# Selecting a Cut-off for Colorectal Cancer Screening With a Fecal Immunochemical Test

Hermann Brenner, MD, MPH[1,2,3] and Simone Werner, PhD[1]

**OBJECTIVES:** Fecal immunochemical tests (FITs) for hemoglobin (Hb) are increasingly used for colorectal cancer (CRC) screening. However, cut-offs for defining test positivity are varying widely. We aimed to evaluate the impact of cut-off selection on key indicators of diagnostic performance in a true screening setting.

**METHODS:** We evaluated diagnostic performance of FOB Gold, a widely used quantitative FIT, for detecting advanced neoplasms (AN) across a wide range of possible cut-offs among 1822 participants of screening colonoscopy aged 50–79 years in Germany.

**RESULTS:** The positive predictive value (PPV) for detecting AN showed a very steep increase with increasing cut-off up to 35.2% (95% CI 29.9–40.9%) at a cut-off of 9 µg Hb/g feces at which sensitivity and specificity were 48.8% (95% CI 42.1–55.6%) and 88.5% (95% CI 86.8–89.9%), respectively. A further moderate increase of PPV up to 56.9% (95% CI 47.8–65.5%), along with a major decrease in sensitivity was observed when gradually increasing the cut-off to 25 µg Hb/g feces at which sensitivity and specificity were 31.9% (95% CI 25.9–38.5%) and 96.9% (95% CI 95.9–97.6%), respectively. Further increases of the cut-off hardly affected PPV and specificity, but went along with further relevant decline in sensitivity.

**CONCLUSIONS:** Our study illustrates delineation of a range of meaningful cut-offs (here: 9–25 µg Hb/g feces) according to expected diagnostic yield in a true screening setting. Selecting a cut-off within or beyond this range should consider characteristics of the specific target population, such as AN prevalence or available colonoscopy capacity.

## INTRODUCTION

Colorectal cancer (CRC) accounts for ~700,000 deaths each year globally.[1] 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.[2–4] The RCTs which were initiated decades ago had used chemical, guaiac based fecal occult blood tests (gFOBTs). In the meantime, fecal immunochemical tests (FITs) for hemoglobin (Hb) have been developed that have been shown to outperform gFOBTs in diagnostic performance[5–8] and adherence.[9] Another advantage of FITs over gFOBTs is that they are specific to human hemoglobin, and hence their application does not require dietary restrictions.[10]

Given these advantages, FITs are meanwhile a broadly recommended option for CRC screening, e.g.,[11,12] and FIT-based CRC screening has recently been introduced or is currently being introduced or expanded in a number of countries.[13] However, a number of features of FIT testing are subject to ongoing debate, such as the number of feces samples to be tested,[5,14] the best cut-off for positivity[5,15] or the time intervals for FIT-based screening.[16,17] For example, in the Netherlands, a cut-off of 10 µg Hb per gram (g) feces was used in the pilot studies for FIT-based screening.[14] The screening program initiated in 2014 started with a cut-off at 15 µg Hb/g feces which was later increased to 47 µg Hb/g feces due to higher than predicted numbers of positive results.[18] In the United States, a cut-off of 20 µg Hb/g feces is commonly used.[15] The aim of this study was to thoroughly evaluate diagnostic performance of a widely used quantitative FIT according to cut-off in a large cohort of participants of screening colonoscopy in Germany.

## 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, blood tests). Up to two screening colonoscopies 10 years apart are offered free of charge to people aged 55 and older (no upper age limit) in Germany since October 2002. Introduction was accompanied by major efforts of quality assurance, and high adenoma detection rates at low levels of complication rates have been achieved on a national level.[19,20] Some health care plans offer a first screening colonoscopy also at younger ages within specific programs.

[1]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; [2]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany and [3]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
Correspondence: H Brenner, MD, MPH, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, Heidelberg 69120, Germany. E-mail: h.brenner@dkfz-heidelberg.de

Details of the design of the BLITZ study have been reported elsewhere.[7,21–25] Briefly, >9,000 participants of screening colonoscopy have been recruited by a network of up to 20 gastroenterologists since the initiation of the study in late 2005. Participants of screening colonoscopy are informed about the study at a pre-colonoscopy practice visit, which is typically scheduled about one week before screening colono-scopy. Upon informed written consent, participants are asked to fill out a short questionnaire, and to give a blood sample. They are also handed out devices for feces collection and asked to collect a fecal sample prior to initiation of the bowel cleansing for colonoscopy. Colonoscopy and histology reports are obtained from the gastroenterologists who are blinded with respect to any blood or stool test results. The study was approved by the Ethics Committees of the University of Heidelberg and of the responsible state physicians' boards. It is registered at the German Clinical Trials Register (Deutsches Register Klinischer Studien, DRKS), DRKS-ID: DRKS00008737.

