

Supplemental material

This supplemental material consists of further explanatory text that complements the article.

“An efficient strategy for evaluating new non-invasive screening tests for colorectal cancer: The guiding principles.”

The numbering system relates to the relevant Principle and the item being addressed in that principle (as pointed to in the main article).

4. Nature of precursor lesions most important to detect

4.1 Relevant precursor lesions

The concept that not all conventional adenomas are of equal relevance has existed for some decades(1) and, coupled with the early observations that adenomas of simple tubular structure and small size (<10 mm) identify a low risk of metachronous CRC(2, 3), the concept of the “advanced adenoma” to identify those adenomas with the greatest risk of progression to CRC if left untreated, has become accepted as a target for screening when the goal is to reduce CRC incidence. The United States Multi-Society Task Force (USMSTF) defines an advanced adenoma as one with “size ≥ 10 mm, with tubulovillous or villous histology, or with high-grade dysplasia in the absence of invasive CRC” (4). Recent studies have shown that the highest risk (for metachronous neoplasia) was associated with high-grade dysplasia or size ≥ 20 mm (5, 6).

More recently, the alternative pathway to CRC via the serrated pathway has been established (7). So-called serrated polyps (SP) include hyperplastic polyps (HP), sessile serrated lesions (SSL) and traditional serrated adenomas or polyps (TSP) but not all serrated lesions pose an obvious risk for CRC. Those SP considered to identify an increased risk are: HP and SP ≥ 10 mm, any SP with dysplasia and any TSP (all collectively referred to here as advanced serrated lesions¹, ASL)(8).

4.2 Non-invasive test sensitivity for precursor lesions

High-quality FIT can detect between 25-42% of conventional advanced adenomas(9) given a single test, although sensitivity compounds to more than 60% with repeated annual testing(10) and the cumulative adenoma yield of five consecutive biennial FIT screening rounds was shown to be equal to the yield of a single colonoscopy(11). Repeated testing at appropriate intervals plays an essential role in detecting precursor lesions. Sensitivity increases (i.e., compounds) with the increasing number of screening rounds(10). This is likely to be due to the intermittency of bleeding and because dwell times before clinical presentation and diagnosis for many of these lesions exceed the intervals between tests. FIT technology is not as sensitive for ASL (12, 13). Detection of advanced precursor lesions is significantly improved, although at the cost of a lower specificity, using the non-invasive multitarget stool test (mtsDNA), which tests for fecal hemoglobin (i.e. incorporates FIT) and neoplasia-derived DNA(12).

6. Outcomes to be estimated in a screening population

¹ The terminologies are not finally settled and some authorities refer to these as serrated polyps even though they are not necessarily polypoid in shape.

6.1 Acceptability

Several demographic, ethnic, cultural, social, health access, and economic variables influence participation in screening. For example, those of lower socio-economic status are less likely to seek screening or participate than the more affluent(14). Thus, there might be an interaction between the test type and the variables that affect participation. It is important to not only determine the overall test participation rate but also the rates in socio-economic, underprivileged and ethnic subgroups underserved populations – to determine if any inequities gap is widening. If Phase 3 studies require that all participants undergo colonoscopy, it will only be possible to undertake a qualitative evaluation of non-invasive test preference. Such will fail to fully assess ease of use and logistics of using the test, since the study's requirement that all undergo colonoscopy will compromise the initial simplicity of the non-invasive test as the design will exclude individuals not willing to undergo primary screening colonoscopy.

6.2 Program goal considerations

As discussed in Principles 4 and 5, because the ability to separate inconsequential from consequential lesions remains limited, we need to better understand the natural history defining those precursor lesions with the highest risk of progression. Detection of advanced precursor lesions is therefore an imperfect surrogate outcome for the impact of screening on CRC incidence, as even advanced precursor lesions are seen in many more people than those who eventually develop CRC(12). It should be noted that an accurate test that differentiates well between advanced and non-advanced precursors, would also serve to identify those in whom colonoscopic surveillance is most efficiently deployed. Ensuring a high quality of colonoscopy in the screening setting is important to minimize overtreatment and reduce the risk associated with (the necessary) polypectomy.(15).

