Preventive Medicine 164 (2022) 107187

Contents lists available at ScienceDirect

Preventive Medicine

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

Guidance for setting international standards on reporting longitudinal adherence to stool-based colorectal cancer screening

Francine van Wifferen [a,*], Marjolein J.E. Greuter [a], Birgit I. Lissenberg-Witte [a], Beatriz Carvalho [b], Gerrit A. Meijer [b], Evelien Dekker [c], Cinzia Campari [d], Montse Garcia [e], Linda Rabeneck [f], Iris Lansdorp-Vogelaar [g], Carlo Senore [h], Veerle M.H. Coupé [a], on behalf of the ICSN Colorectal Cancer Screening Working Group

[a] Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, De Boelelaan 1117, Amsterdam, the Netherlands
[b] Department of Pathology, Netherlands Cancer Institute, Amsterdam, the Netherlands
[c] Department of Gastroenterology and Hepatology, Amsterdam UMC location University of Amsterdam, Meibergdreef 9, Amsterdam, the Netherlands
[d] Screening Unit, Azienda USL-IRCCS di Reggio Emilia, Italy
[e] Cancer Screening Unit, Prevention and Control Programme, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain
[f] Prevention & Cancer Control, Ontario Health (Cancer Care Ontario), University of Toronto, Canada
[g] Department of Public Health, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, the Netherlands
[h] SSD Epidemiology, screening unit – CPO, University Hospital "Città della Salute e della Scienza", Turin, Italy

ARTICLE INFO

Keywords:
Longitudinal adherence
Colorectal cancer
Screening
Stool-based testing

ABSTRACT

Longitudinal adherence to colorectal cancer (CRC) screening is reported using different summarizing measures, which hampers international comparison. We provide evidence to guide recommendations on which longitudinal adherence measure to report.

Using adherence data over four stool-based CRC screening rounds in three countries, we calculated six summarizing adherence measures; *adherence over all rounds, adherence per round, rescreening, full programme adherence (yes/no), regularity (never/inconsistent/consistent screenees)* and *number of times participated*. For each measure, we calculated the accuracy in capturing the *observed* adherence patterns. Using the ASCCA model, we predicted screening effectiveness when using summarizing measures as model input versus the *observed* adherence patterns.

Adherence over all rounds in the Italian, Spanish and Dutch cohorts was 64.9%, 42.8% and 61.5%, respectively, and the proportion of consistent screenees was 50.9%, 26.3% and 45.7%. *Number of times participated* and *regularity* were most accurate and resulted in similar model-predicted screening effectiveness as simulating the *observed* adherence patterns of Italy, Spain and the Netherlands (mortality reductions: 24.4%, 16.9% and 23.5%). *Adherence over all rounds* and *adherence per round* were least accurate. Screening effectiveness was overestimated when using *adherence over all rounds* (mortality reductions: 26.8%, 19.4% and 25.7%) and *adherence per round* (mortality reductions: 26.8%, 19.5% and 25.9%).

To conclude, *number of times participated* and *regularity* were most accurate and resulted in similar model-predicted screening effectiveness as using the *observed* adherence patterns. However they require longitudinal data. To facilitate international comparison of CRC screening programme performance, consensus on an accurate adherence measure to report should be reached.

1. Introduction

Many countries have implemented population-based colorectal cancer (CRC) screening programmes using stool-based tests (Schreuders et al., 2015). Due to the relatively low sensitivity of these tests for CRC precursor lesions, repeated screening every one to two years is recommended (de Wijkerslooth et al., 2012; US Preventive Services Task Force, 2021; Ponti et al., 2017). Therefore, longitudinal adherence, i.e.

* Corresponding author at: Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, De Boelelaan 1117, Amsterdam, the Netherlands.
E-mail address: f.vanwifferen@amsterdamumc.nl (F. van Wifferen).

https://doi.org/10.1016/j.ypmed.2022.107187
Received 7 February 2022; Received in revised form 1 June 2022; Accepted 5 August 2022
Available online 10 August 2022

adherence over multiple rounds, is a critical indicator of programme performance and should be adequately reported (Bulliard et al., 2014).

Longitudinal adherence to a stool-based CRC screening programme can be reported as the proportion of individuals in a closed cohort of the population that have followed each possible 'Yes-No' adherence pattern, in which Yes (No) is assigned when the individual has (not) participated in a round. Stool-based CRC screening programmes consisting of eleven or more rounds are common, resulting in $>2^{11} = 2048$ possible patterns, which means it is not feasible to report adherence as such (Schreuders et al., 2015). Instead, several summarizing measures can be used to report adherence, such as the *adherence over all rounds*, calculated as the overall proportion of individuals participating upon invitation in a given calendar period, or the *rescreening rate*, which is calculated as the probability of participation for those who had participated in their previous round (Bulliard et al., 2014).

No consensus currently exists regarding which measure should be used for reporting longitudinal adherence to population-based CRC screening programmes using biennial fecal immunochemical testing (FIT). Guidelines only recommend to include cross-sectional adherence measures as a performance indicator for CRC screening (Moss et al., 2012). Some CRC screening programmes do report longitudinal adherence measures, but the definition differs across programmes (Binefa et al., 2016; National Institute for Public Health and the Environment (RIVM), Erasmus MC, NKI/Avl, 2018; Connell et al., 2014). To increase comparison of stool-based CRC screening programmes, it is important that the reporting of longitudinal adherence is standardized. This is underpinned by the findings of a recent review that showed that different programmes report different longitudinal adherence measures (Doria-Rose et al., 2021). The authors recommend that screening programmes should collect detailed, longitudinal, individual-level data on adherence, such that at any time, programmes can be compared with respect to different adherence measures depending on the research questions. However, the review did not assess the accuracy of these longitudinal adherence measures in representing the true underlying longitudinal adherence pattern. Summarizing measures vary in the accuracy with which the true underlying longitudinal adherence pattern is captured and it is currently unclear which measure is most accurate. For example, the *adherence rate per round* reflects the proportion of invitees who have participated in a specific round, but does not include longitudinal participation over multiple rounds. Therefore, this measure implicitly assumes that participation in a specific screening round is independent of participation in previous rounds. If this independence would hold, only few individuals in a population would always or never participate. However, it is a fact that there is a non-negligible proportion of structural non-attenders, and likewise, of individuals adhering to the full cancer screening programme (Benito et al., 2019).