Different ways of feces collection have been tested and various tests have been employed during the course of the study. The current analysis is restricted to participants who were recruited between January 2012 and September 2014 when fecal samples were collected for FOB Gold (Sentinel Diagnostics, Milano, Italy), a quantitative FIT which is based on a latex agglutination assay, according to the manufacturer's instructions. Among 2,261 participants recruited during this period we employed the following exclusion criteria to ensure the study participants to represent an average-risk screening population and to minimize the potential of false negative findings of screening colonoscopy (Figure 1): (1) age <50 or age ≥80 years ($n=65$); (2) history of CRC or inflammatory bowel disease ($n=20$); (3) colonoscopy in the previous 5 years ($n=94$); (4) inadequate bowel preparation before colonoscopy ($n=244$); (5) incomplete colonoscopy (cecum not reached, $n=16$). The latter two criteria were not applied for CRC patients with a stenosis caused by the tumor mass. Finally, 1822 remaining participants were included in the analysis.

**Feces sample collection and laboratory analyses.** Participants were asked to collect one fecal sample using the sampling device from Sentinel Diagnostics (Milano, Italy; Ref. 11561H) according to the manufacturer's instructions, i.e., with a sampling stick to be inserted several times at random into one freshly passed whole feces and then re-inserted in a collection tube containing hemoglobin stabilizing buffer (10 mg feces in 1.7 ml extraction buffer). Participants were asked to mail the tubes with buffer-stabilized feces samples in sealed envelopes to the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) at their earliest convenience. At DKFZ, the tubes were kept at 2–8 °C in the refrigerator before transfer in a temperature controlled environment to the central, DIN EN ISO 15189 accredited laboratory (Labor Limbach, Heidelberg, Germany), where they were again stored at 2–8 °C in the refrigerator until FIT analysis.

All FIT analyses were conducted as single measurements in fully automatic manner using Abbott Architect c8000 (Abbott Laboratories, Abbott Park, Illinois, USA). The analytical
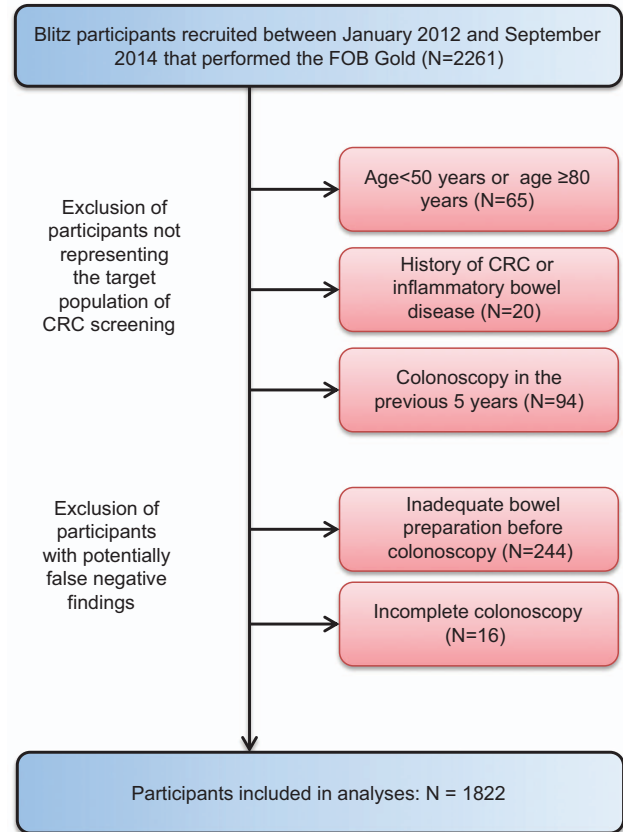


**Figure 1** STAndards for the Reporting of Diagnostic accuracy studies (STARD) diagram of participants of the BLITZ study.

working range was 0.2–132 µg Hb per g feces; specimen with concentrations above the upper limit were not diluted and re-analyzed. Laboratory personnel were blinded with respect to colonoscopy findings. 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 (3–5) days, the median time (IQR) between arrival at DKFZ and laboratory analysis was 2 (1–4) days, and the median overall time (IQR) between fecal sampling and laboratory analysis was 7 (6–9) days.

**Data extraction from colonoscopy and histology reports.** Clinical data were extracted from colonoscopy and histology reports in a standardized manner by trained research assistants who were blinded with respect to results of any blood or stool tests. Participants were classified according to the most advanced finding at screening colonoscopy using the following categories: colorectal cancer, advanced ade-noma, non-advanced adenoma, other (nonneoplastic) polyps, none of above. Advanced adenoma was defined by the presence of at least one adenoma with any of the following features: ≥1 cm in size, tubulovillous or villous components, high grade dysplasia. Information on complete-ness of colonoscopy and quality of bowel cleansing was extracted.

