



An efficient strategy for evaluating new non-invasive screening tests for colorectal cancer: the guiding principles

Robert S Bresalier ¹, Carlo Senore ², Graeme P Young ³, James Allison,⁴ Robert Benamouzig,⁵ Sally Benton,⁶ Patrick M M Bossuyt,⁷ Luis Caro,⁸ Beatriz Carvalho ⁹, Han-Mo Chiu ¹⁰, Veerle M H Coupé,¹¹ Willemijn de Klaver,¹² Clasine Maria de Klerk,¹³ Evelien Dekker ¹⁴, Sunil Dolwani,¹⁵ Callum G Fraser ¹⁶, William Grady ¹⁷, Lydia Guittet,¹⁸ Samir Gupta,¹⁹ Stephen P Halloran,²⁰ Ulrike Haug,²¹ Geir Hoff,^{22,23} Steven Itzkowitz,²⁴ Tim Kortlever ²⁵, Anastasios Koulaouzidis ²⁶, Uri Ladabaum,²⁷ Beatrice Lauby-Secretan,²⁸ Mārcis Leja ²⁹, Bernard Levin,³⁰ Theodore Robert Levin ³¹, Finlay Macrae,³² Gerrit A Meijer ⁹, Joshua Melson,³³ Colm O'Morain,³⁴ Susan Parry,^{35,36} Linda Rabeneck,³⁷ David F Ransohoff,³⁸ Roque Sáenz,³⁹ Hiroshi Saito,⁴⁰ Silvia Sanduleanu-Dascalescu,⁴¹ Robert E Schoen ⁴², Kevin Selby,⁴³ Harminder Singh ⁴⁴, Robert J C Steele ⁴⁵, Joseph J Y Sung ⁴⁶, Erin Leigh Symonds ⁴⁷, Sidney J Winawer,⁴⁸ Members of the World Endoscopy Colorectal Cancer Screening New Test Evaluation Expert Working Group

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2023-329701>).

For numbered affiliations see end of article.

Correspondence to

Professor Graeme P Young, Cancer Research, Flinders Centre for Innovation in Cancer, Flinders University, Adelaide, SA 5001, Australia; graeme.young@flinders.edu.au

RSB, CS and GPY contributed equally.

Received 13 February 2023
Accepted 20 June 2023
Published Online First
18 July 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Bresalier RS, Senore C, Young GP, et al. *Gut* 2023;**72**:1904–1918.

ABSTRACT

Objective New screening tests for colorectal cancer (CRC) are rapidly emerging. Conducting trials with mortality reduction as the end point supporting their adoption is challenging. We re-examined the principles underlying evaluation of new non-invasive tests in view of technological developments and identification of new biomarkers.

Design A formal consensus approach involving a multidisciplinary expert panel revised eight previously established principles.

Results Twelve newly stated principles emerged. Effectiveness of a new test can be evaluated by comparison with a proven comparator non-invasive test. The faecal immunochemical test is now considered the appropriate comparator, while colonoscopy remains the diagnostic standard. For a new test to be able to meet differing screening goals and regulatory requirements, flexibility to adjust its positivity threshold is desirable. A rigorous and efficient four-phased approach is proposed, commencing with small studies assessing the test's ability to discriminate between CRC and non-cancer states (*phase I*), followed by prospective estimation of accuracy across the continuum of neoplastic lesions in neoplasia-enriched populations (*phase II*). If these show promise, a provisional test positivity threshold is set before evaluation in typical screening populations. *Phase III* prospective studies determine single round intention-to-screen programme outcomes and confirm the test positivity threshold. *Phase IV* studies involve evaluation over repeated screening rounds with monitoring for

missed lesions. Phases III and IV findings will provide the real-world data required to model test impact on CRC mortality and incidence.

Conclusion New non-invasive tests can be efficiently evaluated by a rigorous phased comparative approach, generating data from unbiased populations that inform predictions of their health impact.

INTRODUCTION

New non-invasive screening tests for colorectal cancer (CRC) and its precursor lesions are rapidly emerging as novel technologies to identify new biomarkers for detecting these neoplasms, using a range of biological samples.¹ An Expert Working Group (EWG) convened by the World Endoscopy Organisation (WEO) CRC Screening Committee, which published recommendations for practical evaluation of new tests in 2016,² has considered that the principles now warrant updating in view of recent technological developments together with major global differences in the nature and goals of population CRC screening programmes.^{3,4} New sample options, new biomarkers and new approaches to working with a panel of biomarkers provide opportunity to improve CRC screening. These new principles discuss in greater detail the complexities of validation of new non-invasive screening tests and set forth an efficient stepwise strategy for evaluating them and bringing them to clinical practice.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ In 2016, a practical guide for evaluation of new non-invasive screening tests for colorectal cancer (CRC) suggested that comparing test accuracy with a proven comparator such as the guaiac-based faecal occult blood test was the minimum standard to provide evidence for use in practice.
- ⇒ That guide proposed eight principles that underpinned a four-phase test evaluation, with a brief rationale for each.

WHAT THIS STUDY ADDS

- ⇒ This expert-based consensus process has expanded the principles to 12 in view of the necessity to now consider new technology, screening as a multistep process, the use of the quantitative faecal immunochemical test for haemoglobin as the current comparator, to allow for algorithms based on biomarker panels, to undertake prospective evaluation in unbiased intended-use screening populations, to include precursor lesions for CRC as legitimate targets, to provide capacity to adjust a new test's positivity threshold to facilitate screening programme goals and to model findings to provide evidence supporting a likely benefit in reducing CRC mortality and incidence.
- ⇒ A detailed rationale for each principle is now provided.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ By providing a dynamic evaluation framework which is flexible yet rigorous and allows for broad application given the wide global variation in how CRC screening is conducted, it is expected that this will guide researchers, practitioners, regulatory authorities, policy makers and screening programme providers in the development and validation of a new non-invasive test.
- ⇒ It will ensure a test's suitability for the context of its use, no matter how the screening programme is implemented, without requiring randomised trials with CRC mortality as the end point.
- ⇒ New and effective tests with improved capacity to detect advanced CRC precursor lesions should emerge within a manageable timeframe.

The need for this was emphasised by an expert consensus conference held by the American Gastroenterological Association (AGA), which stressed that overcoming the multiple barriers to screening will require efficient use of available screening modalities, continued development of non-invasive screening tests and improved personal risk assessment to best risk-stratify patients.⁵ The AGA Executive Committee on the Screening Continuum has highlighted the need to anticipate new and evolving strategies.⁶

Recently, the WEO CRC Screening Committee has defined two main screening programme contexts³: population-based organised screening (PBOS) based on a WHO-style public health model⁷ and structured opportunistic screening (SOS) based on jurisdictional standards for practitioner practice.⁸ The differences between these contexts demand consideration when establishing recommendations for new test evaluation. There is considerable global variation in test regulatory approval processes and how screening programmes are funded. While CRC is a global disease, a 'one-size-fits-all' approach to CRC screening is not

necessarily practical—flexibility, even within jurisdictions, is required if programmes are to meet public health goals.⁵

What is considered an acceptable level of evidence to justify the use of a new test in population CRC screening varies around the world.^{2–4 8 9} Thus, the guiding principles of test evaluation must be universally applicable and flexible.

The goal of the revision has been to provide an efficient, feasible and rigorous approach to evaluate emerging 'new' non-invasive tests for use in the two main screening contexts of PBOS and SOS. This recognises that using CRC mortality as the end point, while the ultimate goal, may be challenging due to the large study size required, the time involved and cost.² The use of non-invasive tests in non-screening scenarios (eg, surveillance, evaluation of symptomatic individuals) is beyond the scope of the revision and requires a different approach from population screening.

This revision presents revised and expanded guiding principles that emerged from a rigorous consensus process, together with explanatory text addressing each principle's importance, its rationale and what needs to be accomplished. The intention has been to provide a framework that allows a dynamic process that has broad application. It is not bound by any one specific test.

METHODS

To revise the existing guiding principles, we established a consensus process based on the Glaser and Delphi approaches¹⁰ that was adapted to be undertaken by a combination of webinars and voting via virtual platforms due to the constraints of the COVID-19 pandemic.

The *membership* consisted of experts (gastroenterologists, endoscopists, GI surgeons, public health physicians, epidemiologists, clinical biochemists and tumour biologists) with knowledge or experience in practice or research relevant to screening for CRC. Forty-seven experts were involved. Participants confirmed the *problem* being addressed as that of the goal as stated above.

A series of *specific questions* (each of which was a draft principle to be critiqued) was initially expanded from the original 8² to 10 and then, after the first consensus round of voting, further increased to 12. The 12 principles were progressively redrafted in response to specific feedback: webinars, conference seminars addressing specific issues and semi-structured discussions were held and members voted and commented on each principle using a spreadsheet. After four rounds of voting, the consensus goal of >80% agreement (agree or strongly agree on a 5-point scale) was achieved for all 12 principles. The 12 principles were then circulated to a panel of 7 industry representatives (who volunteered for this task from all the industry groups associated with the WEO CRC SC), seeking their feedback on how they would view these principles in light of the regulatory processes that they face. Principles were not altered based on this feedback.

The explanatory text for each principle was developed from the feedback received during the consensus process and from the extensive comments received during the consultation of experts and industry representatives. Multiple drafts of the explanatory text were circulated to the expert panel over a period of 6 months, and feedback has been incorporated into the final manuscript.

THE PRINCIPLES

The topics addressed in each principle are listed in [table 1](#).

Table 1 The topics addressed in each of the principles established by the consensus process

Principle number	Topic
1	Desired outcome of CRC screening.
2	Screening is a multistep process.
3	A screening test identifies individuals with an increased likelihood of CRC and/or advanced precursor lesions.
4	Nature of precursor lesions most important to detect.
5	New biomarkers might detect lesions with a different natural history.
6	Outcomes to be estimated in a screening population.
7	Expectations of a new non-invasive test.
8	An adjustable test positivity threshold accommodates different programme goals.
9	Predicting value by paired comparison with a proven non-invasive screening test.
10	Evaluation proceeds through increasingly complex phases.
11	Accuracy required for evaluation in a screening population.
12	Analytical specifications, standards and performance.