As models, that use summarizing measures as input, are used to estimate long-term screening effectiveness, the accuracy of a summarizing measure in capturing the true underlying adherence pattern may also affect these model-based predictions. Quantification of this impact is important, because screening recommendations and policies, e.g. the screening frequency, the starting or stopping age and the type of test used in screening, are often based on these model predictions (US Preventive Services Task Force, 2021). Therefore, we assessed the accuracy of summarizing measures for longitudinal adherence in capturing the observed adherence pattern and the impact of using summarizing measures rather than detailed longitudinal adherence data on model-predicted CRC screening effectiveness. The results of this study can be used to set international standards on reporting longitudinal adherence that will enhance comparison of biennial FIT-based CRC screening programme outcomes among countries.

2. Material and methods

We evaluated six summarizing measures for longitudinal adherence to biennial FIT-based CRC screening, namely *adherence over all rounds, adherence per round, rescreening, full programme adherence, regularity*, and *number of times participated*. The definitions of all measures are provided in Table 1. In brief, with further details provided below, the following analyses were done. We extracted the relative frequency of occurrence of each possible 'Yes-No' adherence pattern for three closed cohorts of individuals who have all been invited to four rounds of biennial FIT-based CRC screening. For example, an individual who participated in the first three screening rounds but not in the fourth screening round has the following adherence pattern 'Yes-Yes-Yes-No'. Next, using the relative frequencies of these observed adherence patterns, the six summarizing measures were calculated. Subsequently, for each summarizing measure, we constructed a predicted set of relative frequencies of all possible 'Yes-No' adherence patterns over four rounds using only the information provided by that summarizing measure. Next, the deviation of the predicted set of relative frequencies from the observed set of relative frequencies of the longitudinal adherence patterns was calculated to evaluate the accuracy of each summarizing measure. Lastly, we evaluated the deviation in model-based predictions of screening effectiveness when using the predicted set of relative frequencies for each summarizing measure as model input for adherence compared to using the observed longitudinal adherence pattern.

2.1. Observed adherence patterns

The first cohort where detailed longitudinal adherence data was obtained from was from Reggio Emilia in Italy. In this region, biennial FIT screening for residents aged 50 to 69 was implemented in 2005 (Giorgi Rossi et al., 2015). The second cohort was from Catalonia, Spain. Biennial guaiac-based fecal occult blood test (gFOBT) screening for individuals aged 50 to 69 was implemented in this region in 2000 (Binefa et al., 2016). For our study, only the data from 2010 onwards were used, because since then gFOBT was replaced by FIT. The third cohort was from a Dutch population-based CRC screening trial (CORERO), which included individuals between 2006 and 2007. This phase 1 trial was conducted in screen-naïve individuals aged 50–75 in the Netherlands and compared participation and detection rates of gFOBT, FIT and sigmoidoscopy (Hol et al., 2010). Only data from individuals who were randomised to biennial FIT were included in this study.

Table 1
Definitions of the six adherence measures being evaluated.

Adherence measure	Definition
Adherence over all rounds	The proportion of all sent invitations that have led to participation.
Adherence per round	The proportion of all sent invitations in a specific round that have led to participation.
Rescreening	The proportion of invitees that participate among those who have participated in their previous round. Together with the rescreening rate, the adherence over all rounds is reported.
Full programme adherence	The measure full programme adherence classifies individuals in two categories: Full adherers (have completed all screening rounds) and not full adherers (have not completed all screening rounds). The proportion of individuals in each category is reported, as well as the adherence over all rounds.
Regularity	The measure regularity classifies individuals in three categories: Never screenees (individuals who have never participated), inconsistent screenees (individuals who have participated in at least one but not in all rounds) and consistent screenees (individuals who have participated in all rounds). The proportion of individuals in each category is reported, as well as the adherence over all rounds.
Number of times participated	The measure number of times participated classifies individuals based on the number of screening rounds they have participated in (0 rounds, 1 round, etc.). For a programme with m screening rounds, this means that there are m + 1 categories. The proportion of individuals in each category is reported.

2.2. Adherence measures

Using the observed relative frequencies of the longitudinal adherence patterns, we calculated for each cohort the six summarizing adherence measures for longitudinal adherence to FIT screening (Table 1). For example, the *adherence per round* is calculated for round one, two, three and four by summing the relative frequencies of the eight adherence patterns in which individuals have participated in round one, two, three and four, respectively. The calculation of all summarizing measures based on the observed relative frequencies of the longitudinal adherence patterns is provided in the Supplementary Methods. To evaluate how well each summarizing measure approximates the observed adherence pattern in four screening rounds, we derived from each summarizing measure the predicted relative frequency of occurrence of all possible 'Yes-No' adherence patterns. For example, if the adherence per round is 60%, 63%, 65% and 68% in rounds one, two, three and four, respectively, the predicted relative frequency of the 'No-Yes-Yes-No' pattern is calculated as $(1–0.60) \times 0.63 \times 0.65 \times (1–0.68) = 5.2\%$. Next, the accuracy in capturing the observed adherence patterns was quantified for each summarizing measure using the summed chi-square goodness of fit. The Supplementary Methods provides, for all adherence measures, the mathematical functions to calculate the predicted relative frequency of all possible adherence patterns and the summed chi-square goodness of fit.

