing positive in the presence of CRC or advanced adenoma) and specificity (ie, proportion testing negative in the absence of CRC and strongly varied between studies. It is unclear, however, to what extent these differences might be due to differences in study population characteristics, such as sex and age distribution. We assessed sensitivity and specificity of a widely used quantitative FIT according to sex and age in a large cohort of participants of screening colonoscopy ( $n=3211$ ). We found substantially higher sensitivity among older (65–79 years) compared with younger (50–64 years) participants, and higher specificity among women than among men. The observed

Correspondence: Hermann Brenner  
Division of Clinical Epidemiology  
and Aging Research, German  
Cancer Research Center (DKFZ),  
Im Neuenheimer Feld 581, 69120  
Heidelberg, Germany  
Tel +49 62 2142 1301  
Fax +49 62 2142 1302  
Email [h.brenner@dkfz-heidelberg.de](mailto:h.brenner@dkfz-heidelberg.de)

patterns were highly consistent across a broad range of cutoffs of test positivity. They may help to interpret differences in sensitivity and specificity of FITs reported from different studies with different sex and age distributions. The high sensitivity at older age not only for CRC, but also for advanced adenoma, supports the use of FIT as a primary screening method for this age group.

## Introduction

Colorectal cancer (CRC) accounts for approximately 700000 deaths each year globally.<sup>1</sup> A large proportion of these deaths could be prevented by screening. Randomized controlled trials (RCTs) have demonstrated the effectiveness of fecal occult blood testing in reducing CRC incidence and mortality.<sup>2-4</sup> The RCTs that were initiated decades ago used chemical, guaiac-based fecal occult blood tests (gFOBTs). In the meantime, fecal immunochemical tests (FITs) for hemoglobin in stool have been developed that have been shown to outperform gFOBTs in diagnostic accuracy.<sup>5-8</sup> As a result, FITs are meanwhile a broadly recommended option for CRC screening,<sup>9-11</sup> and FIT-based CRC screening has recently been introduced or is currently being introduced in a number of countries.<sup>12</sup>

Despite the consistently better diagnostic performance reported for FITs than for gFOBTs, sensitivities and specificities reported for FITs have also shown substantial variation. For example, sensitivity for detecting CRC varied between 25% and 100%, and specificity varied between 83% and 99%, in 19 studies included in a systematic review on diagnostic accuracy of FITs by Lee et al.<sup>8</sup> Apart from random variation due to small numbers of patients with CRCs and controls in some of the studies, these large differences might be explained by different brands of FITs and different cutoffs for positivity, but might also result from differences in key characteristics of the study populations, such as race, sex, or age. For example, mean age of participants included in the review and meta-analysis by Lee et al ranged from 45.2 years, an age at which CRC screening is typically not yet recommended for the average-risk population, to 62.7 years. The strong heterogeneity in test and population characteristics makes it difficult to disentangle the contributions of single factors to the heterogeneity in diagnostic accuracy.

Few studies have assessed variation of diagnostic accuracy of FITs by sex and age within studies. Although several of them reported a tentatively higher sensitivity and lower specificity among men compared with women, evidence on the role of sex and age for diagnostic accuracy is limited by the large heterogeneity in study designs, outcomes, and positivity cutoffs assessed, as well as sample size and power

limitations of some of the studies.<sup>13-17</sup> The aim of this study was to evaluate sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs) according to age and sex of a quantitative FIT over a range of relevant cutoffs for detecting advanced colorectal neoplasms in a large cohort of screening participants all of whom underwent colonoscopy to confirm absence or presence of colorectal adenomas.

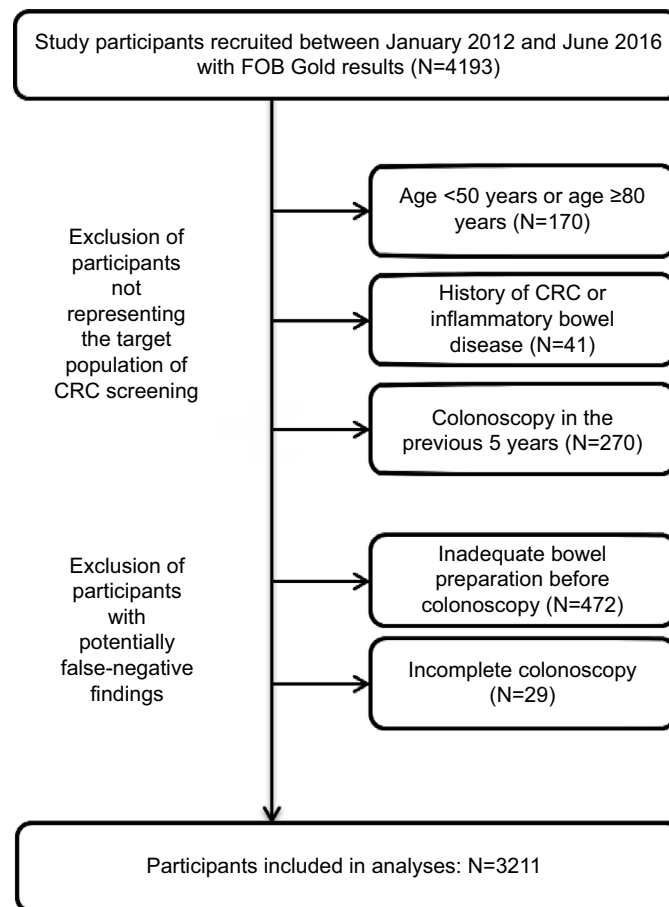
## Materials and methods

### Study design and study population

Our analysis is based on data from the BLITZ study, an ongoing study among participants of screening colonoscopy in Germany aimed to evaluate diagnostic performance of novel noninvasive or minimally invasive CRC screening tests (stool tests and blood tests). Screening colonoscopy has been offered free of charge to people aged  $\geq 55$  years (no upper age limit) in Germany since October 2002. Introduction was accompanied by major efforts toward quality assurance, and high adenoma detection rates at low levels of complication rates have been achieved on a national level.<sup>18,19</sup> Some health-care plans offer screening colonoscopy at younger ages also within specific programs.