6.3 Modelling outcomes

Early-stage modelling based on proxies/surrogate measures for long-term efficacy (mortality and/or incidence) might be an option after Phase 1 and 2 when it has to be decided whether, or not, to proceed to evaluation in a screening population. Such proxies would include lesion detection rates, number needed to colonoscope to detect one individual with either CRC or advanced precursor, and cost to detect one CRC or advanced precursor. Relative cost-effectiveness can be inferred from specificity for the neoplastic outcomes of importance (CRC and advanced precursor lesions). However, as Phase 2 studies will involve enriched populations with consequent biases (see Principle 10), accurate estimates of sensitivity and specificity will not be evident until Phase 3 studies are performed.

It remains to be determined whether these or other potential early-stage quantitative proxies/surrogates are better than a holistic judgment (i.e. evaluation in unbiased intended-use populations) regarding the potential merits of emerging screening technologies, especially as there is considerable risk of bias in Phase 1 and 2 studies. Rigorous exploratory analyses in established models can be pursued and are relatively straightforward to undertake using Phase 3 findings, although they can be time- and labor-intensive. Whether novel proxies or rigorous exploratory modelling are applied using Phase 2 data, the results should not be over-interpreted. Most importantly, such initial explorations (in Phases 1 and 2) should not lead to the abandonment of a test.

7. Expectations of a new non-invasive test.

7.1 The population benefit of screening

gFOBT initially set the standard expected of a new non-invasive test in mass population screening(16). The gFOBT Hemocult, offered biennially, was shown to reduce CRC mortality by 15-20% on an intention-to-screen (ITS) basis (17) (18) (19) (20) (21) and by >40% on a per-protocol test-completion basis(20). ITS mortality benefit increased to 33% with rehydration of Hemocult offered annually(22), although specificity fell considerably. Screening with rehydrated Hemocult was subsequently shown to lead to a 20% ITS reduction in CRC incidence after 18 years of follow-up (19) due to the detection and removal of precursor lesions at follow-up colonoscopy.

7.2 Summary of evidence supporting the use of FIT

CRC mortality and incidence benefits have been observed for FIT in case-control and cohort studies, as well as population-based implementation evaluations(15). A series of case-control studies from Japan showed a reduction in mortality in those who screened with FIT compared to those who were not screened (23). More recently, a large prospective cohort study reported a reduction in mortality comparing screened with unscreened individuals (24). FIT use has also resulted in a shift to an earlier stage at diagnosis (25). Other recent population screening programs using a FIT have confirmed this benefit of CRC incidence reduction in ecological studies in Italy (26, 27) and the USA(28). Several of these published studies have adjusted for lead time, length time and selection biases. A Dutch RCT showed FIT had a higher CRC detection rate than gFOBT(29). While well-designed experimental/formal RCTs are important to demonstrate efficacy, performance in real-world settings as cited above, is best shown in pragmatic comparative effectiveness RCTs in screening practice, supported by observational studies to provide real world evidence, even though adjustment for potential biases requires careful design considerations.

7.3 Efficacy of screening by flexible sigmoidoscopy

Four RCTs evaluating the benefit of flexible sigmoidoscopy (FS) screening consistently demonstrated that endoscopic excision of colorectal adenomas is associated with a substantial and long-lasting (over 15 – 18 years) reduction in CRC incidence (18%-23%) and mortality (26%-31%) on an ITS basis(30-34). Per-protocol analysis showed, among participants, a 33%-35% and a 39%-41% reduction in CRC incidence and mortality respectively (30, 33). Furthermore, even though the observed protective effect of FS screening was mainly limited to the distal colon, with a slight reduction in proximal CRC incidence observed only at 12-year follow-up in some subgroups of the target population (men and women younger than 60 years)(35), such studies clearly show that removal of precursor lesions reduces CRC incidence.