2.3. ASCCA model

To evaluate the impact of using a summarizing adherence measure as CRC screening model input rather than the observed adherence patterns to predict screening effectiveness, we used the Adenoma and Serrated Pathway to Colorectal CAncer (ASCCA) model. The model is extensively described elsewhere (Greuter et al., 2014). In brief, this CRC natural history model incorporates the conventional adenoma-carcinoma pathway and the serrated pathway in CRC development. Individual health trajectories are simulated from the age of 20 to the age of 90 or death, whichever occurs first. During their life, individuals can develop up to ten adenomas and ten serrated polyps. The development of each lesion in terms of growth in size and malignant characteristics is modelled independently. Only advanced adenomas and sessile serrated lesions can progress to CRC. Each year, an asymptomatic tumour can be detected or can progress to a more advanced stage. The natural history model is supplemented with a flexible screening and surveillance component. The model is calibrated to Dutch age- and sex-specific colorectal lesion prevalence rates, and CRC incidence and mortality rates in the absence of screening (Stoop et al., 2012; Cijfers over kanker: Integraal Kankercentrum Nederland, 2017) and is externally validated by replicating several long-term CRC screening trials (Lew et al., 2020). Supplementary Table S1 provides an overview of all model parameters.

2.4. CRC screening and surveillance

To calculate the effectiveness of CRC screening, we simulated two strategies;

1) Neither screening nor surveillance.
2) Biennial FIT screening with colonoscopy surveillance.

2.4.1. No screening and no surveillance

In this strategy, individuals do not undergo screening nor surveillance. Only symptomatic tumours are detected. Once detected, CRC patients have a stage-specific probability of dying due to CRC for ten years. Cancer patients who survive beyond this time are considered cancer survivors in the model and are no longer at risk of dying from CRC.

2.4.2. Biennial FIT screening with colonoscopy surveillance

We set up the model to simulate a cohort of 20,000,000 individuals who underwent four rounds of biennial FIT-based CRC screening with the FOB-Gold (cut-off 47 μg feces/g hemoglobin), starting at age 55. An overview of model parameters related to screening and surveillance is provided in Supplementary Table S1. Adherence was implemented in the model using the observed adherence patterns and the six summarizing adherence measures, resulting in a total of seven approaches to implement adherence to FIT screening. FIT characteristics for detecting colorectal polyps were based on a Dutch FIT screening trial following a previously described calibration procedure (Greuter et al., 2014; van Rossum et al., 2008). Individuals with a positive FIT result are referred for diagnostic colonoscopy. During colonoscopy, all detected lesions are removed by polypectomy, with the exception of small hyperplastic polyps located in the rectosigmoid (Toes-Zoutendijk et al., 2017). We assumed that 92% of FIT positive individuals undergo this procedure (National Institute for Public Health and the Environment (RIVM), Erasmus MC, NKI/Avl, 2018).

Colonoscopy surveillance was modelled according to Dutch guidelines, in which the surveillance interval, i.e. 3 or 5 years, is guided by a risk score based on the number of colorectal lesions as well as their size, location and presence of malignant features (Toes-Zoutendijk et al., 2017). If no adenomas or only one small ($\leq$ 1 cm) tubular adenoma is detected during FIT positive colonoscopy, the individual is not referred to surveillance and returns to screening after 10 years. The participation rate for surveillance colonoscopy was assumed to be equal to that for FIT-positive colonoscopy, i.e. 92%. Based on the surveillance guideline, we assumed that surveillance ends at the age of 75 (Nederlandse Vereniging van Maag-, Darm- en Leverartsen, 2013).

2.5. Model-predicted outcomes

We obtained lifetime model-predicted CRC incidence and mortality rates per 1000 individuals for the no screening strategy and for the FIT screening strategy using the seven approaches to describe adherence. Screening effectiveness was defined as the reduction in model-predicted incidence and mortality due to screening. For each of the six screening strategies using a summarizing measure, this reduction in CRC burden was compared with using the observed adherence patterns as model input. All analyses were performed separately for the Italian, Spanish and Dutch adherence data.

3. Results

3.1. Occurrence of adherence patterns

The Italian, Spanish and Dutch cohorts consisted of 72,980, 56,231 and 11,574 individuals, respectively. The relative frequency of all possible 'Yes-No' adherence patterns over four rounds in the three cohorts is presented in Table 2. The relative frequency of the pattern in which individuals participate in all rounds was slightly higher in the Italian cohort (50.9%) than in the Dutch cohort (45.7%), and considerably lower in the Spanish cohort (26.3%). The six summarizing adherence measures as calculated for the three cohorts are presented in Table 3. The *adherence over all rounds* was highest in the Italian cohort (64.9%), only slightly lower in the Dutch cohort (61.5%) and lowest in the Spanish cohort (42.8%) (Table 3). A high *rescreening rate* was observed in all cohorts (85.3–91.5%).

For each summarizing measure, we calculated the predicted relative frequency of occurrence of all possible 'Yes-No' adherence patterns. To illustrate this, Table 4 shows the calculation of the predicted frequency of one specific pattern, namely 'No-Yes-Yes-No', for all summarizing measures. Supplementary Tables S2, S3 and S4 show the predicted relative frequencies of all adherence patterns based on the summarizing adherence measures and the goodness of fit of the summarizing measures. The predicted relative frequencies using *number of times*

Table 2
The observed relative frequency of occurrence of the adherence patterns in four rounds of FIT-based CRC screening.

