Clinica Chimica Acta 561 (2024) 119809

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Modelling the impact of bias in fecal immunochemical testing on long-term outcomes of colorectal cancer screening

Lucie de Jonge [a,*], Esther Toes-Zoutendijk [a], Brechtje D.M. Koopmann [a], Marith van Schrojenstein Lantman [b,c], Brenda Franken-van Vorsselen [c], Christel Speijers [c], Huub van Ingen [d], Erwin Humer [e], Petra van der Groep [e], Marc Thelen [b,c], Iris Lansdorp-Vogelaar [a], on behalf of the Dutch National Colorectal Cancer Screening Working Group

[a] Department of Public Health, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[b] Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands
[c] Stichting kwaliteitsbewaking Medische Laboratoriumdiagnostiek (SKML), Nijmegen, The Netherlands
[d] Star-shl, Etten-Leur, The Netherlands
[e] Health Screening Organisation Bevolkingsonderzoek Nederland, The Netherlands

ARTICLE INFO

Keywords:
CRC screening
FIT
Allowable performance specifications
Microsimulation modelling

ABSTRACT

Background: As the impact of unmanaged bias (i.e. systematic source of inaccuracy) in fecal immunochemical test (FIT) analytical performance on long-term colorectal cancer (CRC) outcomes is unknown, we assessed the impact bias in FIT performance in an ongoing FIT-based CRC screening program.
Methods: This study consisted of two parts: cross-sectional observational data analysis to estimate change in short-term outcomes and microsimulation modelling to estimate change in long-term outcomes assuming different levels of bias by assuming 15 % lower up to 15 % higher Hemoglobin detected in the stool compared to observed. Two scenarios were considered: bias occurring 1) one-time only, due to the occasional bias associated with the FIT kits used in 2020 and 2) consistently due to a constant bias associated with the FIT kits used from 2020 onwards.
Results: With a hypothetical bias of −15 % to +15 %, we observed a positivity rate ranging from 6.7 % to 7.8 %, and a detection rate for CRC between 0.65 % and 0.68 %. Single biases in FIT performance resulted in less than 0.1 % change in long-term CRC screening outcomes, while consistent biases resulted in a much larger change (up to 1.4 % in CRC cases and CRC-related deaths and up to 2.07 % in total costs). Detecting lower Hemoglobin concentrations resulted in a relatively larger change on long-term CRC outcomes in comparison to positive bias.
Conclusions: Because of the substantial impact of consistent FIT bias, it is important to set evidence-based acceptance criteria of bias on long-term CRC screening outcomes and in particular, the introduction of an asymmetrical or upward shifted tolerance interval for FIT bias.

1. Introduction

Many countries have implemented population-based colorectal cancer (CRC) screening programs using fecal immunochemical testing (FIT), to reduce CRC incidence and mortality [1]. FIT is a stool-based test that allows participants to collect their stool at home using a collection device that preserves a standardized amount of stool in stabilizing buffer and return it by post or through their GP. These kits are then analyzed in laboratories dedicated in CRC screening. A predefined positivity cut-off level determines whether the FIT result is positive (i.e. the hemoglobin (Hb) level is equal to or greater than the cut-off) or negative (i.e. the Hb level is less than the cut-off).

Like all methods in laboratory medicine, FIT testing is subject to laboratory error which consists of random imprecision components and systematic bias components. Both imprecision and bias are recognized as individual analytical performance specifications (APS). However, if the source of bias is accepted, the resulting acceptable bias becomes part of the long-term imprecision, which can be used as measurement uncertainty [2]. An important potential source of unacceptable bias is difference in measurement result caused by changes in reagent/

* Corresponding author at: Department of Public Health, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail address: l.dejonge.3@erasmusmc.nl (L. de Jonge).

https://doi.org/10.1016/j.cca.2024.119809
Received 4 June 2024; Accepted 8 June 2024
Available online 13 June 2024
0009-8981/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

calibrator lot [3]. To determine the level of APS, an internationally widely used guideline has formulated different models, known as the EFLM Milan models [4]. The most preferable rationale to determine an APS is medical outcome, with biological variation and state-of-the-art achievability as alternatives. Currently, the APS for between reagent/calibrator lot variation in the Dutch organized CRC screening program are set at ±7.5 %. These specifications are formulated by experts and based on state-of-the-art achievability [4]. However, systematic variation due to a lot change can be monitored and managed (i.e. by rejecting lots that do not pass acceptance criteria).

A previous study showed variability of the FIT over time on short-term outcomes such as positivity rate and CRC yield, but considered clinically irrelevant [5]. As information on long-term outcomes is lacking [2], we assessed the long-term impact of bias in FIT testing in an ongoing FIT-based CRC screening program to create input that allows formulating APS for allowable between lot variation of FIT testing. Only knowledge of the resulting misclassification rate can determine whether an ambition to reduce the bias is worth the effort.

2. Methods

This study consisted of two parts: cross-sectional observational analysis to estimate the change in short-term outcomes and microsimulation to estimate change in long-term outcomes.

2.1. Data

2.1.1. Dutch population-based CRC screening program

The Dutch national CRC screening program was implemented from 2014 to 2019 with a gradual roll-out by birth cohort. The program uses the Sentinel FOB-Gold method with the Sentinel FOB-Gold-W-calibrator set on Biomajesty analysers from BioSys. The program is supervised by the National Institute of Public Health and Environment and is carried out in four laboratories. These laboratories use identical protocols, instruments, and reagents. National coordination ensures that changes in reagent and calibrator lots are synchronized after verifying their impact on the FIT results within pre-defined tolerances.