**Statistical analyses.** Descriptive statistics were used to present main characteristics (sex, age, most advanced finding at screening colonoscopy, distribution of FIT results) of the study population. We then derived sensitivity, specificity and the positive predictive value (PPV) with 95% confidence intervals (95% CIs) for detecting at least one advanced neoplasm (AN, defined as either CRC or an advanced adenoma), for a broad range of possible FIT cut-offs between 1 and 50 µg Hb/g feces including but not restricted to the cut-off recommended by the manufacturer (17 µg/g) and other widely used cut-offs (10 µg Hb/g feces, 20 µg Hb/g feces). We also derived reciprocal values of PPVs which indicate the numbers of participants needed to undergo follow-up colonoscopy after a positive FIT ("numbers needed to scope", NNS) to detect one AN case. Note that we refer to numbers of AN cases, reflecting numbers of participants with at least one AN which is different from numbers of ANs, as participants may have more than one AN.

In addition to the commonly reported indicators of diagnostic performance for specific cut-offs, we specifically evaluated the expected diagnostic yield of colonoscopies (prevalence of AN, number needed to scope to detect one AN case) conducted among participants with FIT results between defined cut-offs. In supplementary analyses, we repeated evaluation of diagnostic performance using only CRC cases rather than AN cases as the primary endpoint.

All analyses were conducted with R version 3.2.3 (2015–12–10).

## RESULTS

Main characteristics of the study population are shown in Table 1. Men and women were almost equally represented.

**Table 1** Characteristics of the study population

| Characteristic | n | % |
| --- | --- | --- |
| *Sex* | | |
| Male | 894 | 49.1 |
| Female | 928 | 50.1 |
| | | |
| *Age* | | |
| 50–54 years | 50 | 2.7 |
| 55–59 years | 772 | 42.4 |
| 60–64 years | 394 | 21.6 |
| 65–69 years | 288 | 15.8 |
| 70–74 years | 253 | 13.9 |
| 75–79 years | 65 | 3.6 |
| | | |
| *Most advanced finding at screening colonoscopy* | | |
| Colorectal cancer | 14 | 0.8 |
| Advanced adenoma | 193 | 10.6 |
| Non-advanced neoplasm | 347 | 19.0 |
| None of above | 1268 | 69.6 |
| | | |
| Any advanced neoplasm | 207 | 11.4 |
| No advanced neoplasm | 1615 | 88.6 |

The vast majority of participants (93.7%) were between 55 and 74 years old, with the age group 55–59 representing the largest 5-year age group. The most advanced finding at colonoscopy was CRC and advanced adenoma in 0.8 and 10.6% of participants, respectively, resulting in a total proportion of 11.4% of participants with AN. Advanced adenomas were ≥ 1 cm in the majority (76%) of cases.

Figure 2 shows the distribution of FOB Gold results in the entire study population. In 329 (18.1%) of cases, the fecal Hb concentration was below 1 µg/g. Distribution of Hb concentration in fecal samples from the remaining participants was strongly skewed to the left, ranging from 1 to ≥ 132 µg Hb/g feces with the mode at 4 µg Hb/g feces. Distributions strongly varied by most advanced finding at colonoscopy. For example, whereas 11 out of 14 CRC cases (79%) had Hb concentrations ≥ 50 µg Hb/g feces, this proportion was below 2% among those without colorectal neoplasms.

Figure 3 shows key indicators of diagnostic performance for detecting AN cases. Specifically, sensitivity, specificity, and PPV for a wide range of possible cut-offs between 1 and 50 µg Hb/g feces are shown. Specific numeric values for selected cut-offs and their 95% CIs are explicitly given in Table 2. As expected, sensitivity strongly decreased, and specificity strongly increased with increasing cut-offs. The positive predictive value for detecting AN showed a very steep increase with increasing cut-off up to 35.2% (95% CI 29.9–40.9%) at a cut-off of 9 µg Hb/g feces at which sensitivity and specificity were 48.8% (95% CI 42.1–55.6%) and 88.5% (95% CI 86.8–89.9%), respectively. PPV's reciprocal value, the number needed to scope (NNS), showed a corresponding decrease down to 2.8 (95% CI 2.4–3.3) at this cut-off. A further moderate increase of PPV up to 56.9% (95% CI 47.8–65.5%) and decrease of NNS down to 1.8 (95% CI 1.5–2.1), along with a major decrease in sensitivity was observed when gradually increasing the cut-off to 25 µg Hb/g feces at which sensitivity and specificity were 31.9% (95% CI 25.9–38.5%) and 96.9% (95% CI 95.9–97.6%), respectively. Further increases of the cut-off hardly affected PPV, NNS and specificity, but went along with further relevant decline in sensitivity. Stratification of findings by time between collection and arrival at DKFZ of the buffer-stabilized fecal sample (≤4 days/ >4 days), did not reveal any relevant differences in diagnostic test performance.

