Impact of Adenoma Detection on the Benefit of Faecal Testing versus Colonoscopy for Colorectal Cancer

BRIEF TITLE: Impact of Colonoscopy Quality in Faecal Testing for Colorectal Cancer

AUTHORS: Reinier G.S. Meester Ph.D.[1], Chyke A. Doubeni M.D., M.P.H.[2], Ann G. Zauber Ph.D.[3], Marjolein van Ballegooijen M.D., Ph.D.,[1] Douglas A. Corley M.D., Ph.D.[4], and Iris Lansdorp-Vogelaar Ph.D.[1]

AFFILIATIONS: **1** Department of Public Health, Erasmus MC University Medical Center, Rotterdam, Netherlands; **2** Department of Family Medicine and Community Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States; **3** Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, United States; **4** Kaiser Permanente Division of Research, Oakland, CA, United States.

CORRESPONDENCE: Reinier Meester, Department of Public Health, Erasmus MC University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. Email: r.meester@erasmusmc.nl. Tel: +31-6-23 74 79 74. Fax: +31-10-703 84 75.

WORD COUNT: 3377 / 5000

KEY WORDS: Colorectal neoplasms, Early detection, Immunochemical test,

ABBREVIATIONS: ADR = adenoma detection rate; FIT = fecal immunochemical test; PI = probability interval; MISCAN = Microsimulation Screening Analyis.

ARTICLE CATEGORY: Research Article – Cancer Therapy and Prevention

NOVELTY AND IMPACT: This is the first study to have estimated the impact of variation in adenoma detection rates on long-term outcomes of faecal-based screening versus colonoscopy for colorectal cancer. Outcomes of faecal-based screening were inversely associated with adenoma detection, but the association was stronger for colonoscopy screening. With lower adenoma detection rates, annual faecal testing was more effective than ten-yearly colonoscopy. Colonoscopy quality assurance is important for optimising colorectal cancer screening programs regardless of primary screening method.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ijc.30933

ABSTRACT

Colonoscopy quality, as measured by adenoma detection rates, varies widely across providers and is inversely related with patients' post-colonoscopy cancer risk. This has unknown consequences for the benefits of faecal immunochemical testing (FIT) versus primary colonoscopy screening for colorectal cancer. Using an established microsimulation model, we predicted the lifetime colorectal cancer incidence and mortality benefits of annual FIT versus ten-yearly colonoscopy screening at differing ADR levels (quintiles; averages 15.3-38.7%), with colonoscopy performance assumptions estimated from community-based data on physician ADRs and patients' post-colonoscopy risk of cancer. For patients receiving FIT screening with follow-up colonoscopy by physicians from the highest ADR quintile, simulated lifetime cancer incidence and mortality were 28.8 and 5.4 per 1000, respectively, versus 20.6 and 4.4 for primary colonoscopy screening (risk ratios, RR=1.40; 95% probability interval (PI), 1.19-1.71 for incidence, and RR=1.22; 95%PI, 1.02-1.54 for mortality). With every 5% point ADR decrease, lifetime cancer incidence was predicted to increase on average 9.0% for FIT versus 12.3% for colonoscopy, and mortality increased 9.9% versus 13.3%. In ADR quintile 1, simulated mortality was lower for FIT than colonoscopy screening (10.1 versus 11.8; RR=0.85; 95%PI, 0.83-0.90), while incidences were more similar. This suggests that relative cancer incidence and mortality reductions for FIT versus colonoscopy screening may differ by ADR, with fewer predicted deaths with colonoscopy screening in higher ADR settings and fewer deaths with annual FIT screening in lower ADR settings.