Individuals aged between 55 and 75 years were invited biennially to perform FIT using a positivity cut-off of 15 μg Hb per gram (μg Hb/g) feces. After the first half of 2014, the positivity cut-off was increased to 47 μg Hb/g feces because of a higher-than-expected participation rate, positivity rate and lower-than-expected positive predictive values [6]. The initial cut-off was chosen based on a pilot study using a different method, and the lack of standardization between individual FIT methods can explain the need for such large changes in cut-off [7]. In case of a positive FIT, individuals are referred for follow-up colonoscopy. Participants with advanced adenoma (AA) or CRC during colonoscopy, received further treatment and surveillance according to the Dutch surveillance guidelines [8].

2.1.2. Levels of FIT bias and short-term outcomes

Ten different levels of biases ranging between negative and positive bias of 15 % (which is the current APS of plus minus 7.5 %, which would theoretically allow for an occasional or persistent change of 15 % (from plus to minus 7.5 %)) were evaluated:

Measuring fHb concentrations

- −2.5 %, −5.0 %, −7.5 %, −10.0 % and −15.0 %, and
- +2.5 %, +5.0 %, +7.5 %, +10.0 % and +15.0 %

compared to a theoretical situation without any hypothetical level of bias in FIT measurements, i.e. 0 % (comparator). Biases were relative to a standard and unbiased FIT using a positivity cut-off of 47 μg Hb/g feces. A negative/positive FIT bias would mean that the FIT detects x% less/more Hb concentration in an individual's feces.

To assess the impact of FIT biases on short-term CRC screening outcomes, we used data of participants in 2014 in the Dutch CRC screening program obtained from the national screening database (ScreenIT). Participants who opted out for scientific research were not included in this analysis. In our main analysis, we used data from the first half of 2014 in which the lower FIT positivity cut-off (15 μg Hb/g) was used. That allowed us to also evaluate the colonoscopy findings of participants with FIT values below the current cut-off of 47 μg Hb/g feces. We calculated FIT positivity rate and detection rates for non-advanced adenomas (NAA), AA and CRC at every level of bias.

2.2. Microsimulation modelling

2.2.1. MISCAN-Colon model description

The MISCAN-Colon model is a well-established microsimulation model for CRC developed at the Department of Public Health, Erasmus MC University Medical Center and has been extensively described previously [9,10]. In brief, the model simulates the life-histories of a large population of individuals from birth to death. In addition, the model simulates the development of CRC through the adenoma-carcinoma sequence. As each simulated individual ages, one or more adenomas may develop and these adenomas can progress in size from small (≤5 mm) to medium (6–9 mm) to large (≥10 mm). Some adenomas can develop into preclinical cancer, which may progress through cancer stages I to IV. Cancer stages correspond to the American Joint Committee on Cancer/International Union Against Cancer staging system for CRC. At any time during the development of the disease, symptoms may present and CRC may be diagnosed. By introducing screening, the simulated life-histories may be altered through detection and removal of adenomas or CRC at an earlier stage with a more favorable prognosis. By comparing the life-histories of a simulated population undergoing screening to the corresponding life-histories without screening, MISCAN-Colon is able to quantify the effectiveness and costs of screening.

2.2.2. Modelling scenarios

First, we simulated a scenario in which we modelled the Dutch national CRC screening program including the gradual rollout without any hypothetical bias (comparator).

Next, we simulated CRC screening at each level of FIT bias. These biases were applied in two ways:

- One-time occurrence in 2020 (single), for example a single lot-to-lot variation in 2020,
- Consistently at every invitation from 2020 onwards (consistent), for example the vendor of the FIT kits has decided to recalibrate its reagents lots and therefore there exists a bias from 2020 onwards compared to before.

For all scenarios, we simulated a population of 100 individuals born between 1934 and 2013 from birth to death to allow for robust estimates in outcomes. In the 'single' scenario, we only simulated those birth cohorts eligible for screening in 2020. In the 'consistent' scenario, we simulated all birth cohorts eligible for screening from 2020 onwards.

2.2.3. Estimating FIT test characteristics per level of bias

Variability in FIT performance due to bias was translated into lower and higher FIT sensitivities and specificities. Test characteristics for different bias were estimated such that the model predicted positivity and detection rates for NAA, AA and CRC were similar to those observed in the Dutch CRC screening program.

2.2.4. Costs of screening, surveillance and CRC care

The costs of FIT were provided by Dutch National Institute for Public Health and Environment (Table S1). These costs include the test kits, their distribution, return, analysis, and marketing expenses. Costs for colonoscopy, polypectomy and its complications as well as costs for

cancer care were based on retrospective chart reviews (Table S1). We estimated average utilization of CRC patient of specific healthcare products within the Diagnosis and Treatment Combinations system in the Netherlands [11]. This was then multiplied by the average price of all hospitals in the Netherlands for these services based on reimbursement. Estimated future costs were discounted using a 3 % annual rate [12].