7.4 Evidence for benefit of screening colonoscopy

A substantial body of observational data has suggested that colonoscopy screening can reduce CRC incidence and mortality(15). Furthermore, since FS screening reduces CRC mortality and incidence, one would expect a similar, or even larger, benefit from colonoscopy since colonoscopy examines more of the colon. The recently reported findings of the 10-year follow-up of the first RCT of colonoscopy vs no screening (36) provide additional evidence that participation in colonoscopy screening reduces CRC mortality and incidence. The trial found no benefit to population CRC mortality on an ITS basis (screening participation rate of 42%), despite a greater detection of CRC and a significant 18% incidence reduction in the intervention group. However, there was a 50% reduction in CRC mortality and a 31% reduction in CRC incidence was observed in the per protocol analysis (i.e.

those who underwent screening colonoscopy). It should be noted that such benefit is no greater than that achieved by flexible sigmoidoscopy (see above).

8. An adjustable test positivity threshold accommodates different program goals.

8.1 Challenges with endpoint flexibility

Regulatory processes (i.e., those required by government agencies for obtaining approval to market and use a test) may require the stipulation of a single test positivity threshold with corresponding sensitivity, specificity, and “general test utility”² for detecting neoplasia in a screening population. Most jurisdictions require that health claims align with the chosen positivity threshold. But there is a risk that a test approved for use but configured for a specific threshold might have diagnostic accuracy and associated colonoscopy workload that does not align with program goals.

Manufacturers, while needing to comply with local regulatory approval guidelines are encouraged to provide for adjustment of the positivity threshold. This may be achievable by seeking regulatory approval for test performance across the measurement range (i.e., the range of thresholds that meet the analytical quality specifications) of the assay.

This is desirable not only for tests based on a single marker, but also for algorithms that generate a result from multiple factors or markers. Most existing algorithms are fixed, but it is possible that different combinations of weightings of algorithm components could be developed to give some capacity to choose a diagnostic accuracy that best suits a program. Providing transparency in the algorithm would allow programs to make adjustments if desired.

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

9.1 Approach to comparing test accuracy

The approach to undertaking an initial direct head-to-head comparison of a new test with a proven comparator (see Fig. 2(16)), is based on the proposition that the two critical questions concerning diagnostic accuracy are (37, 38):

- 1) *detection* (sensitivity) – a test that is more sensitive in practical terms returns more true-positive results, and
- 2) *the burden associated with detection* (1 minus specificity) – a more specific test, in practical terms, returns fewer false-positive results.

There are several ways to address these two questions, with varying degrees of complexity,

The simplest is when participants perform both tests and every test-positive case (by either or both tests) undergoes colonoscopic diagnostic verification. This distinguishes true- from false-positive results (38–40) as shown in the upper section of Table 2. It identifies *relative* sensitivity and *relative* false-positive rates. The obvious limitation is that it fails to identify the nature of missed lesions because not all cases are colonoscoped.

² This vague and ill-defined concept of test utility is used by some regulatory authorities to refer to the need to understand the implications of false-positive results.

Consequently, to get estimates of absolute sensitivity and specificity, it is necessary that all cases undergo colonoscopy, using either the single-cohort paired testing or the parallel cohort design (see Figure 2 and lower section of Table 2 in main manuscript). While estimates of diagnostic accuracy are verified by colonoscopy in all cases, undertaking these early evaluations remain subject to a range of biases that mean that these initial accuracy estimates might not hold true in an unbiased screening population (see Principle 9).

10. Evaluation proceeds through increasingly complex phases.

10.1 Design considerations for Phases 1 and 2:

The population size for Phase 1 studies is generally 100-200, where approximately 50 should have CRC(16). They should include cases with all relevant stages of colorectal neoplasia (with adequate numbers of those with earlier stage CRC in whom treatment is more likely to be beneficial). Phase 2 studies involve a larger number of individuals, at least several thousand, who will ideally undergo paired (or parallel) testing (new and comparator test) – see Fig. 2 in main article. It is helpful to include the spectrum of non-neoplastic diseases (e.g., inflammatory bowel conditions, diverticular disease, non-neoplastic polyps and non-colorectal neoplasms), especially when evaluating blood tests, since this will expose confounding variables that may influence test results independent of the presence of *colorectal* neoplasia.