EXPLANATORY TEXT FOR EACH PRINCIPLE

1. Desired outcome of CRC screening

Screening for CRC aims to reduce CRC mortality and/or incidence by detecting readily treatable CRC and advanced precursor lesions without adversely affecting the health status or overly burdening those who participate in screening.

Explanatory text

Criteria that justify population-based screening were initially defined by WHO⁷ and revisited in 2008.¹¹

The ultimate goal of a CRC screening programme is to reduce CRC mortality and burden of CRC in the target population⁷ through application of a test^{3,4} that facilitates the detection of CRC and precursor lesions at a sufficiently early stage for treatment to be successful.¹²

The International Agency for Research on Cancer has emphasised that PBOS programmes, as opposed to opportunistic ad hoc screening, provide greater protection against many of the harms of screening, including overtesting, poor quality, complications of screening and poor follow-up of those who test positive.⁴ Ensuring an effective population outcome, dependent initially on test uptake (screening participation), is of considerable importance.^{4,9,13}

Screening programmes vary around the world.^{3,14} PBOS programmes operate in many European countries as well as Canada, Taiwan, Australia and New Zealand,^{3,4,14} while SOS operates in countries such as the USA.³ It needs to be shown how a new test provides benefit in the context for which it is intended without adversely affecting the health status or overly burdening those who participate.^{4,12}

2. Screening is a multistep process

The screening test is just one step in a coordinated multistep process that includes initial and repeated participation by the intended-use population, quality-assured testing, diagnostic follow-up, treatment and referral to high-risk surveillance programmes when appropriate, together with monitoring of key indicators along the screening pathway. Goals for each step in the process should be defined and agreed on by providers.

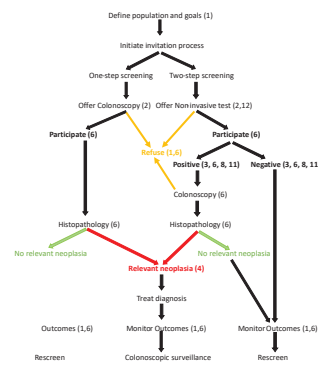


Figure 1 The multistep screening pathway characteristic of organised screening programmes and demonstrating the one-step and two-step strategies³ as discussed in principle 2. How the principles relate to steps and outcomes are identified by the numbers in parentheses. Colours identify important outcomes relevant to detection.

Explanatory text

Performing a screening test is just one event in a complex process that starts with an invitation to get tested and proceeds through diagnostic follow-up, and treatment for identified lesions, with further screening and surveillance as indicated. Key aspects of the multistep screening process necessary to ensure population and individual benefit have been described³ and are shown diagrammatically in figure 1. The main, but not only, test-dependent components of the screening pathway are test accuracy and test participation (willingness of invited individuals to do the test). Demonstrating the value of a screening test must be rigorous and demonstrated at all relevant steps of the screening pathway.¹²

How different countries go about screening and set standards varies widely.³ Established standards will depend on the nature of a healthcare system, what it wants to achieve and what is feasible. For instance, limitations in colonoscopy capacity could lead providers to set the positivity threshold ('analytical specification'), to match colonoscopic capacity, even though test sensitivity might be compromised (see principle 8). Consequently, it will be essential to demonstrate how a new test meets programme goals at each step in these different healthcare settings. The target population should be informed of estimated test accuracy and its expected benefit using well-designed communication strategies to empower individuals to make their own decision.^{3,4} How the principles established in this consensus process relate to steps in the screening pathway are shown in figure 1.

3. A screening test identifies individuals with an increased likelihood of CRC and/or advanced precursor lesions

In two-step screening, based on first performing a non-invasive test followed by colonoscopy if positive, the non-invasive test should identify participants with an increased likelihood of CRC or advanced precursor lesions.

Explanatory text

Globally, the traditional 'two-step screening' approach, where a relatively simple, usually non-invasive screening test is used to identify a subpopulation more likely to have neoplasms of interest^{7,15} is more commonly used.³ Only those participants returning a positive test result are invited to follow-up colonoscopy as they are more likely to have colorectal neoplasia. Identifying the population subgroup more likely to have CRC or advanced precursor lesions substantially reduces the number of

colonoscopies undertaken in individuals unlikely to benefit from colonoscopy.

Neoplasia of interest includes likely curable CRC (stages I and II although 5-year survival in stages IIIA and IIIB is >60% in some countries¹⁶) and advanced precursor lesions (discussed further in principle 4).

It is possible that the number of people to undergo colonoscopy because of a positive test might differ based on the intended-use population. The prevalence of colorectal neoplasia in a population may differ, for example, based on age and gender,¹⁷ and a non-invasive test's performance should be relevant to the population in which it is employed.

In 'one-step screening', more common in jurisdictions undertaking SOS, age is the sole risk-identifying factor that determines who gets colonoscopy.

4. Nature of precursor lesions most important to detect

The precursor lesions currently considered to be of sufficiently high risk to be important to detect are advanced adenomas and advanced serrated lesions. More research is needed to clarify how best to characterise the most important precursor lesions to detect and remove.

Explanatory text

A screening test that identifies individuals with an advanced precursor lesion, that is, a non-invasive neoplasm that carries a risk for progression to CRC, is likely to reduce CRC incidence, as evidenced by a range of studies in which precursor lesions were removed at sigmoidoscopy or colonoscopy.^{18–22} Precursor lesions comprise a range of adenoma types and serrated lesions (polyps), each with different and sometimes uncertain risks for progression to CRC. The characteristics of those currently considered important are detailed in online supplemental material 4.1. How best to detect and characterise these precursor lesions, especially serrated lesions (or polyps) but even conventional adenomas, remains poorly defined.^{23–24} Identifying advanced precursor lesions (advanced adenoma and advanced serrated lesions) depends on subjective characteristics (endoscopist description of size and histopathology) that are open to observer error and subject to variation in professional opinion. In addition, differentiation between advanced serrated lesions and other serrated (hyperplastic) polyps, based on the architectural distortion of crypts, is crucial but often inconsistent between pathologists.²³

While there is a large body of knowledge regarding the natural history (ie, prognosis and responsiveness to treatment) of CRC, the natural history is much less clear for specific morphological features of precursor lesions, because patients with these lesions are usually asymptomatic and the natural history of precursor lesions, when detected, is interrupted by polypectomy. We lack an objective means of identifying risk, and information on the distribution of transition times from small to advanced lesions and, most importantly, from advanced precursors to CRC. Unfortunately, very little research is available that addresses this.²⁵ More research is needed to establish characteristics that objectively identify those characteristics of the lesions that are most important to detect in screening; *useful characteristics could be biomarkers of a molecular nature*^{26–27} rather than the current characterisers based on assessments of morphology with their limited reproducibility. Future research seems likely to lead to a change in definition of what is an advanced precursor.

To evaluate diagnostic accuracy for advanced precursors, colonoscopy is the best reference standard since it is the most sensitive diagnostic procedure for detecting these lesions throughout the

colon and rectum.² The sensitivity of the faecal immunochemical test for haemoglobin (FIT) for precursor lesions is limited as outlined in online supplemental material 4.2. The non-invasive multitarget stool test (mtsDNA), which tests for faecal haemoglobin (ie, incorporates FIT technology) and neoplasia-derived DNA, is more sensitive than FIT.²⁸

5. New biomarkers might detect lesions with a different natural history

Non-invasive tests targeting new biomarkers might detect lesions that differ in their natural history from those detected by established tests. In theory, CRC detected using a new biomarker might be more or less responsive to treatment than those detected by an existing test, although any screen-detected CRC is considered to be worth treating. Differences in risk profiles (genotype or phenotype) of precursor lesions detected solely by new tests are possible. Exploring concordance between the new and comparator test results will identify if differences in neoplasia-related biology should be considered, especially for precursor lesions.

Explanatory text

The analytical target (or analyte or 'measurand') of a new test—especially a test targeting DNA alterations, proteins other than haemoglobin or other molecules—might reflect neoplasia-related biology (clinical phenotype) that is not shared by neoplasms without that analyte. While this does not necessarily mean that the natural history of the new test detected lesion is different from that of those detected by existing tests such as FIT, if such a difference does exist, then the benefit from detection and treatment of a lesion detected solely by the new test might be different.^{2–12}

It is theoretically possible that a CRC detected solely by the new test might be more resistant to current treatment, be more likely to recur or even be relatively indolent (although any screening-detected CRC is considered to be harmful and worth treating). Such a difference or 'shift' in treatability might not be reflected in conventional assessment for the appropriate treatment regimen, based primarily on histopathology and stage. Considering that endoscopic detection of cancers and precursor lesions leads to a reduction in CRC mortality²⁹ regardless of genotype, then a shift in treatability of cancer based on genotype might not be a major concern.

Concerning precursor lesions, rather than a shift in treatability, the new test might identify lesions at lower or higher risk of progressing to CRC even if morphologically similar. It has been suggested that efforts should be made to characterise the phenotypes of lesions detected solely by a new test.¹² However, because of our limited understanding of the natural history and the fact that polypectomy interrupts observations of natural history, we lack a way of determining the relative benefit of detecting one precursor lesion over another. For example, at this time, we do not know the relative CRC incidence-reducing merits of removing advanced serrated lesion compared with advanced conventional adenomas.

Despite these uncertainties, exploring concordance between the new and comparator tests will be informative since detection by one test but not the other will identify whether a shift in treatability might need consideration.

There is a possibility of gender specificity with new biomarkers, for example, lesions detected at older age in females can arise from the progression of a proximal serrated lesion.