2.2.5. Modelling analysis of bias and long-term outcomes

The MISCAN-Colon model for the Dutch population was calibrated to data on age-, stage- and location-specific CRC incidence obtained from the Netherlands Cancer Registry and age-specific prevalence and multiplicity distribution of adenomas from autopsy and colonoscopy studies [13–23]. Age-, and invitation-round-specific FIT and colonoscopy participation rates were based on data from the Dutch CRC screening program [24]. Test characteristics for follow-up and surveillance colonoscopy were based on systemic review of polyp miss rates in tandem colonoscopy studies [25].

Long-term screening outcomes were number of CRC-related deaths and CRC cases compared to no screening and total costs in euros. All outcomes were reported per 100,000 simulated individuals eligible for CRC screening. Additionally, in case of a consistent bias in FIT performance, we showed the annual change between 2030, and 2060.

2.3. Sensitivity analysis

We conducted a sensitivity analysis to validate our results. We assessed the impact of FIT bias using data from individuals who were screened using the current cut-off of 47 μg Hb/g feces. Since we do not have information about the colonoscopy findings below the cut-off of 47 μg Hb/g feces, we only evaluated the impact of a negative bias in the FIT performance.

2.4. Probabilistic sensitivity analysis

In the probabilistic sensitivity analysis, we assessed the uncertainty of the estimated test characteristics at each level of bias. We used the confidence intervals (CI) around the positivity and detection rates to determine relative sensitivities and specificities to compute CIs around the estimated test characteristics. For every level of bias, we performed 1000 simulation runs of 100 million individuals in which we sampled parameter values for the test characteristics from the beta distribution with shape parameters α and β. The means (m) and standard deviations (s) from the 95 %CI of the estimated test characteristics were used to compute $\alpha = m\left(\frac{m(1-m)}{s^2} - 1\right)$ and $\beta = (1-m)\left(\frac{m(1-m)}{s^2} - 1\right)$.

3. Results

3.1. Short-term screening outcomes

Without hypothetical bias in FIT performance, the positivity rate was 7.3 %, the detection rate for NAA 1.10 %, the detection rate for AA 2.78 % and the detection rate for CRC 0.67 % (Table 1). In case the FIT would measure, for example, 2.5 % less hemoglobin in the feces, the positivity rate was 7.2 % resulting in a detection rate for NAA, AA and CRC of 1.08 %, 2.76 % and 0.66 %, respectively. With a bias of −15 % to + 15 %, we observed a positivity rate ranging from 6.7 % to 7.8 %, a detection rate for NAA and AA of 0.97 % to 1.22 % and 2.60 % to 2.93 %, respectively. The detection rate for CRC was between 0.650 % and 0.68 % with biases between −15 % and +15 %.

3.2. Estimated FIT test characteristics per level of bias

We estimated the test characteristics of the FIT at different levels of bias in FIT performance such that the simulated positivity rate and detection rates by MISCAN-Colon match those observed within 0.1 %. The estimated test characteristics showed that, the sensitivity for all types of lesions increased with a positive bias and decreased with a negative bias (Table 2). Conversely, the specificity decreased with a positive bias and increased with a negative bias.

Table 1
FIT biases with positivity rate and detection rate for (non-)advanced adenomas and CRC observed in the first half year of 2014 in the Dutch population-based CRC screening program.

FIT bias (%)	Positivity rate (%; 95 % CI)	Detection rate NAA (%; 95 % CI)	Detection rate AA (%; 95 % CI)	Detection rate CRC (%; 95 % CI)
−15.0	6.7 (6.6–6.8)	0.97 (0.91–1.02)	2.60 (2.51–2.69)	0.65 (0.61–0.69)
−10.0	6.9 (6.7–7.0)	1.01 (0.95–1.06)	2.65 (2.57–2.74)	0.65 (0.61–0.70)
−7.5	7.0 (6.8–7.1)	1.03 (0.97–1.09)	2.69 (2.60–2.78)	0.66 (0.61–0.70)
−5.0	7.1 (6.9–7.2)	1.06 (1.01–1.12)	2.73 (2.64–2.82)	0.66 (0.62–0.71)
−2.5	7.2 (7.0–7.3)	1.08 (1.03–1.14)	2.76 (2.67–2.85)	0.66 (0.62–0.71)
0.0 (comparator)	7.3 (7.1–7.4)	1.10 (1.05–1.16)	2.78 (2.69–2.87)	0.67 (0.62–0.71)
+2.5	7.4 (7.2–7.5)	1.13 (1.07–1.19)	2.81 (2.72—2.90)	0.67 (0.63–0.72)
+5.0	7.5 (7.3–7.6)	1.14 (1.09–1.20)	2.84 (2.75–2.93)	0.67 (0.63–0.72)
+7.5	7.5 (7.4–7.7)	1.16 (1.10–1.22)	2.86 (2.77–2.95)	0.67 (0.63–0.72)
+10.0	7.6 (7.5–7.8)	1.19 (1.13–1.25)	2.88 (2.79–2.97)	0.68 (0.63–0.72)
+15.0	7.8 (7.6–7.9)	1.22 (1.16–1.28)	2.93 (2.84–3.02)	0.68 (0.64–0.73)

Abbreviations: FIT, fecal immunochemical test; CRC, colorectal cancer; μg Hb/g, microgram Hemoglobin per gram; NAA, non-advanced adenoma; AA, advanced adenoma; CI, confidence interval.