Looking at prevalences of AN for specific ranges of Hb concentrations (Table 3), these prevalences were very low for Hb concentrations below 9 µg Hb/g feces. Even for participants with levels between 5 and 9 µg Hb/g feces, prevalence was lower (7.3%, 95% CI 5.3–9.9%) than prevalence in the entire study population (11.4%). Prevalences increased to 20.5% (95% CI 15.1–27.1%) among participants with Hb concentrations between 9 and 25 µg Hb/g feces and >50% among participants with Hb concentrations above 25 µg Hb/g feces.

Results of supplementary analyses using CRC rather than AN as the diagnostic endpoint are shown in Appendix Figure 1. As expected, specificity was very similar to specificity for AN, but sensitivity was much higher. All of the 14 CRC would have been detected, i.e., sensitivity would have been 100% (95% CI 78.5–100%) at cut-offs up to 28 µg Hb/g feces. Even at a cut-off at 50 µg Hb/g feces still 11 CRC would have been
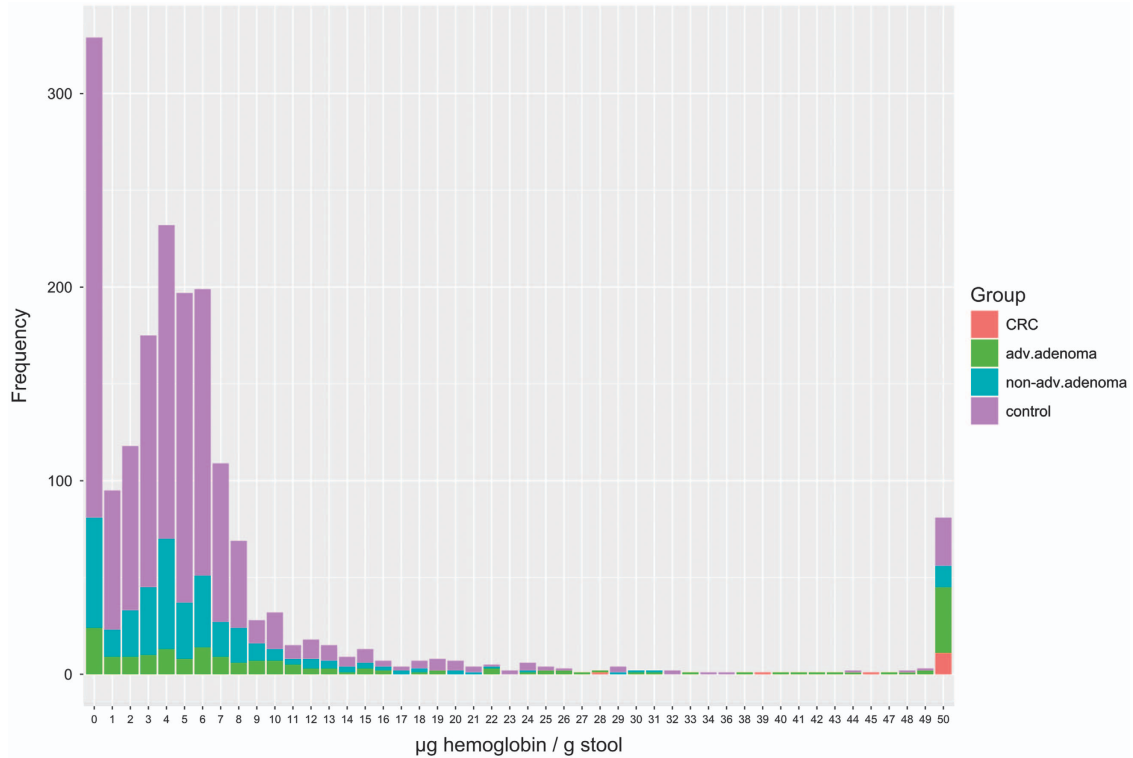
**Figure 2** Distribution of results of FOB Gold.
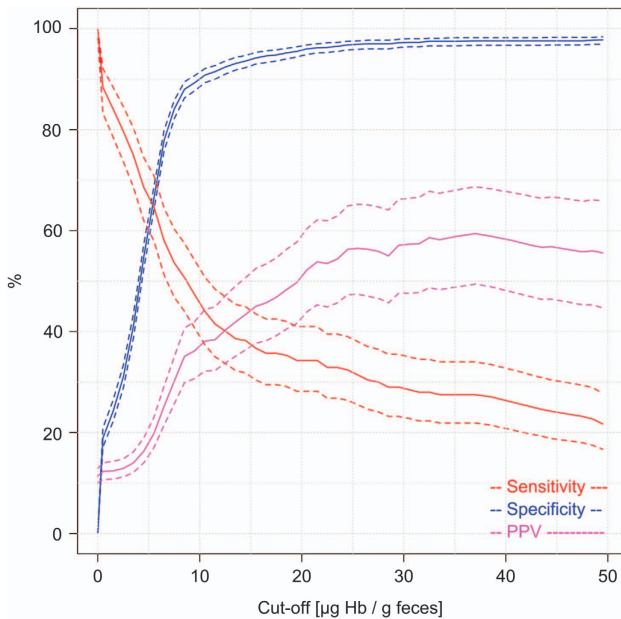


