Liesbet Deprez*, Carolyn Piggott, Eline A.E. van der Hagen, Marieke Frasa and Sally C. Benton, on behalf of the IFCC working group on Fecal Immunochemical Testing (WG FIT)

Comparison and commutability study among four faecal immunochemical tests (FIT) systems

https://doi.org/10.1515/cclm-2023-0278
Received March 17, 2023; accepted June 9, 2023;
published online June 19, 2023

Abstract

Objectives: Faecal immunochemical tests for haemoglobin (FIT) are used in colorectal cancer screening programs around the world and increasingly for triage of symptomatic patients. FIT results are currently not traceable to a common reference standard and results obtained on various FIT systems may not be equivalent. The size of the bias between the systems is difficult to quantify due to the complex pre-analytical aspects of FIT.

Methods: This study aimed to quantify the bias and the correlation between four FIT systems by measuring a panel of 38 faecal samples while limiting the effect of the pre-analytical aspects. In addition, the commutability of seven candidate reference materials (RM) was assessed.

Results: Pairwise method comparisons based on faecal samples demonstrated Pearson correlation coefficients ranging between 0.944 and 0.970 and an average proportional bias of −30 to −35 % for one FIT system compared to the other three. The relative standard deviation among biases of the individual samples was around 20 %. Due to these sample specific differences, no decisive conclusions could be drawn in the commutability study. However, two candidate RMs, prepared in the FIT system-specific storage/extraction buffers, had a better commutable profile than the other five.

Conclusions: The use of a common threshold for all FIT systems is currently not possible due to the presence of a proportional bias. We have identified potential commutable RMs to take to further studies on the production of a common calibrator, with the aim being to reduce the analytical bias observed on different FIT systems.

Keywords: commutability; faecal immunochemical tests; method comparison; reference materials

***Corresponding author: Dr. Liesbet Deprez**, European Commission, Joint Research Centre (JRC), Retieseweg 111, 2440 Geel, Belgium, Phone: +32 14 57 16 34, E-mail: Liesbet.deprez@ec.europa.eu

Carolyn Piggott, NHS Bowel Cancer Screening Programme, South of England Hub, Guildford, UK. https://orcid.org/0000-0002-6343-6202

Eline A.E. van der Hagen, Department of Clinical Chemistry, Queen Beatrix Hospital, Winterswijk, The Netherlands; and Dutch Foundation for Quality Assessment in Medical Laboratories (SKML), Nijmegen, The Netherlands

Marieke Frasa, Department of Clinical Chemistry, Reinier Haga Medical Diagnostic Center, Delft, The Netherlands

Sally C. Benton, NHS Bowel Cancer Screening Programme, South of England Hub, Guildford, UK; and Clinical Biochemistry, Royal Surrey Foundation Trust, Berkshire and Surrey Pathology Services, Guildford, UK. https://orcid.org/0000-0001-9230-9088

Introduction

Colorectal cancer (CRC) is an important health problem in the Western World. In 2020, more than half a million new cases were diagnosed in Europe and about 245,000 patients died due to CRC [1]. It is the second cause of cancer death despite the expert consensus view that CRC is one of the most preventable cancers [2]. Early detection and treatment of colorectal carcinomas and precancerous lesions, like advanced adenomas, can decrease CRC-specific mortality.

Population CRC screening programs are available in many countries worldwide and faecal immunochemical tests (FIT) are often used as they are non-invasive and inexpensive [2–4]. Individuals with a positive FIT result should proceed to further invasive imaging investigations such as colonoscopy. FIT are based on the binding of antibodies specific for human haemoglobin (Hb) and they are available in qualitative or quantitative formats. For the quantitative FIT, different thresholds can be used to define a positive result. In screening programs the threshold is usually set in the range of 15–150 µg Hb/g faeces (also expressed as f-Hb=15–150 µg/g) with the aim to achieve an optimal balance between sensitivity and selectivity and taking into account the available colonoscopy capacity [4].

FIT are also used outside the scope of population screening programs. In some countries, like the UK, FIT are applied to triage patients with lower gastrointestinal symptoms for further investigation [5]. In addition, a large amount of evidence has been produced in the recent years on the use of FIT for the surveillance of high-risk patients after adenoma resection [6]. For these additional purposes, the threshold is usually set lower than in screening

programs (e.g. f-Hb=10 µg/g) as a higher sensitivity is desirable [6, 7].

A number of different quantitative FIT systems are available on the market both for laboratory use and at the point of care. These FIT systems are not calibrated to a common reference point as reference measurement procedures and/or reference materials (RM) are currently lacking. Studies have reported differences in the diagnostic performance of the various FIT systems when the same threshold is used [8, 9]. Quantification of the size of bias between the f-Hb measurements caused by differences in calibration is difficult due to the complexity of the pre-analytical and selectivity aspects of FIT. The sample collection method of FIT is prone to a lot of pre-analytical variability caused by, for example, heterogeneous distribution and instability of the f-Hb in the sample and variations in sampling technique [10]. In addition, differences in the non-selectivity for the measurand can also contribute to the variability. FIT systems use different monoclonal or polyclonal antibodies highly specific for human Hb, but the affinity for Hb variants and significant degradation products can be different [11].