6. Outcomes to be estimated in a screening population

If a non-invasive test is to be widely used in screening programmes and fully supported in guidelines, its application at key points along the multistep screening pathway should be assessed in the intended-use population. In addition to diagnostic sensitivity and specificity, measures would include acceptability to invitees, technical test failure rates, subsequent colonoscopy workloads, cost-effectiveness and surrogate measures for reduction in CRC mortality and incidence. Comparative effectiveness randomised controlled trials (RCTs) are ideal for such purposes. Alternatively, modelling studies mimicking such RCTs based on high-quality observational data can also be informative.

Explanatory text

The value of a non-invasive screening test is determined by how well it detects CRC and advanced precursors and by how well it improves elements of the screening pathway (principle 2) that are directly dependent on the test. Such improvements, especially in screening participation rates, can only be confidently demonstrated in a screening population in which the test is intended to be used.

Screening programme objectives

Three simple, readily determined indicators of screening programme quality³⁰ are:

1. Positivity rate: number of tests positive/number of tests done.
2. Detection rate: number of cases with a specific neoplastic lesion/number of tests done, either relative to those undergoing follow-up colonoscopy or to those invited.
3. Positive predictive value: number with a lesion of interest/number of colonoscopies done for a positive test.

These can be used to compare tests when applied to a screening population but do not comprise all of the measures commonly used in organised screening programmes. Performance indicators and quality measures drawn from existing programmes^{31 32} that are aimed at monitoring each step of the pathway can be used to estimate the benefit of a new test when applied in a screening context (table 2). These variables typically include test-relevant data pertaining to programme type, invitation, test outcomes, lesion detection and adverse events, including test failure rates. These are the basis for deciding whether a new test can meet the goals of a screening programme. Test-independent outcomes, such as the quality of colonoscopy and wait times, are not directly relevant to the current consideration of new test evaluation.

It is acknowledged that the epidemiology of CRC will undoubtedly change over time and alter the nature of the targetted populations as screening becomes more widespread, the population ages and lifestyle risk factors change. Lesion treatability is likely to improve over time and so comparative studies described below (principle 7) must be undertaken in equivalent populations at the same time.

Accuracy estimates are most reliable when all cases undergo colonoscopy, regardless of the test result. But population detection rates can be ascertained on an intention-to-screen (ITS) basis, which take into account differences in test participation rates due to logistic and other considerations inherent in obtaining and performing the test.

Acceptability of the screening test is demonstrated by the participation rate, a major performance indicator for PBOS programmes. Parallel comparison of new and comparator non-invasive tests in phase III (separate randomised cohorts) will demonstrate if the new test is likely to deliver better outcomes.

See online supplemental material 6.1 for further variables that might be test dependent.

Programme goal considerations

With a quantitative test, it will be possible to construct receiver operating characteristic curves to estimate how different positivity thresholds affect sensitivity and specificity.³³ However, screening programmes vary in the degree to which of colonoscopy workload, cost, sensitivity or specificity (for varying combinations of cancer and/or advanced precursor lesions) drive programme goals. Defining specificity can be challenging depending on what classes of neoplastic lesion are considered to be prime targets.³⁴ It might be advantageous to employ a strategy in which the threshold is first set according to test results in a large number of cases that lack cancer and/or advanced precursor lesions but are otherwise typical of the target screening population.³⁵ However, laboratory medicine's population-based reference limits are not well established for faecal haemoglobin levels, and non-neoplastic lesions also bleed.

Modelling the value of fine adjustment of the test positivity threshold will be facilitated by using the lowest threshold to trigger colonoscopy in trials and then modelling the ideal positivity threshold that is feasible for colonoscopy resources while also considering any compromise in sensitivity for cancer and/or advanced precursor lesions. Most screening programmes rely on repeated rounds of screening, so a test of limited sensitivity can be compensated for by acceptable programme detection rates over multiple rounds.³⁶

An improved non-invasive screening test should detect CRC at a curable stage; hence, sensitivity for stages I and II CRC will inform the likelihood of mortality benefit as will demonstration of a shift to earlier stage at diagnosis.³⁷ Comparing the stage distribution of detected cancers between tests will be useful as the test detecting the higher proportion of early cancers may likely lead to improved survival. However, some caution must be exercised when estimating mortality benefits based on stage.³⁸

When aiming for higher diagnostic sensitivity, there is a potential for overdiagnosis (detection of indolent lesions that in the absence of screening would not have led to morbidity during that person's lifetime). The consequences of overdiagnosis are overtreatment and cost. The 20-year follow-up results of the Nottingham RCT³⁹ suggest that overdiagnosis of CRC might not be a major problem, although the guaiac-faecal occult blood test (gFOBT) used was not as diagnostically sensitive as current FIT.

A high diagnostic sensitivity is required when the goal is to increase the detection of precursor lesions, to reduce CRC incidence. Overtreatment then becomes a critical issue as, when increasing diagnostic sensitivity, the increased detection of precursor lesions may lead to the excision of lesions that are not clinically relevant, especially if not advanced (see principle 4). High diagnostic sensitivity may also result in a higher burden of surveillance colonoscopies.⁴⁰ See online supplemental material 6.2 for further discussion on the challenges of predicting incidence reduction from precursor detection.

Modelling outcomes

Not all relevant questions are directly answered by phases I–IV studies (as defined in principle 10). Simulation modelling can estimate the long-term clinical and economic impact of screening in the population based on screening test performance characteristics, testing interval, complication rates, costs and participation rates using data obtained at each step of the screening process in phases III and IV studies. Where modelling is needed

Table 2 Screening programme performance indicators and quality measures that are primarily dependent on the test and which can be estimated at specific phases of the evaluation process

Category of screening outcome	Phase	Measure	Comment
Initial performance indicator	III		
Invitation		Participation (uptake)	Fraction of invitees who complete a test.
		Non-participation	Fraction failing to complete a test. Ascertain reasons why refused.
Sample laboratory analysis		Positivity fraction	Fraction of test results that are positive. Determines colonoscopy workload.
		Test failures	Degree to which samples are unsuitable for, or otherwise fail, in the measurement whether due to collection, storage, transport or measurement difficulties.
Detection rate*		Number of CRCs detected in the population	Fraction of participants doing the test, where CRC is diagnosed at follow-up colonoscopy (per-protocol analysis). OR Fraction of invitees in whom CRC is diagnosed (intention-to-screen analysis). Stage distribution of detected CRCs to also be determined.
		Cancer stage distribution and/or detection rate by stage	Shift to more favourable stage at diagnosis points to a likely benefit on CRC mortality.
		Number of stage I and stage II CRCs detected in the population	
		Advanced precursor lesions detected in the population	Shift to more favourable stage at diagnosis points to a likely benefit on CRC mortality.
		Any adenoma or serrated lesion detected in the population	Shift to more favourable stage at diagnosis points to a likely benefit on CRC mortality.
		Non-neoplastic pathology in the population	This will inform the variables associated with false positivity for CRC or advanced precursor lesions.
Predictive values		Positive predictive value for each category of neoplasia	The efficiency of detection: the proportion of positive tests with CRC or advanced precursor lesions.
Test accuracy		Estimates of sensitivity and specificity	Only obtainable if all participants undergo colonoscopy†
		Estimates of sensitivity, specificity relative to a comparator	Obtainable by comparing true and false positives of new and comparator test (does not require all participants to undergo colonoscopy)‡.
Burden of detection		Number needed to colonoscope to detect one lesion of interest	Inverse of the fraction of all those colonoscoped who have CRC or advanced precursors. A simple indicator of cost-effectiveness.
Subsequent performance indicator	III	Interval cancers ^{31 32}	Best ascertained with follow-up time equal at least to the recommended screening interval. This may include lesions missed by the screening tests (interval lesions among subjects with a negative non-invasive test result) and lesions missed at colonoscopy (ie, among subjects with a negative follow-up colonoscopy). Identify the optimal retest interval.
	IV	Cumulative detection rates over multiple rounds ^{31 32}	Such can also be expressed as a fraction of all test-positive participants.
		Detection and burden of detection according to interval between tests	Further inform the optimal retest interval.
		Cumulative colonoscopy workload	Over subsequent rounds, more and more people will require diagnostic assessment.
	III and/or IV	Modelling of efficacy and effectiveness	CRC mortality and incidence modelled from the above indicators. Stage distribution at diagnosis will be crucial.
		Modelling of cost-effectiveness	Costs will be specific to the jurisdiction of application of the test.
		Unexpected adverse events ³²	Include mortality and 30-day hospital readmission rates following colonoscopy.
Each measure can be compared between the new and a comparator non-invasive test.			
*Detection rates will depend on the threshold chosen for positivity. It will be possible to estimate rates across a range of thresholds depending on the trial design (see principles 8 and 10).			
†Sensitivity is the proportion of cases with a given neoplastic lesion type(s) that return a positive test. Specificity is the proportion of cases without CRC or advanced precursor who return a negative test.			
‡Note that biases referred to in principle 10 will mean that the actual estimates are not necessarily reliable. These are best determined when all cases undergo colonoscopy. CRC, colorectal cancer.			

to inform policy, it is highly desirable to use model inputs that have been ascertained with relatively high confidence in RCTs or large prospective studies. The latter includes the phase III and IV studies outlined under principle 10.

As discussed in online supplemental material 6.3, use of proxies/surrogate measures (eg, accuracy, positive predictive

value) to predict long-term benefit should be interpreted with caution.

7. Expectations of a new non-invasive test

RCTs with CRC mortality and/or incidence reduction as the primary outcome have set the expectations for performance

of new tests. The effectiveness of a new test may be predicted when compared with a standard (ie, comparator or index test) where the comparator's effectiveness has been previously demonstrated.

Explanatory text

Multiple population-based controlled trials of two-step screening analysed on an ITS basis have established the reductions in CRC mortality and/or incidence achievable through early detection of lesions exhibiting a bleeding phenotype (ie, positive gFOBT result) and of those visualised at screening flexible sigmoidoscopy.^{4,29} The gFOBT initially set the standard expected of a new non-invasive test in mass population screening² due to multiple controlled trials demonstrating its efficacy (see online supplemental material 7.1).