3.3. Long-term impact

Without hypothetical bias in FIT performance, MISCAN-Colon estimated 23,350 CRC cases and 9,160 CRC-related deaths per 100,000 simulated individuals eligible for CRC screening during their lifetime (Table 3). For the screening-eligible cohort in 2020, MISCAN-Colon estimated 22,230 CRC cases and 8,910 CRC-related deaths per 100,000 simulated individuals eligible for CRC screening during their life time. The number of CRC cases was estimated to increase up to 0.10 % with a single bias of 15 % lower Hb levels. A single bias of 15 % higher Hb levels showed a smaller decrease in CRC cases of up to 0.08 % (Table 3). If the FIT had a consistent bias, the number of CRC cases was estimated to decrease up to 0.97 % at 15.0 % higher Hb levels and increase up to 1.23 % with 15 % lower Hb levels. The pattern for estimated change in CRC-related deaths was similar compared to CRC cases. However, the relative change compared to no FIT bias is slightly larger for CRC-related deaths, up to ±1.4 % vs 1.2 % in CRC cases (Fig. 1, Table 3). Without hypothetical bias in FIT performance, MISCAN-Colon estimated that the Dutch population-based CRC screening program would cost €18,280,730,- per 100,000 simulated individuals eligible for CRC screening. A single negative or positive FIT bias would result in an estimated costs increase or decrease of less than 0.01 %. The change in total costs with consistent negative bias (increase up to 2.07 %) was much higher compared to consistent positive bias (increase up to 0.05 %) (Fig. 2). Additionally, the annual change in number of CRC cases and CRC-related deaths was relatively higher for negative bias compared to positive bias (Fig. 3).

Table 2
Test characteristics of the FIT at different levels of biases in FIT performance after calibration in MISCAN-Colon.

Bias (%)	Specificity (%; 95 % CI)	Sensitivity[a] (%; 95 % CI)			
		Medium adenomas (6–9 mm)	Large adenoma (≥10 mm)	CRC late preclinical[b]	CRC early preclinical[b]
−15.0	97.81 (97.77–97.85)	4.22 (3.91–4.22)	25.39 (24.06–26.80)	28.11 (25.73–30.70)	64.52 (59.06–70.47)
−10.0	97.75 (97.71–97.79)	4.50 (4.18–4.85)	25.95 (24.61–27.37)	28.24 (25.87–30.83)	64.66 (59.23–70.59)
−7.5	97.71 (97.67–97.75)	4.68 (4.33–5.02)	26.30 24.92–27.70)	28.35 (25.99–30.96)	64.80 (59.37–70.72)
−5.0	97.70 (97.66–97.74)	4.97 (4.62–5.34)	26.68 (25.59–28.40)	28.58 (26.20–31.18)	65.04 (59.62–70-97)
−2.5	97.66 (97.62–97.70)	5.11 (4.75–5.49)	26.96 (25.59–28.40)	28.67 (26.28–31.27)	65.14 (59.71–71.05)
0.0 (comparator)	97.60 (97.56–97.64)	5.18 (4.82–5.56)	27.21 (25.83–28.65)	28.75 (26.36–31.35)	65.23 (59.82–71.13)
+2.5	97.55 (97.51–97.59)	5.35 (4.98–5.74)	27.45 (26.07–28.90)	28.91 (26.52–31.51)	65.41 (60.00–71.30)
+5.0	97.50 (97.46–97.55)	5.38 (5.01–5.77)	27.75 (26.37–29.21)	29.03 (26.64–31.64)	65.54 (60.14–71.42)
+7.5	97.46 (97.42–97.50)	5.47 (5.10–5.86)	27.96 (26.57–29.42)	29.18 (26.78–31.79)	65.70 (60.30–71.58)
+10.0	97.42 (97.38–97.46)	5.65 (5.27–6.05)	28.15 (26.75–29.61)	29.31 (26.91–31.92)	65.85 (60.45–71.72)
+15.0	97.34 (97.29–97.38)	5.85 (5.46–6.26)	28.64 (27.24–30.12)	29.46 (27.06–32.07)	66.01 (60.62–71.87)

Abbreviations: FIT, fecal immunochemical test; CRC, colorectal cancer; CI, confidence interval.
[a] Sensitivity for small adenomas was assumed to be zero.
[b] It was assumed that the probability a CRC bleeds and thus the sensitivity of a FIT for CRC depends on the time until clinical diagnosis [29].

Table 3
Change in long-term CRC screening outcomes per 100,000 simulated individuals eligible for CRC screening at different bias in FIT performance consistently from 2020 onwards and one-time (single) in 2020 in the Dutch CRC screening program.