The lack of result comparability among the various FIT systems limits the transfer of a threshold recommended by an expert group or clinical research studies to screening programmes or to symptomatic services as the established threshold is only valid for a specific FIT system. This is especially important when new FIT systems are introduced into clinical programmes. Enhancement of the result comparability would facilitate the use of common thresholds, reduce health-care costs, improve clinical management and lower the risk of clinical error [10]. In 2017, the International Federation of Clinical Chemistry Working Group on FIT (IFCC WG FIT) was formed with a primary aim being to standardise or harmonise FIT systems [10].

In clinical chemistry, equivalent results among different *in vitro* diagnostic medical devices (IVD-MDs) targeting the same measurand can be achieved by applying the principles of metrological traceability as described in ISO 17511 [12]. A calibration hierarchy needs to be built, in which the assigned values of the calibrators used in the IVD-MDs are linked to higher order measurement standards via an unbroken chain of calibrations. Commutable and certified RM play a key role as they can be used as a common calibrator for the IVD-MDs.

Limitations in achieving comparable results can occur when various IVD-MDs have a different non-selectivity of the measurand or when the RM used as common higher order calibrator lacks commutability [12, 13]. The differences in non-selectivity can be evaluated from the level of correlation obtained by measuring a panel of clinical samples with the IVD-MDs in a split-sample set-up. In the same study, the commutability of a RM can be assessed by comparing the mathematical relationships between the measurement results obtained with various IVD-MDs for the RM and for the clinical samples [14].

As a first step in the process of improving the comparability of the various FIT systems a method comparison study was performed in which four globally used FIT systems with acceptable analytical performance characteristics [15] were compared by measuring 38 faecal samples and at the same time seven candidate RMs were assessed for commutability.

Materials and methods

The preparations of the faecal samples, the dilutions and the reconstitutions of the candidate RM and the FIT measurements were performed by the same laboratory (NHS Bowel Cancer Screening Programme South of England Hub research laboratory, Royal Surrey Foundation Trust, Guildford UK).

Materials

Natural Hb-positive faecal samples (NHF): Fourteen NHF were collected from anonymised residual faecal samples submitted for routine calprotectin measurements with medium to firm consistency. They were shown to be f-Hb positive (>10 µg/g) using the OC-SENSOR PLEDIA (Eiken Chemical Company, Tokyo, Japan). On the day of collection, each sample was sieved through a plastic sieve (1 mm mesh size), homogenised for 5 min using a spatula and immediately loaded into the FIT sampling devices as described below.

Spiked Hb-positive faecal samples (SHF): Twenty-four SHF were prepared from the same sample type as used for the NHF. These samples were shown to be negative for f-Hb (<10 µg/g) with the OC-SENSOR PLEDIA. The selected samples were kept refrigerated for maximum three days or frozen on the day of collection and kept at −20 °C for maximum 16 months. On the day of the preparation, two to five faecal samples were pooled to prepare one SHF and the samples were sieved and homogenised in the same way as the NHF.

For the spiking, four Hb lysates were prepared from anonymised EDTA blood samples with an Hb concentration within the reference range. The Hb concentration in the lysates was measured with the Siemens Advia 2120. The lysates were divided into single-use aliquots and frozen directly after preparation. The spiking was done by adding a calculated volume of the lysate to the faecal sample and homogenising the sample for 5 min using a spatula. Afterwards, the sample was loaded immediately into the FIT sampling devices as described below. The SHF were prepared to reach the predefined target concentrations of f-Hb ensuring coverage of a large part of the measurement range of the FIT systems.

Candidate RM: The characteristics of the seven candidate RM included in this study are summarized in Table 1. The candidate RM A, B and C were prepared in-house from the starting material JCCRM 912-3M

(ReCCS, Yokohama, Japan) which is a certified RM for total Hb in human whole blood with an assigned value of 136.4 g/L. This RM was diluted in the system-specific storage/extraction buffers of the four FIT systems to obtain three different Hb levels. The final Hb concentration in the buffer (ng/mL) was adapted for each FIT system to obtain comparable results in µg/g by taking into account the differences in loading volumes and buffer volumes in the sampling devices. Preliminary studies on dilutions of JCCRM 912 in the storage/extraction buffers showed a good stability for 4 weeks at 3–6 °C (Supplementary data part S1). Due to the COVID-19 crisis, the measurements in this study were spread over a longer period and three sets of candidate RM A-C were prepared.

The candidate RM D and E were FIT EQA samples prepared by the Dutch Foundation for Quality Assessment in Medical Laboratories (SKML, Nijmegen, The Netherlands). These materials consist of lyophilized Hb-spiked faecal extract and two different levels were included in this study. According to the manufacturer of the material, the extraction buffer was chosen as best fit to accommodate usage on the present instruments. These materials were reconstituted in water according to the instructions and measured within 7 h. For each measurement batch, a new vial from the same lot was used.

The candidate RM F and G were FIT EQA samples prepared by Qualimedlab (Genova, Italy). These materials consist of a liquid refrigerated Hb solution but no additional information about the buffer was available. Two different levels were included in this study and the materials were used according to the manufacturer's instructions.