Details of the design of the BLITZ study have been reported elsewhere.<sup>20-24</sup> Briefly, more than 9000 participants of screening colonoscopy have been recruited by a network of up to 20 gastroenterology practices since the initiation of the study in late 2005. The study was approved by the ethics committees of the University of Heidelberg and of the responsible state physicians' boards. Written informed consent was obtained from each participant.

Different ways of stool collection have been tested, and various tests have been employed during the course of the study. Previously, we reported on sensitivity and specificity of one quantitative FIT, RIDASCREEN Hemoglobin, as well as six qualitative FITs, according to sex among 2324 participants recruited between 2005 and 2008.<sup>13</sup> The current analysis reports on FIT performance by both sex and age of 4193 participants recruited between January 2012 and June 2016 when FOB Gold® (Sentinel Diagnostics, Milano, Italy) was employed; a quantitative FIT that is widely used in existing screening programs, such as the National Screening Program of the Netherlands. We employed the following exclusion criteria to ensure the study participants represented an average-risk screening population and to minimize the potential of false-negative findings of screening colonoscopy (Figure 1): 1) Age  $< 50$  years or  $\geq 80$  years (N=170); 2) history of CRC or inflammatory bowel disease



**Figure 1** STARD diagram of study participants.

**Note:** FOB Gold®; Sentinel Diagnostics, Milano, Italy.

**Abbreviations:** CRC, colorectal cancer; STARD, STAndards for the Reporting of Diagnostic accuracy studies.

(N=41); 3) colonoscopy in the previous 5 years (N=270); 4) inadequate bowel preparation before colonoscopy (N=472); 5) incomplete colonoscopy (cecum not reached, N=29). The latter two criteria were not applied for CRC patients as a stenosis caused by the tumor mass might impair bowel cleanse and completion of colonoscopy. Finally, 3211 remaining participants were included in the analysis.

## Data and sample collection

Patients were recruited, and informed consent was obtained at a precolonoscopy visit in the practice. Participants were asked to fill a self-administered questionnaire on factors potentially related to CRC risk, and they were handed out devices for stool collection. There were no specific dietary or medication restrictions before fecal sampling. Participants were asked to collect a stool sample according to routine clinical practice, that is, using a single stool collection tube containing hemoglobin stabilizing buffer (10 mg stool in 1.7 mL extraction buffer; Sentinel Diagnostics, Milano, Italy; Ref. 11561H). The tube was to be sealed in an envelope, which was then mailed to the study center at the German Cancer

Research Center (DKFZ), where it was kept at 2°C–8°C in the refrigerator before transporting in a cold chain to the central, DIN EN ISO 15189 accredited laboratory (Labor Limbach, Heidelberg, Germany) for FIT analysis.

All collection, arrival, and analysis dates of fecal samples were documented. The median time (interquartile range [IQR]) between fecal sampling and arrival in DKFZ was 4 (IQR=3–5) days, and the median time between arrival at DKFZ and laboratory analysis was 3 (IQR=1–5) days.

Clinical data were extracted from colonoscopy and histology reports that were obtained from the gastroenterologists who were blinded with respect to any blood or stool test results. The data extraction was done in a standardized manner by trained research assistants who were blinded with respect to questionnaire data and results of stool tests. Participants were classified according to the most advanced finding at screening colonoscopy using the following categories: CRC, advanced adenoma, nonadvanced adenoma, serrated polyp/adenoma, non-defined polyp, hyperplastic polyp, none of above. Advanced adenoma was defined by the presence of at least one non-serrated adenoma with any of the following

features:  $\geq 1$  cm in size, tubulovillous or villous components, and high-grade dysplasia. Number and location of findings were documented, and information on completeness of colonoscopy and quality of bowel cleansing was extracted.

## Laboratory analyses

FIT reporting and evaluation followed FITTER standards.<sup>25</sup> FOB Gold, which is based on a latex agglutination assay, was used for measuring fecal hemoglobin concentrations. Laboratory personnel were fully blinded with respect to questionnaire data and colonoscopy findings. All FIT analyses were conducted in a fully automated manner using Abbott Architect c8000. The dates of conducting FIT analyses were recorded.