FIT bias (%)	Consistent			Single		
	CRC cases	CRC-related deaths	Costs[a] (€)	CRC cases	CRC-related deaths	Costs[a] (€)
−15.0	290 (1.23 %)	130 (1.42 %)	378,150 (2.07 %)	20 (0.1 %)	10 (0.11 %)	48,650 (0.01 %)
−10.0	200 (0.85 %)	90 (0.99 %)	252,820 (1.38 %)	20 (0.07 %)	10 (0.08 %)	29,300 (0.01 %)
−7.5	150 (0.62 %)	70 (0.73 %)	156,840 (0.86 %)	10 (0.05 %)	10 (0.06 %)	17,200 (<0.001 %)
−5.0	80 (0.35 %)	40 (0.41 %)	−27,030 (−0.15 %)	10 (0.03 %)	3 (0.03 %)	−5,780 (<0.001 %)
−2.5	40 (0.17 %)	20 (0.2 %)	−61,250 (−0.34 %)	3 (0.01 %)	1 (0.01 %)	−7,200 (<0.001 %)
0.0 (comparator)	23,350	9,160	18,280,730	22,230	8,910	521,816,980
+2.5	−50 (−0.2 %)	−20 (−0.25 %)	−8,310 (−0.05 %)	−4 (−0.02 %)	−2 (−0.02 %)	50 (<0.001 %)
+5.0	−80 (−0.35 %)	−40 (−0.42 %)	−470 (0 %)	−10 (−0.03 %)	−3 (−0.03 %)	−940 (<0.001 %)
+7.5	−120 (−0.49 %)	−60 (−0.61 %)	10,330 (0.06 %)	−10 (−0.04 %)	−4 (−0.05 %)	1,980 (<0.001 %)
+10.0	−160 (−0.67 %)	−80 (−0.83 %)	2,690 (0.01 %)	−10 (−0.06 %)	−10 (−0.06 %)	800 (<0.001 %)
+15.0	−230 (−0.97 %)	−110 (−1.17 %)	9,770 (0.05 %)	−20 (−0.08 %)	−10 (−0.09 %)	−820 (<0.001 %)

Abbreviations: FIT, fecal immunochemical test; CRC, colorectal cancer.
[a] Future costs were discounted using a standard annual rate of 3% [12].

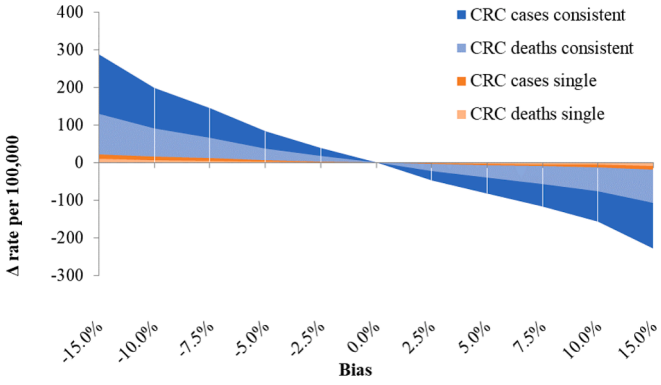

Fig. 1. Change in CRC cases and CRC-related deaths per 100,000 simulated individuals eligible for CRC screening at different bias in FIT performance consistently from 2020 onwards and one-time (single) in 2020 in the Dutch CRC screening program (color).

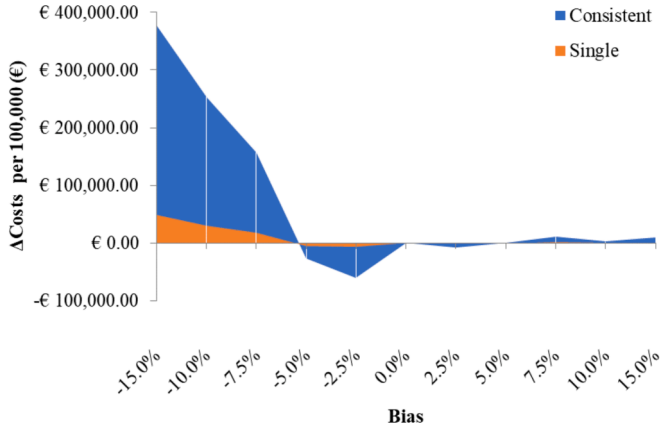

Fig. 2. Change in costs (discounted at 3 %) per 100,000 simulated individuals eligible for CRC screening of the CRC screening program at different levels of bias in FIT performance consistently from 2020 onwards and one-time (single) in 2020 in the Netherlands (color).

3.4. Sensitivity analysis

Our results were valid when using data from participants with the higher FIT positivity cut-off of 47 μg Hb/g feces (Table S2). Using the test characteristics from our validation (Table S3), the estimated change in long-term CRC screening outcomes was comparable to the change in the base case analysis (Table S4).

3.5. Probabilistic sensitivity analysis

Probabilistic sensitivity analysis demonstrated that the prediction intervals around the base case estimates for CRC cases, CRC-related deaths and total costs were small at different bias in FIT performance (Fig. S1). The pattern of prediction intervals over different FIT biases was similar to that of the base case estimates. This indicates that with