				Relative frequency		
Participation in round 1	Participation in round 2	Participation in round 3	Participation in round 4	Italian cohort (%)	Spanish cohort (%)	Dutch cohort (%)
No	No	No	No	23.3	41.1	25.0
No	No	No	Yes	1.9	4.9	4.1
No	No	Yes	No	1.4	2.1	1.6
No	No	Yes	Yes	2.1	3.8	4.4
No	Yes	No	No	1.6	1.8	0.8
No	Yes	No	Yes	1.0	1.3	0.8
No	Yes	Yes	No	1.0	1.1	0.9
No	Yes	Yes	Yes	5.0	6.3	6.3
Yes	No	No	No	2.2	2.4	1.6
Yes	No	No	Yes	0.8	1.1	0.9
Yes	No	Yes	No	0.8	0.7	0.8
Yes	No	Yes	Yes	2.2	2.1	2.8
Yes	Yes	No	No	1.3	1.1	0.7
Yes	Yes	No	Yes	1.7	1.5	1.2
Yes	Yes	Yes	No	2.7	2.1	2.3
Yes	Yes	Yes	Yes	50.9	26.3	45.7

FIT = fecal immunochemical test; CRC = colorectal cancer;
Note: due to rounding, percentages may not always appear to add up to 100%.

Table 3
Summarizing adherence measures to capture longitudinal adherence to four rounds of FIT-based CRC screening.

Summarizing adherence measures	Value (%)		
	Italian cohort	Spanish cohort	Dutch cohort
Adherence over all rounds	64.9	42.8	61.5
Adherence per round			
Round 1	62.7	37.3	56.0
Round 2	65.2	41.7	58.7
Round 3	66.1	44.6	64.8
Round 4	65.7	47.5	66.2
Rescreening rate[1]	91.0	85.3	91.5
Full programme adherence[2]	50.9	26.3	45.7
Regularity			
Consistent screenees	50.9	26.3	45.7
Inconsistent screenees	25.7	32.5	29.3
Never screenees	23.3	41.1	25.0
Number of times participated			
0	23.3	41.1	25.0
1	7.2	11.2	8.1
2	6.9	9.2	8.6
3	11.6	12.1	12.6
4	50.9	26.3	45.7

FIT = fecal immunochemical test; CRC = colorectal cancer;
Note: due to rounding, percentages may not always appear to add up to 100%.
[1] For the rescreening rate it is required that the adherence over all rounds is reported.
[2] The reported proportion is the proportion of individuals who fully adhered to the programme.

participated most closely corresponded with the observed relative frequencies, i.e. this measure led to the lowest chi-square test statistic. Based on the goodness-of-fit, the measure *regularity* also showed good agreement with the observed pattern, followed by the measures *rescreening* and *programme adherence*. For all three cohorts, the predicted relative frequencies of the pattern in which individuals participate in all rounds and the pattern in which individuals never participate in screening were underestimated using the measures *adherence over all rounds* and *adherence per round*, while the predicted relative frequencies of the patterns in which individuals participated in one, two or three rounds were overestimated. The observed relative frequencies of the pattern in which individuals participate in all four rounds were 50.9%, 26.3% and 45.7% in the Italian, Spanish and Dutch cohorts, while the predicted relative frequencies using the measure *adherence over all rounds* were 17.7%, 3.3% and 14.3%. Furthermore, the observed relative frequencies of the pattern in which individuals never participate in the Italian, Spanish and Dutch cohorts were 23.3%, 41.1% and 25.0%, while the predicted frequencies were 1.5%, 10.7% and 2.2%, respectively. Deviations between observed and predicted relative frequencies were similar when using *adherence per round*.

3.2. Impact on model-predicted CRC burden

Model-predicted screening effectiveness is shown in Table 5 and Fig. 1. Using the detailed observed adherence patterns in the Italian, Spanish and Dutch cohort as model input (i.e. the reference), a CRC mortality reduction of 24.4%, 16.9% and 23.5% was predicted for four rounds of CRC screening compared to no screening. The reductions were similar when using the measure *number of times participated*, and also when using the measure *regularity*. Model-predicted screening effectiveness slightly deviated from the predictions under the detailed observed pattern when using the measures *rescreening* and *full programme adherence*. Mortality reduction deviated most for the measures *adherence per round* and *adherence over all rounds*: the predicted mortality reduction was 26.8%, 19.5% and 25.9% using the measure *adherence per round* for the Italian, Spanish and Dutch cohorts, respectively, and 26.8%, 19.4% and 25.7%, using the measure *adherence over all rounds*. Results were similar when screening effectiveness was defined as a reduction in CRC incidence, with slightly smaller deviations between using a summarizing adherence measure and using the observed adherence patterns (Table 5).

4. Discussion

To our knowledge, this is the first study evaluating the accuracy of summarizing measures for longitudinal adherence in capturing the observed adherence patterns and the impact of using different summarizing measures on model-predicted stool-based CRC screening effectiveness. The measure *number of times participated* was most accurate in capturing the observed adherence data, followed by the measures *regularity*, *rescreening* and *full programme adherence*. The measures *adherence per round* and *adherence over all rounds* were least accurate in capturing the observed adherence data. Using a summarizing measure as input for adherence in a CRC screening model instead of the detailed observed adherence patterns led to deviating predictions of screening effectiveness. This deviation was smallest when the measures *number of times participated* or *regularity* were used to derive adherence input for the model. The largest deviation, an absolute overestimation of around 2.5% in CRC mortality reduction, was observed for the measures

Table 4
The observed and predicted relative frequency of the adherence pattern 'No – Yes – Yes – No' for all summarizing adherence measures in the Italian cohort.