In Phase 1 and Phase 2 studies, there is risk of selection bias if enriched for cases with neoplasia. For example, it has been reported that sensitivity for cancer can be higher than that which will subsequently become apparent in a typical screening population(41). When enriching populations, sources of selection bias include factors such as age range not being typical for screening, symptomatic cases not being excluded, cases being at high risk for neoplasia, having returned a recent positive screening test and the timing of recruitment (e.g., tests done after colonoscopy and possibly after biopsy). Sourcing cases from screening programs or using a nested case-control design will reduce the risk of bias, but may reduce ability to determine how other diseases relate to test positivity.

Comparative testing in a paired manner reduces the population size because of improved statistical power for detecting differences. Paired testing can avoid imbalances in non-neoplastic variables that might affect test results when comparing tests. Furthermore, it circumvents selection and verification bias which can be introduced in studies on absolute test accuracy where all subjects undergo colonoscopy, if, for example, adherence to diagnostic verification is higher in test positives than in test negatives when studying parallel cohorts(42).

10.2 Design considerations for Phase 3:

The **first study design** to consider is when a jurisdiction requires absolute estimates of sensitivity and specificity (common in the SOS context). It requires that all participants must perform the non-invasive test and undergo screening colonoscopy regardless of test result.

Here, all participants must agree to screening colonoscopy and undergo at least the new non-invasive test. The new test could be compared with a fecal immunochemical test (FIT) test either in a single cohort undertaking paired testing or else in two cohorts randomized to the new or comparator non-invasive test. However, such studies are subject to limited generalizability since they require an individual to be willing to undergo a screening colonoscopy as well as perform the non-invasive test(s) and assume that colonoscopy is 100% accurate. The comparison to FIT will make

the recruitment process even more complex and the possibility of verification bias has to be considered(43). Subjects might be more or less willing to continue to colonoscopy depending on what they are informed about the new, yet-to-be established test. Furthermore, the population size will need to be in excess of 10,000 to guarantee good estimates of sensitivity for cancer, especially stage I and II cancers(12, 41) (14, 44). While this is costly, the frequency of neoplastic lesions in the population will be apparent, accuracy estimates will be precise, population test positivity will be clear, test performance in all types of precursor lesions will be obtained, differences in the distribution of stage of neoplastic lesions will be apparent for each test, missed lesions will be readily identified, numbers needed to screen to detect a lesion will be apparent. The implications for sensitivity and specificity according to fine adjustment of the test threshold for positivity will be readily described. However, population participation rates cannot be estimated with such a study design.

Researchers need to be aware that the provisional positivity threshold set after Phase 2 studies might not, in Phase 3 studies, deliver the outcomes desired of the screening program. Major risks include inadequate sensitivity, especially for precursor lesions, unacceptably low specificity, or a test positivity rate (and follow-up colonoscopy workload) that is not practical. Pilot studies using a lower positivity threshold may be warranted to enable exploration of performance over a range of thresholds.

Consequently, likely CRC and precursor lesion detection rates according to the willingness to participate (i.e., an ITS basis) will be unobtainable. This will limit modelling the benefit of the new test for mortality and incidence. However, much smaller studies with actual participation (not just the intention to screen) as the sole endpoint can be undertaken(16, 45), provided that other study designs are assessing test performance.

The **second study design** to consider is when a jurisdiction requires intention-to-screen (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. Relativity between the new and comparator non-invasive tests may be the major issue especially assessed by ITS outcomes.

The first is an efficient comparative design where a single cohort performs both tests (better power for comparing accuracy, but relative participation rates are unobtainable) although there may be selection bias given people are required to do both tests. As long as all those positive by either test undergo follow-up colonoscopy for diagnostic assessment, relative detection rates (intention-to-screen and/or participant basis) and associated variables are obtainable from test results as detailed in Principle 4. The relative frequency of neoplastic lesions in the population on an intention-to-screen basis will be apparent. Regarding cancer and advanced precursor detection, the power to detect differences between tests or to demonstrate non-inferiority is obtainable with smaller study groups when paired testing is undertaken, as compared to parallel cohorts. Weaknesses include failure to obtain unbiased absolute estimates of test accuracy and to immediately identify the nature of missed invasive lesions, although this can be assessed if all are followed up for a time duration corresponding at least to the screening interval(46). Follow-up by linkage to population-based cancer registries for all subjects examined can also provide information on test accuracy for CRC without requiring colonoscopy for all participants.