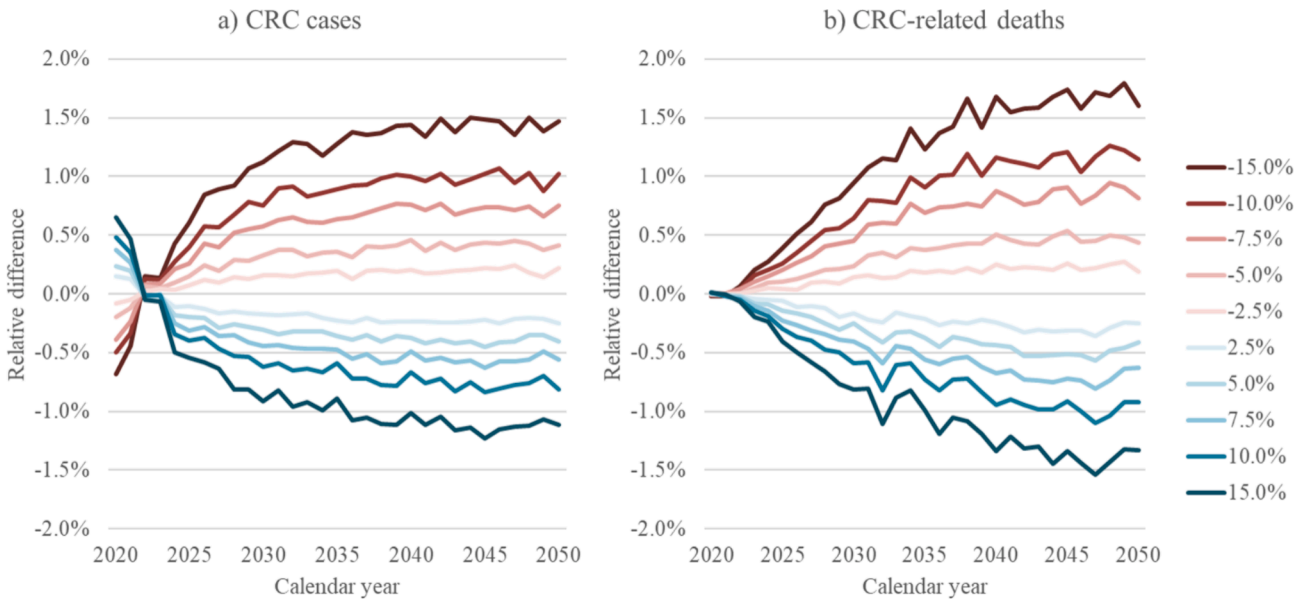

Fig. 3. Annual relative difference in (a) CRC cases and (b) CRC-related deaths in the Dutch CRC screening program at different levels of bias in FIT performance consistently from 2020 onwards (color).

1,000 different sets of parameter values for FIT bias, its impact was quite robust.

4. Discussion

This study demonstrates that a consistent bias in the FIT performance can have an undeniable impact on long-term CRC incidence and CRC-related death. If there is a consistent bias of −15 % from 2020 onwards, the number of CRC cases is estimated to increase with 290 cases per 100,000 simulated individuals eligible for CRC screening. Conversely, with a consistent bias of +15 %, the number CRC cases was estimated to decrease by 230 cases per 100,000 simulated individuals eligible for CRC screening. The impact of a single bias in 2020 is smaller, with an increase of 20 cases and decrease of 20 cases per 100,000 simulated individuals eligible for CRC screening. This study highlights that negative FIT bias has a more significant impact on long-term CRC screening outcomes compared to positive bias, especially in terms of the overall program costs. Although there is a net positive change in outcome and cost – driven by a larger positive change with positive bias compared to the negative change with negative bias – it is preferable to achieve this change by lowering the cut-off, rather than by allowing randomly distributed bias. Alternatively, an asymmetrical or upward-shifted distribution of allowable bias could be considered, with greater tolerance for positive bias (which aids in case finding), than for negative bias (which increases disease prevalence and costs).

Negative FIT biases, which measure a lower Hb-concentration in the stool sample, can be translated into lower test sensitivity and higher test specificity. By definition, lower sensitivity results in an increase in the number of false-negatives, meaning that more cases of CRCs and/or (advanced) adenomas may be missed. As a result, this decreases the chance of preventing a CRC case and potentially a CRC-related death. Higher specificity leads to fewer false-positives, by definition, reducing the number of unnecessary colonoscopies. Conversely, positive FIT biases, which measures a higher Hb-concentration in the stool sample, increase sensitivity but decrease specificity resulting in more unnecessary colonoscopies and higher total costs in the program. This can be explained by the fact that preventing more (advanced) CRCs by colonoscopy is less expensive than treating them. Overall, in this study, negative FIT biases have an estimated greater change on CRC screening outcomes, as we are more likely to miss more relevant findings for higher Hb-concentrations in this case. It is important to minimize FIT biases as much as possible, as this could imply that the detection rate is not affected linearly. Non-symmetrical acceptance criteria should be emphasized.

Two prior studies showed that the positivity rate and detection rate of FIT varied between device and reagent lot and were influenced by the ambient temperature, but clinically irrelevant, which is consistent with our findings in the case of a single bias in the FIT performance [5,26]. In contrast, consistent FIT biases could have a substantial impact as shown in this study, but it is estimated that it would take 10–15 years to be visible in the screening outcomes. The big difference between a single and consistent bias emphasizes the importance of preventing a 'drifting' average bias and this has been addressed before [3]. To prevent these kinds of failures in FIT performance, it is important to closely monitor the outcomes and the quality assurance within CRC screening. However, when setting new acceptance criteria, it is important to consider their feasibility.