The main test-dependent components of the performance of a screening programme are test participation and test accuracy. The decrease in CRC mortality and incidence depends on the participation rate, a test's diagnostic sensitivity and the quality of treatment, while specificity influences feasibility of a programme (colonoscopy capacity and cost-effectiveness). If a new non-invasive screening test leads to improvement in one or more of these components, compared with a screening test with RCT-documented decrease in CRC mortality and/or incidence, then an even more significant population benefit could be obtained with limited risk of error, especially if this test is based on the same biological principle.

For FIT, the analyte is the globin moiety of haemoglobin rather than the chemical properties of haem (as for gFOBT), but both detect the bleeding phenotype. FIT has proved to be superior to gFOBT.^{4,41} The improved diagnostic sensitivity without compromising specificity (depending on the positivity threshold chosen), the availability of automated development and reading of some commercial FIT and their relative ease of use and superior population participation, as well as fewer interval cancers in FIT compared with gFOBT screening^{42,43} have been established in many studies.⁴¹

In view of the evidence supporting the use of FIT (see online supplemental material 7.2), and even though FIT has not been evaluated by RCT against no screening, the FIT technology is now almost universally used in place of gFOBT in PBOS programmes throughout the world.^{3,4} FIT is widely recommended in professional guidelines as the preferred technology to be used in SOS.¹³ FIT has also been the obligatory comparator used in trials designed to obtain regulatory approval of new non-invasive tests.²⁸ The FIT technology, therefore, at this time stands as the main option for a comparator test against which new non-invasive tests can be judged.

A test's capacity to detect precursor lesions considered to be at high risk of progressing to CRC will identify its capacity to reduce incidence of CRC^{12,44} and so estimates of the diagnostic accuracy of a new test for advanced precursors are desirable. However, there is no consensus on what constitutes an adequate sensitivity, except that it should be at least as good, if not better, as a high-quality FIT applied with a low positivity threshold (ie, faecal haemoglobin concentration) (see principle 11). Detection of precursor lesions will require a relatively high positivity rate and hence a higher rate of follow-up colonoscopy⁴⁰ than would be needed if the primary focus were to be on CRC.

Flexible sigmoidoscopy might be an expeditious comparator for evaluating new tests targeting precursor lesions (see online supplemental material 7.3). However, it is not used in many countries as it is relatively invasive and has been supplanted

by colonoscopy as the primary option for an invasive CRC screening test in many countries.

Colonoscopy serves dual purposes in two-step screening: as the best means of diagnostic verification of a positive non-invasive test in two-step screening for CRC and as a therapeutic procedure for removing detected neoplastic lesions. While colonoscopy is an invasive test, high-quality colonoscopy should be used as the standard against which a new non-invasive test is compared, when test accuracy, rather than ITS performance, is the main aim, where it is necessary to provide absolute estimates of diagnostic accuracy and when missed lesions require detection and characterisation.⁴⁵ While a recent RCT documented benefit for screening colonoscopy (see online supplemental material 7.4), few countries apply colonoscopy as the primary screening test in organised screening.³

FIT is a suitable standard by which to judge a new non-invasive screening test in the context of organised two-step population screening, provided that the FIT selected for comparison is one with well-established diagnostic accuracy supported by large reference data sets from screening practice. Screening programme variables associated with FIT, particularly participation rates, diagnostic accuracy (sensitivity and specificity), the efficiency of detection (number needed to colonoscope to detect a lesion and predictive values), the interval cancer rate and earlier stage at diagnosis, all serve to demonstrate if a new test will be likely to meet programme expectations.² Based on these variables, modelling impact on CRC mortality and incidence (see principle 6) will further demonstrate the value of a new non-invasive test.

8. An adjustable test positivity threshold accommodates different programme goals

A non-invasive screening test with an adjustable positivity threshold ('cut-off') or algorithm, enables the choice of test accuracy parameters (diagnostic sensitivity and specificity) and test positivity rate that best match the desired goals of a screening programme. Regulatory approval processes for a new test should consider the capacity for quantitative reporting of test results. This enables screening programme providers or policymakers to choose the desired test performance characteristics.

Explanatory text

The test positivity threshold determines the test positivity rate, which in turn determines the colonoscopy workload, the number needed to colonoscope to detect one case with CRC or an advanced precursor lesion (a potential surrogate measure for cost-effectiveness) (see principle 6), the detection rates of target lesions and the positive predictive value.⁴⁶ Thus, the threshold chosen for test result positivity is a crucial variable in screening for determining the likelihood of neoplasia (related to test accuracy and prevalence) and for the feasibility of performing all the necessary colonoscopies, and to determine diagnostic accuracy, mortality and incidence benefits and cost-effectiveness

Many simple non-invasive screening tests, such as gFOBT or qualitative FIT, have a fixed threshold for positivity set by the manufacturer based on initial clinical studies.^{46,47} Screening programme outcomes using such tests will be constrained to the accuracy determined by the positivity threshold. As programme goals vary around the world, many organised programmes now use quantitative FIT,³ which provide the capacity to adjust the test positivity threshold. If a new test does not possess the desired accuracy in a qualitative format that meets the goals of a programme, then another test that allows the programme to

choose a positivity threshold corresponding to the desired colonoscopy workload capacity, sensitivity, predictive values and/or specificity is likely to be preferred.

Studies based on FIT show that lower thresholds lead to a higher sensitivity and an increased chance of detecting precursor lesions⁴⁰; this will reduce CRC incidence but reduce specificity and increase costs. Screening programme providers vary in which are the most important of these variables. For example, PBOS programmes often choose a test positivity threshold that results in colonoscopy workloads in a single screening round ranging from only a few per cent of the population to as high as 10%–20%.³

Capacity to adjust the test threshold enables management of costs associated with colonoscopy, workforce availability, treatment costs and public expectations inherent in equity-focused programmes. For example, if a target specificity is decided for a programme, then comparing the sensitivity when the threshold for each test is set at the same specificity is simple and informative.⁴⁸ Similar approaches can be undertaken when setting equivalent test positivity rates, when the burden of detection, or the feasibility of colonoscopy workload, is a prime consideration of the screening programme.

Providing an adjustable end point does present challenges for regulatory assessment, test invention and manufacturers. These manageable challenges are discussed further in online supplemental material 8.1.

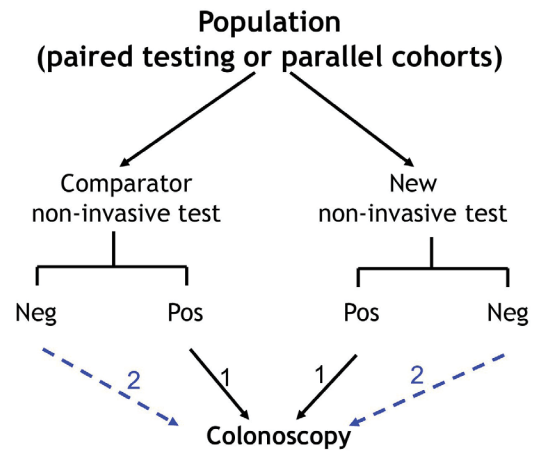
9. Predicting value by paired comparison with a proven non-invasive screening test

The performance of a new non-invasive screening test can be assessed in parallel or paired with an existing non-invasive screening test of proven effectiveness at any step in the screening process, from population engagement to key outcomes. Intermediate end points known to reliably and consistently predict the potential for reducing CRC mortality and/or incidence can be used to compare new with existing screening tests. Such end points include estimates of diagnostic accuracy.

Explanatory text

It has been argued that when an RCT has established that a non-invasive screening test is effective in reducing CRC mortality, then a new non-invasive test does not need to be evaluated in an RCT with CRC mortality as the outcome, provided it is compared with a proven test (an existing non-invasive screening test with known effectiveness)^{2,12} in well-designed studies appropriate to the context of its use. Consequently, comparing a new test with an existing test of proven effectiveness will be highly informative and efficient for initial evaluation, as proposed for phases I and II evaluations (see principle 10 for the phases of evaluation). Proceeding to larger-scale, more definitive studies in typical screening populations will depend on the findings in these early simpler studies. It is acknowledged that the comparator test used may change over time, based on evolving data. Ongoing studies such as the CONFIRM trial^{49,50} will assist in understanding whether projections based on modelling assumptions for tests such as FIT are indeed correct.

How initial studies evaluating a test can be planned in relatively small neoplasia-enriched populations to achieve a direct head-to-head comparison of a new test with a proven comparator using an efficient comparative design is shown in [figure 2](#).² [Table 3](#) shows how to compare test accuracy (see online supplemental material 9.1 for further discussion).



- 1 For comparing true-positive and false-positive fractions.
- 2 For comparing sensitivity and specificity (depending on biases due to population selection).

Figure 2 Diagrammatic outline of a trial design appropriate for comparing non-invasive tests in the initial phases of test evaluation. Paired testing is conducted in a single cohort where an individual does both the new and the comparator test, whereas parallel testing is where study participants are randomised to one or the other test. Neg, negative result; Pos, positive result.

Comparing tests using either approach will identify if the new test has promise, and whether more rigorous and costly evaluation in the unbiased screening context would be worthwhile. Results are unlikely to be sufficient in themselves for acceptance by regulatory authorities, policy-makers, payers and professional guideline-making bodies.

Comparing test sensitivity for CRC and advanced precursor lesions infers mortality and incidence benefits of the new test relative to the known benefits of the comparator test, as sensitivity can be considered an intermediate/surrogate end point for mortality benefit.^{12,44}

Outcomes of importance in the typical screening context (see principle 6) in addition to specificity and sensitivity will be determined in larger-scale unbiased studies with longitudinal follow-up (see phases III and IV studies in principle 10).