An alternative design is for two separate cohorts to be randomized to the new or comparator test (and relative participation and detection rates are determined). This specifically addresses whether a new non-invasive test can replace the comparator test in use in that jurisdiction.

It is also essential to ascertain test technical failure rates of a new test, whether they relate to sample collection, handling and transport, or processing within the laboratory, although it would be hoped that such problems would have become apparent in Phases 1 or 2.

For either instance, screening program outcomes detailed in Principle 11, can be used to determine whether the new test is worthy of application in the applicable screening context.

10.3 Design considerations for Phase 4:

Comparative evaluation is still relevant in Phase 4, particularly where the comparator is the usual screening strategy applicable in that environment. Comparative effectiveness RCTs are ideal for comparing the new to established tests when carried out in a real screening setting(47, 48), whether of single or multiple rounds of screening. Modelling studies mimicking such RCTs and based on high-quality observational data from Phases 3 and 4 can also be informative. The outcomes to be monitored are provided under Principle 11.

New tests might not require the same testing frequency as established tests. The re-test interval can be explored in parallel cohorts and willingness to participate again with the new test. If tests with different re-test intervals are being compared, then cumulative detection rates over time, and associated burden of detection, can be determined. However, methods for reporting participation over multiple screening rounds have yet to be well applied to CRC screening(49).

12. Analytical specifications, standards, and performance

12.1 Sources of Laboratory Standards

The Clinical and Laboratory Standards Institute (CLSI) has developed guidelines documenting a test's required analytical performance characteristics (50, 51). These are global, but, in addition, local guidelines and requirements exist, such as the Quality System Requirements (QSR) of the USA. These specify how the evaluation of analytical performance characteristics should be performed (applying the documented techniques) and the information that test manufacturers should provide in their "Instructions for Use".

The International Organization for Standardization (ISO) has prepared and published a set of standards in ISO 15189 that guide researchers (ISO 15189:2022)(52). These guidelines specify the requirements for competence and quality that are particular to medical laboratories. ISO 15189 also defines all aspects of the performance of a test. Ideally, the sites at which tests are performed should be accredited to ISO 15189 standards. Should ISO standards not be mandated, such as in the USA, then appropriate regional regulatory alternatives must be applied.