A key strength of this study is that we used data from an ongoing national population-based CRC screening program to determine the test characteristics for each bias using the well-established MISCAN-Colon model to identify the long-term impact of FIT biases [9,10]. A limitation is that the biases are modelled constant over time, while it is likely that it is not a fixed value and in practice is accompanied by random imprecision which is bidirectional by nature. However, the combination of a single and consistent bias together with a probabilistic sensitivity analysis shows the minimum and maximum change in long-term outcomes in an on-going national population-based CRC screening program. Although detectable bias, once accepted as less than unacceptable, becomes part of the long-term imprecision, these bias components behave differently than true random sources (of short-term imprecision) and may have a non-symmetrical distribution over time. However, the APS for the Dutch CRC screening program were set such that accepting a positive FIT bias meant accepting a negative FIT bias assuming to result in a symmetric tolerance range. We observe a larger absolute change for negative bias than positive biases. Thus, even with fluctuating bias, there would be a net change and therefore not negligible, but probably to be much smaller than estimated in this study. Another limitation is that we assumed that there is no bias in our data, while this is likely to be incorrect. However, we do not expect the conclusions to change as we compare the impact of bias to a scenario

without hypothetical bias and in this way, we assume equal levels of bias in all scenarios in the data. Moreover, the four laboratories are coordinated using identical protocols instruments, and synchronized reagent and calibrator lot changes. They are also supervised for their within-laboratory and between-laboratory quality control results by a contracted EQA organization. So, we have assumed noise from other sources analytical variation to be acceptable in size. Last, in this study, we only considered AA and CRC and assumed that all cancers develop from the adenoma-carcinoma pathway, while it is known that 15–30 % of the cancers develop from the serrated pathway [27,28]. However, we do not believe this would affect the results of our study, because FIT has very low sensitivity for sessile serrated polyps.

The variability in FIT performance depends not only on bias but also on imprecision (i.e. random source of inaccuracy). Bias primarily occurs at the population level, affecting all FIT kits, whereas imprecision occurs randomly at the individual level. In evaluating the variability in FIT performance within the program, we have focused solely on bias. While imprecision always occurs due to unexpected circumstances, we do not expect it to significantly alter long-term CRC screening outcomes. However, given its random nature, imprecision will likely have a random effect on the direction of the error.

Although the impact of single bias is small, it is important to continue monitoring of screening programs to maintain quality assurance, because in case of consistent bias, the change in CRC cases and CRC-related deaths can be high. In the Dutch CRC screening program, one of the quality control requirements ensures that all reagents should be within acceptance criteria of ±7.5 % of the overall-lot mean. We show that consistent bias of −7.5 % would mean that approximately 150 (0.62 %) more CRC cases per 100,000 individuals will occur in the long term, while consistent bias of +7.5 % would result in approximately 120 (0.49 %) fewer CRC cases per 100,000 individuals. This information can be used by program organizers to set evidence-based acceptance criteria for reagents lots so with acceptable impact on long-term outcomes of the screening programs.

5. Conclusion

Consistent FIT biases could result in substantial change in long-term CRC incidence and CRC-related mortality. Although it is tempting to reason that it is an easy choice to allow less bias because it results in both more survival and less cost, the selected allowable bias also needs to be feasible in practice. An unusual, but feasible modification in the screening program could involve the implementation of an asymmetrical or upward-shifted tolerance interval for FIT bias. Allowing more positive than negative bias could improve healthcare outcome while reducing costs, even without increasing the analytical rejection rate.

CRediT authorship contribution statement

Lucie de Jonge: Conceptualization, Data curation, Formal analysis, Project administration, Software, Visualization, Writing – original draft. **Esther Toes-Zoutendijk:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Brechtje D.M. Koopmann:** Writing – review & editing. **Marith van Schrojenstein Lantman:** Writing – review & editing. **Brenda Franken-van-Vorsselen:** Writing – review & editing. **Christel Speijers:** Writing – review & editing. **Huub van Ingen:** Writing – review & editing. **Erwin Humer:** Writing – review & editing. **Petra van der Groep:** Writing – review & editing. **Marc Thelen:** Writing – review & editing. **Iris Lansdorp-Vogelaar:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Prof. Dr. Iris Lansdorp-Vogelaar: Associate editor at Gastroenterology; expert at the Health Council; panel member of the European Commission Initiative on Colorectal Cancer; visiting scientist at IARC. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was funded by the National Institute for Public Health and the Environment (grant ID: 4410002642). The funder had no role in the data collection, statistical analyses, interpretation of the reuslts and writing the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2024.119809.