10. Evaluation proceeds through increasingly complex phases

Evaluation of a new test should follow a four-phase (sequential) evaluation. This would start with limited-scale cohort or case-control studies in populations with and without neoplasia, possibly enriched for the neoplastic outcomes of interest. Initial estimates of diagnostic accuracy and test positivity threshold will be obtained in phases I and II studies. If results suggest that the test might achieve the desired diagnostic accuracy, evaluation should proceed to screening pathway evaluation requiring larger intended-use screening populations (phases III and IV). The latter studies should be prospective and will identify the most suitable threshold for test result positivity, among other important outcomes.

Explanatory text

Phased (ie, sequential) evaluation in a stepwise manner is an efficient way first to establish the potential value of a new test and then to subsequently gather the evidence that will lead to its acceptance by professionals, healthcare providers and regulatory bodies.^{2,44,51} There are four main phases ([figure 3](#)).

Table 3 The relationship between direct, practical measures of a non-invasive screening test's accuracy, how each measure informs assessment of diagnostic accuracy (sensitivity and specificity) and the consequence of the result for a screening programme

Test result	Diagnostic verification*	Result of diagnostic verification	Related accuracy characteristic	Programme consequence
Positive	Yes	True, hence TP (TP ₁ , TP ₂)†	Relative sensitivity; TP ₁ /TP ₂	Detection of neoplasia
		False, hence FP (FP ₁ , FP ₂)	Positive predictive value; TP ₁ /(TP ₁ +FP ₁) TP ₂ /(TP ₂ +FP ₂) Relative FP rates; FP ₁ /FP ₂	Efficiency of detection Burden associated with detection
Negative	No			
Test result	Diagnostic verification*	Result of diagnostic verification	Related accuracy characteristic	Programme consequence
Positive	Yes	True, hence TP	Sensitivity; TP/(TP+FN)	Detection of neoplasia
		False, hence FP	Positive predictive value; TP/(TP+FP)	Efficiency of detection Burden associated with detection
Negative	Yes	True, hence TN	Specificity; TN/(TN+FP)	Accuracy of detection
		False, hence FN	Negative predictive value; TN/(TN+FN)	Exclusion of neoplasia Miss rate (missed lesions)

The upper section shows relationships when paired testing is undertaken and only test-positive cases undergo colonoscopy while the lower section refers to the situation when all cases undergo colonoscopy.

*Verified by colonoscopy.

†TP₁, TP₂—TP by tests 1 and 2; FP₁, FP₂—FP by tests 1 and 2.

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Prescreening evaluation (phases I and II)

Phases I and II evaluations are intended to be relatively simple and demonstrate whether a test can discriminate between cases with CRC and non-neoplastic states (phase I) before proceeding to gather initial estimates of diagnostic accuracy and identify likely confounders (phase II). Comparative studies using an existing proven non-invasive test, as proposed in principle 9, can provide a strong indication of the potential of a new test and its suitability to advance to phase III studies. Phase II studies are essential for the initial establishment of test positivity criteria that best discriminate between neoplastic and non-neoplastic states. Design strategies are shown in figure 2 (principle 9).

In phases I and II, limited-scale cohort or case-control studies in populations with and without colorectal neoplasia are conducted. Participants in phases I and II studies need not be sourced from typical screening populations and may be enriched for clinical states of interest, especially cancer.² Further important design considerations are presented in online supplemental material 10.1.

Evaluation in the screening context—phase III

Phase III screening trials provide the minimum level of evidence required to justify use in large-scale screening programmes to

Phase	Goal(s)	Context	Approach and measures	Hurdle for progression
1	Main: Differentiates between CRC and non-neoplastic states?	Prescreening cohorts – limited	Distribution of test results in cohorts with and without CRC	• Test result must differ significantly in cancer cases.
2	Main: Detects early cancer and precursor lesions? Others: Initial positivity threshold? Accuracy relative to comparator? Causes of false positives.	Prescreening cohorts - extensive	Distribution of test results in cohorts with CRC relevant precursor lesions, other colorectal diagnoses and no disease. Parallel or paired testing of new and comparator tests will be informative.	• Preliminary (although biased) estimates of accuracy are shown to be promising. • ROC analysis identifies a suitable positivity threshold.
3	Main: Test accuracy in a typical screening evaluation? Test acceptance? Others: Test failure rate? Other variables for modelling effectiveness and cost-effectiveness.	Screening populations – single round	Apply test prospectively to a typical unbiased intended-use population. Choose study design appropriate to program goal and jurisdictional context: e.g., colonoscopy all for estimating test accuracy, parallel testing for comparing non-invasive tests and intention-to-screen outcomes.	• A significant improvement in some aspect of screening. • Non-inferior in accuracy to a comparator test, OR • Accuracy likely delivers benefit. • Feasible colonoscopy workload. • Modeled effectiveness and cost-effectiveness are satisfactory.
4	Main: Missed lesions or adverse events? Others: Participation rates over time and retest intervals?	Screening population – multiple rounds	Apply the test prospectively to an intended-use screening population over multiple rounds, with careful monitoring of population program outcomes.	

Figure 3 Goals, context and approach for each phase of evaluation, together with the hurdle identifying justification to advance to the next phase. Note that clinical status must be ascertained by colonoscopy. CRC, colorectal cancer; ROC, receiver operating characteristic.

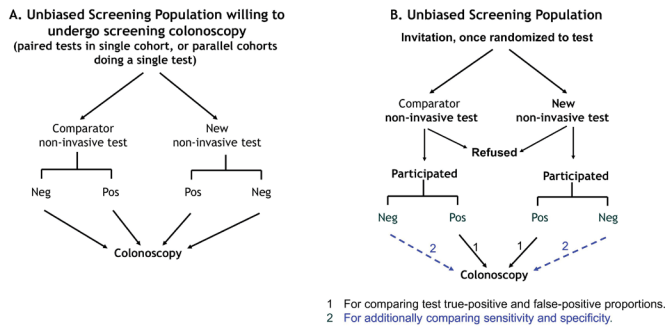


Figure 4 Study design frameworks applicable to phase III studies. (A) Design appropriate to determine test accuracy where all cases undergo colonoscopy, but intention-to-screen outcomes cannot be ascertained (comparison of a comparator with the new test can be paired in a single cohort or parallel in separate cohorts). (B) Design appropriate for estimating intention-to-screen outcomes and where the accuracy of the new test can be compared with that of a non-invasive comparator either when colonoscopying only test-positive individuals (compare true-positive and false-positive fractions) or all participants (sensitivity and specificity). Neg, negative result; Pos, positive result.

satisfy applicable regulatory pathways, any need for health technology assessment and the goals of screening in a jurisdiction. Such studies seek to confirm the value of the new test when applied in the screening context as a one-time event (a single screening round). They must be undertaken in a typical intended-use screening population and be prospective. There is an obligation to have ensured in phase II that diagnostic accuracy is likely to be suitable for screening before offering such tests to a screening population in a research context, and the study population should be appropriately informed. Study design options are shown in figure 4.

Phase III studies will commence using a provisional test positivity threshold. The options for setting this value will be jurisdiction-dependent and are:

1. One that is most discriminatory between those with cancer or advanced precursor lesions and those who are normal or with other pathologies (eg, non-advanced precursors), or
2. One aiming to detect most cancers, or
3. One aiming to detect most advanced precursor lesions or
4. One aiming to minimise the chance of returning a positive result in a person without CRC or an advanced precursor lesion.

Subsequent findings in phase III can be modelled to show how the adjustment of the test positivity threshold would affect colonoscopy workloads, detection efficiency, sensitivity and specificity.

Two main study designs require consideration for phase III. The first is when a jurisdiction requires absolute estimates of sensitivity and specificity (common for the SOS context). It requires that all participants must perform the non-invasive test and undergo screening colonoscopy regardless of test result. The second is when a jurisdiction requires ITS outcomes (expected in PBOS contexts). Here, the new and comparator tests are compared, thus allowing for determining the respective participation rates and how the relative detection of lesions depends on participation with each test. Important design considerations are presented in online supplemental material 10.2 including an approach to achieving the ideal positivity threshold.

These studies, conducted as a one-time event (a single screening round) will confirm the value of the new test in the intended screening context according to what is required in

that jurisdiction. Sensitivity and specificity (or true-positive and false-positive rates if not all cases are colonoscoped) will have been estimated. The positivity threshold options, the stage distribution of detected CRCs as well as the participation rates to be expected will be apparent for the new test. Certain other screening programme outcomes presented under principle 6 will also be apparent.

Evaluation in the screening context—phase IV

Phase IV evaluation involves the application of the new test to a large typical screening population over multiple rounds of screening to allow postimplementation monitoring, including comparative evaluation, particularly where the comparator is the usual screening strategy applicable in that environment. Only test-positive cases undergo follow-up colonoscopy. Referring all persons to colonoscopy irrespective of the test result would lead to a situation that no longer reflects the repeated application of the non-invasive test in many screening settings.

Screening with simple non-invasive tests involves repeated screening at an appropriate interval.⁵² When monitoring cases over multiple rounds, the frequency of unexpected adverse events will be apparent, while the characterisation (stage distribution and biological characteristics) of missed interval cancers and of incident CRC (new CRC at subsequent rounds) will provide additional evidence about the performance of the new test. Results will facilitate the selection of the interval between screening tests and provide real-world data for observational analysis using target trial emulation⁵³ and for inclusion in effectiveness and cost-effectiveness modelling (see principle 11). A critical and unresolved issue after phase III evaluation concerns assumptions about round-to-round test performance. How independent is sensitivity for a given type of lesion in subsequent rounds? Is a missed lesion more likely to be detected or not?

Further design considerations are presented in online supplemental material 10.3.

11. Accuracy required for evaluation in a screening population

The desired diagnostic accuracy considered sufficient to proceed to intended-use population evaluation in phase III will be subject to a range of considerations that vary between jurisdictions. Nonetheless, it is considered ideal if phases I and II studies demonstrate that the accuracy of a new test is at least comparable to that of non-invasive tests accepted for use in existing screening programmes.