## Statistical analyses

We first described the study sample according to sex, age (50–64, 65–79 years), and most advanced finding at colonoscopy (CRC, advanced adenoma, nonadvanced adenoma, serrated polyp/adenoma, and no neoplasm). All further analyses focused on detecting advanced neoplasia (AN), defined as either CRC or advanced adenoma. We chose AN as the main clinical endpoint as the main impact of FIT-based screening is not only detection of CRC in an early, preclinical stage, but also prevention of CRC by detecting and subsequently removing advanced adenoma, the precursors of the vast majority of CRCs, whereas nonadvanced adenoma or serrated adenomas/polyps are not effectively detected by FITs.<sup>7,26,27</sup>

We then determined prevalences of AN, sensitivities, specificities, PPVs and NPVs and their 95% confidence

intervals according to sex and age (50–64, 65–79 years). Differences in these indicators between groups were tested for statistical significance by chi-square tests (two-sided testing at an alpha level of 0.05).

The analyses were conducted using the cutoff for FIT positivity recommended by the manufacturer (17  $\mu\text{g}$  hemoglobin (Hb)/g feces = 100 ng Hb/mL buffer), as well as a broad range of alternative cutoffs between 10 and 50  $\mu\text{g}$  Hb/g feces. Finally, overall diagnostic performance of the FIT according to sex and age across cutoffs yielding levels of specificity between 80% and 100% (which are typically required in population-based screening) was evaluated in receiver operating characteristics (ROC) analysis by partial areas under the curves (pAUCs). The pAUCs were corrected so that a value of 50% indicates a nondiscriminant area under the curve, and a value of 100% indicates the maximum possible value within the included specificity range.

All analyses were conducted with R version 3.2.3.

## Results

Main characteristics of the study population are shown in Table 1. The study population included almost equal proportions of women (N=1652, 51.4%) and men (N=1559, 48.6%). Median age was 60 years, with 2124 participants (66.1%) and 1087 participants (33.9%) in age groups 50–64 and 65–79, respectively. The age distribution was very similar among women and men. Overall, the sample included 311 participants (9.7%) with advanced neoplasia, with CRC and advanced adenoma being the most advanced finding in 25

**Table 1** Distribution of sex and age and findings at colonoscopy

Characteristics of study population	Sex				Age (years)				Total (N=3211)	
	Women (N=1652)		Men (N=1559)		50–64 (N=2124)		65–79 (N=1087)		N	%
	N	%	N	%	N	%	N	%		
Sex										
Women					1106	52.1	546	50.2	1652	51.4
Men					1018	47.9	541	49.8	1559	48.6
Age (years)										
50–64	1106	66.9	1018	65.3					2124	66.1
65–79	546	33.1	541	34.7					1087	33.9
Most advanced finding at screening colonoscopy										
Colorectal cancer	11	0.7	14	0.9	8	0.4	17	1.6	25	0.8
Advanced adenoma	123	7.4	163	10.5	182	8.6	104	9.6	286	8.9
Non-advanced adenoma	200	12.1	333	21.4	329	15.5	204	18.8	533	16.6
Serrated polyp/adenoma	58	3.5	44	2.8	75	3.5	27	2.5	102	3.2
Non-defined polyp	40	2.4	39	2.5	46	2.2	33	3.0	79	2.5
Hyperplastic polyp	127	7.7	138	8.9	193	9.1	72	6.6	265	8.3
No finding	1093	66.2	828	53.1	1291	60.8	630	58.0	1921	59.8
Any advanced neoplasm	134	8.1	177	11.4	190	8.9	121	11.1	311	9.7
No advanced neoplasm	1518	91.9	1382	88.6	1934	91.1	966	88.9	2900	90.3

(0.8%) and 286 (8.9%) of participants. Prevalences of both types of advanced neoplasia were substantially and statistically significantly higher among men than among women (overall prevalence of AN: 11.4% versus 8.1%,  $p=0.002$ ) and among 65–79 year-old participants than among 50–64 year-old participants (11.1% versus 8.9%,  $p=0.047$ ).

Table 2 shows overall sensitivities and specificities for various findings at colonoscopy. At the cutoff recommended by the manufacturer (17  $\mu\text{g}$  Hb/g feces), the overall sensitivity for AN was 41.2% (95% CI 35.6%–46.9%) and reflects a weighted average of very high sensitivity (92.0%) for CRC and a slightly lower sensitivity (36.7%) for advanced adenoma, the most advanced finding in the vast majority of people with AN. The overall specificity for people without AN was 93.6% (95% CI 92.6%–94.4%) and was only slightly lower than the specificity that would be obtained after excluding those with nonadvanced adenoma and serrated adenoma/polyp (94.0%) as the positivity rate in the latter groups (8.3% and 5.9%, respectively) was very similar to the false-positive rate in participants without any neoplasm (6.0%).

Sensitivities and specificities for detecting AN are shown according to sex, age, and their combination in Table 3. At any of the assessed cutoffs, sensitivity was nonsignificantly higher, with differences ranging from 0.9% to 8.1% units, whereas specificity was significantly lower, with differences ranging from 1.5% to 7.1% units among men (sensitivities: 27.7%–55.9%; specificities: 84.3%–96.9%) than among women (sensitivities: 20.9%–47.8%; specificities: 91.4%–98.4%). Even larger differences were found when stratifying for age: sensitivity was much higher, with differences ranging from 10.9% to 19.3% units, whereas specificity was lower, with differences ranging from 0.8% to 6.0% units, in the older age group (sensitivities: 31.4%–60.3%;

specificities: 84.0%–97.1%) than in the younger age group (sensitivities: 20.5%–47.4%; specificities: 90.0%–97.9%). Simultaneous stratification by both sex and age yielded very strong gradients in sensitivity, with differences in sensitivity between younger women and older men consistently around 20% units at all cutoffs, along with differences in specificity between 2% and 13% units.