As the candidate RM D–G were EQA samples dedicated for FIT, their buffers and concentrations were considered to be suitable for the four FIT systems without further dilution in the FIT systems specific storage/extraction buffers. Normalisation factors were used for the conversion of the obtained measurement results in ng/mL to µg/g values, taking into account the differences in loading volumes and buffer volumes in the sampling devices.

FIT systems and measurements

FIT systems: The four FIT systems are summarized in Table 2.

Loading of the sampling devices: Each NHF and SHF was loaded into five sampling devices for each FIT system (i.e. 20 devices in total) by experienced laboratory staff. Unpublished results from previous in-house studies showed that Hb degradation can occur during the loading of the devices and the level of degradation is variable between samples. Therefore, the total loading time of the 20 sampling devices was limited as much as possible and recorded: for the NHF time ranged from 9 to 19 min (mean 12 min) and for the SHF time ranged from 7 to 15 min (mean 10 min). In addition, the sampling devices were loaded in a predefined order eliminating a correlation between the loading time and the specific FIT system. A potential degradation could therefore be differentiated from a bias between the FIT systems (Supplementary data part S2).

After loading, the sampling devices were inverted five times and stored at room temperature in the dark (17–20 °C) for 22–28 h. Afterwards, the devices were inverted five times again and moved to the

Table 1: Specification of the seven candidate reference materials used in this study.

Candidate RM	Content	Expected Hb concentration in buffer	Corresponding f-Hb concentration in faecal samples	Form	Manufacturer product code	Lot no
A	JCCRM 912-3M (Human whole blood) diluted into relevant FIT buffers	For HMJ: 15 ng/mL For NSP/OC: 75 ng/mL For SF: 63.7 ng/mL (assigned value and in-house dilution steps)	For HMJ: 15 µg/g For NSP/OC: 15 µg/g For SF: 10.8 µg/g	Liquid refrigerated for max 4 weeks	In house	N.A.
B	JCCRM 912-3M (Human whole blood) diluted into relevant FIT buffers	For HMJ: 50 ng/mL For NSP/OC: 250 ng/mL For SF: 212.5 ng/mL (assigned value and in-house dilution steps)	For HMJ: 50 µg/g For NSP/OC: 50 µg/g For SF: 36.1 µg/g	Liquid refrigerated for max 4 weeks	In house	N.A.
C	JCCRM 912-3M (Human whole blood) diluted into relevant FIT buffers	For HMJ: 150 ng/mL For NSP/OC: 750 ng/mL For SF: 637.5 ng/mL (assigned value and in-house dilution steps)	For HMJ: 150 µg/g For NSP/OC: 150 µg/g For SF: 108.4 µg/g	Liquid refrigerated for max 4 weeks	In house	N.A.
D	Extract from Hb spiked faecal sample	80 ng/mL (information provided by manufacturer)	For HMJ: 16[b] µg/g For NSP/OC: 16 µg/g For SF: 16[b] µg/g	Lyophilized	SKML	2020.0661 Low
E	Extract from Hb spiked a faecal sample	188 ng/mL (information provided by manufacturer)	For HMJ: 37.6[b] µg/g For NSP/OC: 37.6 µg/g For SF: 37.6[b] µg/g	Lyophilized	SKML	2020.0662 High
F	Hb in buffer[a]	~125–130 ng/mL (in-house measurements with Sentifit 270)	For HMJ: 25–26[b] µg/g For NSP/OC: 25–26 µg/g For SF: 25–26[b] µg/g	Liquid refrigerated	Qualimedlab QML2a	FB_8_20
G	Hb in buffer[a]	~250–260 ng/mL (in-house measurements with Sentifit 270)	For HMJ: 50–52[b] µg/g For NSP/OC: 50–52 µg/g For SF: 50–52[b] µg/g	Liquid refrigerated	Qualimedlab QML2b	FB_8_20

[a]No additional information was disclosed by the material provider. [b]These values were normalized to correct for difference in faecal sample volume and buffer volume between the different FIT systems. N.A., not applicable as prepared in house.

Table 2: The four FIT systems used in this study for the measurement of the faecal samples and the candidate reference materials.

System code	Manufacturer	Sampling device	Analyser	f-Hb measurement range in faecal samples, µg/g
HMJ	Minaris Medical Co. Ltd. (Tokyo, Japan)	EXTEL HEMO AUTO-MC Collection Picker	HM-JACKarc	7–400
NSP	Alfresa Pharma Corp (Osaka, Japan)	Specimen Collection Container A	NS-Prime	10–240
OC	Eiken Chemical Co. Ltd (Tokio, Japan)	OC-Auto Sampling Bottle 3	OC-SENSOR PLEDIA	10–200
SF	Sentinel Diagnostics (Milan, Italy)	SENTiFIT pierceTube	SENTiFIT 270 (FOB Gold Wide Method)	3–170

refrigerator (3–6 °C) for 20–212 h. The total incubation time ranged from 44 to 236 h, which is within the stability period stated by the manufacturers (shortest period: 28 days). The Hb stability in the storage/extraction buffer present in the sampling devices was confirmed by in-house experiments before the start of this study (Supplementary data part S1) and published data [16–18].

FIT analysis: The analyses were spread over eight different days and within one measurement day one group of faecal samples (range: 3–8 samples) were analysed with the four different FIT systems. The sampling devices were brought to room temperature and mixed by inversion at least 1 h before analysis. In each measurement batch the seven candidate RM were measured in four replicates at the beginning and the end.