References

1. Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am*. 2002;12(1):1-9, v. Epub 2002/03/28. doi: 10.1016/s1052-5157(03)00053-9. PubMed PMID: 11916153.
2. Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med*. 1992;326(10):658-62. Epub 1992/03/05. doi: 10.1056/NEJM199203053261002. PubMed PMID: 1736104.
3. Winawer SJ, Zauber AG, O'Brien MJ, Ho MN, Gottlieb L, Sternberg SS, et al. Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. The National Polyp Study Workgroup. *N Engl J Med*. 1993;328(13):901-6. Epub 1993/04/01. doi: 10.1056/NEJM199304013281301. PubMed PMID: 8446136.
4. Gupta S, Lieberman D, Anderson JC, Burke CA, Dominitz JA, Kaltenbach T, et al. Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2020;158(4):1131-53 e5. Epub 2020/02/12. doi: 10.1053/j.gastro.2019.10.026. PubMed PMID: 32044092; PubMed Central PMCID: PMC7672705.
5. Cross AJ, Robbins EC, Pack K, Stenson I, Patel B, Rutter MD, 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(12):2307-20. Epub 2021/03/07. doi: 10.1136/gutjnl-2020-323411. PubMed PMID: 33674342; PubMed Central PMCID: PMC8588296.
6. Wieszczyni P, Kaminski MF, Franczyk R, Loberg M, Kobiela J, Rupinska M, et al. Colorectal Cancer Incidence and Mortality After Removal of Adenomas During Screening Colonoscopies. *Gastroenterology*. 2020;158(4):875-83 e5. Epub 2019/09/30. doi: 10.1053/j.gastro.2019.09.011. PubMed PMID: 31563625.
7. Crockett SD, Nagtegaal ID. Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. *Gastroenterology*. 2019;157(4):949-66 e4. Epub 2019/07/20. doi: 10.1053/j.gastro.2019.06.041. PubMed PMID: 31323292.
8. Trivedi M, Godil S, Demb J, Earles A, Bustamante R, Patterson OV, et al. Baseline Characteristics and Longitudinal Outcomes of Traditional Serrated Adenomas: A Cohort Study. *Clin Gastroenterol Hepatol*. 2022. Epub 2022/10/16. doi: 10.1016/j.cgh.2022.09.030. PubMed PMID: 36243354.
9. Shapiro JA, Bobo JK, Church TR, Rex DK, Chovnick G, Thompson TD, et al. A Comparison of Fecal Immunochemical and High-Sensitivity Guaiac Tests for Colorectal Cancer Screening. *Am J Gastroenterol*. 2017;112(11):1728-35. Epub 2017/10/11. doi: 10.1038/ajg.2017.285. PubMed PMID: 29016558; PubMed Central PMCID: PMC6077997.
10. Lane JM, Chow E, Young GP, Good N, Smith A, Bull J, et al. Interval fecal immunochemical testing in a colonoscopic surveillance program speeds detection of colorectal neoplasia. *Gastroenterology*. 2010;139(6):1918-26. Epub 2010/08/21. doi: 10.1053/j.gastro.2010.08.005. PubMed PMID: 20723544.
11. Zorzi M, Hassan C, Capodaglio G, Narne E, Turrin A, Baracco M, et al. Divergent Long-Term Detection Rates of Proximal and Distal Advanced Neoplasia in Fecal Immunochemical Test Screening Programs: A Retrospective Cohort Study. *Ann Intern Med*. 2018;169(9):602-9. Epub 2018/10/05. doi: 10.7326/M18-0855. PubMed PMID: 30285055.
12. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370:1287-97.
13. Cock C, Anwar S, Byrne SE, Meng R, Pedersen S, Fraser RJL, et al. Low Sensitivity of Fecal Immunochemical Tests and Blood-Based Markers of DNA Hypermethylation for Detection of Sessile

- Serrated Adenomas/Polyps. *Digestive diseases and sciences*. 2019;64(9):2555-62. Epub 2019/03/06. doi: 10.1007/s10620-019-05569-8. PubMed PMID: 30835026.
14. Steele RJ, Kostourou I, McClements P, Watling C, Libby G, Weller D, et al. Effect of gender, age and deprivation on key performance indicators in a FOBT-based colorectal screening programme. *J Med Screen*. 2010;17(2):68-74. Epub 2010/07/28. doi: 10.1258/jms.2010.009120. PubMed PMID: 20660434.
 15. International Agency for Research on Cancer. Colorectal cancer screening. *IARC Handb Cancer Prev*. 2019;17:1-300.
 16. Young GP, Senore C, Mandel JS, Allison JE, Atkin WS, Benamouzig R, et al. Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. *Cancer*. 2016;122(6):826-39. Epub 2016/02/02. doi: 10.1002/cncr.29865. PubMed PMID: 26828588; PubMed Central PMCID: PMC5066737.
 17. Faivre J, Dancourt V, Lejeune C, Tazi MA, Lamour J, Gerard D, et al. Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. *Gastroenterology*. 2004;126(7):1674-80. Epub 2004/06/10. doi: 10.1053/j.gastro.2004.02.018. PubMed PMID: 15188160.
 18. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet*. 1996;348(9040):1472-7. doi: 10.1016/S0140-6736(96)03386-7. PubMed PMID: 8942775.
 19. Mandel JS, Church TR, Bond JH, Ederer F, Geisser MS, Mongin SJ, et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med*. 2000;343(22):1603-7. doi: 10.1056/NEJM200011303432203. PubMed PMID: 11096167.
 20. Kronborg O, Fenger C, Olsen J, Jorgenson OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet*. 1996;348:1467-71.
 21. Mandel JS, Church TR, Ederer F, Bond JH. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst*. 1999;91(5):434-7. PubMed PMID: 10070942.
 22. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med*. 1993;328(19):1365-71. doi: 10.1056/NEJM199305133281901. PubMed PMID: 8474513.
 23. Nakajima M, Saito H, Soma Y, Sobue T, Tanaka M, Munakata A. Prevention of advanced colorectal cancer by screening using the immunochemical faecal occult blood test: a case-control study. *Br J Cancer*. 2003;89(1):23-8. Epub 2003/07/03. doi: 10.1038/sj.bjc.6601002. PubMed PMID: 12838295; PubMed Central PMCID: PMC5066737.
 24. Chiu HM, Chen SL, Yen AM, Chiu SY, Fann JC, Lee YC, et al. Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the One Million Taiwanese Screening Program. *Cancer*. 2015;121(18):3221-9. Epub 2015/05/23. doi: 10.1002/cncr.29462. PubMed PMID: 25995082; PubMed Central PMCID: PMC5066737.
 25. Cole SR, Tucker GR, Osborne JM, Byrne SE, Bampton PA, Fraser RJ, et al. Shift to earlier stage at diagnosis as a consequence of the National Bowel Cancer Screening Program. *Med J Aust*. 2013;198(6):327-30. Epub 2013/04/03. doi: 10.5694/mja12.11357 [pii]. PubMed PMID: 23545032.
 26. Ventura L, Mantellini P, Grazzini G, Castiglione G, Buzzoni C, Rubeca T, et al. The impact of immunochemical faecal occult blood testing on colorectal cancer incidence. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*. 2014;46(1):82-6. Epub 2013/09/10. doi: 10.1016/j.dld.2013.07.017. PubMed PMID: 24011791.
 27. Zorzi M, Fedeli U, Schievano E, Bovo E, Guzzinati S, Baracco S, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. *Gut*. 2015;64(5):784-90. Epub 2014/09/03. doi: 10.1136/gutjnl-2014-307508. PubMed PMID: 25179811.