Overall, PPVs and NPVs for detecting one AN ranged from 31.9% to 53.6% and from 92.4% to 94.5%, respectively, for the various cutoffs. PPVs and NPVs for detecting one AN are shown according to sex, age, and their combination in Table 4. Overall, differences in PPV and NPV by sex and age were much less pronounced and consistent than differences in sensitivity and specificity. Nevertheless, NPV was consistently ~2% units higher for women than for men ( $p<0.05$  at 5 out of 6 cutoffs). Although PPV was consistently higher in the older than in the younger age group, the differences, which ranged from 0.2% to 8.2% units across cutoffs, were not statistically significant.

Despite the major differences in sensitivity and specificity, overall diagnostic performance across cutoffs, as evaluated by ROC analyses, was rather similar for both sexes and age groups. In particular, no difference was seen in pAUC for cutoffs yielding 80%–100% specificity between men and women (0.690 versus 0.697,  $p=0.776$ ). The pAUC was somewhat higher for the older than for the younger age group (0.718 versus 0.687), but the difference was not statistically significant ( $p=0.260$ ).

## Discussion

In this large study, among 3211 participants of screening colonoscopy, we found strong variation in key parameters of diagnostic performance of a quantitative FIT for detecting AN. Regardless of the cutoff of FIT positivity, sensitivity

**Table 2** Sensitivity and specificity of FOB Gold for detecting advanced neoplasia according to cutoff for test positivity

Cutoff ( $\mu\text{g/g}$ )	Positivity rate (95% CI) (%)	Sensitivity (%)						Specificity (%)	
		CRC (N=25)		Advanced adenoma (N=286)		Any advanced neoplasm (N=311)		No advanced neoplasm (N=2900)	
		N <sub>pos</sub>	Sensitivity (95% CI)	N <sub>pos</sub>	Sensitivity (95% CI)	N <sub>pos</sub>	Sensitivity (95% CI)	N <sub>neg</sub>	Specificity (95% CI)
10	15.9 (14.7–17.2)	24	96.0 (79.6–99.9)	139	48.6 (42.7–54.6)	163	52.4 (46.7–58.1)	2552	88.0 (86.8–89.2)
17 <sup>a</sup>	9.8 (8.8–10.9)	23	92.0 (74.0–99.0)	105	36.7 (31.1–42.6)	128	41.2 (35.6–46.9)	2713	93.6 (92.6–94.4)
20	8.4 (7.5–9.5)	23	92.0 (74.0–99.0)	97	33.9 (28.4–39.7)	120	38.6 (33.1–44.2)	2749	94.8 (93.3–95.6)
30	6.3 (5.4–7.2)	22	88.0 (68.8–97.5)	81	28.3 (23.2–33.9)	103	33.1 (27.9–38.7)	2802	96.6 (95.9–97.2)
40	5.2 (4.4–6.0)	20	80.0 (59.3–93.2)	69	24.1 (19.3–29.5)	89	28.6 (23.7–34.0)	2823	97.3 (96.7–97.9)
50	4.5 (3.8–5.3)	18	72.0 (50.6–87.9)	59	20.6 (16.1–25.8)	77	24.8 (20.1–29.9)	2832	97.7 (97.0–98.2)

**Notes:** <sup>a</sup>Cutoff recommended by the manufacturer (see “Materials and methods” section). FOB Gold®; Sentinel Diagnostics, Milano, Italy.

**Abbreviations:** CRC, colorectal cancer; N<sub>pos</sub>, number of participants with positive result; N<sub>neg</sub>, number of participants with negative result.



**Table 3** Sensitivity and specificity of FOB Gold for detecting advanced neoplasia according to sex, age, and cutoff for test positivity