Adherence measure	Required parameters	Adherence pattern: No – Yes – Yes – No	Relative frequency
Observed adherence	*NA*	*NA*	= 1.0%
Adherence over all rounds	$p = 0.649$	$(1-p)\times p \times p \times (1-p)$	= 5.2%
Adherence per round	$p_1 = 0.627$ $p_2 = 0.652$ $p_3 = 0.661$ $p_4 = 0.657$	$(1-p_1) \times p_2 \times p_3 \times (1-p_4)$	= 5.5%
Rescreening	$p = 0.649$ $q_1 = 0.910$ $q_0 = \frac{p \times (1-q_1)}{(1-p)} = 0.1662$	$(1-p)\times q_0 \times q_1 \times (1-q_1)$	= 0.5%
Full programme adherence	$p = 0.649$ $\pi_{\text{full}} = 0.509$ $\widetilde{p} = 0.2898$ is the solution of the equation: $\left(1-\widetilde{p}^4\right) \times \left(\pi_{\text{full}} - p\right) + \left(1-\pi_{\text{full}}\right) \times \widetilde{p} \times \left(1-\widetilde{p}^{4-1}\right) = 0$	$\frac{(1-\widetilde{p}) \times \widetilde{p} \times \widetilde{p} \times (1-\widetilde{p}) \times \left(1-\pi_{\text{full}}\right)}{(1-\widetilde{p}^4)}$	= 2.1%
Regularity	$p = 0.649$ $\pi_{\text{consistent}} = 0.509$ $\pi_{\text{never}} = 0.233$ $\widetilde{p} = 0.5758$ is the solution of the equation: $(1-\widetilde{p}^m - (1-\widetilde{p})^m) \times (\pi_{\text{consistent}} - p) + (1-\pi_{\text{consistent}} - \pi_{\text{never}}) \times \widetilde{p} \times \left(1-\widetilde{p}^{m-1}\right)$	$\frac{(1-\widetilde{p}) \times \widetilde{p} \times \widetilde{p} \times (1-\widetilde{p}) \times (1-\pi_{\text{consistent}} - \pi_{\text{never}})}{1-\widetilde{p}^4 - (1-\widetilde{p})^4}$	= 1.8%
Number of times participated	$\pi_0 = 0.233$ $\pi_1 = 0.072$ $\pi_2 = 0.069$ $\pi_3 = 0.116$ $\pi_4 = 0.509$	$\frac{\pi_2}{6}$	= 1.2%

p = adherence over all rounds.
p_i = adherence in round i.
q_1 = the probability of participation for those who have participated in their previous round.
q_0 = the probability of participation for those who have not participated in their previous round.
π_{full} = proportion of individuals who have completed all screening rounds.
$\pi_{\text{consistent}}$ = proportion of individuals who have completed all screening rounds.
π_{never} = proportion of individuals who have never participated.
$\widetilde{p}$ = the probability of participation among not full adherers or for inconsistent screenees for respectively the measures full programme adherence and regularity.
π_i= proportion of individuals who have participated in i screening rounds.
The mathematical functions of the predicted relative frequency of all possible adherence patterns is provided in the Supplementary Methods.

Table 5
Predicted model-based effectiveness of four rounds of biennial FIT screening starting at age 55 using the detailed observed adherence patterns and the six summarizing adherence measures.

Scenario	In the lifetime of a cohort of 1000 individuals	
	CRC cases (% reduction due to screening)	CRC deaths (% reduction due to screening)
No screening	68.7	28.1
4 rounds of screening – Italian cohort		
Observed adherence	56.3 (18.1)	21.3 (24.4)
Adherence over all rounds	55.2 (19.7)	20.6 (26.8)
Adherence per round	55.2 (19.7)	20.6 (26.8)
Rescreening	56.2 (18.2)	21.2 (24.6)
Full programme adherence	56.0 (18.5)	21.1 (25.0)
Regularity	56.3 (18.1)	21.2 (24.4)
Number of times participated	56.3 (18.1)	21.3 (24.4)
4 rounds of screening – Spanish cohort		
Observed adherence	60.2 (12.4)	23.3 (16.9)
Adherence over all rounds	59.1 (14.1)	22.7 (19.4)
Adherence per round	59.0 (14.1)	22.6 (19.5)
Rescreening	60.2 (12.4)	23.4 (16.8)
Full programme adherence	59.9 (12.9)	23.2 (17.6)
Regularity	60.2 (12.4)	23.4 (16.9)
Number of times participated	60.2 (12.4)	23.4 (16.9)
4 rounds of screening – Dutch cohort		
Observed adherence	56.7 (17.4)	21.5 (23.5)
Adherence over all rounds	55.8 (18.9)	20.9 (25.7)
Adherence per round	55.7 (19.0)	20.8 (25.9)
Rescreening	56.9 (17.3)	21.6 (23.3)
Full programme adherence	56.7 (17.7)	21.4 (24.0)
Regularity	56.8 (17.3)	21.5 (23.4)
Number of times participated	56.8 (17.4)	21.5 (23.4)

FIT = fecal immunochemical test; CRC = colorectal cancer;

adherence per round and *adherence over all rounds*. Results were consistent across the three cohorts that differed in observed longitudinal adherence patterns.

The range of deviations (0.0–2.4%) in predicted CRC mortality reduction when using a summarizing measure instead of the complete enumeration of all possible patterns may seem small, but they are of the same order of magnitude as differences observed in model-based comparisons of different screening strategies (Lew et al., 2018; Knudsen et al., 2021). For example, the recommendation of the US Preventive Services Task Force to start screening at 45 years was based on the modeling study performed by Knudsen et al., in which an additional mortality reduction between 3.7% and 4.2% was found if screening would start at 45 years instead of 50 years (US Preventive Services Task Force, 2021; Knudsen et al., 2021). In addition, we hypothesize that the deviation in screening effectiveness will be larger when evaluating the impact of more than four screening rounds, which is the situation for most screening programmes (Schreuders et al., 2015). As an example of the expected larger impact when evaluating a larger number of rounds, note that the predicted probability that an individual never participates in a programme with a high number of rounds is minimal when the adherence over all rounds is relatively high, while the actual proportion of never screenees could be as large as 60% (Benito et al., 2019; Osborne et al., 2017).

