



Research Article

Evaluation of inter-test agreement and analytical performance of eight fecal immunochemical tests

Nihan Usakligil^{1*}, Sevilay Sezer^{1,2}, Fatma Meric Yilmaz^{1,3}, Turan Turhan^{1,3,4}

¹Medical Biochemistry Laboratory, Ankara Numune Training and Research Hospital, Ankara, Türkiye

²Medical Biochemistry Laboratory, Ankara Bilkent City Hospital, Ankara, Türkiye

³Department of Medical Biochemistry, Ankara Yildirim Beyazıt University Faculty of Medicine, Ankara, Türkiye

⁴Department of Medical Biochemistry, University of Health Sciences, Ankara Bilkent City Hospital Health Application and Research Center, Ankara, Türkiye

*Currently not affiliated with an institution

Abstract

Objectives: Immunochemical fecal occult blood test has been commonly used for community-based colorectal cancer screening. There is a wide variety of fecal immunochemical test (FIT) products available in the market. However, there is limited performance information for many brands, making it essential to assess and compare the performance of these FITs. Therefore, this study aims to assess the level of agreement between eight FIT products with various cut-off values (two quantitative, six qualitative) and to evaluate the analytical performance of FITs through repeatability, interference, and stability analyses.

Methods: This study was conducted using a total of 313 stool specimens of which 211 specimens were obtained from randomly selected patients without any dietary restrictions, and the remainder 101 specimens were spiked with hemoglobin. The fecal occult blood results from all brands were evaluated as positive or negative. The level of agreement of FITs were assessed. Repeatability, stability and interference studies of FITs were also carried out.

Results: FITs were found to have fair to almost perfect agreement with kappa values ranging from 0.28 to 0.94 (all $p < 0.001$) in the pairwise comparisons but statistically significant differences were found among most FITs by McNemar's test with Bonferroni correction (adjusted $\alpha = 0.0018$). Repeatability and interference studies showed consistent results, but stability performance varied among FITs.

Conclusion: This study showed that agreement and analytical performance among FITs vary, that statistically significant differences may be observed between some test pairs, and that agreement measures alone are not sufficient when considering test interchangeability. Test selection should be based on a comprehensive assessment that considers analytical performance together with agreement results and the potential impact on laboratory and clinical management.

Keywords: Agreement, analytical performance, fecal immunochemical test, fecal occult blood, FIT, FOB

How to cite this article: Usakligil N, Sezer S, Yilmaz FM, Turhan T. Evaluation of inter-test agreement and analytical performance of eight fecal immunochemical tests. Int J Med Biochem 2026;9(2):103–110.

Colorectal cancer represents a major public health problem worldwide. According to the 2020 GLOBOCAN statistics, it was among the most frequently diagnosed types of cancer and one of the leading causes of cancer-related deaths. The age-standardized incidence rates of colorectal cancer in Türkiye were 16.2 for women and 26.2 for men per 100,000 popula-

This study is part of a thesis titled "Analytical performance evaluation of immunochemical fecal occult blood tests [İmmünokimyasal gaitada gizli kan testlerinin analitik performans değerlendirilmesi]" (Ankara Numune Training and Research Hospital).

Address for correspondence: Nihan Usakligil, MD. Medical Biochemistry Laboratory, Ankara Numune Training and Research Hospital, Ankara, Türkiye

Phone: +90 532 685 41 64 **E-mail:** nsevketa@gmail.com **ORCID:** 0009-0008-6581-2508

Submitted: November 07, 2025 **Revised:** February 05, 2026 **Accepted:** February 09, 2026 **Available Online:** April 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



tion (world standard population), ranking third among all cancers. Colorectal cancer also ranked third in terms of mortality, with age-standardized mortality rates of 7.8 for women and 13.0 for men per 100,000 population (world standard population) [1]. Therefore, its detection via screening programs in precancerous or early stages remains of paramount importance for prognosis and survival.

While most of the population-based screening programs use a fecal immunochemical test (FIT) or a guaiac-based fecal occult blood test as the screening methodology [2], Türkiye's population-based screening program includes a biennial FIT and a colonoscopy every 10 years for individuals aged 50–70 years [3]. FIT is the preferred approach in Türkiye for detecting human hemoglobin in stool samples because i) food and medicines do not typically interfere with its results [4]; ii) it is specific to human hemoglobin [4]; and iii) it has a high cost-benefit ratio [3].

While a great number of qualitative and quantitative FIT brands are available on the market, this may complicate the selection of an appropriate test for laboratory use. Therefore, studies evaluating both the clinical and analytical performance of these FITs are needed. Most published studies have focused on clinical performance [5–7], while fewer have examined analytical performance through comparative evaluations [8, 9]. However, analytical performance of FITs is essential for reliable test results, which directly affects patient referral and follow-up decisions in colorectal cancer screening.