For some SHF the measured f-Hb result was above the analytical range of a FIT system (n=1 for NSP and OC and n=2 for SF) and the auto-dilution functions of the analysers were used.

Data analysis

The data analysis was performed using Microsoft Excel and the add-in software Analyse-it.

First, the variation among the replicate measurement results were evaluated for both the faecal samples and the candidate RM. For the faecal samples, a regression analysis was performed between the obtained result and the loading sequence of the sampling devices to check for Hb degradation. The variation between eight measurement batches was also checked using the results of candidate RM D–G, which were measured in four replicates at the beginning and end of each batch.

The mathematical relationships between the different FIT systems were evaluated for the NHF and the SHF separately to confirm that both samples types showed equivalent relationships. This evaluation was based on a Deming regression analysis on the ln-transformed measurement results obtained for the NHF only. The mathematical relationships obtained with the SHF were considered equivalent if their data points were within the 95 % prediction interval around the Deming regression line. For the pairwise system comparisons, the non-parametric method of Passing and Bablok was used on the measurements results of both NHF and SHF to obtain the best fitting linear regression and the associated confidence interval (CI). In addition, the Pearson correlation coefficient was calculated. The presence of sample-specific differences was evaluated using the statistical analysis described in the recommendations from the IFCC WG on Commutability in Metrological Traceability [19] (Supplementary data part S3). The standard deviation caused by sample-specific differences only (i.e. s_{ssd}) was calculated based on the ln-transformed data.

The statistical analysis of the commutability of the seven candidate RMs was based on the difference in bias approach [19] with ln-transformed concentrations to obtain a constant bias over the whole measurement range. The commutability criterion was set at 25 %, which means that the maximum allowable difference between the bias of the RM and the average bias of the faecal samples is 25 %. This commutability criterion corresponds to ~2 times the largest expanded uncertainty associated with the difference in bias in this study.

Results

Initial evaluation of results: Hb degradation, repeatability and run-to-run variations

Some NHF had results below the measurement range of a specific FIT system (HMJ and SF n=3, NSP and OC n=1) and these results were excluded from further evaluations. For the faecal samples, the correlation between the obtained result and the loading order on the sampling devices was evaluated (Supplementary data part S4). Based on the results it was decided to remove the results of the last sampled FIT device of each system from further evaluations. The results of the remaining sampling devices (4 for each system, 16 in total) showed only a small significant decline in measured f-Hb concentrations of the SHF which was similar for all systems (slope ranging from −0.035 to −0.049). For the NHF, there was only significant decline for the OC system (slope −0.047). The effect of this degradation on the outcome of the study was considered to be minor. The repeatability among the measured f-Hb concentrations in the four sampling devices of the same FIT system ranged from 7.2 to 14.7 % while the average repeatability among the four consecutive measurements of the liquid candidate RM ranged from 1.8 to 7.6 % (Supplementary data Table S3).