28. Levin TR, Corley DA, Jensen CD, Schottinger JE, Quinn VP, Zauber AG, et al. Effects of Organized Colorectal Cancer Screening on Cancer Incidence and Mortality in a Large Community-Based Population. *Gastroenterology*. 2018;155(5):1383-91 e5. Epub 2018/07/23. doi: 10.1053/j.gastro.2018.07.017. PubMed PMID: 30031768; PubMed Central PMCID: PMC6240353.
29. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, van Dekken H, Reijerink JC, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. *Gut*. 2010;59(1):62-8. Epub 2009/08/13. doi: 10.1136/gut.2009.177089. PubMed PMID: 19671542.
30. Atkin W, Wooldrage K, Parkin DM, Kralj-Hans I, MacRae E, Shah U, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. *Lancet*. 2017;389(10076):1299-311. Epub 2017/02/27. doi: 10.1016/S0140-6736(17)30396-3. PubMed PMID: 28236467; PubMed Central PMCID: PMC6168937.
31. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Long-Term Effectiveness of Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality in Women and Men: A Randomized Trial. *Ann Intern Med*. 2018;168(11):775-82. Epub 2018/05/02. doi: 10.7326/M17-1441. PubMed PMID: 29710125; PubMed Central PMCID: PMC6853067.
32. Miller EA, Pinsky PF, Schoen RE, Prorok PC, Church TR. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: long-term follow-up of the randomised US PLCO cancer screening trial. *Lancet Gastroenterol Hepatol*. 2019;4(2):101-10. Epub 2018/12/07. doi: 10.1016/S2468-1253(18)30358-3. PubMed PMID: 30502933; PubMed Central PMCID: PMC6335177.
33. Senore C, Riggi E, Armaroli P, Bonelli L, Sciallero S, Zappa M, et al. Long-Term Follow-up of the Italian Flexible Sigmoidoscopy Screening Trial. *Ann Intern Med*. 2022;175(1):36-45. Epub 2021/11/09. doi: 10.7326/M21-0977. PubMed PMID: 34748376.
34. Juul FE, Cross AJ, Schoen RE, Senore C, Pinsky P, Miller E, et al. 15-Year Benefits of Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality : A Pooled Analysis of Randomized Trials. *Ann Intern Med*. 2022;175(11):1525-33. Epub 2022/10/11. doi: 10.7326/M22-0835. PubMed PMID: 36215714.
35. Holme O, Schoen RE, Senore C, Segnan N, Hoff G, Loberg M, et al. Effectiveness of flexible sigmoidoscopy screening in men and women and different age groups: pooled analysis of randomised trials. *BMJ*. 2017;356:i6673. Epub 2017/01/15. doi: 10.1136/bmj.i6673. PubMed PMID: 28087510;
36. Bretthauer M, Loberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, et al. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. *N Engl J Med*. 2022;387(17):1547-56. Epub 2022/10/11. doi: 10.1056/NEJMoa2208375. PubMed PMID: 36214590.
37. Cheng H, Macaluso M, Waterbor J. Estimation of relative and absolute test accuracy. *Epidemiology*. 1999;10(5):566-8. Epub 1999/09/01. PubMed PMID: 10468435.
38. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93(14):1054-61. PubMed PMID: 11459866.
39. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics*. 2001;2(3):249-60. Epub 2003/08/23. doi: 10.1093/biostatistics/2.3.249. PubMed PMID: 12933537.
40. Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. Example of automated cytometry for early detection of cervical cancer. *Am J Epidemiol*. 1987;125(4):672-8. Epub 1987/04/01. doi: 10.1093/oxfordjournals.aje.a114580. PubMed PMID: 3826045.
41. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*.