We found that CRC screening effectiveness is overestimated by the measures *adherence over all rounds* and *adherence per round*. Previous studies have argued that this is caused by an overestimation of the number of individuals who participate in all rounds (Gellad et al., 2011; Jacobsen and von Euler-Chelpin, 2012). However, we observed an underestimation instead of an overestimation of the number of individuals

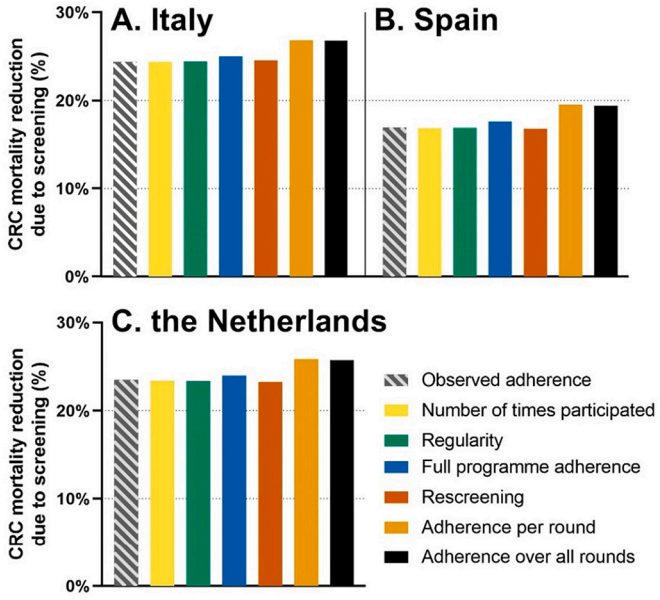

Fig. 1. Model-predicted CRC mortality reduction due to screening using the observed adherence patterns and a summarizing adherence measure as model input. Panel A shows the results for the Italian cohort, panel B for the Spanish cohort and panel C for the Dutch cohort.
CRC = colorectal cancer.

who participate in all rounds. We therefore argue that the overestimation of screening effectiveness is mainly caused by this underestimation. This finding is supported by a modeling study which evaluated CRC screening effectiveness for two types of imperfect adherence at equal adherence over all rounds, namely selective and sporadic adherence (Heisser et al., 2021). With selective adherence, a certain proportion of the population participates in all screening rounds and the remaining proportion of the population never participates. With sporadic adherence, all individuals participate but not at the recommended frequency. This study found higher estimates of CRC screening effectiveness for sporadic adherence, where none of the individuals never participate, compared with selective adherence, where a proportion of the population never participates.

The two best performing adherence measures (*number of times participated* and *regularity*) require longitudinal data for the calculation of these measures. Of these two, the measure *regularity* is most convenient, as it contains the smallest number of parameters to report. The measures *adherence over all rounds* and *adherence per round*, which only require cross-sectional data, were found to be least accurate. Therefore, it is important that longitudinal data on screening adherence over multiple rounds are being collected and reported by CRC screening programmes, although it comes with an increase in complexity and costs. The importance of longitudinal data about screening behaviour and outcomes is increasingly acknowledged by the screening community as it serves multiple purposes including research, monitoring and benchmarking of programmes (Segnan et al., 2022).

Our results can be used to standardize the reporting of longitudinal adherence in FIT-based CRC screening programmes, which is an essential first step to allow international comparison of the performance of programmes. However, to enable comparison between programmes, it is also crucial to incorporate CRC risk factors, particularly when they are related to screening participation, such as age, sex, and sociodemographic status (Klabunde et al., 2015; Deding et al., 2017; Wools et al., 2016). Furthermore, the evaluated measures do not provide information about the number of individuals who drop out and stay out of the programme, which is also relevant for screening programme performances, especially if dropouts are at higher risk of CRC. The fact that screening

protocols differ across programmes, e.g. test and test interval, generates an extra level of complexity. With this in mind, a proposal has been made to share individual-level data on screening events, such as adherence and test results, within a common accessible database that is set up by an international consortium of CRC screening programmes (Segnan et al., 2022).

A limitation of this study is that we did not include the adherence measure proportion of time covered, which is defined as the number of days an individual is compliant with screening divided by the number of days in the cohort (Anderson and Robertson, 2018). This measure may permit the comparison of different screening strategies when intervals are different. As this measure requires detailed information about longitudinal adherence, it is expected that it will quite accurately capture the observed adherence patterns and will result in estimates of model-predicted screening effectiveness that closely approximate those when using the observed adherence patterns as model input.

The population of eligible individuals in a screening programme is a dynamic population. Although we used a closed cohort approach for our analysis, all measures but *number of times participated* can be calculated similarly in a dynamic cohort as in a closed cohort. The *number of times participated* should however be placed in the context of number of times invited and could therefore be expressed as a percentage of number of times offered. However, results may be affected if that definition is used in a dynamic cohort.

Although the ASCCA model has been extensively validated and was found to replicate the findings of long-term randomised controlled trials that compared stool-based screening with no screening, parameter and/or structural uncertainty may have affected our analysis (Lew et al., 2020). However, we expect that the ranking of the measures is not affected, because the bias would be present to the same extent in each analysis, given that each analysis only differed in how adherence is used as input in our model.

To increase the generalizability of our results, multiple sets of real-world adherence data were used in this study. These data originated from different countries with different screening characteristics such as FIT threshold and age of screening. The ASCCA model is calibrated to Dutch CRC incidence and prevalence rates and aspects of CRC screening, such as FIT test characteristics and the surveillance programme were based on the Dutch situation. For example, the compliance to diagnostic colonoscopy was assumed to be 92%, which is higher than observed in Italy (~80%) and Spain (~88%) (Bucchi et al., 2022; Vives et al., 2022). We believe that this does not affect the validity of our results, given that our interest is not in the absolute estimates of screening effectiveness, but in quantifying the difference in screening effectiveness when using different summarizing measures for longitudinal adherence.