Explanatory text

The consensus process considered that the accuracy of a new non-invasive test should be 'at least comparable to that of non-invasive tests accepted for use in existing public-health screening programmes'. Ideally, it would improve on the current FIT standard. Test-dependent outcomes such as higher participation or lesser cost might justify using a new test even if it has no accuracy advantage. These outcomes cannot be assessed in phases I and II. Thus, the key question to be answered by phases I and II studies—before proceeding to phase III—focus on diagnostic accuracy and whether the new test exhibits non-inferiority of sensitivity and specificity for CRC and advanced precursor lesions.

When comparing a new test with a comparator non-invasive test (typically FIT), the configuration of the positivity threshold for the comparator test must be set to deliver an accuracy considered appropriate for the screening programme in which the new

test will be used. There is no universally agreed accuracy standard, but by configuring the comparator test (such as a quantitative FIT) to meet the accuracy recommendations of bodies such as the United States Preventive Services Task Force (USPSTF), comparison studies should give confidence in the *potential* of the new test. Based on published trials, the USPSTF has suggested that an acceptable sensitivity for CRC is at least 70%, and specificity for cancer plus advanced precursor lesions is at least 90%⁸ *using a one-time test*. The US Centers for Medicare Services guidance on the performance of new tests is comparable (74% sensitivity and 90% specificity).⁹ The positivity threshold and sample number combination that achieve these standards are now established for several quantitative FIT.⁴⁶

It must be recognised, however, that while FIT is an appropriate comparator, in recent years, a plethora of new FIT tests have emerged with a wide variation in accuracy.^{54–55} Most are qualitative with the positivity threshold set by the manufacturer and have not been subjected to large population studies,⁵⁶ but several are quantitative, allowing for adjustment of the positivity threshold, and have been extensively studied.^{57–59}

Setting analytical specifications of a new test where the technology incorporates new biomarkers or biomarker combinations will initially depend very much on phases I and II evaluations, where positivity criteria are usually based on the combination that best discriminates between CRC and non-CRC states. Typically, these generate diagnostic sensitivities that are not maintained in validation studies undertaken in phase III studies (typical screening populations) and positivity thresholds might need adjustment. Investigators should clearly state whether reported test performance is derived from discovery or validation data and where feasible, this should be disclosed to study volunteers. It must also be recognised that adjusting the positivity threshold will change sensitivity and specificity.

12. Analytical specifications, standards and performance

Before assessing the diagnostic value of a new test in an ‘intended-use’ population, the analytical performance characteristics of the test must be formally documented according to relevant standards, such as those of the international Clinical and Laboratory Standards Institute (CLSI) or the Quality System Requirements (QSR) of the USA. The test manufacturer should provide this information in their instructions for use. Evaluation of analytical characteristics should ideally conform to recommended protocols described by standards such as those of the CLSI and QSR. Researchers undertaking the development of a new test should abide by such protocols and undertake appropriate verification processes to ensure that prescribed standards are attained or surpassed. Requirements for competence and quality that apply to medical laboratories will be set by bodies such as QSR or the international standard ISO 15189. Laboratories that perform tests should ideally be accredited to applicable standards, such as specified by these bodies

Explanatory text

Recommendations for the evaluation of new screening tests have, so far, focused on diagnostic performance—specifically how well they meet the clinical goals of a screening programme. But it is important that new screening tests need to meet the analytical performance characteristics required for their widespread use.⁶⁰ When an approved clinical laboratory performs a test, it is essential that it provides accurate, reliable and reproducible results under a range of conditions applicable to sampling, handling, transport, measurement and reporting of results.⁶¹ Test failures

must be reported. This applies regardless of whether the specimen is faeces, blood or another biological origin.

It is important that researchers undertaking development and evaluation of a new test are aware of the required standards including those recommended in their region (see online supplemental material 12.1 for specific sources). Researchers should follow the detailed evaluation protocols for test development and undertake the verification processes required in their regulatory and practice context to ensure that standards are attained or surpassed. It is likely that during the early phases of evaluation, the final test format will not have been defined, and measurement procedures and criteria for positivity may not be determined and await completion of phase II and/or III studies. Nonetheless, it is essential that the highest analytical standards are met,⁶⁰ especially when proceeding to phase III evaluations. In the USA, for example, the Food and Drug Administration requires that analytical validation be completed before screening validation, such as would occur in phase III or IV studies.

DISCUSSION

There is no simple universal approach to how evaluation of a new test should be conducted as screening programmes vary significantly around the world and the programme contexts of PBOs or SOS have differing priorities. Some programmes seek to ensure that colonoscopy workloads are feasible in constrained systems and will be satisfied with efficient detection of curable CRC, others will seek to maximise detection rates of CRC and precursor lesions to reduce incidence as well as CRC mortality and yet others will aim for maximising population participation with whatever test is guideline-approved or policy-approved.

Thus, it is desirable that the positivity threshold of a new CRC screening test can be adjusted to facilitate matching a programme’s goals. If regulatory bodies that approve the marketing of tests require that specific claims be made for test performance, then a range of performance variables matched to a range of positivity thresholds could be provided for a threshold-adjustable test. It is recognised, however, that this might provide challenges for proprietary algorithms that generate a test threshold from a marker panel. Doing so in the intended-use population will also provide challenges for screening providers. Pilot studies will be needed to determine if a preliminary positivity threshold remains suitable for the programme’s goals

Over the past two decades, there has been an increasing desire to reduce CRC mortality and its incidence. This requires detection of advanced precursor lesions, and a clearer understanding of which precursor lesions are most important. The consensus process agreed that precursor lesions are a legitimate goal of screening. But it noted that efforts to improve precursor lesion detection will very likely compromise specificity and increase colonoscopy workloads, so screening programme goals should consider what workload is feasible. As yet, there is no agreement on a minimum detection standard for precursor lesions, and more research is needed to better understand the natural history of precursor lesions, particularly differences in their progressive potential.

Current RCT evidence regarding CRC incidence reduction based on non-invasive screening tests is dependent on detection of the bleeding phenotype, notwithstanding serendipitous detection of precursor lesions with lower specificity rehydrated Haemoccult.⁶² On the other hand, the sensitivity of FIT and the mtsDNA test—both technologies that meet performance standard for detection of colorectal neoplasia in the USA⁹—seem to be less than satisfactory for the detection of serrated precursors and could be improved for

adenomas. New biomarkers and technologies, that is, alternatives to those identifying the bleeding phenotype, bear the potential to improve precursor lesion detection. It is likely that biomarker panels, and the algorithms that target the best combination of such markers, will be needed.

The revised and expanded principles presented here address the evaluation of new non-invasive CRC screening tests, and detail an efficient feasible and rigorous strategy for gathering the evidence that would justify a test's use in screening programmes, without the need to undertake RCTs with mortality as the main outcome measure. They provide options for how to consider and address the practicalities and demands of the PBOS or SOS contexts and so accommodate the considerable differences, approaches, policies and regulatory requirements that apply to screening programmes worldwide. Yet, it is a demanding strategy as many jurisdictions require more than evaluation of test accuracy as they see screening as a pathway with a wide range of measured outcomes (principle 6). This is because the screening test is offered to healthy people, where its acceptability, feasibility within healthcare resources and cost-effectiveness must be demonstrated.^{7 12}

What is desired of a new test? It can be seen from the discussion of these principles that a new non-invasive test should achieve at least some of the following, when compared with an established test that has been demonstrated to have a positive impact on CRC-specific mortality:

- ▶ Be flexible, thus enabling providers to achieve the desired goals of a screening programme according to the demands of the healthcare environment.
- ▶ Improve sensitivity for relevant neoplasia (curable CRC and advanced precursor lesions) while maintaining acceptable specificity. Fewer false negatives will enable increasing the interval between tests.
- ▶ Improve precursor lesion detection and hence reduce CRC incidence.
- ▶ Improve participation rates over initial and subsequent rounds.

Thus, the phased approach to test evaluation is essential. Promising initial results will justify more complex and costly evaluation in intended-use screening populations. The degree to which evaluation proceeds will depend to a large extent on policymakers' perspectives and regulatory processes that apply to PBOS and SOS contexts. Test developers will initially be focused on the results of phases I and II studies. In contrast, phase III data as a minimum will be needed for regulatory and marketing approval applications and to enable health policymakers, screening programme managers/funders and those developing professional guidelines to decide on its usefulness. By recognising that screening is a complex process that starts with the invitation and proceeds through the delivery of the screening test, the diagnostic follow-up and treatment of identified lesions, with further screening and surveillance, when indicated, this revision of the recommendations shifts the adequacy of the evidence in support of a new test from phase II to phase III and phase IV and it provides more detail on the expected outcomes assessed in these phases.

A quantitative FIT is currently used by most two-step organised screening programmes around the world.^{3 4} The consensus process considered FIT to be the ideal comparator for a non-invasive test as its accuracy is preferable to gFOBt (the previous standard). It sets the standard to which a new non-invasive test should be compared.

It will be important to be confident that a new test can reduce CRC mortality and incidence. As conducting trials with mortality reduction as the end point, modelling these outcomes is possible provided that the parameters are ascertained in intended-use populations (phase III and phase IV studies). Measurement of missed cancer rates and modelling of cost-effectiveness will also be necessary as

well as test failure rates. Most if not all PBOS contexts do not see screening as a one-time event, given the importance of subsequent screening rounds in the detection of missed and newly developed lesions during the two to three decades of life during which population screening is considered appropriate. While phase IV studies might not be needed for regulatory approval, they will influence acceptance by healthcare jurisdictions.