2014;63(2):317-25. Epub 2013/02/15. doi: 10.1136/gutjnl-2012-304149. PubMed PMID: 23408352; PubMed Central PMCID: PMC3913123.

42. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*. 1983;39(1):207-15. Epub 1983/03/01. PubMed PMID: 6871349.

43. Harel O, Zhou XH. Multiple imputation for correcting verification bias. *Statistics in medicine*. 2006;25(22):3769-86. Epub 2006/01/26. doi: 10.1002/sim.2494. PubMed PMID: 16435337.

44. Shaikat A, Marsh TL, Crockett SD, Syngal S, Bresalier RS, Brenner DE. Low Prevalence of Screen-Detected Colorectal Cancer in an Average-Risk Population: The New Normal. *Clin Gastroenterol Hepatol*. 2021. Epub 2021/09/22. doi: 10.1016/j.cgh.2021.09.013. PubMed PMID: 34547437; PubMed Central PMCID: PMC8934307.

45. Cole SR, Young GP, Esterman A, Cadd B, Morcom J. A randomised trial of the impact of new faecal haemoglobin test technologies on population participation in screening for colorectal cancer. *J Med Screen*. 2003;10(3):117-22. Epub 2003/10/17. doi: 10.1177/096914130301000304. PubMed PMID: 14561262.

46. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. *N Engl J Med*. 1996;334(3):155-9. Epub 1996/01/18. doi: 10.1056/NEJM199601183340304. PubMed PMID: 8531970.

47. Mushlin AI, Ghomrawi H. Health care reform and the need for comparative-effectiveness research. *N Engl J Med*. 2010;362(3):e6. Epub 2010/01/08. doi: 10.1056/NEJMp0912651. PubMed PMID: 20054035.

48. Passamonti B, Malaspina M, Fraser CG, Tintori B, Cariani A, D'Angelo V, et al. A comparative effectiveness trial of two faecal immunochemical tests for haemoglobin (FIT). Assessment of test performance and adherence in a single round of a population-based screening programme for colorectal cancer. *Gut*. 2018;67(3):485-96. Epub 2016/12/16. doi: 10.1136/gutjnl-2016-312716. PubMed PMID: 27974550.

49. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-64. Epub 2016/03/20. doi: 10.1093/aje/kwv254. PubMed PMID: 26994063; PubMed Central PMCID: PMC4832051.

50. Clinical Laboratory Standards Institute. Defining, establishing and verifying reference intervals in the clinical laboratory; Approved Guideline--third edition. Wayne, PA: Clinical Laboratory Standards Institute, 2008 Contract No.: 30.

51. International Organization for Standardization. Medical laboratories - Requirements for quality and competence. 2012.

52. International Organization for Standardization. Medical laboratories — Requirements for quality and competence. ISO 15189:2022022. p. 62.