Although clinical performance of FITs should ideally be assessed using colonoscopy as the reference method, this may not always be practical due to limited availability of colonoscopy, difficulties in obtaining an appropriate patient population for clinical performance evaluation, and challenges related to clinical workload management. Therefore, clinical performance data for test brands may not always be available. In such cases, comparative agreement studies between screening tests may still provide valuable information for test selection and laboratory practice.

In addition, the cut-off values of FITs may vary significantly [10]. The variation in cut-off values used by different FIT brands for the same specimen may lead to different interpretations (i.e., positive or negative) of test results, which play an important role in decision-making in colorectal cancer screening and may also lead to variability in positivity rates. FITs with lower cut-off values are expected to yield more positive results, which require further diagnostic evaluation and thus greater effort in managing the clinical workload [11]. However, using higher cut-off values may result in missed diagnoses of the disease. Therefore, the potential impacts of cut-off values and available clinical capacity should be considered together when selecting a test.

Although several studies have been conducted, performance data for many brands is still insufficient, making it difficult for health professionals to make informed choices [12]. Accordingly, the primary aim of this study was to compare the agreement between seven FITs (one automated quantitative

and six qualitative) and the automated quantitative FIT analyzer currently used in our laboratory and previously evaluated in various studies [13–15], as well as the agreement among the seven FITs, using overall percent agreement, kappa statistics, and McNemar analyses. The secondary aim of this study was to evaluate the analytical performance of the FITs through interference testing, repeatability assessment, and stability analyses.

Materials and Methods

The study was approved by the Ankara Numune Training and Research Hospital Ethics Committee (No: 178/2014, Date: 07/05/2014).

This study was conducted using a total of 313 stool specimens in accordance with the principles of the Declaration of Helsinki. Of the 313 specimens, 211 were randomly collected from patients without dietary restrictions, while the remaining 102 were spiked to achieve a hemoglobin concentration of 50 ± 10 ng/mL by mixing negative fecal occult blood (FOB) samples with hemolysate, as most of the FITs evaluated in the study had a cut-off value of 50 ng/mL. The FOB results of the mixtures were determined based on the values measured by the NS-Plus C15 analyzer (Alfresa Pharma Corporation, Japan) currently used in our laboratory.

Hemolysate was prepared by osmotic (hypotonic) lysis of erythrocytes using distilled water, supported by mechanical mixing. For this purpose, blood samples collected from healthy volunteers into K2 EDTA blood collection tubes were used. The tubes were first centrifuged, and the supernatant was removed. The remaining pellet was washed with isotonic saline and subsequently resuspended in distilled water at a 1:1 ratio. After mechanical mixing, the mixture was centrifuged, and the supernatant (hemolysate) was separated for later use.

Quality control of the quantitative analyzers was performed daily, and all FITs were carried out in accordance with the users' instructions. The stool samples used in the tests were collected from the same or nearby regions of stool specimens.

All samples collected from three different parts of the stool specimens were first analyzed using the NS-Plus C15 analyzer that works on the basis of a colloidal gold agglutination immunoassay method with a cut-off value of 100 ng/mL. The same samples were then tested by using i-Chroma (Boditech Med, Korea, cut-off 50 ng/mL), a quantitative FIT brand that works on the basis of a fluorescence immunoassay method, and the following six qualitative FIT brands that work on the basis of a lateral flow immunoassay method: Certest (Biotec S.L., Spain, cut-off 16 ng/mL), True Line (Biocare Diagnostics Ltd., China, cut-off 50 ng/mL), SD (Standard Diagnostics Inc., Korea, cut-off 10 ng/mL), Rapidan Tester (Türklab Tıbbi Malz. San. Tic. A.Ş., Türkiye, cut-off 50 ng/mL), Laboquick (Koroğlu Tıbbi Malz. San. ve Tic. Ltd. Şti., Türkiye, cut-off 50 ng/mL) and Innovacon (Innovacon Inc., USA, cut-off 50 ng/mL).

During the study, differences were observed within the same test brands in the buffer volume among sample collection

Table 1. Agreement values between NS-Plus C15 analyzer (cut-off 100 ng/mL) and different fecal immunochemical tests

Tests	Positivity rate (%)	OPA (95% CI)	PPA (95% CI)	NPA (95% CI)	κ (95% CI)
i-Chroma	45	79.9 (75.1–83.9)	95.3 (88.6–98.2)	74.0 (67.9–79.3)	0.58 (0.49–0.67)*
Certest	42	77.6 (72.7–82.9)	86.0 (77.2–91.8)	74.4 (68.4–79.7)	0.52 (0.42–0.61)*
SD	64	61.7 (56.2–66.9)	97.7 (91.9–99.4)	48.0 (41.6–54.0)	0.32 (0.25–0.39)*
Rapidan tester	57	67.1 (61.7–72.1)	94.2 (87.1–97.5)	56.8 (50.3–63.1)	0.38 (0.30–0.46)*
Laboquick	67	58.5 (52.9–63.8)	97.7 (91.9–99.4)	43.6 (37.3–50.1)	0.28 (0.22–0.35)*
Innovacon	59	66.1 (60.7–71.2)	95.3 (88.6–98.2)	55.1 (48.6–61.4)	0.37 (0.29–0.45)*
True line	65	60.4 (54.9–65.7)	97.7 (91.9–99.4)	46.3 (39.9–52.8)	0.31 (0.24–0.38)*

All analyses were based on 313 samples; the positivity rate of NS-Plus C15 was 27. *: Statistically significant after Bonferroni correction (McNemar's test; adjusted $\alpha = 0.0018$). OPA: Overall percent agreement; PPA: Positive percent agreement; NPA: Negative percent agreement; κ : Kappa coefficient; CI: Confidence interval.

tubes and in the amount of stool collected by different sampling sticks. In addition, NS-Plus C15, SD, and Certest sample collection tubes were equipped with a filter system to remove excess stool.

The FOB results from all brands were evaluated as positive or negative. In qualitative tests, the presence of faint test lines was considered a positive result. Results from all brands were compared with each other. Further, the same comparison was conducted with the NS-Plus C15 analyzer by setting the cut-off value to 50 ng/mL, consistent with that of most FITs evaluated in this study. Then the compatibility of the tests was evaluated.

Interference study

In order to examine whether FIT tests would cross-react with hemoglobins other than human hemoglobin, hemolysate prepared from blood samples collected from sheep, goats, and cattle using the same hemolysis procedure described above was mixed with stool samples with negative FOB results. The samples were run across all FIT brands, and the results were recorded.

Repeatability study

A total of three samples with negative (14 ng/mL), positive (371 ng/mL), and a concentration of 50 ± 10 ng/mL (53 ng/mL) were analyzed 10 times in all FITs.

The coefficient of variation (CV) was calculated for quantitative tests, and the results were evaluated according to the total allowable error defined by the analytical performance specifications of the Royal College of Pathologists of Australasia (RCPA) Quality Assurance Program.

Stability study

The users' manual of all FITs (except for Laboquick and Certest brands) we used in this study indicated that the samples taken into the sample collection tubes should remain stable at room temperature for three days. In order to assess the stability period indicated above, 20 stool specimens with different FOB concentrations, including 10 FOB_{negative} and 10 FOB_{positive} samples, were selected.

A sample of each stool specimen was collected into a sample collection tube of each FIT brand, and the FOB result for the first day was recorded. Over the following three days, FOB was analyzed daily using samples from sample collection tubes kept at room temperature (20–24°C), and the results were recorded.

Statistical analysis

In the comparison of the tests, agreements (e.g., overall percent agreement; OPA, positive percent agreement; PPA and negative percent agreement; NPA) were determined at a 95% confidence interval according to the User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline (EP12-A2) [16].

In addition, agreements were also evaluated using kappa statistics. For this purpose, SPSS V23 (SPSS Inc., Chicago, Illinois, USA) statistical software was used. Due to the inherent nature of the kappa statistic being influenced by the potential bias between tests, Cochran's Q and McNemar's tests were used in addition to the kappa statistic to determine whether the tests produced statistically different results. Statistical significance for McNemar's test was defined using a Bonferroni-adjusted α level (adjusted $\alpha = 0.0018$). The kappa coefficients (κ) values were evaluated according to Landis and Koch's classification [17].

Results

Fecal occult blood study

The agreement among all FITs included in this study ranged from fair to almost perfect based on their kappa values. Cochran's Q test, used to compare the FITs with each other, showed statistically significant differences ($p < 0.001$). Therefore, pairwise comparisons among all FITs were performed using McNemar's test with Bonferroni correction (adjusted $\alpha = 0.0018$).

The results from the statistical analyses demonstrating the agreement between the NS-Plus C15 analyzer and other FITs for two cut-off values of the NS-Plus C15 analyzer were summarized in Table 1 (cut-off value 100 ng/mL) and Table 2 (cut-off value 50 ng/mL). The agreement between the NS-Plus C15 (cut-

Table 2. Agreement values between NS-Plus C15 analyzer (cut-off 50 ng/mL) and different fecal immunochemical tests

Tests	Positivity rate (%)	OPA (95% CI)	PPA (95% CI)	NPA (95% CI)	κ (95% CI)
i-Chroma	45	76.9 (72.2–81.1)	71.9 (68.8–74.1)	82.1 (73.1–89.5)	0.54 (0.45–0.63)
Certest	42	75.4 (70.5–79.7)	67.5 (63.0–71.2)	83.3 (74.4–90.7)	0.51 (0.41–0.60)
SD	64	77.3 (71.6–82.3)	91.7 (88.6–92.8)	62.8 (53.2–71.8)	0.55 (0.46–0.63)*
Rapidan tester	57	76.9 (71.5–81.8)	84.1 (80.4–86.2)	69.9 (60.2–78.6)	0.54 (0.45–0.63)
Laboquick	67	76.0 (70.3–81.1)	93.6 (90.5–94.7)	58.3 (48.9–67.4)	0.52 (0.43–0.61)*
Innovacon	59	77.9 (72.4–82.8)	86.6 (83.1–88.4)	69.2 (59.5–78.0)	0.56 (0.47–0.65)
True line	65	76.0 (70.3–81.1)	91.7 (88.6–92.8)	60.3 (50.7–69.3)	0.52 (0.43–0.61)*

All analyses were based on 313 samples; the positivity rate of NS-Plus C15 was 50%. *: Statistically significant after Bonferroni correction (McNemar's test; adjusted $\alpha=0.0018$). OPA: Overall percent agreement; PPA: Positive percent agreement; NPA: Negative percent agreement; κ : Kappa coefficient; CI: Confidence interval.

off 100 ng/mL, positivity rate 27%) analyzer and the other FITs ranged from fair to moderate with κ values of 0.28 to 0.58 across pairwise comparisons (all $p<0.001$), and statistically significant differences were observed in all comparisons by McNemar's test with Bonferroni correction (adjusted $\alpha=0.0018$). For the cut-off value of 50 ng/mL (positivity rate 50%), the agreements were moderate, with κ values ranging from 0.51 to 0.56 across pairwise comparisons (all $p<0.001$). Using McNemar's test with Bonferroni correction (adjusted $\alpha=0.0018$), only three test pairs, NS-Plus C15 vs. SD, NS-Plus C15 vs. Laboquick, and NS-Plus C15 vs. True Line, showed statistically significant differences.

The statistical results of the pairwise comparisons among the remaining FITs were summarized in Table 3. According to these results, there was moderate to almost perfect agreement between the tests, with κ values ranging from 0.52 to 0.94 across pairwise comparisons (all $p<0.001$). Using McNemar's test with Bonferroni correction (adjusted $\alpha=0.0018$), statistically significant differences were observed in most test pairs, except for the i-Chroma vs. Certest, Innovacon vs. Rapidan Tester, True Line vs. SD, SD vs. Laboquick, and True Line vs. Laboquick test brands.

Interference study

Fecal occult blood tests using stool samples spiked with sheep, goat, and cattle blood yielded negative results in all FITs, indicating no cross-reactivity.

Repeatability study

In the repeatability study of qualitative tests performed with negative stool specimens, only the Innovacon test yielded positive results in one of the ten replicates. For all other stool specimens used in repeatability testing, all tests yielded positive results in all replicates.

Among the quantitative analyzers, the CV% values of NS-Plus C15 were 15.8%, 9.8%, and 23.8% for negative, 53 ng/mL hemoglobin concentration, and positive samples, respectively, while these values for i-Chroma were found to be 10.2%, 10.3%, and 18.9%, respectively. For the positive sample, the CV values of NS-Plus C15 and i-Chroma were higher than the total allowable error, which is defined as ± 15 ng/mL or 15% if the concentration of hemoglobin in the stool sample is higher than 100 ng/mL.

Stability study

The evaluation of the stability analysis, based on the proportion of samples that remained stable on each measurement day, is summarized in Table 4.

In the stability analysis of negative samples, positive results were observed in five samples with the i-Chroma test and in one sample each with the Certest, True Line, and Innovacon tests. Of the five samples that became positive in the i-Chroma test, one yielded a positive result on Day 1 but reverted to negative on Days 2 and 3. Among the remaining four samples, two became positive on Day 1, one on Day 2, and one on Day 3. In the Certest and True Line tests, the sample that became positive on Day 1 reverted to negative on Days 2 and 3. In the Innovacon test, the sample that became positive on Day 1 remained positive on Days 2 and 3. The remaining negative samples maintained their stability across all test brands.

In the stability analysis of positive samples, negative results were observed in three samples with the Certest test, two samples each with the Innovacon, True Line, and Laboquick tests, and one sample each with the i-Chroma, NS-Plus C15, Rapidan Tester, and SD tests. The remaining positive samples maintained their stability across all brands.

Discussion

In this study, the NS-Plus C15 analyzer, with a cut-off value of 100 ng/mL, produced different results compared with the other quantitative and qualitative tests used for comparison, both in terms of agreement levels and the statistical significance of differences (Table 1). The evaluation at a cut-off value of 50 ng/mL (Table 2) showed that the NS-Plus C15 analyzer had better agreement with several tests. As also discussed by Brenner et al. [18] in their comparative study, differences in cut-off values between tests appeared to have influenced the level of agreement observed among them.

However, despite a higher but still moderate level of agreement, statistically significant differences were still observed between the NS-Plus C15 analyzer and some of the other test brands. Therefore, careful evaluation is required when considering the interchangeability of the tests.

Table 3. Inter-test agreement of paired fecal immunochemical tests

	i-Chroma	Certest	SD	Rapidan tester	Laboquick	Innovacon	True line
i-Chroma	OPA	81	81	83	77	82	79
	κ	0.62	0.62*	0.66*	0.56*	0.66*	0.59*
	(95%CI)	(0.53–0.71)	(0.54–0.70)	(0.58–0.74)	(0.48–0.64)	(0.57–0.74)	(0.50–0.67)
Certest	OPA	78	81	74	83	76	76
	κ	0.57*	0.63*	0.52*	0.66*	0.55*	0.55*
	(95%CI)	(0.49–0.65)	(0.55–0.71)	(0.44–0.60)	(0.59–0.74)	(0.47–0.63)	(0.47–0.63)
SD	OPA	91	96	92	92	96	96
	κ	0.81*	0.92	0.84*	0.84*	0.92	0.92
	(95%CI)	(0.74–0.87)	(0.87–0.96)	(0.78–0.90)	(0.78–0.90)	(0.87–0.96)	(0.87–0.96)
Rapidan tester	OPA	88	93	89	89	89	89
	κ	0.75*	0.85	0.77*	0.77*	0.77*	0.77*
	(95%CI)	(0.68–0.82)	(0.79–0.91)	(0.69–0.84)	(0.69–0.84)	(0.69–0.84)	(0.69–0.84)
Laboquick	OPA	91	97	97	97	97	97
	κ	0.81*	0.94	0.94	0.94	0.94	0.94
	(95%CI)	(0.74–0.88)	(0.90–0.98)	(0.90–0.98)	(0.90–0.98)	(0.90–0.98)	(0.90–0.98)
Innovacon	OPA	92	92	92	92	92	92
	κ	0.82*	0.82*	0.82*	0.82*	0.82*	0.82*
	(95%CI)	(0.76–0.89)	(0.76–0.89)	(0.76–0.89)	(0.76–0.89)	(0.76–0.89)	(0.76–0.89)
True line							

All analyses were based on 313 samples; values in parentheses indicate 95% confidence intervals for κ. *: Statistically significant after Bonferroni correction (McNemar's test; adjusted $\alpha=0.0018$). OPA: Overall percent agreement; κ: Kappa coefficient.

Table 4. Stability assessment for fecal immunochemical tests

Tests	Day 1 (%)^{b,c}		Day 2 (%)^{b,c}		Day 3 (%)^{b,c}	
	Negative^a samples	Positive^a samples	Negative^a samples	Positive^a samples	Negative^a samples	Positive^a samples
NS-Plus C15	100	100	100	90	100	90
i-Chroma	70	90	70	90	60	90
Certest	90	90	100	70	100	70
SD	100	100	100	90	100	90
Rapidan tester	100	100	100	90	100	90
Laboquick	100	100	100	100	100	80
Innovacon	90	100	90	90	90	80
True line	90	90	100	80	100	80

^a: Analyses were based on 20 samples, including 10 negative and 10 positive samples; ^b: The percentages in the table show the proportion of samples that remained stable on each measurement day; ^c: Day 0 was defined as the initial measurement day; Day 1–3 refer to measurements performed 1–3 days after Day 0.

Although kappa statistics and McNemar analysis provide information on agreement and statistical differences or bias between tests, they do not directly reflect the clinical impact of discordant results. Therefore, when choosing between tests, it is important to consider not only agreement measures but also analytical performance of the tests, positivity rates and their possible effects on laboratory and clinical management.

The comparisons among the tests shown in Table 3 revealed heterogeneous results in terms of agreement levels and statistical differences. Some test pairs showed high agreement and non-significant McNemar results, suggesting that these

tests may be used interchangeably in routine laboratory practice. However, significant McNemar results in other comparisons, despite moderate or high agreement, indicated the need for more cautious interpretation regarding interchangeability. Therefore, for screening purposes, agreement results should be interpreted together with analytical performance, positivity rates, and laboratory and clinical considerations, as mentioned above.

However, despite the moderate to almost perfect agreement found from pairwise comparisons of the seven tests (one quantitative, six qualitative) with same or different cut-off

values and with same or different measurement methods, the systematic differences found in all tests in the current study, except for five pairwise tests (i.e., i-Chroma vs. Certest, SD vs. True Line, Laboquick vs. True Line, SD vs. Laboquick, Rapidan Tester vs. Innovacon) is an indication to interpret the agreement with caution, and to consider the impact of preanalytical and analytical processes.

In terms of the preanalytical processes, differences in the amount of buffer observed among different tubes of the same test brand during the study may lead to different FOB test results, indicating the need to ensure a standard buffer volume across all tubes. Similarly, the test results may have been impacted by other preanalytical factors, including, but not limited to, the characteristics of the buffer solution that differed among different tests [19]. In addition, differences observed in the amount of stool collected by the sampling stick within the same test brands may also lead to different FOB test results. Among the tests used for comparison, the filter system in the sample collection tubes of the NS-Plus C15, SD, and Certest brands may have a beneficial effect in standardizing the amount of stool collected.

In addition to the preanalytical factors, analytical factors may also be influencing these results. For example, similar positivity rates and the substantial agreement observed in the test pair of Certest and i-Chroma (McNemar's test, $p=0.298$) with different cut-off values and measurement methods, or the systematic differences observed in the test pair Innovacon and True Line (McNemar's test, $p<0.001$) with almost perfect agreement and the same cut-off values and measurement methods, demonstrated that different analytical processes may be involved. This interpretation appears to be aligned with the reasoning of Chiang et al. [20], who compared two FITs with the same cut-off value and reported differences in positive predictive values despite similar positivity rates.

The analytical factors leading to different results may originate from differences in the antibodies used in antigen-antibody binding, for example, antibodies produced from different animal species, targeting different epitopes, or being monoclonal or polyclonal [8, 21]. While the use of monoclonal mouse antibodies to bind hemoglobin in all tests except for NS-Plus C15, Innovacon, and True Line, and the polyclonal antibodies in the NS-Plus C15 analyzer may not be interpreted as the direct factor influencing the different results in tests with the same cut-off values, it still cannot be ruled out. In addition, the form of hemoglobin in the feces plays an important role in antigen-antibody binding. In this respect, the breakdown products of hemoglobin may have interfered with the analytical processes [8, 20].

The repeatability results of this study showed that although the CV values obtained from quantitative analyzers for positive samples appear to be high according to the RCPA, the results resembled the CV values published in repeatability studies using stool samples [8, 22, 23]. In these studies, the CV values calculated from stool samples at various hemoglo-

bin concentrations range from 7.2% to 49.5% and were higher than the CV values (0.6% – 8.5%) obtained from repeatability studies using quality control solutions and hemoglobin solutions prepared in buffer solution [13, 22–26]. The higher CVs observed in fecal samples may be attributed to incomplete homogenization and particle sedimentation, which may occur through interactions with components of the fecal matrix in the buffer [27, 28]. In addition, the obvious CV differences between the two quantitative analyzers in this study, which used samples prepared from the same stool specimens, suggested the potential influence of differences in buffer, sampling, and sample collection tubes [22, 29].

The evaluation of the stability results of the study suggested that the unexpected positivity observed in some negative stool samples may potentially be attributable to changes in the buffer or to substances extracted from the stool into the buffer over time, which may have interfered with test performance [29, 30]. It is also suggested that the increasingly high positive results (>1000 ng/mL) observed in a sample measured by the i-Chroma analyzer from Day 1 onward may indicate a prozone effect [22, 26].

Even though manufacturers reported varying stability periods for hemoglobin in stool samples suspended in buffer, generally not less than three days, this study observed negative results in some previously positive samples from Day 1 onward. Furthermore, studies by van Rossum LG et al. [31], van Roon AH et al. [32], Gies et al. [33], and Guittet et al. [34] have shown that while hemoglobin levels tend to decrease over time, test results may either remain stable or become negative. These findings show the importance of performing a FOB test with fresh stool samples and immediately after taking the specimen for reliable results.

While this study reports important findings from evaluated FITs, it also has some limitations: the specimens prepared in a concentration of 50 ± 10 ng/mL in laboratory conditions were not natural fecal samples and may have unforeseeable effects in depicting the actual performance. In addition, since most of the tests used in the comparison were qualitative, ROC-based normalization could not be performed. Therefore, the variability of the cut-off values might have affected the study outcomes. In addition, kappa statistics are sensitive to the distribution of positive and negative results, and their interpretation may differ according to the outcome distribution. In this study, agreement was evaluated across tests with different positivity rates. Therefore, kappa results should be interpreted in the context of the observed result distribution. Furthermore, the absence of sufficient colonoscopy data did not allow for evaluating the FITs for diagnostic sensitivity and specificity. As a result, it was not possible to provide further clarification on the systematic differences observed between test pairs in pairwise comparisons, for example, whether one test yielded more accurate results or produced disproportionately more positive or negative outcomes than the other.

Conclusion

This study showed that agreement and analytical performance among FITs vary, that statistically significant differences may be observed between some test pairs, and that agreement measures alone are not sufficient when considering test interchangeability.

Although the FITs were not compared with a reference method, the use of multiple complementary agreement measures (kappa, McNemar's test, overall percentage agreement, and positivity rates) supports a more reliable comparison of test performance.

Beyond agreement measures, analytical characteristics that play an important role in the evaluation and selection of FITs, such as consistency, stability, and positivity rates, may influence procurement decisions, routine laboratory practice, and the implementation of population-based screening programs. Analytical variability in a test may compromise the reliability of results and increase laboratory costs and workload by requiring procedures such as repeat testing and calibration checks. It may also lead to errors in patient follow-up and referral processes during the implementation of the screening program, thereby affecting clinical workload, confidence in the screening program, and resource allocation. Therefore, test selection should be based on a comprehensive assessment that considers analytical performance together with agreement results and the potential impact on laboratory and clinical management.

However, further studies assessing the analytical and clinical performance of commercially available test brands are warranted to better determine the reliability and overall performance of the test.

Disclosures

Ethics Committee Approval: The study was approved by the Ankara Numune Training and Research Hospital Ethics Committee (no: 178/2014, date: 07/05/2014).

Informed Consent: Informed consent was obtained from all participants.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Funding: Financial support provided by the Ankara Numune Training And Research Hospital Scientific Research Fund.

Use of AI for Writing Assistance: No AI technologies utilized.

Authorship Contributions: Concept – F.M.Y., N.U.; Design – F.M.Y., N.U.; Supervision- T.T.; Materials – N.U.; Data collection and/or processing – N.U.; Data analysis and/or interpretation – N.U.; Literature search – N.U.; Writing– N.U., S.S.; Critical review – S.S.

Acknowledgments: The authors would like to thank the staff of the biochemistry laboratory at Ankara Numune Training and Research Hospital for their support.

Peer-review: Externally peer-reviewed.

References

- Eser S, Orun H, Hamavioglu E, Lofti F. Cancer incidence, mortality and survival in Türkiye as of 2020. *Basic Clin Cancer Res* 2022;14:1–13. [CrossRef]
- Chen Y, Zhang Y, Yan Y, Han J, Zhang L, Cheng X, et al. Global colorectal cancer screening programs and coverage rate estimation: An evidence synthesis. *J Transl Med* 2025;23:811. [CrossRef]
- Republic of Türkiye Ministry of Health, General Directorate of Public Health, Cancer Department. Türkiye cancer control program 2021. Available at: https://hsgm.saglik.gov.tr/depo/birimler/kanser-db/Dokumanlar/Raporlar/17.Agustos_2021_Kanser_Kontrol_Programi_versiyon-1.pdf. Accessed Oct 1, 2025.
- Young GP, Symonds EL, Allison JE, Cole SR, Fraser CG, Halloran SP, et al. Advances in fecal occult blood tests: The FIT revolution. *Dig Dis Sci* 2015;60:609–22. [CrossRef]
- Tao S, Seiler CM, Ronellenfitsch U, Brenner H. Comparative evaluation of nine faecal immunochemical tests for the detection of colorectal cancer. *Acta Oncol* 2013;52:1667–75. [CrossRef]
- Raginel T, Puvinel J, Ferrand O, Bouvier V, Levillain R, Ruiz A, et al. A population-based comparison of immunochemical fecal occult blood tests for colorectal cancer screening. *Gastroenterology* 2013;144:918–25. [CrossRef]
- Gies A, Cuk K, Schrotz-King P, Brenner H. Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. *Gastroenterology* 2018;154:93–104. [CrossRef]
- Deprez L, Piggott C, van der Hagen EAE, Frasa M, Benton SC; IFCC Working Group on Fecal Immunochemical Testing (WG FIT). Comparison and commutability study among four faecal immunochemical tests (FIT) systems. *Clin Chem Lab Med* 2023;62:50–9. [CrossRef]
- Catomeris P, Baxter NN, Boss SC, Paszat LF, Rabeneck L, Randell E, et al. Effect of temperature and time on fecal hemoglobin stability in 5 fecal immunochemical test methods and one guaiac method. *Arch Pathol Lab Med* 2018;142:75–82. [CrossRef]
- Young GP, Benton SC, Bresalier RS, Chiu HM, Dekker E, Fraser CG, et al. Fecal immunochemical test positivity thresholds: An international survey of population-based screening programs. *Dig Dis Sci* 2025;70:1637–45. [CrossRef]
- Sen B, Okcu O, Arpa M, Akdogan RA. FOBTs with lower detection limits increase the colonoscopy rates. *J Turk Clin Biochem* 2023;21:145–54. [CrossRef]
- Daly JM, Xu Y, Levy BT. Which fecal immunochemical test should I choose? *J Prim Care Community Health* 2017;8:264–77. [CrossRef]
- Randell E, Kennell M, Taher A, Antle S, Bursey F, Tavenor T, et al. Evaluation of hemo techt NS-Plus system for use in a province-wide colorectal cancer screening program. *Clin Biochem* 2013;46:365–8. [CrossRef]
- Lee YH, Hur M, Kim H, Jeon KN, Yun CH, Lee CH, et al. Optimal cut-off concentration for a faecal immunochemical test for haemoglobin by Hemo Techt NS-Plus C15 system for the colorectal cancer screening. *Clin Chem Lab Med* 2015;53:e69–71. [CrossRef]

15. Oono Y, Iriguchi Y, Doi Y, Tomino Y, Kishi D, Oda J, et al. A retrospective study of immunochemical fecal occult blood testing for colorectal cancer detection. *Clin Chim Acta* 2010;411:802–5. [CrossRef]
16. CLSI. User protocol for evaluation of qualitative test performance: Approved guideline-second edition. CLSI document EP12-A2. Wayne: Clinical and Laboratory Standards Institute; 2008.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [CrossRef]
18. Brenner H, Haug U, Hundt S. Inter-test agreement and quantitative cross-validation of immunochromatographical fecal occult blood tests. *Int J Cancer* 2010;127:1643–9. [CrossRef]
19. Fraser CG, Allison JE, Halloran SP, Young GP; Expert Working Group on Faecal Immunochemical Tests for Hemoglobin, Colorectal Cancer Screening Committee, World Endoscopy Organization. A proposal to standardize reporting units for fecal immunochemical tests for hemoglobin. *J Natl Cancer Inst* 2012;104:810–4. [CrossRef]
20. Chiang TH, Chuang SL, Chen SL, Chiu HM, Yen AM, Chiu SY, et al. Difference in performance of fecal immunochemical tests with the same hemoglobin cutoff concentration in a nationwide colorectal cancer screening program. *Gastroenterology* 2014;147:1317–26. [CrossRef]
21. Levin TR. Optimizing colorectal cancer screening by getting FIT right. *Gastroenterology* 2011;141:1551–5. [CrossRef]
22. Guildford Medical Device Evaluation Centre (GMEC). Evaluation of quantitative faecal immunochemical tests for haemoglobin. Available at: https://www.faecal-immunochemical-test.co.uk/wp2/wp-content/uploads/2018/06/FIT_007-Evaluation-of-Quantitative-Faecal-Immunochemical-Tests-for-Haemoglobin.pdf. Accessed Oct 1, 2025.
23. Lamph SA, Bennitt WE, Brannon CR, Halloran SP. Evaluation report: Immunochemical faecal occult blood tests (report no. CEP09042) [archived report]. NHS Purchasing & Supply Agency, Centre for Evidence-based Purchasing; 2009.
24. Piggott C, Carroll MRR, John C, O'Driscoll S, Benton SC. Analytical evaluation of four faecal immunochemistry tests for haemoglobin. *Clin Chem Lab Med* 2020;59:173–8. [CrossRef]
25. Ahn A, Kim J, Ko YJ, Sung H, Kim MN. Performance evaluation of two automated quantitative fecal occult blood tests. *Korean J Lab Med* 2016;6:233–9. [CrossRef]
26. O'Driscoll S, Carroll M, Maclean W, Piggott C, Jourdan I, Benton SC. Assessment of the analytical performance of point-of-care faecal immunochemical tests for haemoglobin. *Ann Clin Biochem* 2021;58:181–9. [CrossRef]
27. Devan SS, Ramli R, Alshehade SA, Lim SYM, Mamat N. Comparison of the techniques for isolating immunoassay-suitable proteins from heterogeneous fecal samples. *Anal Biochem* 2025;698:115748. [CrossRef]
28. Benton SC, Symonds E, Djedovic N, Jones S, Deprez L, Kocna P, et al.; International Federation of Clinical Chemistry Faecal Immunochemical Test Working Group (IFCC FIT-WG). Faecal immunochemical tests for haemoglobin: Analytical challenges and potential solutions. *Clin Chim Acta* 2021;517:60–5. [CrossRef]
29. Rubeca T, Cellai F, Confortini M, Fraser CG, Rapi S. Impact of preanalytical factors on fecal immunochemical tests: need for new strategies in comparison of methods. *Int J Biol Markers*. 2015 Jul 22;30(3):e269-74. [CrossRef]
30. Deda O, Gika HG, Wilson ID, Theodoridis GA. An overview of fecal sample preparation for global metabolic profiling. *J Pharm Biomed Anal*. 2015 Sep 10;113:137-50. [CrossRef]
31. van Rossum LG, van Rijn AF, van Oijen MG, Fockens P, Laheij RJ, Verbeek AL, et al. False negative fecal occult blood tests due to delayed sample return in colorectal cancer screening. *Int J Cancer* 2009;125:746–50. [CrossRef]
32. van Roon AH, Hol L, van Vuuren AJ, Francke J, Ouwendijk M, Heijens A, et al. Are fecal immunochemical test characteristics influenced by sample return time? A population-based colorectal cancer screening trial. *Am J Gastroenterol* 2012;107:99–107. [CrossRef]
33. Gies A, Cuk K, Schrotz-King P, Brenner H. Direct comparison of ten quantitative fecal immunochemical tests for hemoglobin stability in colorectal cancer screening. *Clin Transl Gastroenterol*. 2018 Jul 6;9(7):168. [CrossRef]
34. Guittet L, Guillaume E, Levillain R, Beley P, Tichet J, Lantieri O, et al. Analytical comparison of three quantitative immunochemical fecal occult blood tests for colorectal cancer screening. *Cancer Epidemiol Biomarkers Prev* 2011;20:1492–501. [CrossRef]