**Figure 3** Sensitivity, specificity and positive predictive value (PPV) for detecting participants with advanced neoplasms according to cut-off (solid lines: point estimates; dashed lines: 95% confidence intervals).

detected (sensitivity 78.6%, 95% CI 52.4–92.4%). Because of the much lower prevalence of CRC compared with AN, the PPVs for CRC were substantially lower for CRC than for AN, with maximum levels below 15% even for cut-offs ≥ 25 µg Hb/g

feces. Despite the differences in absolute levels of PPV, the shape of the PPV-cut-off relationship was rather similar to the one seen for AN in that the increase of PPV with cut-offs was steepest between ~5 and 9 µg Hb/g feces, and essentially leveled off at cut-offs above 25 µg Hb/g feces.

## DISCUSSION

In CRC screening practice, quantitative FITs are commonly used as dichotomous tests, using cut-offs for test positivity recommended by the manufacturers. The basis for determining such cut-offs is commonly not known to the users of quantitative FITs, and simply adopting a specific recommended cut-off may not be the best choice for application of FITs in many situations. In this article, we provide relevant information for selecting a cut-off using FOB Gold, a widely used quantitative FIT.

The cut-off recommended by the manufacturer for this test, 17 µg Hb/g feces, seems to be reasonably high to ensure high specificity (95%) and a PPV to detect at least one AN of 47% in subsequent colonoscopy. Nevertheless, our analyses suggest that lowering this cut-off to 9 µg Hb/g feces may be worthwhile in screening settings where sufficient colonoscopy capacities are available, as this would substantially increase the sensitivity for detecting AN from 36% to almost 50%, albeit at a modest loss of specificity and PPV. Our data also clearly show, however, that cut-offs below 9 µg Hb/g feces may not be meaningful, as prevalence of AN for people with fecal Hb concentrations between 5 and 9 µg Hb/g feces was even lower (7%) than prevalence in the entire screening population

**Table 2** Indicators of test performance for detecting advanced neoplasms according to test cut-off

| Indicator of test performance | Cut-off for test positivity (µg Hb/g feces) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **5** | **9** | **10** | **17**[a] | **20** | **25** | **50** |
| *Positivity rate* | | | | | | | |
| N | 782/1822 | 287/1822 | 253/1822 | 159/1822 | 138/1822 | 116/1822 | 79/1822 |
| % | 42.9 | 15.8 | 13.9 | 8.7 | 7.6 | 6.4 | 4.3 |
| 95% CI | 40.7–45.2 | 14.2–17.5 | 12.4–15.5 | 7.5–10.1 | 6.4–8.9 | 5.3–7.6 | 3.5–5.4 |
| *Sensitivity* | | | | | | | |
| N | 137/207 | 101/207 | 93/207 | 74/207 | 71/207 | 66/207 | 43/207 |
| % | 66.2 | 48.8 | 44.9 | 35.7 | 34.3 | 31.9 | 20.8 |
| 95% CI | 59.5–72.3 | 42.1–55.6 | 38.3–51.7 | 29.5–42.5 | 28.2–41.0 | 25.9–38.5 | 15.8–26.8 |
| *Specificity* | | | | | | | |
| N | 970/1615 | 1429/1615 | 1455/1615 | 1530/1615 | 1548/1615 | 1565/1615 | 1579/1615 |
| % | 60.1 | 88.5 | 90.1 | 94.7 | 95.9 | 96.9 | 97.8 |
| 95% CI | 57.7–62.4 | 86.8–89.9 | 88.5–91.5 | 93.5–95.7 | 94.8–96.7 | 95.9–97.6 | 96.9–98.4 |
| *Positive predictive value* | | | | | | | |
| N | 137/782 | 101/287 | 93/253 | 74/159 | 71/138 | 66/116 | 43/79 |
| % | 17.5 | 35.2 | 36.8 | 46.5 | 51.4 | 56.9 | 54.4 |
| 95% CI | 15.0–20.3 | 29.9–40.9 | 31.1–42.9 | 39.0–54.3 | 43.2–59.6 | 47.8–65.5 | 43.5–65.0 |
| *Number needed to scope to detect one advanced neoplasm* | | | | | | | |
| N | 5.7 | 2.8 | 2.7 | 2.2 | 1.9 | 1.8 | 1.8 |
| 95% CI | 4.9–6.7 | 2.4–3.3 | 2.3–3.2 | 1.8–2.6 | 1.7–2.3 | 1.5–2.1 | 1.5–2.3 |

CI, confidence interval; Hb, hemoglobin.
[a]cut-off recommended by the manufacturer
.