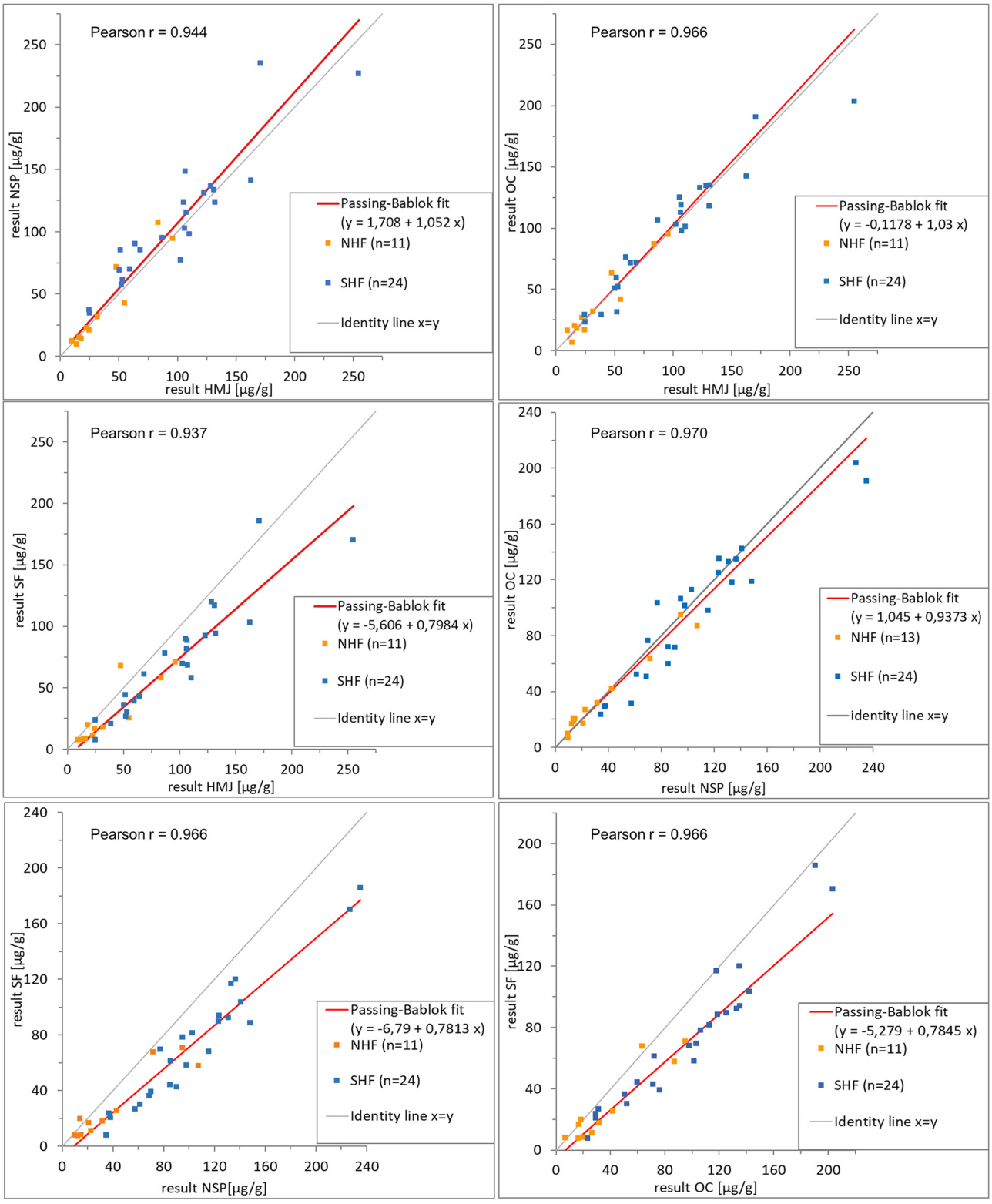

Figure 1: Pairwise correlation plots between the four FIT systems. The measurement results obtained for 35–37 faecal samples (natural Hb-positive faecal samples [NHF] and spiked Hb-positive faecal samples [SHF] combined). The non-parametric method of Passing and Bablok was used to calculate the best fitting linear regression and the Pearson correlation coefficient r to evaluate the correlation.

The variation between the eight measurement batches was also checked and corrected for using a correction factor (range: 0.91–1.12) based on the results obtained for the RM D–G.

Pairwise comparison of the FIT systems

The evaluation of the mathematical relationships between the different FIT systems for the NHF and the SHF separately confirmed that both samples types showed equivalent relationships (Supplementary data part S5). Therefore, the measurement results of both NHF and SHF were combined for the method comparison using the Passing Bablok regression analysis (Figure 1). The analysis showed that for the four pair wise comparisons among the systems HMJ, NSP and OC the slopes of the regression lines were not significantly different from 1 and the intercepts were not significantly different from 0 (Table 3). For the three pair wise comparisons including the SF system the obtained Passing Bablok regression lines had a slope below 1 (CI 95 %) indicating the presence of a proportional bias. The intercepts were also negative. The bias plots based on the ln-transformed results confirmed a negative bias with the average bias ranging between −0.36 and −0.43 (Table 4) which corresponds to relative bias of −30 to −35 % on the non-transformed concentrations.

The correlation among the systems was assessed by calculation of the Pearson r and the values ranged between 0.937 and 0.970 (Figure 1). The sample specific differences were evaluated by calculating the standard deviation s_{ssd} on the ln-transformed concentrations (Table 4). For each system comparison, s_{ssd} was significant ($p<0.01$) and ranged between 0.19 and 0.28 corresponding to relative standard deviation of 17–25 % on the non-transformed concentrations.

Table 4: Evaluation of the bias and the samples specific differences among the FIT systems using bias plots based on the ln-transformed results obtained for 35–37 faecal samples (natural Hb-positive faecal samples [NHF] and spiked Hb-positive faecal samples [SHF] combined).

Average bias on ln-transformed results ± expanded uncertainty[a]		**FIT system**		
Standard deviation due to sample specific differences (s_{ssd})[b]		**NSP**	**OC**	**SF**
FIT system	HMJ	0.08 ± 0.08 0.19	0.00 ± 0.08 0.27	−0.36 ± 0.10 0.28
	NSP		−0.06 ± 0.08 0.20	−0.43 ± 0.11 0.28
	OC			−0.36 ± 0.10 0.25

[a]The expanded uncertainty on the average bias is calculated as the standard deviations of de biases divided by the square root of the number on faecal samples. [b]The s_{ssd} is the standard deviation among the faecal samples which is caused be samples specific effects alone, excluding the variations caused by the repeatability of the measurements. The equations used to calculate this s_{ssd} are described in the Supplementary data part S3.

Commutability assessment of the candidate RM

Commutability assessment was based on bias plots on the ln-transformed concentrations of the faecal samples and the candidate RMs (Figure 2). The results of the commutability assessment of each candidate RM are summarized in Figure 3. The candidate RM A, D, E, F and G are non-commutable for some combinations for FIT systems. RM A, RM F and RM G are not commutable for two or three pair-wise combinations including the HMJ. The RM D and E show non-commutability for two combinations including the

Table 3: Slope and intercepts of the Passing and Bablok regression lines obtained for the pair wise comparison of the FIT systems based on the measurement results of 35–37 faecal samples (natural Hb-positive faecal samples [NHF] and spiked Hb-positive faecal samples [SHF] combined).