INTRODUCTION

Colorectal cancer is a leading cause of cancer deaths that is preventable through screening.[1,2] Colonoscopy is indispensable for colorectal cancer screening, as either a primary screening test or for diagnostic follow-up of positive tests results from other screening methods. Colonoscopy quality, as measured by adenoma detection rate (ADR), or the proportion of a physician's screening exams detecting adenomas, varies widely across providers. ADR has been shown to be inversely related to subsequent cancer incidence and mortality risks among patients undergoing screening colonoscopy.[3,4] In a previous modelling study, we predicted that the observed ADR variation may translate to 50-60% differences in lifetime colorectal cancer outcomes for primary colonoscopy screening; however, few data exist regarding the potential influence of ADR variation on faecal-based CRC screening.[4]

Annual faecal immunochemical testing (FIT) is increasingly used worldwide as either a primary colorectal cancer screening method or as an adjunct to colonoscopy-based screening programs to increase overall population screening rates.[5] FIT and colonoscopy screening strategies each have their advantages and disadvantages. Colonoscopy screening is more sensitive for cancers and adenomas and has a long screening interval. FIT may be more acceptable to patients because of the non-invasive nature, the lack of dietary restrictions, and lower risk of complications.[6] Although FIT screening requires diagnostic colonoscopy follow-up of positive results, the overall effectiveness of FIT-based screening may also be affected less by lower ADR levels than primary colonoscopy screening given FIT primarily detects more advanced lesions.[7-9] However, no data exist to compare the benefits of colonoscopy and FIT screening at different ADR levels. Modelling studies used to inform screening recommendations on comparative benefits of colonoscopy and FIT have assumed constant colonoscopy quality levels up to this point.

The purpose of this study is to use a modified microsimulation model informed by community-based data,[3] to compare the benefits of a program of annual FIT versus colonoscopy every ten years at various ADR levels.

MATERIALS AND METHODS

Microsimulation Screening Analysis

This study used the Microsimulation Screening Analysis-Colon (MISCAN-Colon) model, developed by the Erasmus MC University Medical Center, Rotterdam, Netherlands. The model, its main assumptions, and results for colonoscopy screening have been published.[4] The model has been used to inform the US Preventive Services Task Force recommendations for colorectal cancer screening, as well as national screening programs.[10,11]

A detailed description of the model is included as a *Supplementary Appendix*. In brief, MISCAN-colon simulates an average-risk screening population similar to the United States population in terms of life expectancy and cancer risk to evaluate the effect of screening. Colorectal cancer is modelled as developing through the adenoma-carcinoma pathway.[12] Every simulated individual begins in a health state without colorectal cancer or precursor lesions. Throughout their simulated lifetime, individuals may develop one or more adenoma, which may develop progressively from small size (≤5mm in diameter), to medium size (6-9mm) to large size (≥10mm), and to stage I-IV cancer. Cancer may progress from stage I through IV without symptoms, or may be clinically diagnosed in any of the states. Patients may die from cancer or may die from other causes first. In the model, screening has the potential to alter patient's life histories by detecting cancer in an earlier, more treatable stage, or by detecting and removing precursor adenomas.

The effectiveness of screening follows from the natural history assumptions and a screening test's assumed ability to detect adenomatous lesions. Effects of screening have been shown previously to be concordant with results from randomised controlled trials of faecal-occult blood testing.[13] Results have also been validated against the estimated mortality reduction from the UK flexible sigmoidoscopy screening trial (Supplementary Figure 3).[14,15] Although no experimental data were available to directly validate the modelled effectiveness of FIT and colonoscopy screening, effect sizes were concordant with

observational data (Supplementary Figure 4-5).[16,17] Predicted mortality risks after polypectomy were validated against long-term National Polyp Study observations (Supplementary Figure 6).[18]