**Table 3** Indicators of test performance for detecting advanced neoplasms between cut-offs

| Results and findings within range | Test result range (µg Hb/g feces) | | | | |
|---|---|---|---|---|---|
| | **<5** | **5–<9** | **9–<25** | **≥25–<50** | **≥50** |
| *Number of results* | | | | | |
| N | 1040 | 495 | 171 | 37 | 79 |
| *Advanced neoplasms* | | | | | |
| N | 70 | 36 | 35 | 23 | 43 |
| % | 6.7 | 7.3 | 20.5 | 62.2 | 54.4 |
| 95% CI | 5.4–8.4 | 5.3–9.9 | 15.1–27.1 | 46.1–75.9 | 43.5–65.0 |
| *Number needed to scope to detect one advanced neoplasm* | | | | | |
| N | 14.9 | 13.7 | 4.9 | 1.6 | 1.8 |
| 95% CI | 11.9–18.5 | 10.1–18.9 | 3.7–6.6 | 1.3–2.2 | 1.5–2.3 |

CI, confidence interval; Hb, hemoglobin.

(11.4%). Declaring a FIT value in the 5–9 µg Hb/g feces range as positive would appear counterintuitive, as it would imply recommending follow-up colonoscopy to people who have undergone a screening test to determine the need of such colonoscopy but whose post-test probability of having AN was lower than their pre-test probability.

On the other hand, increasing the cut-off from 17 to 25 µg Hb/g feces would increase the PPV from 47 to 57% and decrease the NNS from 2.1 to 1.8 at only a modest loss of sensitivity for AN detection (from 36 to 32%) and might be a cost-effective alternative that might be of special interest in case of limited colonoscopy capacities. Further increasing the cut-off might not be advisable, however, as it would go along with further reduction in sensitivity without any relevant gain in

specificity or PPV or reduction of NNS. Our results therefore suggest that a range of cut-offs from 9 to 25 µg Hb/g feces would be meaningful from an epidemiological perspective for the target population of screening in Germany, with the best cut-off within this range depending on priorities given to either high sensitivity or high PPV and specificity.

Even though a number of studies have evaluated diagnostic performance of FITs at their predefined cut-offs[8,12] or at a few alternative cut-offs, e.g.,[26] comprehensive evaluations over the full range of potential cut-offs have been sparse.[27] Although the more commonly presented receiver operating characteristic curves and their areas under the curve also reflect performance of FITs over a broad range of cut-offs, e.g.,[23,27] the specific cut-offs yielding the pairs of sensitivity

and specificity are not commonly depicted in these curves which limits their use for selecting cut-offs.

To our knowledge, our study provides the first comprehensive clinical evaluation of selection of the cut-off and its implications for FOB Gold, a widely used screening test, in a true screening setting. In the Netherlands' pilot study, use of this FIT in the fourth round of biennial screening with a cut-off at 10 µg Hb/g feces yielded a PPV of 32%,[28] which is close to the value of 37% for this cut-off found in our study. Another recent study from the Netherlands pointed to the decline of the PPV of FIT based testing using the same cut-off over multiple rounds of screening, most likely as a result of declining AN prevalence after detecting large proportions of AN at initial FIT screenings,[29] suggesting that adaptations of cut-offs to higher values might be warranted in the longer run in population-based screening. In fact, the initial cut-off of 15 µg Hb/g feces in the official Dutch CRC screening program was later increased to 47 µg Hb/g feces due to higher than predicted participation rates and lower than predicted positive predictive value (PPV).[18]

Other previous work evaluating FOB Gold focused on technical and operational performance.[30] In our study, we focused on indicators of diagnostic performance across different FIT cut-offs which are the main indicators of interest in clinical care. From a societal perspective, additional factors, such as colonoscopy capacity as well as cost-effectiveness of FIT based screening require additional consideration. For example, in countries with limited colonoscopy capacity, a cut-off at the upper end of or even above the range that appears most meaningful from a clinical perspective might be the only way to go.[18,31] Conversely, analyses like ours might be helpful to suggest a lower limit of reasonable cut-offs in situations where colonoscopy capacity is not a limitation. In such cases, modeling effectiveness and cost-effectiveness of FIT based screening by microsimulation models, e.g.,[32,33] using various cut-offs within the clinically relevant range, might be useful to select the "best" cut-off from a societal perspective. Results on diagnostic performance according to various cut-offs as derived in our study may be most helpful to inform such modeling. Such modeling should though also take additional considerations into account, such as the age range in which screening is offered, time intervals between tests, possible adaptations of cut-offs at repeated testing, costs of the FITs and colonoscopies, as well as potential harms of screening.