Slope (95 % CI)		**FIT system**		
Intercept (95 % CI)		**NSP**	**OC**	**SF**
FIT system	HMJ	1.05 (0.91; 1.26) 1.71 (−6.72; 12.74)	1.03 (0.90; 1.12) −0.12 (−5.64; 6.26)	0.80 (0.69; 0.95) −5.61 (−14.51; 0.25)
	NSP		0.94 (0.83;1.03) 1.05 (−6.73; 5.92)	0.78 (0.68; 0.88) −6.79 (−16.43; −1.27)
	OC			0.78 (0.71; 0.91) −5.28 (−12.19; 0.67)

CI, confidence interval.

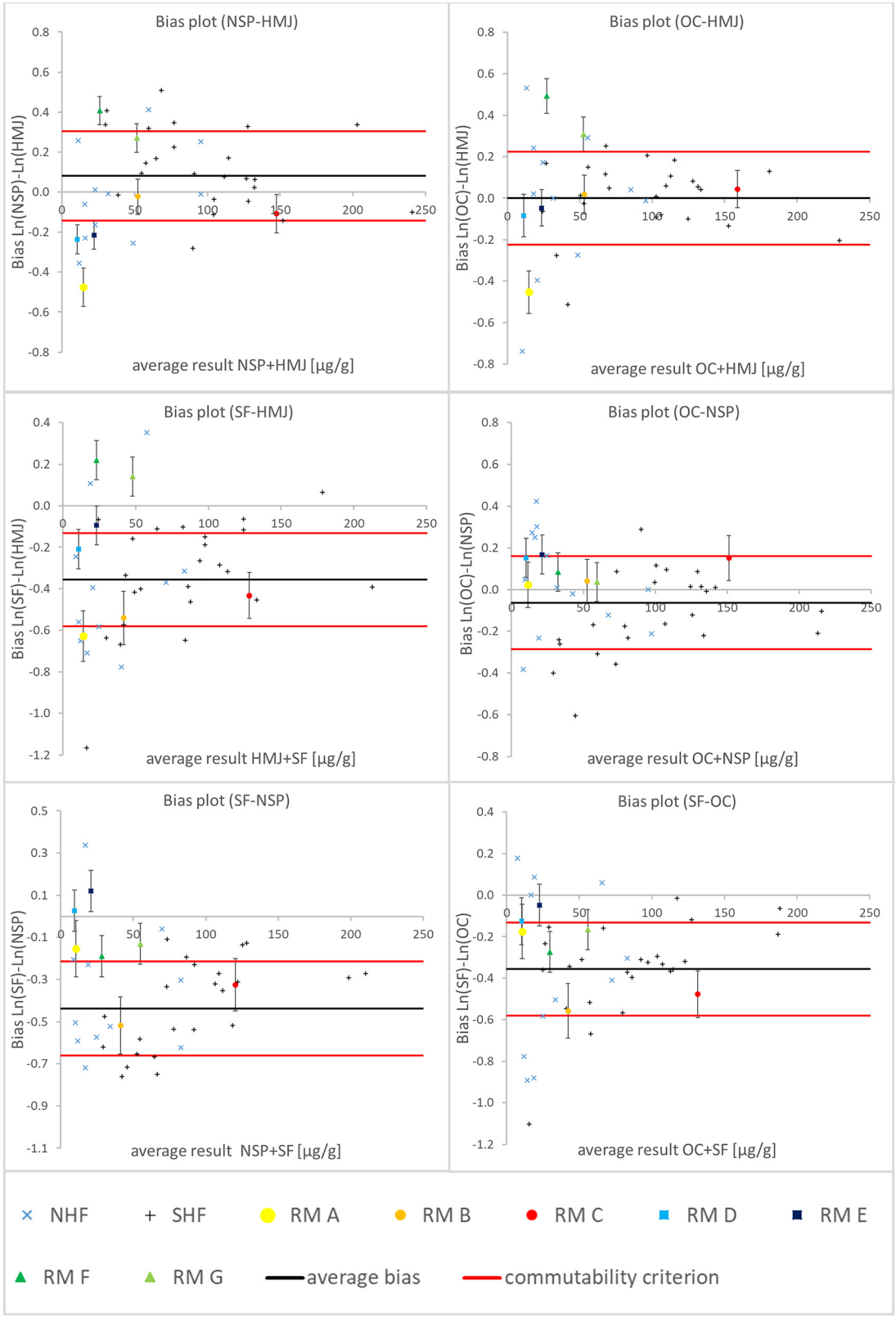

Figure 2: Commutability assessment of seven candidate reference materials for FIT according to a difference in bias approach. This assessment was based on the results of 35–37 faecal samples (natural Hb-positive faecal samples [NHF] and spiked Hb-positive faecal samples [SHF] combined) and seven candidate RM. The relative bias was calculated as the difference between the ln-transformed mean concentration measured with the one FIT system and the ln-transformed mean concentration measured with the other FIT system. The error bars on the bias of individual candidate RM represent the expanded uncertainty associated with the estimated difference in bias.