Cutoff ( $\mu\text{g/g}$ )	Sensitivity according to sex and age					Specificity according to sex and age				
	Women		Men		Difference (95% CI) (%)	Women		Men		Difference (95% CI) (%)
	$N_{TP}/N_{AN}$	%	$N_{TP}/N_{AN}$	%		$N_{TN}/N_{noAN}$	%	$N_{TN}/N_{noAN}$	%	
10	64/134	47.8	99/177	55.9	-8.1 (-19.5, 3.3)	1387/1518	91.4	1165/1382	84.3	7.1 (4.7, 9.5)
17 <sup>a</sup>	54/134	40.3	74/177	41.8	-1.5 (-12.8, 9.8)	1438/1518	94.7	1275/1382	92.3	2.4 (0.6, 4.2)
20	51/134	38.1	69/177	39.0	-0.9 (-12.0, 10.2)	1452/1518	95.7	1297/1382	93.8	1.9 (0.2, 3.6)
30	41/134	30.6	62/177	35.0	-4.4 (-15.1, 6.3)	1481/1518	97.6	1321/1382	95.6	2.0 (0.7, 3.4)
40	33/134	24.6	56/177	31.6	-7.0 (-17.2, 3.2)	1490/1518	98.2	1333/1382	96.5	1.7 (0.5, 2.9)
50	28/134	20.9	49/177	27.7	-6.8 (-16.5, 2.9)	1493/1518	98.4	1339/1382	96.9	1.5 (0.4, 2.6)
	50–64 years		65–79 years		Difference (95% CI) (%)	50–64 years		65–79 years		Difference (95% CI) (%)
	$N_{TP}/N_{AN}$	%	$N_{TP}/N_{AN}$	%		$N_{TN}/N_{noAN}$	%	$N_{TN}/N_{noAN}$	%	
10	90/190	47.4	73/121	60.3	-12.9 (-24.4, -1.4)	1741/1934	90.0	811/966	84.0	6.0 (3.3, 8.7)
17 <sup>a</sup>	66/190	34.7	62/121	51.2	-16.5 (-28.4, -4.6)	1834/1934	94.8	879/966	91.0	3.8 (1.7, 5.9)
20	59/190	31.1	61/121	50.4	-19.3 (-31.1, -7.5)	1848/1934	95.6	901/966	93.3	2.3 (0.4, 4.2)
30	49/190	25.8	54/121	44.6	-18.8 (-30.2, -7.4)	1883/1934	97.4	919/966	95.1	2.3 (0.7, 3.9)
40	44/190	23.2	45/121	37.2	-14.0 (-25.1, -2.9)	1891/1934	97.8	932/966	96.5	1.3 (-0.1, 2.7)
50	39/190	20.5	38/121	31.4	-10.9 (-21.5, -0.3)	1894/1934	97.9	938/966	97.1	0.8 (-0.5, 2.1)
	Women 50–64 years		Men 65–79 years		Difference (95% CI) (%)	Women 50–64 years		Men 65–79 years		Difference (95% CI) (%)
	$N_{TP}/N_{AN}$	%	$N_{TP}/N_{AN}$	%		$N_{TN}/N_{noAN}$	%	$N_{TN}/N_{noAN}$	%	
10	37/85	43.5	46/72	63.9	-20.4 (-36.0, -4.8)	953/1021	93.3	377/469	80.4	12.9 (8.9, 16.9)
17 <sup>a</sup>	29/85	34.1	37/72	51.4	-17.3 (-32.9, -1.7)	979/1021	95.9	420/469	89.6	6.3 (3.2, 9.4)
20	26/85	30.6	36/72	50.0	-19.4 (-34.9, -3.9)	985/1021	96.5	434/469	92.5	4.0 (1.3, 6.7)
30	20/85	23.5	33/72	45.8	-22.3 (-37.2, -7.4)	1003/1021	98.2	441/469	94.0	4.2 (1.9, 6.5)
40	17/85	20.0	29/72	40.3	-20.3 (-34.8, -5.8)	1006/1021	98.5	448/469	95.5	3.0 (0.9, 5.1)
50	15/85	17.6	25/72	34.7	-17.1 (-31.0, -3.2)	1006/1021	98.5	451/469	96.2	2.3 (0.4, 4.2)

**Notes:** <sup>a</sup>Cutoff recommended by the manufacturer (see "Materials and methods" section). FOB Gold®; Sentinel Diagnostics, Milano, Italy.

**Abbreviations:**  $N_{TP}/N_{AN}$ , number of participants with true positive result among participants with advanced neoplasia;  $N_{TN}/N_{noAN}$ , number of participants with true negative result among participants with no advanced neoplasia.

was consistently much higher and specificity was consistently somewhat lower in older participants than in younger participants. In women specificity and NPV were higher than in men. Sensitivity was higher in men even though this difference did not reach statistical significance. Across a broad range of cutoffs from 10 to 50  $\mu\text{g Hb/g}$  feces, sensitivity for detecting AN was ~20% units higher and specificity was between 2% and 13% units lower among 65–79 year-old men compared with 50–64 year-old women. Overall diagnostic performance was similar for both sexes and age groups.

Few previous studies have assessed sex and age differences in diagnostic performance of FITs. Our results regarding higher specificities and NPVs in women are consistent with findings in an earlier analysis among 2324 participants recruited in the initial years of the BLITZ study, when a different quantitative FIT (RIDASCREEN Hemoglobin; R-Biopharm, Darmstadt, Germany; analyzed on Tecan Freedom Evolyzer) had been used. In that analysis, which had not addressed age differences, we had additionally found signifi-

cantly higher sensitivities and PPVs for detecting advanced colorectal neoplasms in men compared with women. Like in the current study, sex differences were consistently seen at any FIT cutoff, and the same sex differences were also seen for all of six qualitative FITs that were assessed in frozen stool samples of the same participants.<sup>13</sup> Similar albeit statistically not significant sex differences in sensitivity and specificity for advanced colorectal neoplasms were later reported in 1112 screening participants from the Netherlands undergoing colonoscopy in addition to a quantitative FIT (OC Sensor, Eiken Chemical, Tokyo, Japan).<sup>14</sup> A more recent study from the Netherlands confirmed a significantly increased rate of false-positive results for OC Sensor Micro (Eiken Chemical, Tokyo, Japan) among men.<sup>15</sup> Studies from Italy<sup>16</sup> and Korea<sup>17</sup> also derived higher sensitivities and lower specificities for detecting CRC among males compared with females from registry-based follow-up of participants in FIT-based screening programs that used different qualitative and quantitative FITs.