This consensus process has reaffirmed the view that comparative evaluation of a new with a proven non-invasive comparator test (ideally an appropriately configured quantitative FIT) with well-established benefits for incidence and mortality reduction is a powerful approach for new test evaluation. The choice of the most appropriate study design for conducting such comparisons depends on screening context and need. In the early phases (I and II), this is achieved in a preliminary fashion by having participants complete both the new and comparator tests, even when only test-positive cases undergo colonoscopy. However, while the assessment of the absolute accuracy of the new test might represent a requirement for market approval in some jurisdictions, the results of early phase testing must undergo rigorous scrutiny by undertaking unbiased designs in intended-use populations (phase III studies), before the new test can be considered for adoption in the context of the screening pathway. The latter studies should be designed in order to provide more accurate estimates of test accuracy, intention to screen outcomes (where required), willingness to undergo follow-up colonoscopy for positive results and information about missed lesions, detection rates at subsequent rounds and adverse events (including test failure). They will also identify if a test's positivity threshold needs to be adjusted.

In conclusion, the revised and expanded principles presented here outline a phased strategy for evaluation of new non-invasive CRC screening tests, irrespective of the type of biological sample studied. They provide an efficient and manageable method of reaching a clear understanding of the potential benefits and liabilities of new screening tests essential to furthering the goal of minimising the disease burden of CRC. The approach provides flexibility, allowing for global variation in screening programmes, their conduct (such as PBOS and SOS contexts) and goals. It also highlights challenges, such as inclusion of adjustable test end points in order to achieve a test configuration best suited for a range of programme goals, uncertainty about how to define those precursor lesions most at risk and the need for modelling parameters that enable prediction of test impact on CRC mortality and incidence. Existing knowledge gaps will ultimately need to be addressed through research aimed at understanding the natural history of precursor lesions, what influences uptake and adherence to individual screening tests, improved risk stratification, performance of screening tests over multiple rounds of programmatic screening and factors impacting incidence and mortality in specific intended-use populations.

It is expected that the information provided will guide researchers, practitioners, regulatory authorities, policy makers and screening programme providers in the development and validation of a new test, as well as a test's suitability for the context of its use.

Author affiliations

¹Department of Gastroenterology, Hepatology, and Nutrition, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

²Epidemiology and screening unit, Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica in Piemonte, Turin, Italy

³Flinders Centre for Innovation in Cancer, Flinders University, Adelaide, South Australia, Australia

⁴Internal Medicine/Division of Gastroenterology, University of California San Francisco (UCSF), San Francisco, California, USA

⁵Gastroenterology & Digestive Oncology Department, Hôpital Avicenne University Paris Nord La Sorbonne, Bobigny, France

⁶Department of Clinical Biochemistry and NHS Bowel Cancer Screening South of England Hub, Royal Surrey County Hospital, Guildford, Surrey, UK

⁷Department of Epidemiology & Data Science, Amsterdam Public Health, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

⁸Carrera de especialista de Endoscopia Digestiva, Institución GEDYT (Gastroenterología diagnóstico y terapéutica), Buenos Aires, Argentina

⁹Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

¹⁰Department of Internal Medicine, College of Medicine, National Taiwan University Hospital, Taipei, Taiwan

¹¹Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, The Netherlands

¹²Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

¹³Department of Gastroenterology and Hepatology C2-310, Amsterdam UMC University of Amsterdam, Amsterdam, The Netherlands

¹⁴Department of Gastroenterology and Hepatology C2-115, Amsterdam University Medical Centres, Duivendrecht, The Netherlands

¹⁵Dept of Gastroenterology, Division of Population Medicine, Cardiff University School of Medicine, Cardiff, UK

¹⁶Centre for Research into Cancer Prevention and Screening, University of Dundee School of Medicine, Dundee, UK

¹⁷Division of Translational Science and Therapeutics, Fred Hutchinson Cancer Center, Seattle, Washington, USA

¹⁸ER13 Cancers & Populations, Normandie University, UNICAEN, Caen, France

¹⁹Department of Medicine, University of California San Diego, La Jolla, CA, USA

²⁰Clinical Biochemistry, University of Surrey, Guildford, UK

²¹Division of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology, Bremen, Germany

²²Department of Research, Telemark Hospital, Skien, Norway

²³Department of CRC screening, Cancer Registry of Norway, Oslo, Norway

²⁴Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

²⁵Gastroenterology and Hepatology, Amsterdam University Medical Centres, Duivendrecht, The Netherlands

²⁶Department of Clinical Research, University of Southern Denmark, Odense, Denmark

²⁷Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Stanford, California, USA

²⁸Section of Evidence Synthesis and Classification, International Agency for Research on Cancer, Lyon, France

²⁹Institute of Clinical and Preventive Medicine, Faculty of Medicine, University of Latvia, Riga, Latvia

³⁰Division of Cancer Prevention and Population Sciences, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

³¹Gastroenterology, Kaiser Permanente Division of Research, Oakland, California, USA

³²Colorectal Medicine and Genetics, The University of Melbourne Department of Medicine Royal Melbourne Hospital, Parkville, Victoria, Australia

³³High-Risk Clinic for Gastrointestinal Cancers, University of Arizona Cancer Center Division of Gastroenterology, Banner University, Tucson, Arizona, USA

³⁴Gastroenterology, Trinity College Dublin Faculty of Health Sciences, Dublin, Ireland

³⁵National Bowel Screening Programme, National Screening Unit, Te Whatu Ora Health New Zealand, Auckland, New Zealand

³⁶Department of Medicine, The University of Auckland, Auckland, New Zealand

³⁷Department of Medicine, University of Toronto Dalla Lana School of Public Health, Toronto, Ontario, Canada

³⁸Department of Medicine, University of North Carolina, Chapel Hill, NC, USA

³⁹Clinica Alemana de Santiago, Universidad del Desarrollo, Santiago de Chile, Chile

⁴⁰Department of Gastroenterology, Aomori Prefectural Central Hospital, Aomori, Japan

⁴¹GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands

⁴²Departments of Medicine and Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

⁴³Department of ambulatory Care, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

⁴⁴Internal Medicine, University of Manitoba Max Rady College of Medicine, Winnipeg, Manitoba, Canada

⁴⁵Surgery and Oncology, University of Dundee, Dundee, UK

⁴⁶Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

⁴⁷Department of Gastroenterology, Flinders Medical Centre, Adelaide, South Australia, Australia

⁴⁸Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Twitter Sunil Dolwani @sdolwani, Tim Kortlever @tim_kortlever, Theodore Robert Levin @tr_levin, Robert E Schoen @Rockscho and Robert J C Steele @BobSteele6

Acknowledgements We are grateful to the members of the industry Advisory Panel (L LaPointe, R Schoengold, H Landicho, R Bruce, G Davis, G Putcha) who

provided comments on the principles (see 'Methods' section for detail) and to the World Endoscopy Organisation secretariat for managing the consensus process, collating all feedback and organising webinars and seminars which addressed the issues arising from feedback.

Contributors GPY, CS and RSB were responsible for conceiving the process, drafting the initial 10 principles, overseeing the consensus process and establishing the original design. All authors then contributed to the process of revising and expanding the principles and reaching consensus (over four rounds) on their wording. RSB, CS and GPY then drafted the explanatory text and circulated this to all authors for comment and revision. All authors contributed to the revision of the text, including choice of references over six revisions of the text. During revision of the text, three workshops were held in which the principles were presented to attendees (available authors plus other members of the WEO Screening Committee) and further debated. Speakers at these workshops were RBS, CS, GPY, PMMB, BC, VMHC, SG, TK, UL, LG, DFR, RS, RJCS and SB. Text was further edited as a result by GPY who also extracted text suitable for the supplemental material. The resultant text was again circulated to all authors and then finalised according to feedback by GPY, RSB and CS. GPY, RSB and CS are collectively responsible for the overall content. GPY takes full responsibility as guarantor of the work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer Dr Lauby-Secretan asserts that she alone is responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer /World Health Organization.

Competing interests Board membership: TRL, RES, LG, FM, CS, RS, H-MC, ED, AK, HS, GAM, SI. Consultancy: LG, UL, GPY, FM, JM, SG, ED, AK, HS, SI. Expert testimony: FM. Grants or contract research: RSB, TRL, RES, FM, RS, FM, ED, ML, GAM, LC. Lectures/Other education events: LG, FM, H-MC, ED, AK. Patents: GPY, RSB, BC, AK, GAM. Receipt of equipment or supplies: LG, RES, ED, ML, GAM. Stock/Stock options: GPY, UL, JM, SG, ED, AK, GAM. Other professional relationships: GPY, SG.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

Author note GPY takes full responsibility as guarantor of the work.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Robert S Bresalier <http://orcid.org/0000-0002-9740-281X>
 Carlo Senore <http://orcid.org/0000-0003-1023-7477>
 Graeme P Young <http://orcid.org/0000-0001-9458-8383>
 Beatriz Carvalho <http://orcid.org/0000-0003-4903-399X>
 Han-Mo Chiu <http://orcid.org/0000-0003-2786-8056>
 Evelien Dekker <http://orcid.org/0000-0002-4363-0745>
 Callum G Fraser <http://orcid.org/0000-0002-1333-7994>
 William Grady <http://orcid.org/0000-0002-0129-7809>
 Tim Kortlever <http://orcid.org/0000-0001-5416-9571>
 Anastasios Koulaouzidis <http://orcid.org/0000-0002-2248-489X>
 Mărcis Leja <http://orcid.org/0000-0002-0319-8855>
 Theodore Robert Levin <http://orcid.org/0000-0002-0842-2377>
 Gerrit A Meijer <http://orcid.org/0000-0003-0330-3130>
 Robert E Schoen <http://orcid.org/0000-0001-7153-2766>
 Harminder Singh <http://orcid.org/0000-0002-9354-2356>
 Robert J C Steele <http://orcid.org/0000-0003-4248-6785>

Joseph J Y Sung <http://orcid.org/0000-0003-3125-5199>
Erin Leigh Symonds <http://orcid.org/0000-0003-2451-0358>