4.1. Conclusion

The measures *number of times participated* and *regularity* are most accurate in capturing the observed adherence patterns in CRC FIT-based screening whereas the measures *adherence over all rounds* and *adherence per round* the least. Model-predicted effectiveness of biennial FIT-based CRC screening using the measure *number of times participated* led to comparable model-predicted screening effectiveness using the observed adherence patterns as input, shortly followed by the measure *regularity*. If cross-sectional instead of longitudinal data are available, only measures resulting in deviating predictions of screening effectiveness, i.e. *overall adherence* and *adherence per round*, can be used. This study provides guidance for setting international standards on reporting longitudinal adherence in FIT-based CRC screening, which is an essential first step to allow international comparison of screening programmes.

Data availability

All data that support the findings of this study are included in the manuscript and Supplementary Materials of this study. Detailed data are available from the corresponding author upon reasonable request.

Role of the funding source

No funding to declare.

Ethical compliance

This study was solely based on aggregated data, and thus exempt from ethical approval.

CRediT authorship contribution statement

Francine van Wifferen: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization. **Marjolein J.E. Greuter:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Birgit I. Lissenberg-Witte:** Methodology, Writing – review & editing. **Beatriz Carvalho:** Writing – review & editing. **Gerrit A. Meijer:** Writing – review & editing. **Evelien Dekker:** Writing – review & editing. **Cinzia Campari:** Writing – review & editing. **Montse Garcia:** Data curation, Writing – review & editing. **Linda Rabeneck:** Writing – review & editing. **Iris Lansdorp-Vogelaar:** Data curation, Writing – review & editing. **Carlo Senore:** Data curation, Writing – review & editing. **Veerle M.H. Coupé:** Conceptualization, Methodology, Writing – review & editing, Supervision.

ICSN Colorectal Cancer Screening Working Group

Nereo Segnan[1], Sharon McCarthy[2], Douglas M. Puricelli-Perin[3], Isabel Portillo[4,5], Beate Jahn[6]

[1] Centre for Cancer Prevention, CPO, Piedmonte, Turin, Italy; [2] Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA; [3] Frederick National Laboratory for Cancer Research, Frederick, MD, USA; [4] Osakidetza Basque Health Service, Basque Country Colorectal Cancer Screening Programme, 48011 Bilbao, Spain; [5] Biocruces Health Research Institute, Cancer Biomarker Area, 48903 Barakaldo, Spain; [6] Department of Public Health, Health Services Research and Health Technology Assessment, Institute of Public Health, Medical Decision Making and Health Technology Assessment, UMIT-University for Health Sciences, Medical Informatics and Technology, Eduard-Wallnoefer Zentrum 1, A-6060 Hall in Tirol, Austria

Declaration of Competing Interest

Beatriz Carvalho is inventor on several biomarker patents pending.

GA Meijer is co-founder and board member (CSO) of CRCbioscreen BV, he has a research collaboration with CZ Health Insurances (cash matching to ZonMW grant) and he has research collaborations with Exact Sciences, Sysmex, Sentinel Ch. SpA, Personal Genome Diagnostics (PGDX), DELFI and Hartwig Medical Foundation; these companies provide materials, equipment and/or sample/genomic analyses.

Evelien Dekker has endoscopic equipment on loan of FujiFilm and Olympus, received a research grant from FujiFilm, and received honorarium for consultancy from FujiFilm, Olympus, GI Supply, CPP-FAP, PAION and Ambu, and speakers' fees from Olympus, GI Supply, Norgine, IPSEN, PAION and FujiFilm.

Acknowledgments

No Acknowledgements.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ypmed.2022.107187.

References

Anderson, J.C., Robertson, D.J., 2018. Monitoring compliance with colorectal cancer screening: Do we have it covered? Gastrointest. Endosc. 88, 332–334.

Benito, L., Travier, N., Binefa, G., Vidal, C., Espinosa, J., Milà, N., et al., 2019. Longitudinal Adherence to immunochemical fecal occult blood testing vs Guaiac-based FOBT in an organized colorectal cancer screening program. Cancer Prev. Res. (Phila.) 12, 327–334.

Binefa, G., Garcia, M., Milà, N., Fernández, E., Rodríguez-Moranta, F., Gonzalo, N., et al., 2016. Colorectal cancer screening programme in Spain: Results of key performance indicators after five rounds (2000−2012). Sci Rep. Nat. Publ. Group 6, 19532–19,532.

Bucchi, L., Mancini, S., Baldacchini, F., Ravaioli, A., Giuliani, O., Vattiato, R., et al., 2022. How a fecal immunochemical test screening programme changes annual colorectal cancer incidence rates: an Italian intention-to-screen study. Br. J. Cancer. 127, 541–548.

Bulliard, J.-L., Garcia, M., Blom, J., Senore, C., Mai, V., Klabunde, C., 2014. Sorting out measures and definitions of screening participation to improve comparability: the example of colorectal cancer. Eur. J. Cancer 50, 434–446.

National Monitor National Colorectal Cancer Screening Programme 2017, 2018. National Institute for Public Health and the Environment (RIVM), Erasmus MC, NKI/AvL.

Cijfers over kanker: Integraal Kankercentrum Nederland [Internet]. Available from: https://www.cijfersoverkanker.nl/.

Connell, E., Negrello, T., Meere, D., Australian Institute of Health and Welfare, 2014. Key Performance Indicators for the National Bowel Cancer Screening Program: Technical Report [Internet] [cited 2020 Aug 13]. Available from: https://www.aihw.gov.au/reports/cancer-screening/key-performance-indicators-for-the-national-bowel/contents/table-of-contents.

de Wijkerslooth, T.R., Stoop, E.M., Bossuyt, P.M., Meijer, G.A., van Ballegooijen, M., van Roon, A.H.C., et al., 2012. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am. J. Gastroenterol. 107, 1570–1578.

Deding, U., Henig, A.S., Salling, A., Torp-Pedersen, C., Bøggild, H., 2017. Sociodemographic predictors of participation in colorectal cancer screening. Int. J. Color. Dis. 32, 1117–1124.