Our study was conducted in a country with relatively high CRC incidence and no previous FIT-based screening (even though screening with guaiac based fecal occult blood tests has been in place in Germany for decades, albeit in an opportunistic manner and with very low uptake rates). Prevalences of AN may be lower in countries with lower CRC incidence or with more intensive previous screening activities. Although such differences in prevalence should not affect sensitivities and specificities, PPVs would be lower under such circumstances. Nevertheless, when we repeated our analyses assuming identical sensitivities and specificities but 50% lower prevalence of AN, very similar cut-off dependency of PPVs emerged, with a major increase of PPV up to 38% at a cut-off of 25 µg Hb/g feces and no relevant changes of PPV by further increases of the cut-off. Our

diagnostic performance derived in a country with middle European climate may not necessarily hold for countries with hotter climate, where increased degradation of Hb in routine practice might lead to lower Hb values, and therefore might require more stringent requirements for fecal sample handling, shipment and processing. Variation between FITs in the ability of the stabilizing buffer to prevent Hb degration might be a particularly important factor in this context. Such additional factors therefore require careful attention when determining and adapting cut-offs for specific applications. Again, modeling taking such factors into account may help in the selection of a "best" context-specific cut-off.

Our study has specific strengths and limitations. Strengths include the possibility to compare FIT results with results from screening colonoscopy in all participants of a large study conducted in a true screening setting. A limitation which is though shared with most evaluations of FITs reported to date is that just one specific FIT, FOB Gold, was evaluated. Different brands of FITs may differ in their diagnostic performance even at comparable cut-offs.[34] Diagnostic performance may vary between different quantitative FITs for a variety of reasons, such as differences in quality of the feces collection device, composition of buffer or detected epitopes which may be more or less affected by Hb degradation.[34] Results for FOB Gold may therefore not necessarily hold for other FIT brands. In previous comparative studies diagnostic performance of FOB Gold was though roughly similar to diagnostic performance of OC Sensor, another very commonly used FIT.[6,28,35,36] However, diagnostic performance may even vary for FOB Gold itself according to the analyzer used, suggesting that the cut-off might need to be determined taking this additional factor into account by the laboratories in charge of the measurements.[37]

Although colonoscopy is commonly used as gold standard for evaluating diagnostic performance of noninvasive CRC screening tests, it is not perfect and may miss some proportion of AN.[38,39] However, the screening colonoscopy program in Germany includes comprehensive efforts of quality assurance, and quality of screening colonoscopies is generally very high according to commonly employed criteria, such as the adenoma detection rate.[20] To further minimize miss rates, we excluded colonoscopies with imperfect bowel preparation and incomplete colonoscopies. Despite the overall large size of our study sample, some of our interval-specific estimates of diagnostic performance and the estimates of sensitivity and PPV for CRC only were based on relatively small numbers. For this reason, we also refrained from presenting subsite specific sensitivities. Variation in sensitivity according to colorectal subsites have recently been addressed in detail elsewhere.[40]

Despite its limitations, our study provides information that may be useful for cut-off selection in FIT–based screening. We suggest that similar evaluations should be reported, along with commonly reported sensitivities and specificities for a single pre-defined cut-off or receiver operating characteristic curves, in future evaluations of quantitative FITs in order to enhance the empirical evidence for cut-off selection in FIT-based screening in different populations, different screening settings

and for different FITs. Future studies should also explore the implications of selecting different cut-offs for effectiveness and cost-effectiveness of screening programs on the societal level. Such evaluations may best be done by microsimulation models for which results such as ours may provide important evidence-based input parameters.

## Study Highlights

### WHAT IS CURRENT KNOWLEDGE
✓ Fecal immunochemical tests for hemoglobin are increasingly used for colorectal cancer screening.
✓ Cut-offs for defining test positivity vary widely.

### WHAT IS NEW HERE
✓ Detailed quantitative information on the impact of cut-off choice on key diagnostic performance characteristics is provided.
✓ Results enable evidence-based cut-off choices tailored to the specific needs of the respective screening program.

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**: E359–E386.
2. Hewitson P, Glasziou P, Watson E et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol* 2008; **103**: 1541–1549.
3. Scholefield JH, Moss SM, Mangham CM et al. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut* 2012; **61**: 1036–1040.
4. Shaukat A, Mongin SJ, Geisser MS et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med* 2013; **369**: 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**: 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**: 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**: 3049–3054.
8. Lee JK, Liles EG, Bent S et al. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med* 2014; **160**: 171.
9. Hol L, van Leerdam ME, van Ballegooijen M et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. *Gut* 2010; **59**: 62–68.
10. Tinmouth J, Lansdorp-Vogelaar I, Allison JE. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. *Gut* 2015; **64**: 1327–1337.
11. Sung JJ, Ng SC, Chan FK et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. *Gut* 2015; **64**: 121–132.
12. 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. *Gastroenterology* 2017; **112**: 37–53.
13. Schreuders EH, Ruco A, Rabeneck L et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015; **64**: 1637–1649.
14. Kapidzic A, van Roon AH, van Leerdam ME et al. Attendance and diagnostic yield of repeated two-sample faecal immunochemical test screening for colorectal cancer. *Gut* 2017; **66**: 118–123.
15. Allison JE, Graser CG, Halloran SP et al. 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**: 117–130.
16. Haug U, Grobbee EJ, Lansdorp-Vogelaar I et al. Immunochemical faecal occult blood testing to screen for colorectal cancer: can the screening interval be extended? *Gut* 2016; **66**: 1262–1267.
17. Jensen CD, Corley DA, Quinn VP et al. Fecal immunochemical test program performance over 4 rounds of annual screening: a retrospective cohort study. *Ann Intern Med* 2016; **164**: 456–463.
18. Toes-Zoutendijk E, van Leerdam ME, Dekker E et al. Real-time monitoring of results during first year of Dutch colorectal cancer screening program and optimization by altering fecal immunochemical test cut-off levels. *Gastroenterology* 2016; **152**: 767–775.e2.
19. Pox CP, Altenhofen L, Brenner H et al. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. *Gastroenterology* 2012; **142**: 1460–1467.
20. 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**: 356–366.
21. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med* 2009; **150**: 162–169.
22. Brenner H, Tao S, Haug U. Low-dose aspirin use and performance of immunochemical fecal occult blood tests. *JAMA* 2010; **304**: 2513–2520.
23. 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**: 682–690.
24. Brenner H, Haug U, Hundt S. Sex differences in performance of fecal occult blood testing. *Am J Gastroenterol* 2010; **105**: 2457–2464.
25. 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* 2016 [Epub ahead of print].
26. Rozen P, Comaneshter D, Levi Z et al. Cumulative evaluation of a quantitative immunochemical fecal occult blood test to determine its optimal clinical use. *Cancer* 2010; **116**: 2115–2125.
27. Hernandez V, Cubiella J, Gonzalez-Mao MC et al. Fecal immunochemical test accuracy in average-risk colorectal cancer screening. *World J Gastroenterol* 2014; **20**: 1038–1047.
28. Grobbee EJ, van der Vlugt M, van Vuuren AJ et al. A randomised comparison of two faecal immunochemical tests in population-based colorectal cancer screening. *Gut* 2016 [Epub ahead of print].
29. Stegeman I, van Doorn SC, Mundt MW et al. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. *Cancer Epidemiol* 2015; **39**: 388–393.
30. NHS Bowel Cancer Screening Southern Programme Hub. Evaluation of quantitative faecal immunochemical tests for haemoglobin, available at http://194.97.148.137/assets/downloads/pdf/activities/fit_reports/gmec_fit_evaluation_report.pdf. Accessed 2 April 2017.
31. Moss S, Mathews C, Day TJ et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. *Gut* 2016 [Epub ahead of print].
32. 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**: 1985–1997.
33. Knudsen AB, Zauber AG, Rutter CM et al. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US Preventive Services Task Force. *J Am Med Assoc* 2016; **315**: 2595–2609.
34. Chiang TH, Chuang SL, Chen SL et al. Difference in performance of fecal immunochemical tests with the same hemoglobin cutoff concentration in a nationwide colorectal cancer screening program. *Gastroenterology* 2014; **147**: 1317–1326.
35. Rubeca T, Rapi S, Confortini M et al. Evaluation of diagnostic accuracy of screening by fecal occult blood testing (FOBT). Comparison of FOB Gold and OC Sensor

assays in a consecutive prospective screening setting. *Int J Biol Markers* 2006; **21**: 157–161.

36. Zubero MB, Arana-Arri E, Pijoan JI *et al.* Population-based colorectal cancer screening: comparison of two fecal occult blood tests. *Front Pharmacol* 2014; **4**: 175.

37. National Institute for Health and Care Excellence Overview—Quantitative faecal immunochemical tests to assess people with symptoms, who are at low risk of colorectal cancer, in primary care. Available at https://www.nice.org.uk/guidance/GID-DG10005/documents/overview. Accessed 7 April 2017.

38. 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**: 284–290.

39. Ahn SB, Han DS, Bae JH *et al.* The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. *Gut Liver* 2012; **6**: 64–70.

40. Brenner H, Niedermaier T, Chen H. Strong subsite-specific variation in detecting advanced adenomas by fecal immunochemical testing for hemoglobin. *Int J Cancer* 2017; **140**: 2015–2022.

### Appendix Figure 1

Sensitivity, specificity and positive predictive value (PPV) for detecting participants with CRC according to cut-off (solid lines: point estimates; dashed lines: 95% confidence intervals).