REFERENCES

- Bresalier RS, Grady WM, Markowitz SD, et al. Biomarkers for early detection of colorectal cancer: the early detection research network, a framework for clinical translation. *Cancer Epidemiol Biomarkers Prev* 2020;29:2431–40.
- Young GP, Senore C, Mandel JS, et al. Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. *Cancer* 2016;122:826–39.
- Young GP, Rabeneck L, Winawer SJ. The global paradigm shift in screening for colorectal cancer. *Gastroenterology* 2019;156:843–51.
- International Agency for Research on Cancer. Colorectal cancer screening. *IARC Handb Cancer Prev* 2019;17:1–300.
- Melson JE, Imperiale TF, Itzkowitz SH, et al. AGA white paper: roadmap for the future of colorectal cancer screening in the United States. *Clin Gastroenterol Hepatol* 2020;18:2667–78.
- Lieberman D, Ladabaum U, Brill JV, et al. Reducing the burden of colorectal cancer: AGA position statements. *Gastroenterology* 2022;163:520–6.
- Wilson JM, Jungner YG. Principles and practice of mass screening for disease. *Bol Oficina Sanit Panam* 1968;65:281–393.
- US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for colorectal cancer: US preventive services task force recommendation statement. *JAMA* 2016;315:2564–75.
- Centers for Medicare and Medicaid Services. National coverage determination (NCD) 210.3 - screening for colorectal cancer (CRC): CMS.Gov. 2021. Available: <https://www.cms.gov/files/document/R10818NCD.pdf>
- Fink A, Koseoff J, Chassin M, et al. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–83.
- Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ* 2008;86:317–9.
- Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850–5.
- Lin JS, Perdue LA, Henrikson NB, et al. Screening for colorectal cancer: updated evidence report and systematic review for the US preventive services task force. *JAMA* 2021;325:1978–98.
- Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;64:1637–49.
- Young GP, Macrae FA, St John DJB. Clinical methods of early detection: basis, use and evaluation. In: Young GP, Rozen P, eds. *Prevention and early detection of colorectal cancer*. 5th ed. London: Saunders, 1996: 241–70.
- Canadian Cancer Statistics Advisory Committee. *Canadian cancer statistics 2019*. Toronto, ON: Canadian Cancer Society, 2019. Available: <https://cancer.ca/en/research/cancer-statistics/past-editions>
- Heisser T, Kretschmann J, Hagen B, et al. Prevalence of colorectal neoplasia 10 or more years after a negative screening colonoscopy in 120 000 repeated screening colonoscopies. *JAMA Intern Med* 2023;183:183–90.
- Brethauer M, Loberg M, Wieszczy P, et al. Effect of colonoscopy screening on risks of colorectal cancer and related death. *N Engl J Med* 2022;387:1547–56.
- Elmunzer BJ, Hayward RA, Schoenfeld PS, et al. Effect of flexible sigmoidoscopy-based screening on incidence and mortality of colorectal cancer: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* 2012;9:e1001352.
- Levin TR, Corley DA, Jensen CD, et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. *Gastroenterology* 2018;155:1383–91.
- Zorzi M, Fedeli U, Schievano E, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal Immunochemical test. *Gut* 2015;64:784–90.
- Pinsky PF, Loberg M, Senore C, et al. Number of adenomas removed and colorectal cancers prevented in randomized trials of flexible sigmoidoscopy screening. *Gastroenterology* 2018;155:1059–68.
- Crockett SD, Nagtegaal ID. Terminology, molecular features, epidemiology, and management of serrated colorectal neoplasia. *Gastroenterology* 2019;157:949–66.
- Cross AJ, Robbins EC, Pack K, et al. Colorectal cancer risk following polypectomy in a Multicentre, retrospective, cohort study: an evaluation of the 2020 UK post-polypectomy surveillance guidelines. *Gut* 2021;70:2307–20.
- Brenner H, Hoffmeister M, Stegmaier C, et al. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut* 2007;56:1585–9.
- Carvalho B, Diosdado B, Terhaar Sive Droste JS, et al. Evaluation of cancer-associated DNA copy number events in colorectal (advanced) adenomas. *Cancer Prev Res (Phila)* 2018;11:403–12.
- Carvalho B, Postma C, Mongera S, et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* 2009;58:79–89.
- Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;371:187–8.
- Juul FE, Cross AJ, Schoen RE, 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.
- Senore C, Basu P, Anttila A, et al. Performance of colorectal cancer screening in the European Union member states: data from the second European screening report. *Gut* 2019;68:1232–44.
- Australian Institute of Health and Welfare. *Key performance indicators for the National Bowel Cancer Screening Program: technical report. Contract No.: Cat. no.CAN 84*. Canberra: AIHW, 2014.
- Moss S, Ancelle-Park R, Brenner H, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First edition—evaluation and interpretation of screening outcomes. *Endoscopy* 2012;44 Suppl 3:SE49–64.
- Moons KGM, de Groot JAH, Linnet K, et al. Quantifying the added value of a diagnostic test or marker. *Clin Chem* 2012;58:1408–17.
- Ladabaum U, Church TR, Feng Z, et al. Counting advanced precancerous lesions as true positives when determining colorectal cancer screening test specificity. *J Natl Cancer Inst* 2022;114:1040–3.
- Clinical Laboratory Standards Institute. *Defining, establishing and verifying reference intervals in the clinical laboratory approved guideline. Contract No.: 30*. third edition. Wayne, PA: Clinical Laboratory Standards Institute, 2008.
- Lane JM, Chow E, Young GP, et al. Interval fecal Immunochemical testing in a colonoscopic surveillance program speeds detection of colorectal neoplasia. *Gastroenterology* 2010;139:1918–26.
- Cole SR, Tucker GR, Osborne JM, et al. Shift to earlier stage at diagnosis as a consequence of the National bowel cancer screening program. *Med J Aust* 2013;198:327–30.
- Owens L, Gulati R, Etzioni R. Stage shift as an endpoint in cancer screening trials: implications for evaluating multicancer early detection tests. *Cancer Epidemiol Biomarkers Prev* 2022;31:1298–304.
- Scholefield JH, Moss SM, Mangham CM, et al. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut* 2012;61:1036–40.
- Young GP, Woodman RJ, Symonds E. Detection of advanced colorectal neoplasia and relative colonoscopy workloads using quantitative faecal Immunochemical tests: an observational study exploring the effects of simultaneous adjustment of both sample number and test positivity threshold. *BMJ Open Gastroenterol* 2020;7:e000517.
- Robertson DJ, Lee JK, Boland CR, et al. Recommendations on fecal Immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US multi-society task force on colorectal cancer. *Gastroenterology* 2017;152:1217–37.
- Bretagne J-F, Carlo A, Piette C, et al. Significant decrease in interval colorectal cancer incidence after implementing immunochemical testing in a multiple-round guaiac-based screening programme. *Br J Cancer* 2021;125:1494–502.
- Wieten E, Schreuders EH, Grobbee EJ, et al. Incidence of faecal occult blood test interval cancers in population-based colorectal cancer screening: a systematic review and meta-analysis. *Gut* 2019;68:873–81.
- Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;332:1089–92.
- Kaminski MF, Robertson DJ, Senore C, et al. Optimizing the quality of colorectal cancer screening worldwide. *Gastroenterology* 2020;158:404–17.
- Young GP, Symonds EL, Allison JE, et al. Advances in fecal occult blood tests: the FIT revolution. *Dig Dis Sci* 2015;60:609–22.
- Young GP, Woodman RJ, Ang FLI, et al. Both sample number and test positivity threshold determine colonoscopy efficiency in detection of colorectal cancer with quantitative fecal immunochemical tests. *Gastroenterology* 2020;159:1561–3.
- Gies A, Cuk K, Schrotz-King P, et al. Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. *Gastroenterology* 2018;154:93–104.
- Dominitz JA, Robertson DJ, Ahnen DJ, et al. Colonoscopy vs. fecal immunochemical test in reducing mortality from colorectal cancer (CONFIRM): rationale for study design. *Am J Gastroenterol* 2017;112:1736–46.
- Castells A, Quintero E. Programmatic screening for colorectal cancer: the COLONPREV study. *Dig Dis Sci* 2015;60:672–80.
- Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001;93:1054–61.
- van Roon AHC, Goede SL, van Ballegoijen M, et al. Random comparison of repeated faecal immunochemical testing at different intervals for population-based colorectal cancer screening. *Gut* 2013;62:409–15.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
- Ding H, Lin J, Xu Z, et al. A global evaluation of the performance indicators of colorectal cancer screening with fecal Immunochemical tests and colonoscopy: a systematic review and meta-analysis. *Cancers (Basel)* 2022;14:1073.
- Meklin J, Syrjänen K, Eskelinen M. Fecal occult blood tests in colorectal cancer screening: systematic review and meta-analysis of traditional and new-generation fecal immunochemical tests. *Anticancer Res* 2020;40:3591–604.
- Daly JM, Xu Y, Levy BT. Which fecal immunochemical test should I choose? *J Prim Care Community Health* 2017;8:264–77.

- 57 Gies A, Niedermaier T, Alwers E, *et al.* Consistent major differences in sex- and age-specific diagnostic performance among nine faecal immunochemical tests used for colorectal cancer screening. *Cancers (Basel)* 2021;13:3574.
- 58 Peng L, Balavarca Y, Niedermaier T, *et al.* Risk-adapted cutoffs in colorectal cancer screening by fecal immunochemical tests. *Am J Gastroenterol* 2020;115:1110–6.
- 59 Piggott C, Carroll MRR, John C, *et al.* Analytical evaluation of four faecal immunochemistry tests for haemoglobin. *Clin Chem Lab Med* 2020;59:173–8.
- 60 Sandberg S, Fraser CG, Horvath AR, *et al.* Defining analytical performance specifications: consensus statement from the 1ST strategic conference of the European Federation of clinical chemistry and laboratory medicine. *Clin Chem Lab Med* 2015;53:833–5.
- 61 Horvath AR, Lord SJ, StJohn A, *et al.* From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta* 2014;427:49–57.
- 62 Ransohoff DF, Lang CA. Small adenomas detected during fecal occult blood test screening for colorectal cancer. The impact of serendipity. *JAMA* 1990;264:76–8.