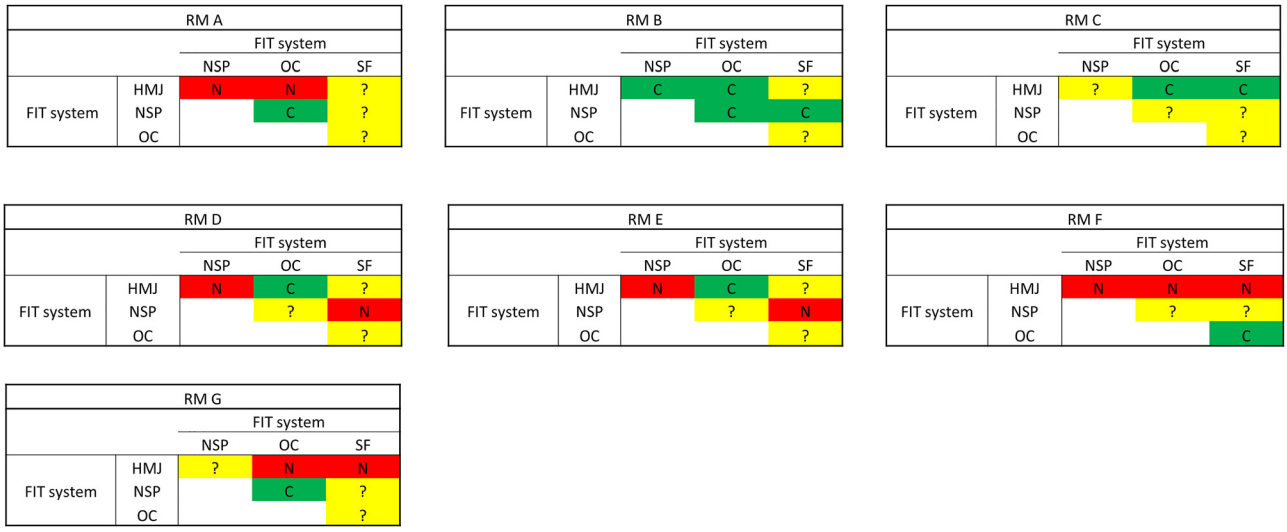

Figure 3: Outcome of the commutability assessment of seven candidate reference materials for FIT according to a difference in bias approach. Commutability of candidate RM was assessed according to the difference in bias analysis (as described in [19]) between the FIT systems. Commutability letter code, C, commutable; N, non-commutable; ?, inconclusive.

NSP. For the candidate RM B and C the outcome of the commutability study is commutable or inconclusive with the data point within the range of the commutability criterion for all six combinations.

Discussion

Setting up a robust comparison/commutability study for FIT systems is very challenging. A large number of pre-examination variables can affect the FIT result, including the sampling technique, the consistency of the faecal sample, the uneven distribution of f-Hb in the faecal sample, and the stability of f-Hb [10]. The design of this study was focused on limiting the effects of the pre-examination variables as much as possible. The repeatability among the four replicate analyses (including the sampling and the f-Hb measurement) of the faecal samples ranged between 7.2 and 14.7 % for the four FIT systems. This repeatability is lower than the repeatability reported for artificial faecal-like EQA materials [20].

The study shows that for the four FIT systems evaluated, the SF showed a proportional bias of −30 to −35 % compared with the other three FIT systems. As there is currently no reference point available for f-Hb, it is not possible to evaluate the real bias of each FIT system. The bias can have an impact on the diagnostic sensitivity in situations where a predefined general threshold is applied for example the triage of patients with lower gastrointestinal symptoms in the UK (f-Hb=10 µg/g). The results of a previous study, in which 233 patients sampled the same bowel movement with four FIT systems, also indicated a lower referral rate for this specific FIT system at a threshold of 10 µg/g [9].

The pair-wise comparisons of the four FIT systems showed correlation coefficients ranged between 0.944 and 0.970. In clinical chemistry, a correlation coefficient >0.975 is often considered as an acceptable correlation between IVD-MDs [21]. The correlation between FIT systems obtained in this study is therefore suboptimal and quite large sample specific effects have been identified. The relative standard deviation related to the sample specific effects alone (excluding the effect of the repeatability) is estimated to be around 20 %. These sample specific effects could be linked to non-selectivity differences due to the use of different antibodies. The antibodies can be derived from different species (e.g., goat, mouse, chicken), and raised against different epitopes in the target protein. Therefore, these antibodies might react differently against degraded forms of Hb and/or interfering substances present in the faecal samples. The faecal matrix background could also contribute to the sample specific effects as the composition and pH is much more variable for faecal samples than for blood or serum samples. The observed sample-specific effects had a negative effect on the commutability study by increasing the uncertainties associated with the commutability assessments.

The seven different candidate RM tested in this study can be grouped into three different types. RM A–C consists of human Hb spiked in the specific storage/extraction buffers of each FIT system. RM D–E are lyophilized extracts from human faecal samples spiked with human Hb and the same extraction buffer was used for all FIT systems. RM F–G are

liquid solutions of human Hb in a buffer that was also identical for all FIT systems. The candidate RM D, E, F and G are non-commutable for some combinations for FIT systems. The commutability profile of candidate RM B and C is better, however, the outcome is also inconclusive for some system combinations. RM A, which is produced in the same way as RM B and RM C but at a lower concentration, shows non-commutability in two pair-wise system combinations. This could be due to the lower concentration level of RM A which is close to the lower end of the measurement range of the FIT systems and the bias could be different in this region. This study is only a preliminary commutability study but the results suggest that the specific storage/extraction buffer of the FIT systems can have an impact on the measurement result and that should be taken into account during the future studies on the development of a commutable common calibrator for FIT systems.

In conclusion, the results of the study show that there is a significant proportional bias among the four FIT systems evaluated. The availability of a commutable certified RM could help to reduce this bias. The future RM could be used as common calibrator to recalculate the measurement results of the different FIT systems and to obtain more harmonized results compared to what is currently available. Further work should continue to assess the possibility of producing this commutable RM. The large sample specific differences seen in this study could also be investigated further as they can remain as unpredictable biased results on the individual samples after re-calibration. This might have a negative impact on the analytical performance goals that can be achieved for the FIT systems.

Acknowledgements: We are thankful to Shane O' Driscoll for his support to the preparation of the faecal samples and the measurements. We thank also Berkshire and Surrey Pathology Services for supplying the lysate and faecal samples. We acknowledge FIT system suppliers HM-JACKarc: Alpha Laboratories, Eastleigh, Hampshire, UK; OC-SENSOR PLEDIA: Mast Diagnostics; Division, Bootle, Merseyside, UK; SENTiFIT 270: Sysmex UK Ltd, Milton; Keynes, Buckinghamshire, UK; NS-Prime: Alfresa Pharma Corp., Osaka, Japan for supplying the analysers and consumables and the SKML and Qualimedlab for providing their quality control materials.

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References

1. European Cancer Information System (ECIS). https://ecis.jrc.ec.europa.eu [Accessed 16 Mar 2023].
2. Young GP, Rabeneck L, Winawer SJ. The global paradigm shift in screening for colorectal cancer. Gastroenterology 2019;156:843–51.
3. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015;64:1637–49.
4. Lauby-Secretan B, Vilahur N, Bianchini F, Guha N, Straif K, International Agency for Research on Cancer Handbook Working G. The IARC perspective on colorectal cancer screening. N Engl J Med 2018;378:1734–40.
5. National Institute for Health and Care Excellence. Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care. Diagnostics guidance DG30. https://www.nice.org.uk/guidance/dg30 [Accessed 16 Mar 2023].
6. Pin-Vieito N, Puga M, Fernandez-de-Castro D, Cubiella J. Faecal immunochemical test outside colorectal cancer screening? World J Gastroenterol 2021;27:6415–29.
7. Westwood M, Lang S, Armstrong N, van Turenhout S, Cubiella J, Stirk L, et al. Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. BMC Med 2017;15:189.
8. Gies A, Niedermaier T, Alwers E, Hielscher T, Weigl K, Heisser T, et al. Consistent major differences in sex- and age-specific diagnostic performance among nine faecal immunochemical tests used for colorectal cancer screening. Cancers (Basel) 2021;13:3574.
9. Benton SC, Piggott C, Zahoor Z, O'Driscoll S, Fraser CG, D'Souza N, et al. A comparison of the faecal haemoglobin concentrations and diagnostic accuracy in patients suspected with colorectal cancer and serious bowel disease as reported on four different faecal immunochemical test systems. Clin Chem Lab Med 2022;60:1278–86.
10. Benton SC, Symonds E, Djedovic N, Jones S, Deprez L, Kocna P, et al. Faecal immunochemical tests for haemoglobin: analytical challenges and potential solutions. Clin Chim Acta 2021;517:60–5.
11. Carroll MR, John C, Mantio D, Djedovic NK, Benton SC. An assessment of the effect of haemoglobin variants on detection by faecal immunochemical tests. Ann Clin Biochem 2018;55:706–9.
12. ISO. 17511: In vitro diagnostic medical devices – requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples, 2nd ed. Geneva: International Organisation for Standardisation (ISO); 2020.
13. Schimmel H, Zegers I. Performance criteria for reference measurement procedures and reference materials. Clin Chem Lab Med 2015;53:899–904.
14. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC working group recommendations for assessing commutability part 1: general experimental design. Clin Chem 2018;64:447–54.
15. 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.
16. 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;9:168.

17. Fujimura K, Kurita Y, Nishimura E, Yamaguchi J, Mori S, Satho H. Comparison study of automated immunochemical fecal occult blood test analyser performance. J Clin Lab Instrum Reagents 2013;36: 679–85.
18. Gnatta E, Zaninotto M, Epifani MG, Padoan A, Gjini R, Plebani M. A new sampling device for faecal immunochemical testing: haemoglobin stability is still an open issue. Clin Chem Lab Med 2014;52:1203–9.
19. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC working group recommendations for assessing commutability part 2: using the difference in bias between a reference material and clinical samples. Clin Chem 2018;64:455–64.
20. O'Driscoll SP, Piggott C, Bruce H, Benton SC. An evaluation of ten external quality assurance scheme (EQAS) materials for the faecal immunochemical test (FIT) for haemoglobin. Clin Chem Lab Med 2020; 59:307–13.
21. CLSI. EP09 measurement procedure comparison and bias estimation using patient samples, 3rd ed. Wayne, PA, USA: Clinical and Laboratory Standards Institute; 2018.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2023-0278).