**Table 4** Positive and negative predictive value of FOB Gold for detecting advanced neoplasia according to sex, age, and cutoff for test positivity

Cutoff (µg/g)	Positive predictive value according to sex and age						Negative predictive value according to sex and age					
	Women		Men		Difference (95% CI) (%)	Women		Men		Difference (95% CI) (%)		
	N <sub>TP</sub> /N <sub>P</sub>	%	N <sub>TP</sub> /N <sub>P</sub>	%		N <sub>TN</sub> /N <sub>N</sub>	%	N <sub>TN</sub> /N <sub>N</sub>	%			
10	64/195	32.8	99/316	31.3	1.5 (-7.0, 10.0)	1387/1457	95.2	1165/1243	93.7	1.5 (-0.3, 3.3)		
17 <sup>a</sup>	54/134	40.3	74/181	40.9	0.6 (-10.6, 11.8)	1438/1518	94.7	1275/1378	92.5	2.2 (0.4, 4.0)		
20	51/117	43.6	69/154	44.8	-1.2 (-13.4, 11.0)	1452/1535	94.6	1297/1405	92.3	2.3 (0.5, 4.1)		
30	41/78	52.6	62/123	50.4	2.2 (-12.3, 16.7)	1481/1574	94.1	1321/1436	92.0	2.1 (0.2, 4.0)		
40	33/61	54.1	56/105	53.3	0.8 (-15.3, 16.9)	1490/1591	93.7	1333/1454	91.7	2.0 (0.1, 3.9)		
50	28/53	52.8	49/92	53.3	-0.5 (-17.7, 16.7)	1493/1599	93.4	1339/1467	91.3	2.1 (0.2, 4.0)		
	50–64 years		65–79 years		Difference (95% CI) (%)	50–64 years		65–79 years		Difference (95% CI) (%)		
	N <sub>TP</sub> /N <sub>P</sub>	%	N <sub>TP</sub> /N <sub>P</sub>	%		N <sub>TN</sub> /N <sub>N</sub>	%	N <sub>TN</sub> /N <sub>N</sub>	%			
10	90/283	31.8	73/228	32.0	-0.2 (-8.5, 8.1)	1741/1841	94.6	811/859	94.4	0.2 (-1.7, 2.1)		
17 <sup>a</sup>	66/166	39.8	62/149	41.6	-1.8 (-12.9, 9.3)	1834/1958	93.7	879/938	93.7	0.0 (-1.9, 1.9)		
20	59/145	40.7	61/126	48.4	-7.7 (-19.8, 4.4)	1848/1979	93.4	901/961	93.8	-0.4 (-2.3, 1.5)		
30	49/100	49.0	54/101	53.5	-4.5 (-18.6, 9.6)	1883/2024	93.0	919/986	93.2	-0.2 (-2.2, 1.8)		
40	44/87	50.6	45/79	57.0	-6.4 (-21.9, 9.1)	1891/2037	92.8	932/1008	92.5	0.3 (-1.7, 2.3)		
50	39/79	49.4	38/66	57.6	-8.2 (-24.8, 8.4)	1894/2045	92.6	938/1021	91.9	0.7 (-1.4, 2.8)		
	Women		Men		Difference (95% CI) (%)	Women		Men		Difference (95% CI) (%)		
	50–64 years		65–79 years			50–64 years		65–79 years				
	N <sub>TP</sub> /N <sub>P</sub>	%	N <sub>TP</sub> /N <sub>P</sub>	%	N <sub>TN</sub> /N <sub>N</sub>	%	N <sub>TN</sub> /N <sub>N</sub>	%				
10	37/105	35.2	46/138	33.3	1.9 (-10.4, 14.2)	953/1001	95.2	377/403	93.5	1.7 (-1.1, 4.5)		
17 <sup>a</sup>	29/71	40.8	37/86	43.0	-2.2 (-18.0, 13.6)	979/1035	94.6	855/923	92.3	2.3 (-0.2, 4.2)		
20	26/62	41.9	36/71	50.7	-8.8 (-26.1, 8.5)	985/1044	94.3	434/470	92.3	2.0 (-0.8, 4.8)		
30	20/38	52.6	33/61	54.1	-1.5 (-22.1, 19.1)	1003/1068	93.9	441/480	91.9	2.0 (-0.9, 4.9)		
40	17/32	53.1	29/50	58.0	-4.9 (-27.4, 17.6)	1006/1074	93.7	448/491	91.2	2.5 (-0.5, 5.5)		
50	15/30	50.0	25/43	58.1	-8.1 (-31.8, 15.6)	1006/1076	93.5	451/498	90.6	2.9 (-0.1, 5.9)		

**Note:** <sup>a</sup>Cutoff recommended by the manufacturer (see "Materials and methods" section). FOB Gold®; Sentinel Diagnostics, Milano, Italy.

**Abbreviations:** N<sub>TP</sub>/N<sub>P</sub>, number of participants with true-positive result (ie, with advanced neoplasia) among participants with positive result; N<sub>TN</sub>/N<sub>N</sub>, number of participants with true-negative result (ie, without advanced neoplasia) among participants with negative result.

The aforementioned study from the Netherlands also reported tentatively, but not significantly, lower specificity for detecting advanced neoplasms at older ages,<sup>14</sup> and in the study from Korea, both sensitivity and specificity for detecting CRC were estimated to be lower at older ages,<sup>17</sup> whereas no age differences were observed in the Italian study.<sup>16</sup> To our knowledge, no previous study has simultaneously assessed diagnostic performance after joint stratification by sex and age.