Test performance assumptions

In this study, assumed variation in colonoscopy performance was based on previously published data from Kaiser Permanente Northern California, an integrated healthcare delivery system in the United States with a well-defined denominator population.[3,4] In Corley et al, ADR quintile averages (ranges) varied: 15.3% (7.35-19.05) for quintile 1; 21.3% (19.06-23.85) for quintile 2; 25.6% (23.86-28.40) for quintile 3; 30.9% (28.41-33.5.50) for quintile 4; and 38.7% (33.5.51-52.5.51) for quintile 5.[3] Corresponding estimates of per-lesion sensitivity of colonoscopy varied from quintile 1-5: 14.7-98% for adenomas of 0-5mm in diameter, 39.6-98% for adenomas of 6-9mm, and 88.0-98% for adenomas of ≥10mm (Table 1; see Supplementary Appendix for the methodology).[4] The assumed rate of colonoscopy completeness was fixed at 98% for all ADR quintiles. Complication rates for colonoscopy with polypectomy were assumed to increase exponentially with age, from 0.4 per 1000 at age 50 years to 8.5 per 1000 at age 100 years.[19]

The modelled effectiveness of FIT-based screening (OC Sensor test with a positivity cut-off of 20 μg/g cut-off) is based both on the sensitivity and specificity of FIT and the sensitivity and completeness of the colonoscopy exam used for follow-up of positive FIT results. Colonoscopy performance assumptions were varied according to ADR level as described above. Assumed per-lesion sensitivity of FIT was derived from recently published observational data, and was 4.9% for adenomas of 6-9mm, 16.2% for adenomas ≥10mm, and 64-89% for cancer (Table 1).[7]

Analysis

For this study, MISCAN-Colon was used to generate an average-risk screening population of ten million men and women born on January 1, 1965. Patients received annual FIT between the ages 50-75 years.[20] Patients with a positive FIT received follow-up colonoscopy. Patients with adenomas detected in screening received colonoscopy surveillance according to the most current guidelines.[21] We compared colorectal cancer outcomes for FIT according to level of adenoma detection. For reference, we also simulated outcomes with colonoscopy screening and without any screening.

Primary study outcomes were simulated lifetime colorectal cancer incidence and mortality rates according to ADR quintile (undiscounted). We also predicted the continuous change in outcomes per 5% point increase in ADR using linear regression. Further, we also simulated the burden of screening according to ADR quintile, as measured by the required number of follow-up and surveillance colonoscopies, and the associated complications. Multivariate probabilistic sensitivity analysis was used to derive 95% probability intervals (95% PI) for all model outcomes. In 1000 simulation runs of 10 million persons we varied 13 key parameters along uniform, beta, or lognormal distributions.[4] FIT performance assumptions were also varied (Table 1).

We conducted additional sensitivity analyses repeating our estimation of the continuous change in outcomes for every 5% point lower ADR, assuming 5-15% point lower or higher FIT sensitivity, 1.25% point lower or higher FIT specificity, assuming dependency in FIT performance in individuals across study rounds (50% correlated false positive or false negative results), varying the colonoscopy completion rate 75-98% in association with observed ADR variation, and also varying the extent to which ADR variation was attributed by the model to colonoscopy sensitivity for diminutive lesions. In the one extreme, all ADR variation was attributed to variation in small adenoma miss rates; in the other extreme, physicians were assumed to miss all sizes of lesions with equal probability. In addition, we evaluated the impact of ADR for biennial FIT screening, which is recommended and used in many settings.

Role of the funding source

This work was funded by the United States National Cancer Institute. The funder played no role in the design of the study, the analysis and interpretation of the data, and the writing and submission of this report.

RESULTS

Among unscreened patients, the simulated lifetime risk of colorectal cancer was 66.8 (95%PI, 50.7-85.1) per 1000, and the simulated risk of colorectal cancer mortality was 27.8 (95%PI, 20.8-36.5) per 1000 (Figure 1, Table 2).

Screening effectiveness

Among patients screened with annual FIT (with colonoscopy follow-up for positive results), the average simulated colorectal cancer incidence and mortality risks were 37.9 (95%PI, 27.9-50.7) and 7.4 (95%PI, 5.3-10.2) per 1000, respectively (Table 2). Among patients screened with colonoscopy, the average simulated colorectal cancer incidence and mortality risks across all ADR quintiles were 33.4 (95%PI, 24.8-42.8) and 7.7 (95%PI, 5.6-10.2) per 1000, respectively.