References

[1] E.H. Schreuders, A. Ruco, L. Rabeneck, R.E. Schoen, J.J.Y. Sung, G.P. Young, et al., Colorectal cancer screening: a global overview of existing programmes, Gut 64 (10) (2015) 1637–1649.
[2] M.H.M. Thelen, L.M. van Schrojenstein, When bias becomes part of imprecision: how to use analytical performance specifications to determine acceptability of lot-lot variation and other sources of possibly unacceptable bias, Clin. Chem. Lab. Med. (CCLM) (2024).
[3] L.M. van Schrojenstein, H.C. Çubukçu, G. Boursier, M. Panteghini, F.A. Bernabeu-Andreu, N. Milinkovic, et al., An approach for determining allowable between reagent lot variation, Clin. Chem. Lab. Med. (CCLM) 60 (5) (2022) 681–688.
[4] F. Ceriotti, P. Fernandez-Calle, G.G. Klee, G. Nordin, S. Sandberg, T. Streichert, et al., Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference, Clin. Chem. Lab. Med. (CCLM) 55 (2) (2017) 189–194.
[5] E. Toes-Zoutendijk, J.M.G. Bonfrer, C. Ramakers, M. Thelen, M.C.W. Spaander, E. Dekker, et al., Quality monitoring of a FIT-based colorectal cancer screening program, Clin. Chem. 65 (3) (2019) 419–426.
[6] E. Toes-Zoutendijk, M.E. van Leerdam, E. Dekker, F. van Hees, C. Penning, I. Nagtegaal, 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 152 (4) (2017) 767–75 e2.
[7] L. Deprez, C. Piggott, E.A.E. van der Hagen, M. Frasa, S.C. Benton, Testing IWGoFI. Comparison and commutability study among four faecal immunochemical tests (FIT) systems, Clin. Chem. Lab. Med. (CCLM) 62 (1) (2024) 50–59.
[8] E. Dekker, M.E. van Leerdam, Y. Hazewinkel, S. Sanduleanu, H.F. Vasen, I. Lansdorp-Vogelaar, Nederlandse richtlijn coloscopie surveillance, 2016.
[9] F. van Hees, A.G. Zauber, H. van Veldhuizen, M.-L.-A. Heijnen, C. Penning, H.J. de Koning, et al., The value of models in informing resource allocation in colorectal cancer screening: the case of the Netherlands, Gut 64 (12) (2015) 1985–1997.
[10] F. Loeve, R. Boer, G.J. van Oortmarssen, M. van Ballegooijen, J.D. Habbema, The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening, Comput. Biomed. Res. 32 (1) (1999) 13–33.
[11] A.-H. Omidvari, C.A.M. Roumans, S.K. Naber, S. Kroep, B.P.L. Wijnhoven, A. van der Gaast, et al., The impact of the policy-practice gap on costs and benefits of Barrett's esophagus management, Am. J. Gastroenterol. 115 (7) (2020) 1026.
[12] G.D. Sanders, P.J. Neumann, A. Basu, D.W. Brock, D. Feeny, M. Krahn, et al., Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine, J. Am. Med. Assoc. 316 (10) (2016) 1093–1103.
[13] T.C. Arminski, D.W. McLean, Incidence and distribution of adenomatous polyps of the colon and rectum based on 1,000 autopsy examinations, Dis. Colon Rectum 7 (4) (1964) 249–261.
[14] L.J. Blatt, Polyps of the colon and rectum, Dis. Colon Rectum 4 (4) (1961) 277–282.
[15] J.A. Bombi, Polyps of the colon in Barcelona, Spain. An autopsy study, Cancer 61 (7) (1988) 1472–1476.
[16] I. Chapman, Adenomatous polypi of large intestine: incidence and distribution, Ann. Surg. 157 (2) (1963) 223.
[17] J.C. Clark, Y. Collan, T.J. Eide, J. Estève, S. Ewen, N.M. Gibbs, et al., Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer, Int. J. Cancer 36 (2) (1985) 179–186.

[18] J.R. Jass, P.J. Young, E.M. Robinson, Predictors of presence, multiplicity, size and dysplasia of colorectal adenomas. A necropsy study in New Zealand, Gut 33 (11) (1992) 1508–1514.
[19] L.G.K. Johannsen, O. Momsen, N.O. Jacobsen, Polyps of the large intestine in Aarhus, Denmark, Scand. J. Gastroenterol. 24 (7) (1989) 799–806.
[20] R.R. Rickert, O. Auerbach, L. Garfinkel, E.C. Hammond, J.M. Frasca, Adenomatous lesions of the large bowel: an autopsy survey, Cancer 43 (5) (1979) 1847–1857.
[21] E.M. Stoop, M.C. de Haan, T.R. de Wijkerslooth, P.M. Bossuyt, M. van Ballegooijen, C.Y. Nio, et al., Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial, Lancet Oncol. 13 (1) (2012) 55–64.
[22] M.H. Vatn, H. Stalsberg, The prevalence of polyps of the large intestine in Oslo: an autopsy study, Cancer 49 (4) (1982) 819–825.
[23] A.R. Williams, B.A. Balasooriya, D.W. Day, Polyps and cancer of the large bowel: a necropsy study in Liverpool, Gut 23 (10) (1982) 835–842.
[24] Landelijk Evaluatie Team voor Colorectaal Kanker Bevolkingsonderzoek (LECO), Landelijke Monitoring En Evaluatie van Het Bevolkingsonderzoek Naar Darmkanker in Nederland; 2014–2017, 2019.
[25] J.C. van Rijn, J.B. Reitsma, J. Stoker, P.M. Bossuyt, S.J. van Deventer, E. Dekker, Polyp miss rate determined by tandem colonoscopy: a systematic review, Am. J. Gastroenterol. 101 (2) (2006) 343–350.
[26] G. De Girolamo, C.A. Goldoni, R. Corradini, O. Giuliani, F. Falcini, P. Sassoli De'Bianchi, et al., Ambient temperature and FIT performance in the Emilia-Romagna colorectal cancer screening programme, J. Med. Screen. 23 (4) (2016) 186–191.
[27] J.E. East, W.S. Atkin, A.C. Bateman, S.K. Clark, S. Dolwani, S.N. Ket, et al., British Society of Gastroenterology position statement on serrated polyps in the colon and rectum, Gut 66 (7) (2017) 1181–1196.
[28] E. Dekker, J.E.G. Ijspeert, Serrated pathway: a paradigm shift in CRC prevention, Gut 67 (10) (2018) 1751–1752.
[29] I. Lansdorp-Vogelaar, M. van Ballegooijen, R. Boer, A. Zauber, J.D.F. Habbema, A novel hypothesis on the sensitivity of the fecal occult blood test: results of a joint analysis of 3 randomized controlled trials, Cancer 115 (11) (2009) 2410–2419.