There might be several plausible explanations for the lower sensitivity and higher specificity of FITs among women compared with men, such as the higher proportion of AN located in the proximal colon which are more difficult to detect by FIT,<sup>28</sup> lower rates of aspirin use for cardioprotection,<sup>21</sup> and the longer colonic transit time that may favor Hb degradation prior to defecation.<sup>29</sup> Higher rates of aspirin use and potentially larger sizes of adenomas might also explain the higher sensitivity among older compared with younger screening participants. Sex- and age-specific variations in PPVs and NPVs are furthermore codetermined

by major differences in the prevalence of AN according to sex and age.

The large differences in parameters of diagnostic performance according to sex and age require careful consideration when diagnostic performance of various FITs reported in different studies or different screening programs is compared. For example, in 19 studies assessing sensitivity and specificity of different FITs for detecting CRC included in the systematic review and meta-analysis by Lee et al,<sup>8</sup> mean age of study participants ranged from 45.2 to 62.7 years. Although the outcome in this review was restricted to CRC, differences in sensitivity and specificity between studies might also partly be due to differences in the age distribution besides other factors, such as differences in study populations, FIT brands, and cutoffs used.

In Germany, a decision has been made in 2016 to include FITs among the CRC screening offers covered by the statutory health insurance. Only FITs for which minimum levels of sensitivity (25%) and specificity (90%) for detecting AN

have been demonstrated in validation studies conducted in screening settings shall be covered. Our results underline the importance of such validation studies to be representative of the age and sex distribution for the target population of screening in order for the results to be meaningful and comparable. In case of nonrepresentative age and sex distributions, the possibility of achieving comparability by established epidemiological techniques, such as age-standardized analysis, should be considered at the very least.

Another question of practical relevance is whether the differences in diagnostic performance according to sex and age should prompt use of sex- and age-specific cutoffs for FIT positivity. For example, a lower cutoff might be considered for women than for men in order to ensure comparable levels of sensitivity and specificity. However, as can be seen from our results presented in Table 4 and pointed out by Grobbee et al,<sup>30</sup> such an approach would rather introduce or increase sex differences in positive and negative predictive values. Our results, therefore, support conclusions that use of sex-specific cutoffs might not be warranted.<sup>15</sup> Nevertheless, potential optimization of effectiveness and cost-effectiveness of CRC screening by taking sex differences in diagnostic performance parameters as well as in incidence and prevalence of colorectal neoplasms into account should be carefully explored by microsimulation models<sup>31</sup> for which our study may provide valuable input parameters. The same would apply to potential variation of screening modalities according to age.

Our study has specific strengths and limitations. Apart from its large sample size, a major strength is conducting of the study in the target population of screening, with screening colonoscopy results being available for evaluating presence or absence of AN for all participants, not only for FIT-positive individuals. Despite the overall large sample size, some of the subgroup-specific estimates of sensitivity and PPV were based on rather limited numbers of participants which may, for example, explain lack of statistical significance of the rather consistent but modest sex differences in sensitivity. Replication in further, even larger studies would be desirable. Sample size limitations also hindered further evaluation of diagnostic performance for specific subgroups of AN, such as CRC or adenomas of various size or location. Given that the most advanced finding among participants with advanced neoplasia was advanced adenoma in the vast majority (~92%) of cases, our results essentially reflect diagnostic performance parameters in this group. Although sensitivities are substantially higher among participants with CRC (ranging between 72%

and 96% overall for the various cutoffs in our sample), the small overall (n=25) and stratum-specific numbers of CRC cases did not allow for meaningful comparison of sex- and age-specific performance with respect to this outcome.

Although colonoscopy is commonly considered as the gold standard for evaluating diagnostic performance of noninvasive CRC screening tests, it is not perfect and may miss some, albeit a small proportion, of AN.<sup>32,33</sup> In order to minimize the potential impact of missed AN, we excluded participants with inadequate bowel preparation or incomplete colonoscopy. In the light of the overall high adenoma detection rate achieved in our study (>30% among men, ~20% among women), relevant bias due to missed AN seems unlikely.

Despite its limitations, our study provides valuable novel insights into the variation of diagnostic performance parameters of FIT-based screening for AN. Such a variation should receive careful consideration in the design and interpretation of studies evaluating diagnostic performance of FITs, as well as in modeling studies aiming to evaluate effectiveness and cost-effectiveness of FIT-based screening strategies in various populations or population subgroups. The high sensitivity at reasonable levels of specificity among the older age group is reassuring and supports the use of FIT as a primary screening method for this age group.

## Acknowledgments

We gratefully acknowledge the excellent cooperation of gastroenterology practices and clinics in patient recruitment and of Labor Limbach in sample collection. We also gratefully acknowledge Isabel Lerch, Susanne Köhler, Utz Benschaid, and Jason Hochhaus, for their contribution to data collection, monitoring, and documentation. This work was partly funded by grants from the German Research Council (DFG, grant number BR1704/16-1) and the German Federal Ministry of Education and Research (grant number 01GL1712). The sponsors had no role in the study design or in the collection, analysis, and interpretation of data.

## Author contributions

HB designed and supervised the study, conceived the analyses, and wrote and finalized the manuscript. JQ and SW contributed to conduction of the BLITZ study, conducted the statistical analyses, and reviewed the manuscript. All authors contributed toward data analysis, drafting and critically revising the paper, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.



## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–E386.
2. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemocult): an update. *Am J Gastroenterol*. 2008;103(6):1541–1549.
3. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut*. 2012;61(7):1036–1040.
4. Shaikat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med*. 2013;369(12):1106–1114.
5. Park DI, Ryu S, Kim YH, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. *Am J Gastroenterol*. 2010;105(9):2017–2025.
6. Faivre J, Dancourt V, Denis B, et al. Comparison between a guaiac and three immunochemical faecal occult blood tests in screening for colorectal cancer. *Eur J Cancer*. 2012;48(16):2969–2976.
7. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *Eur J Cancer*. 2013;49(14):3049–3054.
8. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med*. 2014;160(3):171.
9. Pox C, Aretz S, Bischoff SC, et al; Leitlinienprogramm Onkologie der AWMF; Deutschen Krebsgesellschaft e.V.; Deutschen Krebshilfe e.V. [S3-guideline colorectal cancer version 1.0]. *Z Gastroenterol*. 2013;51(8):753–854. German.
10. Armaroli P, Villain P, Suonio E, et al. European code against cancer, 4th Edition: cancer screening. *Cancer Epidemiol*. 2015;39(Suppl 1):S139–S152.
11. Robertson DJ, Lee JK, Boland CR, et al. Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US Multi-Society Task Force on colorectal cancer. *Am J Gastroenterol*. 2017;112(1):37–53.
12. Allison JE, Fraser CG, Halloran SP, Young GP. Population screening for colorectal cancer means getting FIT: the past, present, and future of colorectal cancer screening using the fecal immunochemical test for hemoglobin (FIT). *Gut Liver*. 2014;8(2):117–130.
13. Brenner H, Haug U, Hundt S. Sex differences in performance of fecal occult blood testing. *Am J Gastroenterol*. 2010;105(11):2457–2464.
14. Stegeman I, de Wijkerslooth TR, Stoop EM, et al. Risk factors for false positive and for false negative test results in screening with fecal occult blood testing. *Int J Cancer*. 2013;133(10):2408–2414.
15. Kapidzic A, van der Meulen MP, Hol L, et al. Gender differences in fecal immunochemical test performance for early detection of colorectal neoplasia. *Clin Gastroenterol Hepatol*. 2015;13(8):1464–1471.
16. Zorzi M, Fedato C, Grazzini G, et al. High sensitivity of five colorectal screening programmes with faecal immunochemical test in the Veneto Region, Italy. *Gut*. 2011;60(7):944–949.
17. Shin A, Choi KS, Jun JK, et al. Validity of fecal occult blood test in the national cancer screening program, Korea. *PLoS One*. 2013;8(11):e79292.
18. Pox CP, Altenhofen L, Brenner H, Theilmeier A, von Stillfried D, Schmiegel W. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. *Gastroenterology*. 2012;142(7):1460–1467.
19. Brenner H, Altenhofen L, Kretschmann J, et al. Trends in adenoma detection rates during the first 10 years of the German screening colonoscopy program. *Gastroenterology*. 2015;149(2):356–366.
20. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med*. 2009;150(3):162–169.
21. Brenner H, Tao S, Haug U. Low-dose aspirin use and performance of immunochemical fecal occult blood tests. *JAMA*. 2010;304(22):2513–2520.
22. Haug U, Hundt S, Brenner H. Quantitative immunochemical fecal occult blood testing for colorectal adenoma detection: evaluation in the target population of screening and comparison with qualitative tests. *Am J Gastroenterol*. 2010;105(3):682–690.
23. Chen H, Werner S, Brenner H. Fresh vs frozen samples and ambient temperature have little effect on detection of colorectal cancer or adenomas by a fecal immunochemical test in a colorectal cancer screening cohort in Germany. *Clin Gastroenterol Hepatol*. 2017;15(10):1547–1556.
24. Chen H, Qian J, Werner S, Cuk K, Knebel P, Brenner H. Development and validation of a panel of five proteins as blood biomarkers for early detection of colorectal cancer. *Clin Epidemiol*. 2017;9(10):517–526.
25. Fraser CG, Allison JE, Young GP, Halloran SP, Seaman H. A standard for faecal immunochemical tests for haemoglobin evaluation reporting (FITTER). *Ann Clin Biochem*. 2014;51(Pt 2):301–302.
26. Heigh RI, Yab TC, Taylor WR, et al. Detection of colorectal serrated polyps by stool DNA testing: comparison with fecal immunochemical testing for occult blood (FIT). *PLoS ONE*. 2014;9(1):e85659.
27. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370(14):1287–1297.
28. Haug U, Kuntz KM, Knudsen AB, Hundt S, Brenner H. Sensitivity of immunochemical faecal occult blood testing for detecting left- vs right-sided colorectal neoplasia. *Br J Cancer*. 2011;104(11):1779–1785.
29. Sadik R, Abrahamsson H, Stotzer PO. Gender differences in gut transit shown with a newly developed radiological procedure. *Scand J Gastroenterol*. 2003;38(1):36–42.
30. Grobbee EJ, Wieten E, Hansen BE, et al. Fecal immunochemical test-based colorectal cancer screening: The gender dilemma. *United Eur Gastroenterol J*. 2017;5(3):448–454.
31. van Hees F, Zauber AG, van Veldhuizen H, et al. The value of models in informing resource allocation in colorectal cancer screening: the case of The Netherlands. *Gut*. 2015;64(12):1985–1997.
32. Heresbach D, Barrioz T, Lapalus MG, et al. Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy*. 2008;40(4):284–290.
33. Ahn SB, Han DS, Bae JH, Byun TJ, Kim JP, Eun CS. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. *Gut Liver*. 2012;6(1):64–70.

### Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress