BMC Medicine

RESEARCH

Open Access

Single-round performance of colorectal cancer screening programs: a network meta-analysis of randomized clinical trials

Teruhiko Terasawa[1*], Toshihiro Tadano[2], Koichiro Abe[3], Seiju Sasaki[4], Satoyo Hosono[5], Takafumi Katayama[6], Keika Hoshi[7], Tomio Nakayama[5] and Chisato Hamashima[8]

Abstract

Background Demonstrating mortality reduction in new colorectal cancer (CRC) screening programs through randomized clinical trials (RCTs) is challenging. We systematically reviewed single-round program performance outcomes using a stepwise approach proposed by the World Endoscopy Organization CRC Screening Committee framework.

Methods The MEDLINE, EMBASE, Central, and Ichushi Web databases were searched until October 28, 2024, to find RCTs comparing guaiac-based and immunochemical fecal occult blood testing (gFOBT and FIT), flexible sigmoidoscopy (FS), computed tomographic colonography (CTC), and total colonoscopy (TCS). Paired reviewers screened studies, extracted data, and assessed bias risk. A Bayesian random-effects network meta-analysis was conducted, and the certainty of evidence was evaluated using the Grading of Recommendations Assessment, Development and Evaluation approach. The primary outcome was advanced neoplasia (AN) detection, and the secondary outcomes were participation and colorectal cancer (CRC) detection, all during the first screening round.

Results Eighteen RCTs (437,072 invitees) were included. The risk of bias was low or raised some concerns for screening participation, but it was high for detection outcomes. In the network meta-analysis of 15 RCTs not allowing crossover, the FIT-based program had a higher AN detection rate than the gFOBT-based program (relative risk [RR] 2.48; 95% credible interval [CrI] 1.52–4.21; moderate certainty). AN detection rates were not different in the CTC- (RR 1.01; CrI 0.43–2.23; very low certainty) and TCS-based (RR 1.03; CrI 0.54–1.78; low certainty) programs compared with the FS-based program. All the visualization modality programs had higher AN detection rates than the FIT-based program (FS: RR 2.13 [CrI 1.38–3.77]; CTC 2.16 [1.11–4.51]; and TCS 2.19 [1.43–3.48]; all with low certainty). Low event rates precluded definitive conclusions regarding CRC detection (very low to low certainty). The TCS-based program had the worst participation rate (very low to low certainty). Comparative data allowing crossover were limited.

Conclusions This is the first network meta-analysis that evaluates program-level initial performance indicators. FIT-based programs likely detect more AN cases than gFOBT-based programs, while FS-, CTC-, and TCS-based programs may outperform FIT. Due to limitations in first-round results, long-term outcomes should be assessed after 10–15 years.

Keywords Computed tomographic colonography, Colorectal cancer screening, Fecal immunochemical test, Flexible sigmoidoscopy, Total colonoscopy

*Correspondence:
Teruhiko Terasawa
terasawa@fujita-hu.ac.jp
Full list of author information is available at the end of the article

Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer-specific mortality worldwide [1]. CRC screening reduces the incidence of CRC by detecting and removing advanced adenomas before they progress to cancer [2]. Currently, several screening methods are available, including the guaiac-based fecal occult blood test (gFOBT), fecal immunochemical test (FIT), flexible sigmoidoscopy (FS), total colonoscopy (TCS), and computed tomographic colonography (CTC) among others [3].

Cancer screening is a test-triggered, multistep interventional program that includes subsequent treatments and high-risk surveillance. Demonstrating program efficacy requires large randomized clinical trials (RCTs) with long-term follow-up. Currently, only screening programs with gFOBT or FS have demonstrated reduced CRC incidence and mortality compared with standard care without screening in population-based RCTs [4, 5]. Therefore, to recommend other screening modalities, guideline developers must use evidence from observational studies or microsimulation modeling [6–8].

The World Endoscopy Organization (WEO) CRC Screening Committee recently proposed a four-phased, stepwise approach to efficiently evaluate new and promising screening tests against an established reference test [9, 10]. For a new test successfully assessed in phase I and II studies, in addition to diagnostic accuracy in a screening population, this approach evaluates program performance indicators in population-based RCTs using a standard test that has demonstrated benefits in reducing CRC mortality and incidence as the reference. For RCTs assessing single-round outcomes (phase III), screening participation and detection rates for advanced neoplasia (AN) and CRC are assessed as the initial performance indicators, respectively, for invitation alone and both invitation and the performance of the test itself.

Currently, randomized evidence of reductions in CRC mortality and incidence is insufficient in population-based programs based on primary screening with FIT, CTC, or TCS. Additionally, randomized comparative evidence of multicycle fecal test-based programs versus single-cycle endoscopy-based programs, although beginning to emerge [11–17], is still sparse for a comprehensive evaluation, which is required for the proposed phase IV evaluation [9, 10]. Therefore, formally assessing and synthesizing the available evidence from phase III program performance indicator outcomes compared to those of gFOBT- or FS-based programs can be an informative intermediate step to complement comparative effectiveness. However, previous systematic reviews failed to address the totality of the comparative evidence and only conducted pairwise meta-analyses [18–20]. Given the limited comparisons among all available screening modalities and the clinical heterogeneity in the employed screening algorithms, a random-effects network meta-analysis is the preferred approach to better account for the direct and indirect comparative data from all available randomized evidence in a single analysis [21]. This study aimed to conduct a systematic review and network meta-analysis on the comparative effectiveness of population-based CRC screening RCTs in average-risk adults. We focused on the single-round program performance indicator outcomes using the WEO approach.

Methods

This study was based on an updated evidence review to revise the Japanese National Guidelines for Colorectal Cancer Screening [22]. Similarly to the previous systematic reviews, the evidence review evaluated the comparative evidence among screening programs based on stool-based and endoscopic tests only and performed pairwise meta-analysis of direct data. The protocol for this extended study was not prospectively registered. This report followed the PRISMA Network Meta-Analysis Extension Statement (see Additional file 1) [23]. For complete details on the methods, see Additional file 2: Supplementary Methods.

Information sources

The MEDLINE, EMBASE, Central, and Ichushi Web databases were searched for publications up to October 14, 2023, with no language restrictions. We updated the search using these four databases to identify pertinent reports published between January 1, 2023, and October 28, 2024 (December 14, 2024, for the Ichushi Web only), without language restrictions. Search strategies are detailed in the Additional file 2: Supplementary Methods. Reference lists of eligible studies and relevant review articles were also screened for potentially pertinent studies.

Study selection and data collection

Paired reviewers independently double-screened and determined the eligibility of the selected full-text articles, with discrepancies adjudicated by discussion. We included RCTs with designs that compared screening programs using gFOBT, FIT, FS, CTC, or TCS as the primary test to detect advanced adenomas and invasive cancers in average-risk adults. We included RCTs with any designs irrespective of the unit of randomization (i.e., individual vs cluster), the number of compared screening program groups, or crossover to alternative screening programs. We included RCTs that compared screening programs based on a single primary screening test

(followed by TCS as the post-triage diagnostic test for positive results) or upfront primary screening with TCS.

We excluded RCTs that evaluated screening programs based on sequential testing (e.g., gFOBT triage followed by FIT), screening programs based on parallel testing with two primary screening modalities (e.g., primary screening with combined gFOBT and FS) before diagnostic TCS, or risk-adapted screening programs based on the application of a risk prediction score and risk-based selection of screening tests (e.g., TCS for those with high risk and FIT for those with low risk). When these ineligible screening programs were additionally compared with eligible single-modality screening programs in an RCT, we included the RCT but excluded the ineligible screening programs.

One main data extractor with another data verifier in pairs extracted descriptive data (see Additional file 2: Supplementary Methods for complete details). Paired reviewers independently double-extracted numerical data. Any discrepancies were resolved by consensus. Two attempts at email correspondence with authors of RCTs without extractable data failed to salvage potentially eligible RCTs.

Risk of bias

Paired independent reviewers double-rated the risk of bias using the revised risk of bias tool [24]. Operationally, a study was rated as having appropriately dealt with missing outcome data only when the screen-positive samples had been histologically verified in all invitees or results were statistically corrected for the potential for bias caused by missing outcome data. For screening participation, we considered any reported methods of outcome measurement to be appropriate if they were clinically relevant for the assigned specific modalities (e.g., attendance and receipt for endoscopy and CTC, and direct submission or mailing back of the test samples for any fecal tests), even if the methods differed across the assigned groups. Any discrepancies were resolved via a consensus.

Data synthesis

Our primary outcome was the detection of AN at the program level (i.e., the intention-to-screen [ITS] effect). Secondary outcomes included screening participation and CRC detection as the ITS effect. We separately analyzed RCTs that allowed crossover of the invitees across alternative screening modalities because our focus was on a program-level effect in which screening invitees are exposed to the invitation of a specific screening modality. Relative risks (RRs) were used as the effect measure. As the outcomes were positive, an $RR > 1$ corresponded to the beneficial effects of a comparator *vs.* a reference program. Additionally, these outcomes were assessed in the screening participants only as the per-protocol (PP) effect. In the PP analysis, the rates for screening positivity, positive predictive value (PPV), and number needed to screen (NNTS) for detecting one screening-positive participant with at least one AN lesion, calculated as the inverse of the PPV were assessed as additional outcomes.

To reflect the heterogeneity and uncertainty in the true effects estimated by each RCT, all meta-analyses were based on random-effects models and performed in a fully Bayesian framework [25] using JAGS 4.3.1 (Plummer M. JAGS: Just Another Gibbs Sampler [Software]. https://mcmc-jags.sourceforge.net) from R version 4.4.2 (R Core Team. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/) or StataNow 18.5/SE (StataCorp. Texas, USA) [26]. Summary RRs and their 95% credible intervals (CrIs), the surface under the cumulative ranking curve (SUCRA), and the probability of ranking [27, 28] were calculated with a Bayesian random-effects network meta-analysis based on binomial likelihood and log link [29] using the GEMTC package [30] and MetaInsight [31]. For cluster RCTs, the variance of an effect estimate was conventionally increased using the C adjustment [32, 33]. We conducted a Bayesian random-effects meta-analysis based on binomial likelihood and probit link to calculate summary estimates of proportions, including screening positivity and PPV [34]. For details of the C adjustment [32, 35–37], model fitting [38], and selection of prior distributions for parameters assessed [39], see Additional file 2: Supplementary Methods.

The across-study heterogeneity of direct evidence was first visually assessed using forest plots and then quantified as the between-study standard deviation, *tau*. Next, inconsistencies between the direct and indirect evidence were evaluated globally by comparing the mean residual deviance (MRD) and deviance information criterion (DIC) statistics between the primary network meta-analysis model (i.e., the consistency model) and an unrelated mean effect (UME) inconsistency model [40] and locally by performing node-splitting models [41, 42]. Funnel plot analysis was not conducted since no contrasts comprised ≥ 10 studies [43]. Few RCTs without CRC detection data precluded the risk of bias due to missing evidence [44].

We performed a network meta-analysis for a subgroup of RCTs conducted in Europe only to address clinical heterogeneity arising from geographical and cultural variability. For age (59 years and younger vs. 60 years and older) and sex, the paucity of data on these individual-level subgroup outcomes precluded a full network meta-analysis; only a pairwise random-effects meta-analysis of the interaction was performed [45]. Additionally, to evaluate the stability of the C adjustment and the inclusion of

cRCTs, we performed sensitivity network meta-analyses using an alternative, conventional C value or excluding all cRCTs from the analysis.

Assessing the certainty of the evidence

To evaluate the certainty of the evidence, we highlighted the comparative effectiveness of selected comparison pairs on the ITS effects following the WEO-proposed phase III framework [10] and applied the Grading of Recommendation Assessment, Development, and Evaluation (GRADE) tool for network meta-analysis [46]. First, we used gFOBT and FS as reference screening tests for FOB-based modalities (defined as gFOBT and FIT) and visualization modalities (defined as FS and TCS [as direct visualization methods] and CTC [as indirect visualization methods]), respectively [10]. Second, we extended the comparative effectiveness of all visualization modalities being compared with FIT as the reference test since this was the most commonly recommended modality worldwide [47, 48]. In each comparison, we evaluated the risk of bias, inconsistency, indirectness, and intransitivity of direct and indirect evidence and then assessed incoherence between direct and indirect evidence and imprecision of network evidence to rate the certainty of the evidence according to the Pragmatic Guide [46]. The indirectness of this outcome was operationally downgraded by default because participation in screening is an intermediate outcome before the detection of AN and CRC. One reviewer (T.T.) assessed each domain for each comparison per outcome, with uncertainties resolved via a discussion with an internal cancer screening expert (C.H.). To obtain illustrative estimated absolute risk differences (RDs), we assumed an average-risk screening population of 10,000 individuals, where the absolute event rate of the reference program was calculated based on the relevant event data from all included RCTs (e.g., event data from all gFOBT arms in case the comparison between gFOBT- and FIT-based programs were combined) using random-effects meta-analysis of proportions. For all other comparisons, we did not formally apply the GRADE approach and only explored the comparative effectiveness based on the SUCRA and ranking analysis [27].

Results

Study selection

Our literature search identified 11,274 citations, which included 18 eligible RCTs (14 from Europe, 2 from the US, and 1 each from Israel and China) [13–15, 49–63], involving a total of 437,072 average-risk screening invitees (Fig. 1; Additional file 2: Fig. S1). The details of the excluded studies are provided in Additional file 2: List of Excluded Study Reports [11, 16, 17, 64–93].

Study characteristics

Eleven RCTs had a standard, two-armed individual randomization design [13, 14, 49, 51, 54, 58–63], three had a three-armed individual randomization design [15, 53, 56], and the remaining four had a two-armed, cluster design [50, 52, 55, 57] (Table 1). Three RCTs that included FIT- *vs.* gFOBT- [50], FIT- *vs.* FS- [52], and gFOBT- *vs.* TCS-based programs [55] allowed the participants to crossover to an alternative screening test program. The most assessed programs were FIT-based, which were commonly compared with gFOBT-based programs (five trials; 51,614 invitees), TCS-based programs (six trials; 180,101 invitees), and FS-based programs (four trials; 183,603 invitees). Included RCTs consistently excluded high-risk individuals and similarly defined AN (Additional file 2: Table S1).

The most common target age group was 50–75 years. The median number of invitees was 12,166 (range, 1513–154,743), with a median percentage of male participants of 47% (range, 36–51%).

The interventions for gFOBT- and FIT-based programs differed for the applied test kits, positive cutoff values of hemoglobin concentrations, and delivery methods (Additional file 2: Table S2). Test kits were mailed directly with the screening invitation letter in approximately 50% (4/8 for gFOBT and 6/13 for FIT) of the programs. Conversely, the bowel preparation methods and positivity criteria were relatively similar within and across the FS-, CTC-, and TCS-based programs (Additional file 2: Table S3).

Risk of bias

For screening participation, the overall risk of bias was generally rated low, or some concerns due to inappropriate randomization, the possibility of selective reporting, or both (Additional file 2: Fig. S2–S4). For detection outcomes, all RCTs were deemed to have a high risk of bias due to missing outcome data, inappropriate measurement of outcomes for evaluating ITS effects, and deviations from intended interventions for evaluating PP effects.

RCTs not allowing crossover

Screening participation

Fourteen RCTs involving 357,576 screening invitees, provided data (Fig. 2A; Additional file 2: Table S4; and Additional file 2: Fig. S5). No evidence of inconsistency was detected globally (MRD = 31.20 and DIC = 62.06 in the consistency model versus MRD = 31.21 and DIC = 62.16 in the UME model) or locally for all eight contrasts (Bayesian *P*-value range, 0.40–0.99). Median participation rates for programs based on gFOBT, FIT, FS, CTC, and TCS were 31.6% (range, 18.5–49.5%), 54.3%

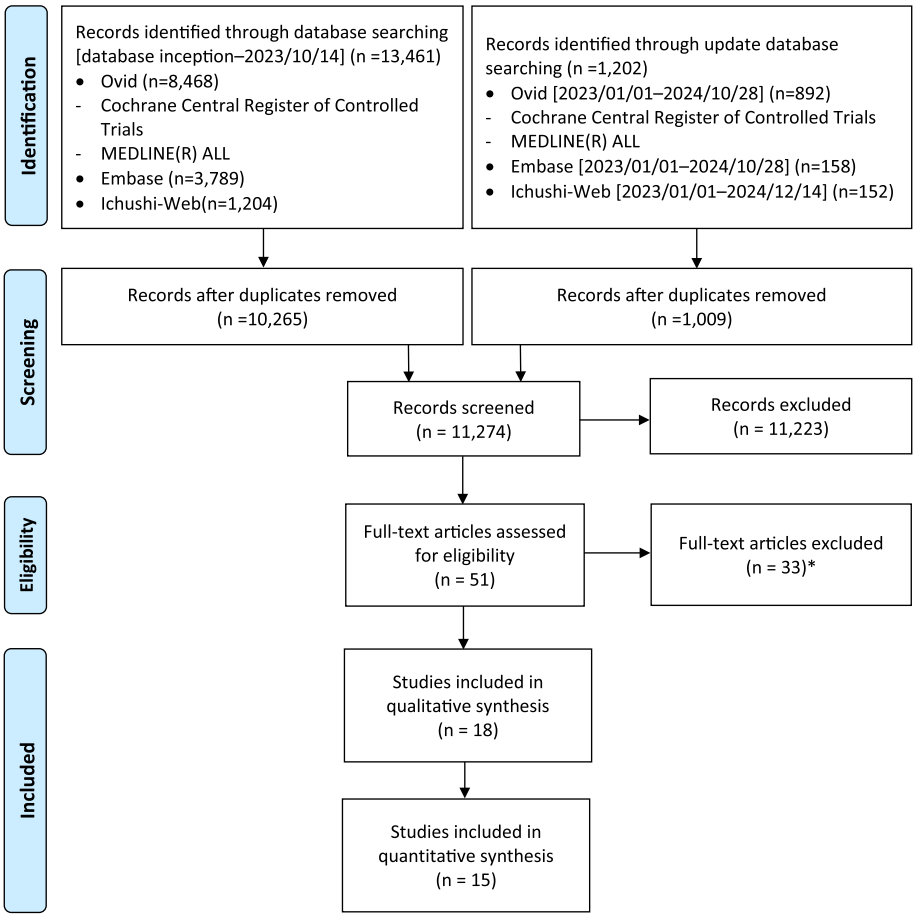

Fig. 1 Study flow diagram. *See Supplementary documents for excluded study reports and reasons for exclusion

(28.6–96.1%), 32.3% (7.9–52.1%), 30.1% (26.7–33.6%), and 24.0% (10.0–42.5%), respectively.

In the network meta-analysis, the participation rate for the FIT-based program was similar to that for the gFOBT-based program (RR 0.98; CrI 0.69–1.40), with an estimated RD of 72 fewer participations per 10,000 invitees, ranging from 1120 fewer to 1445 more per 10,000. This wide credibility interval indicated that there could still be clinically significant increases and decreases in screening participation (very low certainty) (Fig. 3 and Table 2). Compared with that in the FS-based program, the participation rate was not significantly lower in the CTC-based program (RR 0.91; CrI 0.45–1.83), with an estimated RD of 306 fewer participations per 10,000 invitees, ranging from 1869 fewer to 2820 more per 10,000. This wide credibility interval suggested that there could still be clinically significant increases and decreases in screening participation (very low certainty). In contrast, the participation rate was marginally significantly lower in the TCS-based program (RR 0.62; CrI 0.39–1.01), resulting in an RD of 1291 fewer participations per 10,000 screening invitees, ranging from 2073 fewer to 34 more per 10,000. This indicated that the TCS-based program could decrease participation rates (low certainty). When compared with the FIT-based program as the reference, the participation rate was significantly lower in the TCS-based program only (RR 0.49; CrI 0.34–0.73), resulting in an RD of 2627 fewer participations per 10,000 screening invitees, ranging from 3400 to 1391 fewer per 10,000. This suggested that the TCS-based program could cause a clinically significant decrease in participation rates (low certainty). In contrast, the participation rate was not significantly lower in the FS-based program (RR 0.79; CrI 0.55–1.15), with an estimated RD of 1082 fewer participations per 10,000 screening invitees, ranging from 2318 fewer to 773 more per 10,000. Similarly, the participation rate was not significantly lower in the CTC-based program (RR 0.72; CrI 0.39–1.35), with an estimated RD of 1442 fewer participations per 10,000 invitees, ranging from 3142 fewer to 1803 more per 10,000. These wide credibility intervals

Table 1 Study and participant characteristics

Author and year	Study location	Study years	Unit of randomization	Target age, years	Mean (SD) or median [IQR] age, years	% Male	Screening test	Sample size, *n*
Studies not allowing crossover								
Verne 1998 [49]	UK, East Hertfordshire	ND	Individual	50–75	ND	ND	FS *vs.* gFOBT	2494
Federici 2005 [50]	Italy, Lazio	ND	Cluster[a]	50–74	ND	46	FIT *vs.* gFOBT	7320
Segnan 2005 [51]	Italy, nationwide	1999–2001	Individual	55–64	ND	47	FIT *vs.* FS	23,051
Federici 2006 [52]	Italy, Rome	ND	Cluster[a]	50–74	ND	44	FS *vs.* gFOBT	2987
Segnan 2007 [53]	Italy, nationwide	2002–2004	Individual	55–64	ND	48	FIT *vs.* FS *vs.* TCS	18,114
van Rossum 2008 [54]	Netherlands, multiple cities	2006–2007	Individual	50–75	60.7 (7.1)	48	FIT *vs.* gFOBT	20,623
Lisi 2009 [55]	Italy, nationwide	2003–2006	Cluster[a]	55–64	ND	ND	TCS *vs.* gFOBT	9889
Hol 2010 [56]	Netherlands, Rijnmond	ND	Individual	50–74	ND	ND	FIT *vs.* FS *vs.* gFOBT	15,011
Levi 2011 [57]	Israel, Tel Aviv	ND	Cluster[b]	50–75	61.0 (7.6)	44	FIT *vs.* gFOBT	12,537
Stoop 2012 [59]	Netherlands, Amsterdam and Rotterdam	2009–2010	Individual	50–75	60.8 (6.6)	50	CTC *vs.* TCS	8844
Sali 2016 [15]	Italy, Florence	2012–2014	Individual	54–65	59.0 (3.6)	46	CTC *vs.* FIT *vs.* TCS	16,087
Regge 2017 [62]	Italy, Piedmont	2012–2013	Individual	58	58.6 (1.08)	51	CTC *vs.* FS	5412
Chen 2020 [63]	China, five provinces	2018–2019	Individual	50–74	60.5 (6.5)	42	FIT *vs.* TCS	11,795
Randel 2021 [14]	Norway, two southeast regions	2012–2019	Individual	50–74	60 [54-66]	49	FIT *vs.* FS	154,743
Forsberg 2022 [13]	Sweden, 18 regions	2014–2020	Individual	60	60	50	FIT *vs.* TCS	91,440
Studies allowing crossover								
Quintero 2012 [58]	Spain, eight regions	2009–2011	Individual	50–69	59.2 (5.6)	46	FIT *vs.* TCS	57,404
Chubak 2013 [60]	USA, Washington	2010–2011	Individual	50–74	58.5 (6.0)	43	FIT *vs.* gFOBT	1513
Gupta 2013 [61]	USA, Texas	2011–2012	Individual	50–64	59 (3)	36	FIT *vs.* TCS	2080

CTC Computed tomography colonography, *FIT* Fecal immunochemical test, *FS* Flexible sigmoidoscopy, *gFOBT* Guaiac fecal occult blood test, *ND* No data, *TCS* Total colonoscopy

[a] General practitioners were randomized

[b] Primary care clinics were randomized

indicated that there could still be clinically significant increases and decreases in screening participation (low certainty).

The ranking analysis and SUCRA plots showed that the gFOBT- and FIT-based programs were the two best screening programs, with the probability of being the best at 48 and 39%, respectively (Additional file 2: Table S5 and Additional file 2: Fig. S6). Conversely, the TCS-based program was the worst with the probability of being the worst at 94%.

AN detection

Thirteen RCTs involving 334,900 screening invitees provided data on the ITS effect (Fig. 2B, Additional file 2: Table S6 and Additional file 2: Fig. S7). No inconsistency was suggested globally (MRD = 29.68 and DIC = 54.21 in the consistency model versus MRD = 29.34 and DIC = 55.65 in the UME model) or locally for all eight contrasts (*P*-value range, 0.15–0.99). The median AN detections per 100,000 invitees for programs based on gFOBT, FIT, FS, CTC, and TCS were 279 (range,

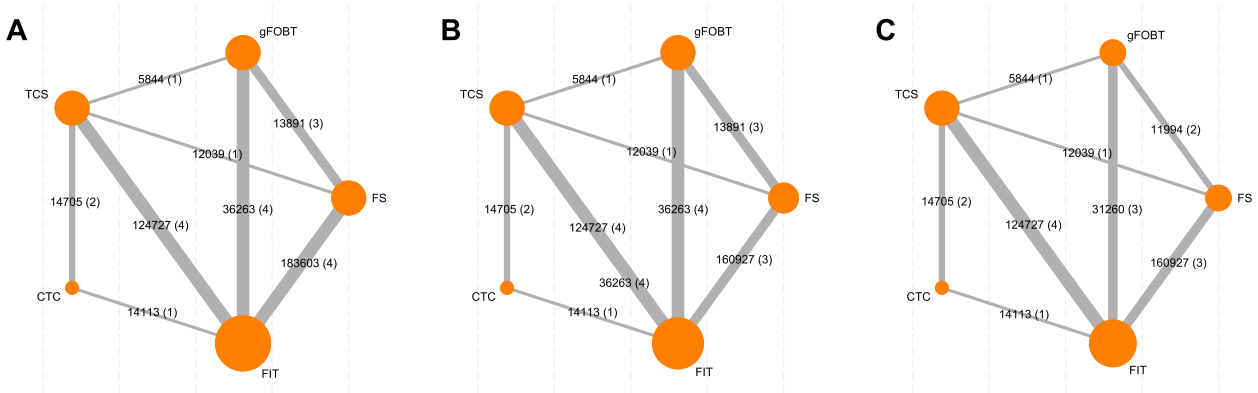

Fig. 2 Geometry of the network of included randomized clinical trials without crossover. Network diagrams showing included randomized clinical trials (RCTs) for screening participation (A), and advanced neoplasia and colorectal cancer detection (B and C , respectively) evaluated for the intention-to-screen effect. Numbers indicate the number of trial participants, followed by the number of RCTs that compared specific screening test programs in parentheses. The size of the nodes and the thickness of the links represent the number of trial participants and the number of RCTs evaluated, respectively. CTC = computed tomography colonography; FIT = fecal immunochemical test; FS = flexible sigmoidoscopy; gFOBT = guaiac-based fecal occult blood test; TCS = total colonoscopy

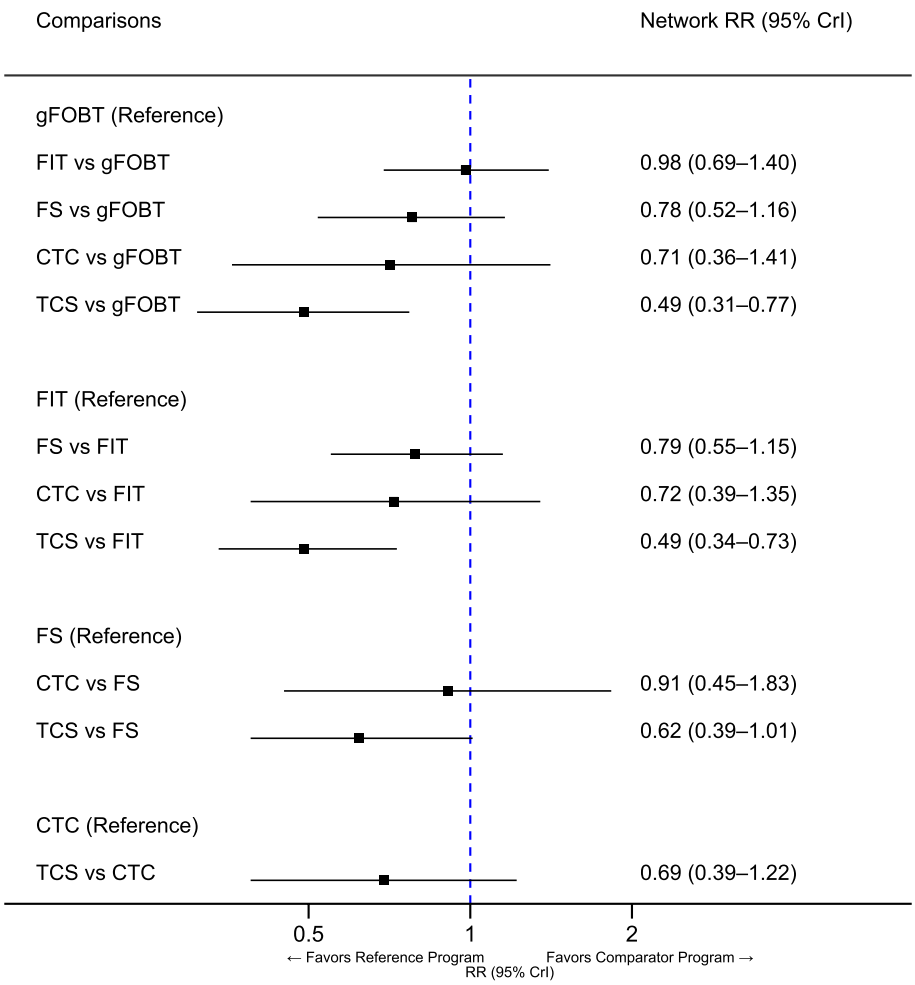

Fig. 3 Network meta-analysis of screening participation. Squares and solid horizontal lines represent the summary risk ratios and their 95% credible interval (CrI), respectively, estimated by network meta-analysis. CTC = computed tomography colonography; FIT = fecal immunochemical test; FS = flexible sigmoidoscopy; gFOBT = guaiac-based fecal occult blood test; TCS = total colonoscopy

Table 2 GRADE assessment of single-round program-level performance indicators of colorectal cancer screening programs

Outcomes; **referent and target screening programs**	**Assessment of direct (upper row) and indirect (lower row) evidence**						**Assessment of network evidence**		**GRADE**	**Summary of network findings**	**Absolute event rate of reference**	**Risk difference,** ***per 10,000 screening invitees***
	Participants in direct evidence (studies), ***n***	**Risk of bias**	**Inconsistency**	**Indirectness**	**Publication bias**	**Intransitivity (Indirect evidence only)**	**Incoherence**	**Imprecision**		**Risk ratio (95% CrI)**	**Average,[a]** ***per 10,000 screening invitees***	
AN detection												
gFOBT												
vs FIT	36,263 (4)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	No serious limitations	Moderate	2.48 (1.52–4.21)	38	56 more (20 more to 122 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
FS												
vs CTC	—	—	—	—	—	—	No serious limitations	Serious limitations[e,l]	Very low	1.01 (0.43–2.23)	265	3 more (151 fewer to 326 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
vs TCS	12,039 (1)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e]	Low	1.03 (0.54–1.78)	265	8 more (122 fewer to 207 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
FIT												
vs FS	160,927 (3)	Serious limitations[b]	Serious limitations[c]	No serious limitations	Not assessed	—	No serious limitations	No serious limitations	Low	2.13 (1.38–3.77)	109	123 more (41 more to 302 more)
		Serious limitations[b]	No serious limitations	Serious limitations[d]	Not assessed	No serious limitations						
vs CTC	14,113 (1)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	No serious limitations	Low[f]	2.16 (1.11–4.51)	109	126 more (12 more to 383 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
vs TCS	124,727 (4)	Serious limitations[b]	Serious limitations[c]	No serious limitations	Not assessed	—	No serious limitations	No serious limitations	Low	2.19 (1.43–3.48)	109	130 more (47 more to 270 more)
		Serious limitations[b]	No serious limitations	Serious limitations[d]	Not assessed	No serious limitations						
CRC detection												
gFOBT												
vs FIT	31,260 (3)[g]	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e]	Low	2.25 (0.94–6.92)	7	9 more (0 to 41 more)
		Serious limitations[b]	No serious limitations	Serious limitations[d]	Not assessed	No serious limitations						
FS												
vs CTC	—	—	—	—	—	—	No serious limitations	Serious limitations[e]	Very low	2.11 (0.52–8.76)	26	29 more (12 fewer to 202 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						

Table 2 (continued)

Outcomes; referent and target screening programs	Assessment of direct (upper row) and indirect (lower row) evidence						Assessment of network evidence		GRADE	Summary of network findings	Absolute event rate of reference	Risk difference, *per 10,000 screening invitees*
	Participants in direct evidence (studies), *n*	Risk of bias	Inconsistency	Indirectness	Publication bias	Intransitivity (Indirect evidence only)	Incoherence	Imprecision		Risk ratio (95% CrI)	Average,[a] *per 10,000 screening invitees*	
vs TCS	12,039 (1)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e,l]	Low	1.35 (0.49–3.42)	26	9 more (13 fewer to 63 more)
		Serious limitations	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
FIT												
vs FS	160,927 (3)	Serious limitations[b]	Serious limitations[c]	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e,l]	Very low	1.25 (0.64–3.15)	17	4 more (6 fewer to 37 more)
		Serious limitations[b]	No serious limitations	Serious limitations[d]	Not assessed	No serious limitations						
vs CTC	14,113 (1)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e,l]	Low	2.69 (0.85–10.34)	17	29 more (3 fewer to 159 more)
		Serious limitations[b]	No serious limitations	Serious limitations[d]	Not assessed	No serious limitations						
vs TCS	124,727 (4)	Serious limitations[b]	No serious limitations	No serious limitations	Not assessed	—	No serious limitations	Serious limitations[e,l]	Low	1.70 (0.85–3.87)	17	12 more (3 fewer to 49 more)
		Serious limitations[b]	Serious limitations[c]	Serious limitations[d]	Not assessed	No serious limitations						
Screening participation												
gFOBT												
vs FIT	36,263 (4)	No serious limitations	Serious limitations[c]	Serious limitations[h]	Not assessed	—	No serious limitations	Serious limitations[e,l]	Very low	0.98 (0.69–1.40)	3613	72 fewer (1120 fewer to 1445 more)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed	No serious limitations						
FS												
vs CTC	—	—	—	—	—	—	No serious limitations	Serious limitations[e,l]	Very low	0.91 (0.45–1.83)	3398	306 fewer (1869 fewer to 2820 more)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed							
vs TCS	12,039 (1)	No serious limitations	No serious limitations	Serious limitations[h]	Not assessed	—	No serious limitations	Serious limitations[e,l]	Low	0.62 (0.39–1.01)	3398	1291 fewer (2073 fewer to 34 more)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed	No serious limitations						
FIT												
vs FS	183,603 (4)	No serious limitations	Serious limitations[c]	Serious limitations[h]	Not assessed	—	No serious limitations	Serious limitations[e,l]	Very low	0.79 (0.55–1.15)	5151	1082 fewer (2318 fewer to 773 more)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed	No serious limitations						

Table 2 (continued)

Outcomes; referent and target screening programs	Assessment of direct (upper row) and indirect (lower row) evidence						Assessment of network evidence		GRADE	Summary of network findings	Absolute event rate of reference	Risk difference, *per 10,000 screening invitees*
	Participants in direct evidence (studies), *n*	Risk of bias	Inconsistency	Indirectness	Publication bias	Intransitivity (Indirect evidence only)	Incoherence	Imprecision		Risk ratio (95% CrI)	Average,[a] *per 10,000 screening invitees*	
vs CTC	14,113 (1)	No serious limitations	No serious limitations	Serious limitations[h]	Not assessed	—	No serious limitations	Serious limitations[e,l]	Low	0.72 (0.39–1.35)	5151	1442 fewer (3142 fewer to 1803 more)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed	No serious limitations						
vs TCS	124,727 (4)	No serious limitations	Serious limitations[c]	Serious limitations[h]	Not assessed	—	No serious limitations	No serious limitations	Low	0.49 (0.34–0.73)	5151	2627 fewer (3400 fewer to 1391 fewer)
		No serious limitations	Serious limitations[c]	Serious limitations[d, h]	Not assessed	No serious limitations						

CrI Credible interval, *CTC* Computed tomography colonography, *FIT* Fecal immunochemical test, *FS* Flexible sigmoidoscopy, *gFOBT* Guaiac fecal occult blood test, *GRADE* Grading of Recommendations Assessment, Development and Evaluation, *RCT* Randomized clinical trial, *TCS* Total colonoscopy

[a] Average event rates were estimated based on a random-effects meta-analysis of proportions except for the CRC detection rate in the gFOBT-based program, where a fixed-effect meta-analysis for rare events was performed

[b] Due to high risk of bias

[c] Statistical heterogeneity was suggested by forest plots

[d] Indirect evidence was downgraded by default

[e] The summary estimates crossed the null effect

[f] Rated down because direct evidence was based on a single RCT

[g] One study, a cluster-RCT [57], was excluded due to zero events in both compared programs

[h] Screening participation, an intermediate outcome of AN and CRC detection, was downgraded by default

141–584), 1133 (379–2193), 2362 (219–2747), 1722 (1389–2055), and 1875 (677–2400), respectively.

In the network meta-analysis, the FIT-based program had a higher AN detection rate than the gFOBT-based program (RR 2.48; CrI), resulting in an increase in the RD of 56 more detections per 10,000 screening invitees, ranging from 20 to 122 more per 10,000. This indicated that the FIT-based program could significantly improve the detection rates (moderate certainty) (Fig. 4 and Table 2). As compared with the FS-based program, there was no significant difference in the AN detection rates in the CTC-based program (RR 1.01; CrI 0.43–2.23), with an estimated RD of 3 more detections per 10,000 screening invitees, ranging from 151 fewer to 326 more per 10,000. Similarly, the AN detection rates did not show a significant difference in the TCS-based program (RR 1.03; CrI 0.54–1.78), with an estimated RD of 8 more detections per 10,000 invitees, ranging from 122 fewer to 207 more per 10,000. These wide credibility intervals indicated that there could still be clinically significant increases and decreases in AN detections (very low to low certainty). Compared with those in the FIT-based program as a reference, the detection rates of all the visualization modality programs were significantly higher. The RRs of AN detection were 2.13 (CrI 1.38–3.77) for FS, 2.16 (CrI 1.11–4.51) for CTC, and 2.19 (CrI 1.43–3.48) for TCS. These results indicate clinically important and comparable increases in AN detections in all visualization modality programs: 123 more, ranging from 41 to 302 more for FS; 126 more, ranging from 12 to 383 more for CTC; and 130 more, ranging from 47 to 270 more for TCS, all per 10,000 screening invitees.

In the ranking analysis, the gFOBT-based program was the worst in detecting AN with the probability of being the worst of 100%. However, no program was

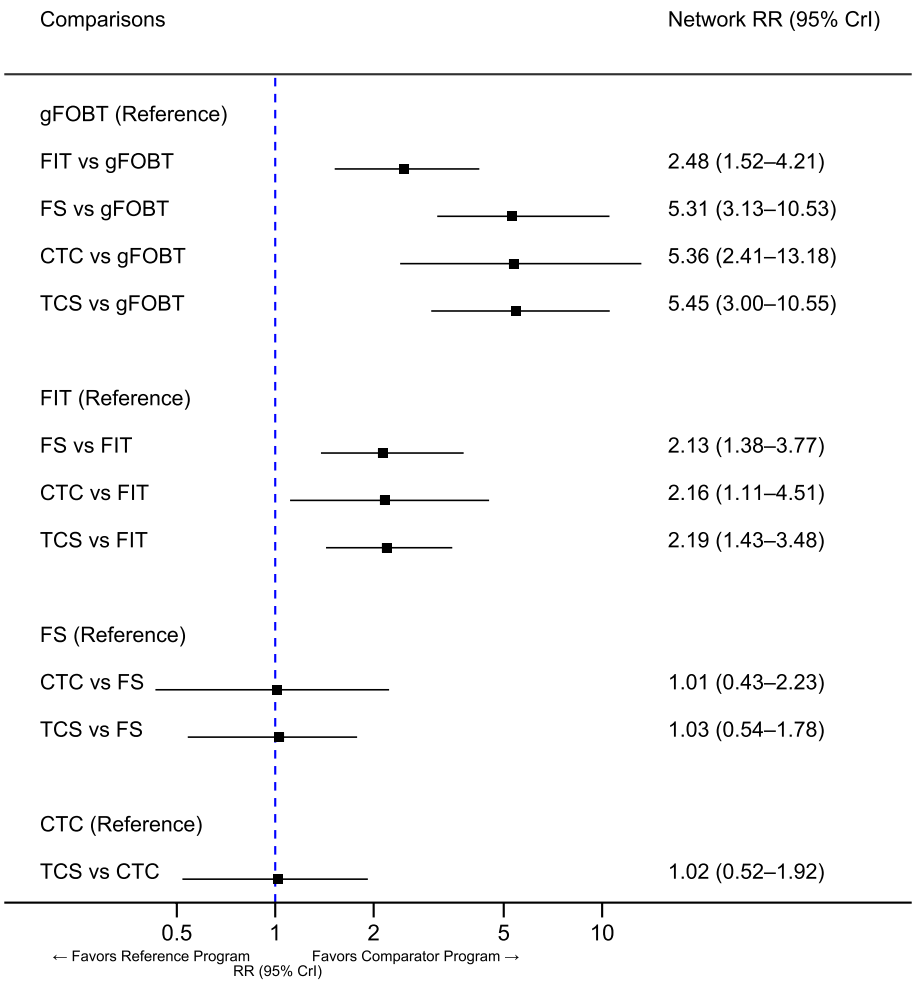

Fig. 4 Network meta-analysis of AN detection for evaluating the intention-to-screen effect. Squares and solid horizontal lines represent the summary risk ratios and their 95% credible interval (CrI), respectively, estimated by network meta-analysis. AN = advanced neoplasia; CTC = computed tomography colonography; FIT = fecal immunochemical test; FS = flexible sigmoidoscopy; gFOBT = guaiac-based fecal occult blood test; TCS = total colonoscopy

superior to any other among the FS-, CTC-, and TCS-based programs, with the best program of these three modalities ranging from 31 to 36% (Additional file 2: Table S7 and Additional file 2: Fig. S8).

In 14 RCTs evaluating the PP effect, TCS was the best screening modality with a mean ranking of 1 and the probability of being the best modality at 99% (Additional file 2: Tables S8–S9). The detection rate of AN lesions was higher among TCS recipients than among those of all other screening modalities other than CTC, CTC (RR 1.37; CrI 0.91–2.09), FS (RR 1.44; CrI 1.00–2.10), FIT (RR 4.06; CrI 3.05–5.79), and gFOBT (RR 8.71; CrI 5.61–14.53) (Additional file 2: Fig. S9–S11).

CRC detection

Ten RCTs involving 327,005 screening invitees provided data on the ITS effect (Fig. 2C, Additional file 2: Table S10, and Additional file 2: Fig. S12). One cluster design RCT with no events in all arms was excluded from the analysis [57]. No inconsistency was suggested globally (MRD = 26.04 and DIC = 43.62 in the consistency model versus MRD = 25.27 and DIC = 44.50 in the UME model) or locally for all eight contrasts (*P*-value range, 0.06–0.87). Median CRC detections per 100,000 invitees for programs based on gFOBT, FIT, FS, CTC, and TCS were 102 (range, 0–125), 179 (33–289), 220 (170–292), 158 (145–171), and 140 (0–230), respectively.

In contrast to AN detection, the CrIs of all summary estimates were wide (Fig. 5 and Table 2). The CRC detection rate for the FIT-based program was marginally significantly higher than that for the gFOBT-based program (RR 2.25; CrI 0.94–6.92), resulting in an increase in the RD of 9 more per 10,000 screening invitees, ranging from almost zero to 41 more per 10,000 screening

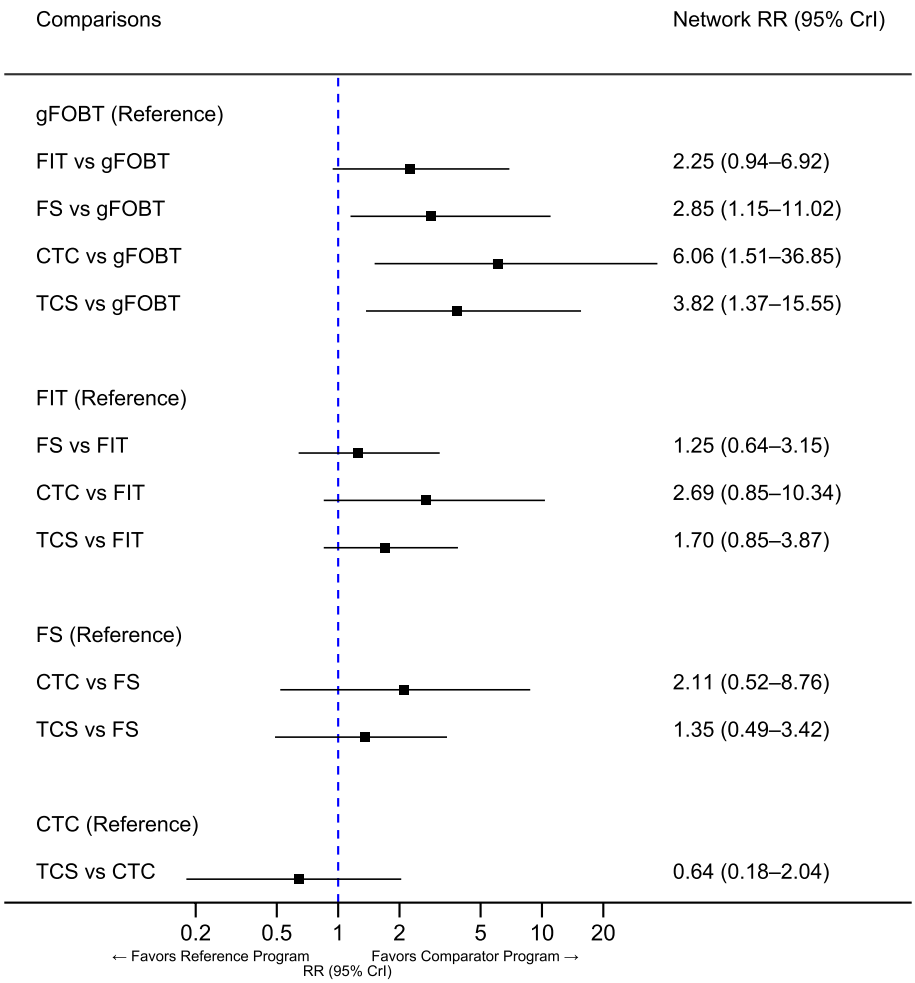

Fig. 5 Network meta-analysis of CRC detection for evaluating the intention-to-screen effect. Squares and solid horizontal lines represent the summary risk ratios and their 95% credible interval (CrI), respectively, estimated by network meta-analysis. CRC = colorectal cancer; CTC = computed tomography colonography; FIT = fecal immunochemical test; FS = flexible sigmoidoscopy; gFOBT = guaiac-based fecal occult blood test; TCS = total colonoscopy

invitees. This suggested that the FIT-based program could increase the detection of CRC (low certainty). As compared with the FS-based program, the rates of CRC detection did not show a significant difference in the CTC-based program (RR 2.11; CrI 0.52–8.76), with an estimated RD of 29 more detections per 10,000 screening invitees, ranging from 12 fewer to 202 more per 10,000. Similarly, CRC detection rates were not significantly different in the TCS-based program (RR 1.35; CrI 0.49–3.42), with an estimated RD of 9 more detections per 10,000 invitees, ranging from 13 fewer to 63 more per 10,000. These wide credibility intervals suggested that there could still be clinically significant increases and decreases in CRC detections (very low to low certainty). Compared with those in the FIT-based program as a reference, the detection rates of all the visualization modality programs were not significantly higher. The RRs of AN detection were 1.25 (CrI 0.64–3.15) for FS, 2.69 (CrI 0.85–10.34) for CTC, and 1.70 (CrI 0.85–3.87) for TCS. These results similarly included clinically significant increases and the possibility of decreases in CRC detections in all the visualization modality programs: 4 more, ranging from 6 fewer to 37 more for FS; 29 more, ranging from 3 fewer to 159 more for CTC; and 12 more, ranging from 3 fewer to 270 more for TCS, all per 10,000 screening invitees.

In the ranking analysis, the gFOBT-based program was, again, the poorest at detecting CRC, with a 96% probability of being the worst (Additional file 2: Table S11 and Additional file 2: Fig. S13). Conversely, the CTC- and TCS-based programs were the two best screening programs: the probabilities of being the best modality for the CTC- and TCS-based programs were 76% and 17%, respectively.

In 12 RCTs evaluating the PP effect, TCS and CTC were the two best screening modalities, with probabilities of being the best modality of 76 and 17%, respectively (Additional file 2: Tables S12–S13). The detection rate of CRC was higher among TCS recipients than in those of FIT (RR 2.54; CrI 1.63–5.13) and gFOBT (RR 4.89; CrI 2.12–14.34), whereas no difference was found in the CRC detection among recipients between TCS and FS (RR 1.69; CrI 0.89–3.36) and between TCS and CTC (RR 0.98; CrI 0.41–2.40) (Additional file 2: Fig. S14–S16).

Screening positivity, PPV, and NNTS

Although the summary point estimates of screening positivity and PPV were ordered from lowest to highest for gFOBT, FIT, FS, and CTC, statistical heterogeneity across studies was suggested, with widely reported percentages of individual studies and wide CrIs for summary estimates (Table 3 and Additional file 2: Fig. S17–S18). The summary NNTS for identifying AN for gFOBT, FIT, FS, and CTC was 3.7 (CrI 2.3–6.3), 3.5 (CrI 2.5–5.0), 1.7 (CrI: 1.3–2.7), and 1.6 (CrI 1.2–2.8), respectively.

Subgroup and sensitivity analyses

Only seven RCTs provided subgroup aggregate data on age (< 60 vs > 60 years) or sex (Additional file 2: Tables S14–S18) [15, 51, 53, 54, 59, 62, 63]. Data were generally sparse, and subgroup evidence of all available comparisons was based on only one or two RCTs, except for the comparison between FIT- and TCS-based programs, which were based on three RCTs. Multiple RCTs consistently reported significant results only in the following three contexts: men were less likely to participate in the FIT-based program than women in the comparison between FIT- and FS-based programs, individuals < 60 years were less likely to participate in the FIT-based program than individuals > 60 years in the comparison between FIT- and FS-based programs, and individuals < 60 years were less likely to participate in FIT-based program than individuals > 60 years in the comparison between FIT- and TCS-based programs (Additional file 2: Table S14). However, the random-effects meta-analysis of the limited data calculated significant results only for the interaction between sex and the comparison between FIT- and FS-based programs (relative RR 0.78; CrI 0.62–0.97).

In the subgroup network meta-analyses, where the analysis was restricted to European RCTs excluding two RCTs conducted outside Europe [57, 63] or to individual randomization design RCTs excluding four cluster design RCTs [50, 52, 55, 57], the summary results were similar to the results in the main analysis (Additional

Table 3 Meta-analysis of screening positivity and PPV and NNTS for detecting one advanced neoplasia in screened populations

Modality	Positivity, % (95% CrI [95% PI])	PPV, % (95% CrI [95% PI])	NNTS (95% CrI)
gFOBT	4.9 (2.9–10.1 [0.9–21.8])	27.0 (15.8–42.8 [3.7–73.4])	3.7·(2.3–6.3)
FIT	8.2 (6.0–12.0 [2.5–22.1])	28.3 (19.9–39.5 [6.2–66.9])	3.5 (2.5–5.0)
FS	9.2 (6.9–15.0 [4.1–21.7])	59.5 (36.5–76.7 [11.3–94.0])	1.7 (1.3–2.7)
CTC	9.6 (5.3–23.2 [2.4–35.9])	60.8 (35.6–81.1 [16.5–93.1])	1.6 (1.2–2.8)

CrI Credible interval, *CTC* Computed tomography colonography, *FS* Flexible sigmoidoscopy, *FIT* Fecal immunochemical testing, *gFOBT* Guaiac-based fecal occult blood testing, *NE* Not estimable, *NNTS* Number needed to scope, *PI* Prediction interval, *PPV* Positive predictive value

file 2: Table S19). Furthermore, using an alternative C adjustment for cluster design RCTs as sensitivity analysis did not materially change the results (Additional file 2: Table S19).

When the FOB-based modality programs were subcategorized by kit delivery method (i.e., direct mailing *vs.* conventional receipt), participation rates across delivery methods did not significantly differ in the gFOBT- (RR 0.54; CrI 0.28–1.03) and FIT-based (RR 0.87; CrI 0.54–1.39) programs (Additional file 2: Table S20). This potential over-stratification widened the CrIs of the comparative effects by reducing the statistical power, precluding valuable interpretation of specific comparisons. Therefore, we did not perform similar analyses for AN and CRC detection rates.

RCTs allowing crossover

An RCT comparing gFOBT- and FIT-based programs [60] and another two RCTs comparing FIT- and TCS-based programs [58, 61] allowed participants to crossover to screening modalities different from those originally allocated and provided data on ITS effects. Similar to the RCTs that did not allow crossover, the overall risk of bias was generally rated low or some concerns for screening participation and high for detection outcomes regardless of the ITS or PP effects (Additional file 2: Fig. S19).

Data on the crossover to different screening modalities were extractable from only two RCTs [58, 60]. Overall participation rates were increased by 9.4% (from 53.4 to 62.9%; RR 1.18; 95% confidence interval [CI] 1.13–1.23) [60] in the gFOBT-based program, 0.4% (from 35.2 to 35.6%; RR 1.01; CI 1.01–1.02) [58] and 6.2% (from 64.0 to 70.2%; RR 1.10; CI 1.07–1.13) [60] in the FIT-based program, and 6.3% (from 21.2 to 27.5%; RR 1.30; CI 1.29–1.32) [58] in the TCS-based program.

Participation rates were significantly higher in the FIT-based program than in the gFOBT- (RR point estimate 1.12) [60] or TCS-based programs (RR point estimate range 1.30–1.64) [58, 61]. However, sparse data coupled with no consistent increase or decrease in detection rates of AN and CRC between groups precluded a conclusion (Additional file 2: Table S21).

Discussion

This systematic review and network meta-analysis follows the WEO approach [9, 10] and provides, for the first time, a comprehensive overview of the comparative evidence of single-round program performance indicators in average-risk screening invitees, reported primarily from Europe. The strengths of this study include a comprehensive literature search involving multiple databases, dual-screen, dual-selection of eligible studies, dual-extraction of data, dual-assessment of risk of bias, and Bayesian random-effects model network meta-analysis with multiple subgroup and sensitivity analyses to evaluate the result stability.

First, the FIT-based program, with moderate certainty, appears to detect more AN cases (an average of 56 more cases per 10,000 screening invitees) than the gFOBT-based program. This finding supports the current guidelines recommending FIT-based programs alongside evidence-based gFOBT-based programs as the reference standard [6–8]. Second, all the visualization modality programs, FS-, CTC-, and TCS-based programs, with low certainty, appear to detect more AN cases (averages of 123–130 cases per 10,000 screening invitees) than FIT-based programs. Among the three visualization modality programs, none is shown to be superior to any other, and the CrIs of RD in these comparisons include clinically important differences favoring either of the compared programs (very low to low certainty). These findings do not deny but rather support the current guidelines recommending CTC-based and TCS-based programs alongside evidence-based FS-based programs as the reference standard [6–8]. As indicated by the ranking analysis, the TCS-based program's lower participation rate could partly explain why this program did not outperform the FS-based or CTC-based programs in detecting AN. Third, although the ranking analysis demonstrates results similar to those of AN detection, due to the low event rates, which specific program is superior to any other in detecting CRC cases remains uncertain; in most comparisons, the wide CrIs of RD include clinically important differences favoring either of the compared programs (very low to low certainty).

Fourth, when results were limited to participants who had received invited modalities, TCS showed optimal performance in detecting AN and CRC, which agreed with theoretical expectations. Notably, this observation is based on highly selective populations and lacks the across-group comparability created by random allocations. Fifth, randomized evidence allowing crossover to other screening modalities was limited.

Although three previous pairwise meta-analyses have evaluated single-round screening results similar to this study, none formally addressed the WEO approach, used the GRADE approach, or performed a network meta-analysis [18–20] (Additional file 2: Table S22). Furthermore, other methodological remarks in these meta-analyses include old literature searches [18], lack of adjustment for the intercluster correlations in cluster RCTs [18], use of odds ratio as the outcome measure [19], joint meta-analysis of RCTs that did not allow crossover and RCTs that did [18–20], and inclusion of RCTs that only assessed screening participation [18].

Despite these differences, the summary results were generally congruent with our results.

Our meta-analysis has limitations. All included RCTs were designed with varying degrees of pragmatism. Hence, all studies lacked blinding and were subject to bias due to deviations from intended interventions and biased measurement of the outcome. Moreover, the effects observed in the ITS could be influenced by several factors with various levels of diversity, including participation rates and screening algorithms. However, the unadjusted PP effects presented here cannot replace the program-level comparisons among different screening modalities. Sophisticated analytical approaches addressing the bias due to deviations from the intended interventions [94] and advanced data imputation methods for missing outcome data [95] may have merit in addressing the issues in analyzing pragmatic RCTs. Given the absence of these statistical approaches, the single-round ITS effects, such as the results in our study, are particularly subject to selection bias resulting from participation in screening, including participation in diagnostic colonoscopy that follows the screening results, which highly affects AN and CRC detection, potentially resulting in biased estimates in long-term outcomes, including CRC mortality.

Furthermore, our systematic review exclusively focused on the results observed in the first-round screening only, that is the introductory phase results in an entire screening program. Screening programs must be evaluated at the entire program level to consider the effects of multiple screening rounds, including a crossover between modalities, if allowed, as subsequent performance indicators, including interval cancers, cumulative detection rates, and cumulative colonoscopy workload, all of which are proposed as the phase IV outcomes [10]. In this line of thinking, although our single-round results may facilitate the design and/or modification of a future program or be applied to simulation models to estimate long-term clinical and economic impact [10, 96], they should never be viewed as the final outcome. Thus, CRC mortality and/or incidence, the long-term outcomes that comprehensively account for these effects, remain the ultimate and desired goal. Additionally, several countries and regions offer multiple alternative screening modalities, generating comparative data that account for crossover between these alternatives.

Lastly, our findings were based on an extended, exploratory analysis of an evidence review conducted to update the national guidelines [22], which did not formally follow a prospectively registered protocol. Although we followed the PRISMA extension for network meta-analysis [23] to maintain transparency of the methodologies used, the results need to be validated in future studies.

Recently, several RCTs utilizing multicycle fecal testing programs compared with a single-cycle FS or TCS have shown that, as theoretically expected, a higher number of screening cycles produced higher screening participation and higher AN detection rates in the fecal testing group [11–17]. Nevertheless, challenges remain in rigorously assessing the comparative effectiveness, including diversity in targeted populations, cumulative participation rates, and screening and follow-up algorithms. These points should be addressed in future research.

Conclusions

This is the first network meta-analysis that comprehensively evaluated program-level initial performance indicator outcomes formally following the approaches proposed by the WEO and GRADE for network meta-analysis. In the first round, moderate certainty of evidence shows that FIT-based programs are likely to detect more AN cases than the gFOBT-based programs. FS-, CTC-, and TCS-based programs may detect more AN cases than FIT-based programs, although the certainty of evidence is low. Low certainty evidence fails to show that any one of the FS-, CTC-, and TCS-based programs is superior or inferior to any others. Importantly, these single-round results lack long-term outcomes. More research is warranted to evaluate these performance indicator outcomes cumulatively after 10–15 years and, more ideally, CRC mortality and incidence.

Abbreviations

AN	Advanced neoplasia
CI	Confidence interval
CRC	Colorectal cancer
CrI	Credible interval
CTC	Computed tomographic colonography
DIC	Deviance information criteria
FIT	Fecal immunochemical test
FS	Flexible sigmoidoscopy
gFOBT	Guaiac-based fecal occult blood test
ITS	Intention-to-screen
MRD	Mean residual deviance
NNTS	Number needed to screen
PP	Per-protocol
PPV	Positive predictive value
RCT	Randomized clinical trial
RR	Relative risk
SUCRA	Surface under the cumulative ranking curve
TCS	Total colonoscopy
WEO	World Endoscopy Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-03948-9.

Additional file 1. PRISMA network meta-analysis checklist.

Additional file 2. Supplementary Methods. List of Excluded Study Reports. Table S1-S22, Fig. S1-S19. Table S1. Exclusion criteria and histological

definitions of neoplastic outcomes. Table S2. Screening test and adherence characteristics. Table S3. Test characteristics of flexible sigmoidoscopy, computed tomography colonography, and total colonoscopy. Table S4. Network Table S for screening participation. Table S5. Ranking analysis of screening participation. Table S6. Network Table S for advanced neoplasia detection in the intention-to-screen analysis. Table S7. Ranking analysis of advanced neoplasia detection in the intention-to-screen analysis. Table S8. Network Table S for advanced neoplasia detection in the per-protocol analysis. Table S9. Ranking analysis of advanced neoplasia detection in the per-protocol analysis. Table S10. Network Table S for colorectal cancer detection in the intention-to-screen analysis. Table S11. Ranking analysis of colorectal cancer detection in the intention-to-screen analysis. Table S12. Network Table S for colorectal cancer detection in the per-protocol analysis. Table S13. Ranking analysis of colorectal cancer detection in the per-protocol analysis. Table S14. Meta-analysis of subgroup interactions for screening participation. Table S15. Meta-analysis of subgroup interactions for advanced neoplasia detection in the intention-to-screen analysis. Table S16. Meta-analysis of subgroup interactions for colorectal cancer detection in the intention-to-screen analysis. Table S17. Meta-analysis of subgroup interactions for advanced neoplasia detection in the per-protocol analysis. Table S18. Meta-analysis of subgroup interactions for colorectal cancer detection in the per-protocol analysis. Table S19. Subgroup and sensitivity analysis results. Table S20. Network meta-analysis of participation that subcategorized the FOB-based method programs by kit delivery methods. Table S21. Randomized clinical trials that allowed crossover to different modalities. Table S22. Previous systematic reviews of RCTs that evaluated direct evidence on program-level outcomes. Fig. S1. Geometry of the network of included randomized clinical trials. Fig. S2. The risk of bias assessment of studies that evaluated screening participation in all screening invitees. Fig. S3. The risk of bias assessment of studies that evaluated the intention-to-screen effect of the detection of advanced neoplasia, colorectal cancer, or both in all screening invitees. Fig. S4. The risk of bias assessment of studies that evaluated the per-protocol effect of the detection of advanced neoplasia, colorectal cancer, or both in screening participants only. Fig. S5. Forest plot of individual study results included in the network meta-analysis of screening participation. Fig. S6. SUCRA plot and Litmus Rank-O-Gram of network meta-analysis of screening participation. Fig. S7. Forest plot of individual study results included in the network meta-analysis of advanced neoplasia detection for evaluating the intention-to-screen effect. Fig. S8. SUCRA plot and Litmus Rank-O-Gram of network meta-analysis of advanced neoplasia detection for evaluating the intention-to-screen effect. Fig. S9. Forest plot of individual study results included in the network meta-analysis of advanced neoplasia detection for evaluating the per-protocol effect. Fig. S10. Network meta-analysis of advanced neoplasia detection for evaluating the per-protocol effect. Fig. S11. SUCRA plot and Litmus Rank-O-Gram of network meta-analysis of advanced neoplasia detection for evaluating the per-protocol effect. Fig. S12. Forest plot of individual study results included in the network meta-analysis of colorectal cancer detection for evaluating the intention-to-screen analysis. Fig. S13. SUCRA plot and Litmus Rank-O-Gram of network meta-analysis of colorectal cancer detection for evaluating the intention-to-screen effect. Fig. S14. Forest plot of individual study results included in the network meta-analysis of colorectal cancer detection for evaluating the per-protocol effect. Fig. S15. Network meta-analysis of colorectal cancer detection for evaluating the per-protocol effect. Fig. S16. SUCRA plot and Litmus Rank-O-Gram of network meta-analysis of colorectal cancer detection for evaluating the per-protocol effect. Fig. S17. Meta-analysis of screening positivity. Fig. S18. Meta-analysis of PPV. Fig. S19. The risk of bias assessment of randomized clinical trials that allowed crossover.

Acknowledgements
The English language editing was provided by MARUZEN-YUSHODO Co., Ltd. (https://kw.maruzen.co.jp/kousei-honyaku/).

Authors' contributions
TT1 (Conceptualization: Equal; Data curation: Lead; Formal analysis: Lead; Investigation: Lead; Methodology: Equal; Project administration: Supporting; Resources: Equal; Software: Lead; Validation: Lead, Visualization: Lead, Writing – original draft: Lead; Writing – review & editing: Equal). TT2 (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). KA (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). SS (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). SH (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). TK (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). KH: (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Investigation: Supporting, Validation: Supporting; Writing – review & editing: Supporting). TN (Conceptualization: Supporting; Formal analysis: Supporting; Funding: Equal; Supervision: Supporting, Writing – review & editing: Supporting). CH (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Lead; Resources: Equal; Supervision: Lead, Validation: Supporting; Writing – review & editing: Lead). All authors read and approved the final manuscript.

Funding
This work was supported by The National Cancer Center Research and Development Fund (grant number A-21, 2023). The funder had no role in developing the protocol and did not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

Data availability
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Ethics approval is not applicable as this is a secondary analysis of publicly available data.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Section of General Internal Medicine, Department of Emergency Medicine and General Internal Medicine, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-Cho, Toyoake, Aichi 470-1192, Japan. [2]Cancer Detection Center, Miyagi Cancer Society, Sendai, Japan. [3]Department of Medicine, Teikyo University School of Medicine, Tokyo, Japan. [4]Center for Preventive Medicine, St. Luke's International Hospital Affiliated Clinic, Tokyo, Japan. [5]Division of Cancer Screening Assessment and Management, National Cancer Center Institute for Cancer Control, Tokyo, Japan. [6]Department of Statistics and Computer Science, College of Nursing Art and Science, University of Hyogo, Hyogo, Japan. [7]Center for Health Informatics Policy, National Institute of Public Health, Wako, Japan. [8]Department of Nursing, Faculty of Medical Technology, Teikyo University, Tokyo, Japan.

Received: 3 October 2024 Accepted: 13 February 2025
Published online: 21 February 2025

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.

2. Brenner H, Heisser T, Cardoso R, Hoffmeister M. Reduction in colorectal cancer incidence by screening endoscopy. Nat Rev Gastroenterol Hepatol. 2024;21:125–33. https://doi.org/10.1038/s41575-023-00847-3.
3. Lauby-Secretan B, Vilahur N, Bianchini F, Guha N, Straif K. International agency for research on cancer handbook working group. The IARC perspective on colorectal cancer screening. N Engl J Med. 2018;378:1734–40. https://doi.org/10.1056/NEJMsr1714643.
4. Jodal HC, Helsingen LM, Anderson JC, Lytvyn L, Vandvik PO, Emilsson L. Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. BMJ Open. 2019;9: e032773. https://doi.org/10.1136/bmjopen-2019-032773.
5. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA. 2021;325:1978–98. https://doi.org/10.1001/jama.2021.4417.
6. United States Preventive Services Task Force, Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA. 2021;325:1965–77. https://doi.org/10.1001/jama.2021.6238.
7. Helsingen LM, Vandvik PO, Jodal HC, Agoritsas T, Lytvyn L, Anderson JC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. BMJ. 2019;367: l5515. https://doi.org/10.1136/bmj.l5515.
8. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018;68:250–81. https://doi.org/10.3322/caac.21457.
9. Young GP, Senore C, Mandel JS, Allison JE, Atkin WS, Benamouzig R, et al. Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. Cancer. 2016;122:826–39. https://doi.org/10.1002/cncr.29865.
10. Bresalier RS, Senore C, Young GP, Allison J, Benamouzig R, Benton S, et al. An efficient strategy for evaluating new non-invasive screening tests for colorectal cancer: the guiding principles. Gut. 2023;72:1904–18. https://doi.org/10.1136/gutjnl-2023-329701.
11. Zauber AG, Winawer SJ, O'Brien MJ, Mills GM, Allen JI, Feld AD, et al. Randomized trial of facilitated adherence to screening colonoscopy vs sequential fecal-based blood test. Gastroenterology. 2023;165:252–66. https://doi.org/10.1053/j.gastro.2023.03.206.
12. Singal AG, Gupta S, Skinner CS, Ahn C, Santini NO, Agrawal D, et al. Effect of colonoscopy outreach vs fecal immunochemical test outreach on colorectal cancer screening completion: A randomized clinical trial. JAMA. 2017;318:806–15. https://doi.org/10.1001/jama.2017.11389.
13. Forsberg A, Westerberg M, Metcalfe C, Steele R, Blom J, Engstrand L, et al. Once-only colonoscopy or two rounds of faecal immunochemical testing 2 years apart for colorectal cancer screening (SCREESCO): preliminary report of a randomised controlled trial. Lancet Gastroenterol Hepatol. 2022;7:513–21. https://doi.org/10.1016/S2468-1253(21)00473-8.
14. Randel KR, Schult AL, Botteri E, Hoff G, Bretthauer M, Ursin G, et al. Colorectal cancer screening with repeated fecal immunochemical test versus sigmoidoscopy: baseline results from a randomized trial. Gastroenterology. 2021;160:1085-1096.e5. https://doi.org/10.1053/j.gastro.2020.11.037.
15. Sali L, Mascalchi M, Falchini M, Ventura L, Carozzi F, Castiglione G, et al. Reduced and full-preparation CT colonography, fecal immunochemical test, and colonoscopy for population screening of colorectal cancer: A randomized trial. J Natl Cancer Inst. 2016;108: djv319. https://doi.org/10.1093/jnci/djv319.
16. Sali L, Ventura L, Mascalchi M, Falchini M, Mallardi B, Carozzi F, et al. Single CT colonography versus three rounds of faecal immunochemical test for population-based screening of colorectal cancer (SAVE): a randomised controlled trial. Lancet Gastroenterol Hepatol. 2022;7:1016–23. https://doi.org/10.1016/S2468-1253(22)00269-2.
17. Grobbee EJ, van der Vlugt M, van Vuuren AJ, Stroobants AK, Mallant-Hent RC, Lansdorp-Vogelaar I, et al. Diagnostic yield of one-time colonoscopy vs one-time flexible sigmoidoscopy vs multiple rounds of mailed fecal immunohistochemical tests in colorectal cancer screening. Clin Gastroenterol Hepatol. 2020;18:667-675.e1. https://doi.org/10.1016/j.cgh.2019.08.015.
18. Hassan C, Giorgi Rossi P, Camilloni L, Rex DK, Jimenez-Cendales B, Ferroni E, et al. Meta-analysis: adherence to colorectal cancer screening and the detection rate for advanced neoplasia, according to the type of screening test. Aliment Pharmacol Ther. 2012;36:929–40. https://doi.org/10.1111/apt.12071.
19. Mutneja H, Agrawal R, Bhurwal A, Arora S, Go A, Attar B. Comparative Effectiveness of Fecal Immunochemical Tests versus Flexible sigmoidoscopy for Colorectal Cancer Screening: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. J Gastrointestin Liver Dis. 2021;30:267–73. https://doi.org/10.15403/jgld-3432.
20. Zhu H, Li F, Tao K, Wang J, Scurlock C, Zhang X, et al. Comparison of the participation rate between CT colonography and colonoscopy in screening population: a systematic review and meta-analysis of randomized controlled trials. Br J Radiol. 2020;93: 20190240. https://doi.org/10.1259/bjr.20190240.
21. Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: continuing controversies and recent developments. Res Synth Methods. 2024;15:702–27. https://doi.org/10.1002/jrsm.1700.
22. Colorectal Cancer Screening Literature Review Committee. Colorectal cancer screening evidence report. 2023rd ed. Tokyo: National Cancer Center; 2024.
23. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162:777–84. https://doi.org/10.7326/M14-2385.
24. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366: l4898. https://doi.org/10.1136/bmj.l4898.
25. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172:137–59. https://doi.org/10.1111/j.1467-985X.2008.00552.x.
26. White IR. Network meta-analysis. STATA J. 2015;15:951-85. https://doi.org/10.1177/1536867X1501500403.
27. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011;64:163–71. https://doi.org/10.1016/j.jclinepi.2010.03.016.
28. Nevill CR, Cooper NJ, Sutton AJ. A multifaceted graphical display, including treatment ranking, was developed to aid interpretation of network meta-analysis. J Clin Epidemiol. 2023;157:83–91. https://doi.org/10.1016/j.jclinepi.2023.02.016.
29. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. Stat Med. 2002;21:1601–23. https://doi.org/10.1002/sim.1189.
30. van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Res Synth Methods. 2012;3:285–99. https://doi.org/10.1002/jrsm.1054.
31. Owen RK, Bradbury N, Xin Y, Cooper N, Sutton A. MetaInsight: an interactive web-based tool for analyzing, interrogating, and visualizing network meta-analyses using R-shiny and netmeta. Res Synth Methods. 2019;10:569–81. https://doi.org/10.1002/jrsm.1373.
32. Uhlmann L, Jensen K, Kieser M. Bayesian network meta-analysis for cluster randomized trials with binary outcomes. Res Synth Methods. 2017;8:236–50. https://doi.org/10.1002/jrsm.1210.
33. Higgins JP, Eldridge S, Li T. Including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions. 2nd ed. Hoboken: Wiley; 2019. p. 569–93. https://doi.org/10.1002/9781119536604.ch23.
34. Siegel L, Rudser K, Sutcliffe S, Markland A, Brubaker L, Gahagan S, et al. A Bayesian multivariate meta-analysis of prevalence data. Stat Med. 2020;39:3105–19. https://doi.org/10.1002/sim.8593.
35. Hade EM, Murray DM, Pennell ML, Rhoda D, Paskett ED, Champion VL, et al. Intraclass correlation estimates for cancer screening outcomes: estimates and applications in the design of group-randomized cancer screening studies. J Natl Cancer Inst Monogr. 2010;2010:97–103. https://doi.org/10.1093/jncimonographs/lgq011.
36. Crespi CM, Maxwell AE, Wu S. Cluster randomized trials of cancer screening interventions: are appropriate statistical methods being used? Contemp Clin Trials. 2011;32:477–84. https://doi.org/10.1016/j.cct.2011.03.001.

37. Crespi CM, Ziehl K. Cluster-randomized trials of cancer screening interventions: has use of appropriate statistical methods increased over time? Contemp Clin Trials. 2022;123: 106974. https://doi.org/10.1016/j.cct.2022.106974.
38. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7:434–55. https://doi.org/10.1080/10618600.1998.10474787.
39. Röver C, Bender R, Dias S, Schmid CH, Schmidli H, Sturtz S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Res Synth Methods. 2021;12:448–74. https://doi.org/10.1002/jrsm.1475.
40. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making. 2013;33:641–56. https://doi.org/10.1177/0272989X12455847.
41. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29:932–44. https://doi.org/10.1002/sim.3767.
42. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Res Synth Methods. 2016;7:80–93. https://doi.org/10.1002/jrsm.1167.
43. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343: d4002. https://doi.org/10.1136/bmj.d4002.
44. Chiocchia V, Holloway A, Salanti G. Semi-automated assessment of the risk of bias due to missing evidence in network meta-analysis: a guidance paper for the ROB-MEN web-application. BMC Med Res Methodol. 2023;23:223. https://doi.org/10.1186/s12874-023-02038-9.
45. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ. 2017;356: j573. https://doi.org/10.1136/bmj.j573.
46. Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. BMJ. 2023;381: e074495. https://doi.org/10.1136/bmj-2022-074495.
47. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. Public Health Rev. 2018;39:7. https://doi.org/10.1186/s40985-018-0080-0.
48. Bénard F, Barkun AN, Martel M, von Renteln D. Systematic review of colorectal cancer screening guidelines for average-risk adults: summarizing the current global recommendations. World J Gastroenterol. 2018;24:124–38. https://doi.org/10.3748/wjg.v24.i1.124.
49. Verne JE, Aubrey R, Love SB, Talbot IC, Northover JM. Population based randomized study of uptake and yield of screening by flexible sigmoidoscopy compared with screening by faecal occult blood testing. BMJ. 1998;317:182–5. https://doi.org/10.1136/bmj.317.7152.182.
50. Federici A, Giorgi Rossi P, Borgia P, Bartolozzi F, Farchi S, Gausticchi G. The immunochemical faecal occult blood test leads to higher compliance than the guaiac for colorectal cancer screening programmes: a cluster randomized controlled trial. J Med Screen. 2005;12:83–8. https://doi.org/10.1258/0969141053908357.
51. Segnan N, Senore C, Andreoni B, Arrigoni A, Bisanti L, Cardelli A, et al. Randomized trial of different screening strategies for colorectal cancer: patient response and detection rates. J Natl Cancer Inst. 2005;97:347–57. https://doi.org/10.1093/jnci/dji050.
52. Federici A, Marinacci C, Mangia M, Borgia P, Giorgi Rossi P, Guasticchi G. Is the type of test used for mass colorectal cancer screening a determinant of compliance? A cluster-randomized controlled trial comparing fecal occult blood testing with flexible sigmoidoscopy. Cancer Detect Prev. 2006;30:347–53. https://doi.org/10.1016/j.cdp.2006.03.009.
53. Segnan N, Senore C, Andreoni B, Azzoni A, Bisanti L, Cardelli A, et al. Comparing attendance and detection rate of colonoscopy with sigmoidoscopy and FIT for colorectal cancer screening. Gastroenterology. 2007;132:2304–12. https://doi.org/10.1053/j.gastro.2007.03.030.
54. van Rossum LG, van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, van Krieken HH, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology. 2008;135:82–90. https://doi.org/10.1053/j.gastro.2008.03.040.
55. Lisi D, Hassan C, Crespi M, AMOD Study Group. Participation in colorectal cancer screening with FOBT and colonoscopy: an Italian, multicentre, randomized population study. Dig Liver Dis. 2010;42:371–6. https://doi.org/10.1016/j.dld.2009.07.019.
56. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, van Dekken H, Reijerink JC, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut. 2010;59:62–8. https://doi.org/10.1136/gut.2009.177089.
57. Levi Z, Birkenfeld S, Vilkin A, Bar-Chana M, Lifshitz I, Chared M, et al. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. Int J Cancer. 2011;128:2415–24. https://doi.org/10.1002/ijc.25574.
58. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366:697–706. https://doi.org/10.1056/NEJMoa1108895.
59. Stoop EM, de Haan MC, de Wijkerslooth TR, Bossuyt PM, van Ballegooijen M, Nio CY, 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. 2012;13:55–64. https://doi.org/10.1016/S1470-2045(11)70283-2.
60. Chubak J, Bogart A, Fuller S, Laing SS, Green BB. Uptake and positive predictive value of fecal occult blood tests: A randomized controlled trial. Prev Med. 2013;57:671–8. https://doi.org/10.1016/j.ypmed.2013.08.032.
61. Gupta S, Halm EA, Rockey DC, Hammons M, Koch M, Carter E, et al. Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized clinical trial. JAMA Intern Med. 2013;173:1725–32. https://doi.org/10.1001/jamainternmed.2013.9294.
62. Regge D, Iussich G, Segnan N, Correale L, Hassan C, Arrigoni A, et al. Comparing CT colonography and flexible sigmoidoscopy: a randomised trial within a population-based screening programme. Gut. 2017;66:1434–40. https://doi.org/10.1136/gutjnl-2015-311278.
63. Chen H, Lu M, Liu C, Zou S, Du L, Liao X, et al. Comparative evaluation of participation and diagnostic yield of colonoscopy vs fecal immunochemical test vs risk-adapted screening in colorectal cancer screening: interim analysis of a multicenter randomized controlled trial (TARGET-C). Am J Gastroenterol. 2020;115:1264–74. https://doi.org/10.14309/ajg.0000000000000624.
64. Hoffman RM, Steel S, Yee EFT, Massie L, Schrader RM, Murata GH. Colorectal cancer screening adherence is higher with fecal immunochemical tests than guaiac-based fecal occult blood tests: a randomized, controlled trial. Prev Med. 2010;50:297–9. https://doi.org/10.1016/j.ypmed.2010.03.010.
65. Levy BT, Daly JM, Xu Y, Ely JW. Mailed fecal immunochemical tests plus educational materials to improve colon cancer screening rates in Iowa Research Network (IRENE) practices. J Am Board Fam Med. 2012;25:73–82. https://doi.org/10.3122/jabfm.2012.01.110055.
66. Levy BT, Xu Y, Daly JM, Ely JW. A randomized controlled trial to improve colon cancer screening in rural family medicine: an Iowa Research Network (IRENE) study. J Am Board Fam Med. 2013;26:486–97. https://doi.org/10.3122/jabfm.2013.05.130041.
67. Greiner KA, Daley CM, Epp A, James A, Yeh HW, Geana M, et al. Implementation intentions and colorectal screening: a randomized trial in safety-net clinics. Am J Prev Med. 2014;47:703–14. https://doi.org/10.1016/j.amepre.2014.08.005.
68. Charlton ME, Mengeling MA, Halfdanarson TR, Makki NM, Malhotra A, Klutts JS, et al. Evaluation of a home-based colorectal cancer screening intervention in a rural state. J Rural Health. 2014;30:322–32. https://doi.org/10.1111/jrh.12052.
69. Murphy CC, Ahn C, Pruitt SL, Hughes AE, Halm EA, Gupta S, et al. Screening initiation with FIT or colonoscopy: post-hoc analysis of a pragmatic, randomized trial. Prev Med. 2019;118:332–5. https://doi.org/10.1016/j.ypmed.2018.11.020.
70. Gruner LF, Hoffmeister M, Ludwig L, Meny S, Brenner H. The effects of differing invitation models on the uptake of immunological fecal occult blood testing. Dtsch Ärztebl Int. 2020;117:423–30. https://doi.org/10.3238/arztebl.2020.0423.

71. Robertson DJ, Dominitz JA, Beed A, Boardman KD, Del Curto BJ, Guarino PD, et al. Baseline features and reasons for nonparticipation in the colonoscopy versus fecal immunochemical test in reducing mortality from colorectal cancer (CONFIRM) study, a colorectal cancer screening trial. JAMA Netw Open. 2023;6: e2321730. https://doi.org/10.1001/jamanetworkopen.2023.21730.
72. Salas D, Vanaclocha M, Ibáñez J, Molina-Barceló A, Hernández V, Cubiella J, et al. Participation and detection rates by age and sex for colonoscopy versus fecal immunochemical testing in colorectal cancer screening. Cancer Causes Control. 2014;25:985–97. https://doi.org/10.1007/s10552-014-0398-y.
73. Castells A, Quintero E, Álvarez C, Bujanda L, Cubiella J, Salas D, et al. Rate of detection of advanced neoplasms in proximal colon by simulated sigmoidoscopy vs fecal immunochemical tests. Clin Gastroenterol Hepatol. 2014;12:1708-16.e4. https://doi.org/10.1016/j.cgh.2014.03.022.
74. Denters MJ, Deutekom M, Bossuyt PM, Stroobants AK, Fockens P, Dekker E. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology. 2012;142:497–504. https://doi.org/10.1053/j.gastro.2011.11.024.
75. Hol L, Wilschut JA, van Ballegooijen M, van Vuuren AJ, van der Valk H, Reijerink JC, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer. 2009;100:1103–10. https://doi.org/10.1038/sj.bjc.6604961.
76. Chen H, Shi J, Lu M, Li Y, Du L, Liao X, et al. Comparison of colonoscopy, fecal immunochemical test, and risk-adapted approach in a colorectal cancer screening trial (TARGET-C). Clin Gastroenterol Hepatol. 2023;21:808–18. https://doi.org/10.1016/j.cgh.2022.08.003.
77. Chen H, Liao X, Du L, Dong D, Wei D, Gao Y, et al. Evaluation of the effectiveness of colonoscopy, fecal immunochemical test and risk-adapted screening strategies in population-based colorectal cancer screening: A multicenter randomized controlled trial. Med J Peking Union Med Coll Hosp. 2023;14:114–23.
78. Xu M, Yang JY, Meng T. Effectiveness of colonoscopy, immune fecal occult blood testing, and risk-graded screening strategies in colorectal cancer screening. World J Gastrointest Surg. 2024;16:2270–80. https://doi.org/10.4240/wjgs.v16.i7.2270.
79. Bonander C, Westerberg M, Chauca Strand G, Forsberg A, Strömberg U. Colorectal cancer screening with fecal immunochemical testing or primary colonoscopy: inequities in diagnostic yield. JNCI Cancer Spectr. 2024;8:pkae043. https://doi.org/10.1093/jncics/pkae043.
80. Berry DP, Clarke P, Hardcastle JD, Vellacott KD. Randomized trial of the addition of flexible sigmoidoscopy to faecal occult blood testing for colorectal neoplasia population screening. Br J Surg. 1997;84:1274–6.
81. Rasmussen M, Kronborg O, Fenger C, Jørgensen OD. Possible advantages and drawbacks of adding flexible sigmoidoscopy to hemoccult-II in screening for colorectal cancer A randomized study. Scand J Gastroenterol. 1999;34:73–8. https://doi.org/10.1080/00365529950172862.
82. Steele RJ, Carey FA, Stanners G, Lang J, Brand J, Brownlee LA, et al. Randomized controlled trial: flexible sigmoidoscopy as an adjunct to faecal occult blood testing in population screening. J Med Screen. 2020;27:59–67. https://doi.org/10.1177/0969141319879955.
83. Bretthauer M, Løberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, et al. Effect of colonoscopy screening on risks of colorectal cancer and related death. N Engl J Med. 2022;387:1547–56. https://doi.org/10.1056/NEJMoa2208375.
84. Senore C, Riggi E, Armaroli P, Bonelli L, Sciallero S, Zappa M, et al. Long-term follow-up of the Italian flexible sigmoidoscopy screening trial. Ann Intern Med. 2022;175:36–45. https://doi.org/10.7326/M21-0977.
85. Juul FE, Cross AJ, Schoen RE, Senore C, Pinsky P, Miller E, et al. 15-year benefits of sigmoidoscopy screening on colorectal cancer incidence and mortality: A pooled analysis of randomized trials. Ann Intern Med. 2022;175:1525–33. https://doi.org/10.7326/M22-0835.
86. MACS. A comparison of colorectal neoplasia screening tests: a multicentre community-based study of the impact of consumer choice. Med J Aust. 2006;184:546–50. https://doi.org/10.5694/j.1326-5377.2006.tb00377.x.
87. Scott RG, Edwards JT, Fritschi L, Foster NM, Mendelson RM, Forbes GM. Community-based screening by colonoscopy or computed tomographic colonography in asymptomatic average-risk subjects. Am J Gastroenterol. 2004;99:1145–51. https://doi.org/10.1111/j.1572-0241.2004.30253.x.
88. You JJ, Liu Y, Kirby J, Vora P, Moayyedi P. Virtual colonoscopy, optical colonoscopy, or fecal occult blood testing for colorectal cancer screening: results of a pilot randomized controlled trial. Trials. 2015;16:296. https://doi.org/10.1186/s13063-015-0826-7.
89. Sekiguchi M, Westerberg M, Ekbom A, Hultcrantz R, Forsberg A. Detection rates of colorectal neoplasia during colonoscopies and their associated factors in the SCREESCO study. J Gastroenterol Hepatol. 2022;37:2120–30. https://doi.org/10.1111/jgh.15990.
90. Bretthauer M. Long-term colorectal cancer incidence and mortality after screening colonoscopy: the NordICC randomized trial. UEG Week. 2022.
91. Sali LP, D, Mascalchi M, Mallardi B, Zappa M. Incidence of advanced adenoma and colorectal cancer after screening CTC in the SAVE randomised trial. ESGAR. 2023 – 34th Annual Meeting And Postgraduate Course, 2023.
92. Brevinge H, Lindholm E, Buntzen S, Kewenter J. Screening for colorectal neoplasia with faecal occult blood testing compared with flexible sigmoidoscopy directly in a 55–56 years' old population. Int J Colorectal Dis. 1997;12:291–5. https://doi.org/10.1007/s003840050108.
93. Álvarez-Delgado A, García MLP, García-González JM, de Sena HI, Chamorro AJ, Gómez MFL, et al. Improvements in the effectiveness of early detection in colorectal cancer with open-label randomised study. J Clin Med. 2021;10: 5072. https://doi.org/10.3390/jcm10215072.
94. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med. 2017;377:1391–8. https://doi.org/10.1056/NEJMsm1605385.
95. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol. 2019;48:1294–304. https://doi.org/10.1093/ije/dyz032.
96. Dahabreh IJ, Trikalinos TA, Balk EM, Wong JB. Recommendations for the conduct and reporting of modeling and simulation studies in health technology assessment. Ann Intern Med. 2016;165:575–81. https://doi.org/10.7326/M16-0161.

Publisher's Note