Doria-Rose, V.P., Lansdorp-Vogelaar, I., McCarthy, S., Puricelli-Perin, D.M., Butera, V., Segnan, N., et al., 2021. Measures of longitudinal adherence to fecal-based colorectal cancer screening: Literature review and recommended approaches. Int. J. Cancer United States 149, 316–326.

Gellad, Z.F., Stechuchak, K.M., Fisher, D.A., Olsen, M.K., McDuffie, J.R., Ostbye, T., et al., 2011. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am. J. Gastroenterol. 106, 1125–1134.

Giorgi Rossi, P., Vicentini, M., Sacchettini, C., Di Felice, E., Caroli, S., Ferrari, F., et al., 2015. Impact of screening program on incidence of colorectal cancer: a cohort study in Italy. Am. J. Gastroenterol. 110, 1359–1366.

Greuter, M.J.E., Xu, X.-M., Lew, J.-B., Dekker, E., Kuipers, E.J., Canfell, K., et al., 2014. Modeling the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA). Risk Anal. 34, 889–910.

Heisser, T., Cardoso, R., Guo, F., Moellers, T., Hoffmeister, M., Brenner, H., 2021. Strongly divergent impact of adherence patterns on efficacy of colorectal cancer screening: the need to refine adherence statistics. Clin. Transl. Gastroenterol. 12, e00399.

Hol, L., van Leerdam, M.E., van Ballegooijen, M., van Vuuren, A.J., van Dekken, H., Reijerink, J.C.I.Y., et al., 2010. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical fecal occult blood testing and flexible sigmoidoscopy. Gut. England 59, 62–68.

Jacobsen, K.K., von Euler-Chelpin, M., 2012. Performance indicators for participation in organized mammography screening. J. Public Health (Oxf). 34, 272–278.

Klabunde, C., Blom, J., Bulliard, J.-L., Garcia, M., Hagoel, L., Mai, V., et al., 2015. Participation rates for organized colorectal cancer screening programmes: an international comparison. J. Med. Screen. 22, 119–126.

Knudsen, A.B., Rutter, C.M., Peterse, E.F.P., Lietz, A.P., Seguin, C.L., Meester, R.G.S., et al., 2021. Colorectal cancer screening: an updated modeling study for the US preventive services task force. JAMA. 325, 1998–2011.

Lew, J.-B., St John, D.J.B., Macrae, F.A., Emery, J.D., Ee, H.C., Jenkins, M.A., et al., 2018. Benefits, harms, and cost-effectiveness of potential age extensions to the national bowel cancer screening program in Australia. Cancer Epidemiol. Biomark. Prev. 27, 1450–1461.

Lew, J.B., Greuter, M.J., Caruana, M., He, E., Worthington, J., St John, D.J., et al., 2020. Validation of microsimulation models against alternative model predictions and long-term colorectal cancer incidence and mortality outcomes of randomized controlled trials. Med. Dec. Making 40, 815–829.

Moss, S., Ancelle-Park, R., Brenner, H., 2012. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition–Evaluation and interpretation of screening outcomes. Endoscopy Germany 44 (Suppl. 3) (SE49–64).

Nederlandse Vereniging van Maag-, Darm- en Leverartsen. Nederlandse Richtlijn Coloscopie Surveillance 2013, 2013 (Accessed at https://www.mdl.nl/kwaliteitszaken/richtlijnen on 1 December 2020).

Osborne, J.M., Wilson, C., Duncan, A., Cole, S.R., Flight, I., Turnbull, D., et al., 2017. Patterns of participation over four rounds of annual fecal immunochemical test-based screening for colorectal cancer: what predicts rescreening? BMC Public Health 18, 81.

Ponti, A., Anttilla, A., Ronco, G., Senore, C., Basu, P., Segnan, N., et al., 2017 May. Cancer Screening in the European Union (2017): Report on the Implementation of the Council Recommendation on Cancer Screening [Internet]. International Agency for Research on Cancer, Lyon, France (Available from: https://ec.europa.eu/health/sites/health/files/major_chronic_diseases/docs/2017_cancerscreening_2ndreportimplementation_en.pdf).

Schreuders, E.H., Ruco, A., Rabeneck, L., Schoen, R.E., Sung, J.J.Y., Young, G.P., et al., 2015. Colorectal cancer screening: a global overview of existing programmes. Gut. 64, 1637–1649.

Segnan, N., Dekker, E., Doria-Rose, V.P., Senore, C., Rabeneck, L., Lansdorp-Vogelaar, I., International Cancer Screening Network Colorectal Cancer Screening Working Group, 2022. Comparing colorectal cancer screening outcomes in the International Cancer Screening Network (ICSN): A Consortium proposal. Gastroenterology United States 162, 668–674.

Stoop, E.M., de Haan, M.C., de Wijkerslooth, T.R., Bossuyt, P.M., van Ballegooijen, M., Nio, C.Y., et al., 2012. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. Lancet Oncol. 13, 55–64.

Toes-Zoutendijk, E., van Leerdam, M.E., Dekker, E., van Hees, F., Penning, C., Nagtegaal, I., et al., 2017. 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, 767–775.e2.

US Preventive Services Task Force, 2021. Screening for colorectal cancer: US preventive services task force recommendation statement. JAMA. 325, 1965–1977.

van Rossum, L.G., van Rijn, A.F., Laheij, R.J., van Oijen, M.G., Fockens, P., van Krieken, H.H., et al., 2008. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology. 135, 82–90.

Vives, N., Binefa, G., Vidal, C., Milà, N., Muñoz, R., Guardiola, V., et al., 2022. Short-term impact of the COVID-19 pandemic on a population-based screening program for colorectal cancer in Catalonia (Spain). Prev. Med. 155, 106929.

Wools, A., Dapper, E.A., de Leeuw, J.R.J., 2016. Colorectal cancer screening participation: a systematic review. Eur. J. Pub. Health 26, 158–168.