The outcomes of FIT screening and primary colonoscopy screening varied according to level of adenoma detection. Among patients receiving annual FIT screening with potential follow-up colonoscopy from providers in the highest ADR quintile, incidence was 28.8 (95%PI, 20.9-40.8) and mortality 5.4 (95%PI, 3.7-7.8) (Figure 1). In contrast, for patients receiving colonoscopy screening from the highest ADR quintile providers, the simulated lifetime cancer incidence and mortality were 20.6 (95%PI, 15.4-27.1) and 4.4 (95%PI, 3.2-5.9) per 1000, respectively (risk ratios for FIT versus colonoscopy, RR=1.40; 95%PI, 1.19-1.71, and RR=1.22; 95%PI, 1.02-1.54) (Figure 2). For every 5% point decrease in ADR, incidence was predicted to increase on average 9.0% (95%PI, 6.7-10.5) for FIT screening and 12.3% (95%PI, 11.1-12.9) for colonoscopy screening (Table 3**,** Figure 3). Thus, in ADR quintile 1, simulated lifetime cancer incidences were more similar, at 49.7 (95%PI, 37.0-65.2) per 1000 for FIT screening and 48.1 (95%PI, 36.1-62.2) per 1000 for colonoscopy screening (RR=1.03; 95%PI, 1.00-1.09) (Figure 1-2). For every 5% point decrease in ADR, simulated mortality increased by an amount similar to cancer incidence: by 9.9% (95%PI, 7.3-11.7) for FIT screening and 13.3% (95%PI, 11.8-14.2) for colonoscopy screening (Table 3**,** Figure 3). Simulated mortality in quintile 1 was lower with primary FIT than with primary colonoscopy,

at 10.1 per 1,000 (95%PI, 7.3-13.6) versus 11.8 per 1000 (95%PI, 8.6-15.8), respectively (RR=0.85; 95%PI,0.83-0.90) (Figure 1-2).

The ratio of simulated mortality risks in each quintile compared to the average simulated risk across quintiles, varied 0.73-1.37 for FIT screening (from quintile 5 to quintile 1), and 0.57-1.53 for colonoscopy screening (Table 2). Simulated variation in incidence ratios across quintiles was similar.

Screening burden

Among patients receiving FIT screening, the number of required follow-up and surveillance colonoscopies varied from 2,408 per 1,000 patients (95%PI, 1,612-3,062) from providers in the highest ADR quintile, to 2,195 per 1,000 patients (95% PI, 1,478-2,806) receiving colonoscopies from providers in the lowest ADR quintile; the associated complications varied from 12.0 (95% PI, 10.2-17.8) to 13.1 per 1,000 patients (95% PI, 7.0-16.4), respectively. In comparison, the number of required colonoscopies in a colonoscopy screening setting varied from 4,645 per 1000 patients (95% PI, 4,151-5,135) in the highest ADR quintile, to 3,756 per 1,000 patients (95% PI, 3,505-4,029) in the lowest ADR quintile, and complications varied from 15.8 (95% PI, 10.8-21.5) to 10.9 per 1,000 (95% PI, 7.3-15.4), respectively.

Sensitivity analysis

Outcomes were sensitive to the assumed test characteristics for FIT and colonoscopy (Table 3). The relative increase in cancer mortality per 5% point lower ADR was smaller for FIT screening when assuming lower FIT sensitivity (8.7%), higher assumed FIT specificity (7.6%), or with dependency of false test results (7.9%), and larger when assuming higher FIT sensitivity (10.8%) or lower specificity (11.3%). When ADR variation was attributed predominantly to small adenomas, the mortality change was lower for both colonoscopy (11.3%) and FIT (7.7%), while with variation in colonoscopy completion or more variation in detection of larger adenomas, mortality changes were larger than the base-case (13.5-14.1% for colonoscopy and 11.1-12.6% for FIT). In all scenarios, ADR variation influenced outcomes more for colonoscopy screening than FIT screening.

In a simulated scenario of biennial FIT screening, which was less effective on average than annual FIT screening, colorectal cancer incidence and mortality were relatively more stable to ADR variations, with predicted relative increases in incidence and mortality of 6.1% per 5% point lower ADR.

DISCUSSION

Using advanced modelling techniques, we showed that there is an inverse relationship between physicians' observed ADR and colorectal cancer screening outcomes that may be stronger when primary screening is performed with colonoscopy than with FIT. Although FIT-based and colonoscopy-based screening strategies had similar average predicted mortality reductions in line with previous estimates,[10] with varying ADRs, the outcomes may differ. For providers from the highest ADR quintile, the model suggested that primary colonoscopy screening would result in fewer colorectal cancer cases and deaths than FIT screening, while conversely, FIT screening outperformed colonoscopy in terms of mortality reductions when physician ADRs levels were <20% (male and female patients combined).

The simulated outcome differences between FIT and colonoscopy screening can be explained by the different test characteristics. While colonoscopy, with relatively long screening intervals, provides long-term protection through removal of most existing lesions at the time of screening,[22] the more frequent FIT screening with follow-up colonoscopy of positive results may primarily detect large adenomas and early-stage cancers before they progress to more advanced-stages.[7] The model assumed that physicians with lower detection rates have a higher proclivity for missing small rather than large adenomas.[4,23] Therefore, FIT outcomes were relatively more stable to varying ADRs than primary screening with colonoscopy (9.4% versus 13.3% predicted increase in disease-related mortality per 5% point ADR decrease). In an alternative model with more assumed variation in sensitivity of colonoscopy for large adenomas (Table 2), outcomes still remained more stable for FIT than colonoscopy, but the differences were smaller (12.6% versus 14.1% increase in mortality per 5% point ADR decrease).

Another consequence of the different test characteristics of colonoscopy and FIT is that, although annual FIT was predicted to be more effective for preventing colorectal cancer deaths than low-quality screening colonoscopy, primary colonoscopy screening resulted in lower simulated colorectal cancer incidences across all ADR quintiles. This is a potential advantage for colonoscopy screening, which has induced some expert panels to favour colonoscopy over other less invasive modalities for colorectal cancer

screening.[24] In contrast, given the different risk and benefit profiles of different screening strategies, the most recent recommendation by the US Preventive Services Task Force puts more emphasis on patient preferences and shared decision-making.[2] European experts prefer faecal testing for screening given the strength of evidence from randomised trials to support these strategies, and the lower burden of testing.

To our knowledge, the present study is the first to have looked at the influence of ADRs on screening outcomes for a stool-based screening setting. Previous empirical studies have found inverse associations between physician ADR levels and post-colonoscopy cancer risk.[25-27] In the largest study to this date, Corley and colleagues found associations between ADR and interval cancer risk that were similar for screening, diagnostic, and surveillance exams.[3] A prior modelling study predicted that commonly observed ADR variation may translate to 50-60% differences in lifetime colorectal cancer outcomes for primary colonoscopy screening.[4] The present study suggests that variations in ADR may have less influence on the outcomes of FIT, with maximum predicted outcome differences of approximately 45%.

The current study may overestimate the outcome differences for faecal testing. First, many settings use biennial rather than annual FIT screening. Our sensitivity analyses found that outcomes of screening were more stable to ADR variation with less frequent (biennial) FIT screening than annual screening, which may be explained by a larger proportion of adenomas presenting as advanced adenomas. Further, while multiple studies have shown that there is substantial variation in ADRs from screening exams,[3,28,29] there are less data available on variation in adenoma detection during colonoscopies after positive faecal colorectal cancer screening test results. Physicians may examine a patient more carefully with evidence of gastrointestinal blood loss, which may improve adenoma detection, even for small adenomas that are unlikely to have caused the positive test result. Although adenoma detection rates in diagnostic examinations are not directly comparable to those in screening exams, wide variation in detection rates across physicians within population-based FIT screening settings suggest that miss rates may also vary substantially.[8,9] High observed risks of cancer after positive FITs followed by negative colonoscopies (for adenomas) could also be indicative of suboptimal quality.

A limitation of this study is the lack of direct experimental data to inform the model on efficacy of FIT and colonoscopy screening.[6] We modelled the average efficacy of FIT using an established approach used before in a decision analysis to inform the US Preventive Services Task Force.[10] This approach combines evidence from guaiac faecal occult blood testing trial data with observational data on FIT's diagnostic performance.[7,30] Colonoscopy efficacy predictions were derived similarly using flexible sigmoidoscopy trial results. The predicted effects in this study for colonoscopy and FIT were substantively larger than the effects found for guaiac-based faecal testing and sigmoidoscopy, which may be partly due to better performance characteristics, and to assumed high patient adherence in this study. The simulated mortality effects of FIT are comparable to results from a recent major population-based study (for exposed vs unexposed people),[16] and colonoscopy effects are within the outcome range of observational studies,[17] supporting the use of this approach for the present study.

Our study focused on the influence of observed ADR variation on screening effectiveness. There are other important outcomes, such as cost and cost-effectiveness, and modifiable outcome determinants for colonoscopy and FIT, including ambient FIT temperature,[31] adherence with and time from a positive FIT to diagnostic colonoscopy follow-up,[32] and particularly, patient adherence with annual or biennial FIT testing.[33] Our assumption of 100% adherence with both colonoscopy and FIT screening does not represent actual practice, but provides maximum achievable effects for both screening modalities, in line with previous decision analyses. In actual practice, adherence may differ for colonoscopy and FIT. With more data on patient adherence over time, future modelling studies could inform priorities of quality-related interventions by assessing and ranking the relative contribution of quality in comparison with adherence.

A strength of this study is that we based our estimates for variable colonoscopy performance characteristics on empirical data regarding interval cancer incidence rates after colonoscopy screening according to physician ADR.[3] Our assumptions have been shown to match well with the observed decreasing incidence pattern from lower to higher ADRs.[3,4] Alternative models with relatively more or less emphasis on variation in detection of diminutive lesions and variable colonoscopy completion rates,

as evaluated in sensitivity analyses, matched the data less well, which suggests that our base-case assumptions are reasonable. However, we cannot rule out other possible explanations for the observed incidence pattern, such as an association of ADR with adequate polyp management or serrated polyp detection rates.

Our study has three main implications. First, the study results are consistent with the results of the most recent decision analysis for the US Preventive Services Task Force in finding similar overall estimated mortality reductions for annual FIT screening and colonoscopy screening strategies. This confirms that if no information is available on ADRs, both strategies are potentially equivalent in terms of the predicted mortality reduction. The average predicted benefits across ADR quintiles in this study were somewhat more conservative (20 deaths averted per 1000 patients) than the estimated benefits in the USPSTF analysis (23 and 24 death averted, respectively). The difference is mainly due to the alternative diagnostic performance assumptions for colonoscopy as used in this study.

Second, physician ADR is an important indicator for colorectal cancer screening performance, irrespective of whether the primary screening modality is colonoscopy or FIT. This underscores the importance of ongoing efforts to measure and benchmark physicians' ADR scores,[34] as formalised in some quality assurance programs.[35] Research suggests that endoscopist training programs may be effective at increasing ADR levels.[36] If large population-based studies confirm that such programs also have a favourable impact on cancer-related outcomes, other screening programs should consider offering similar trainings to improve screening effectiveness.

Finally, our results suggest that ADR may be useful not only as a quality indicator for screening, but may also affect comparative screening program performance and outcomes. We found that the benefits of FIT relative to colonoscopy screening may differ depending on the quality of colonoscopy achieved in a particular program. This observation suggests that available colonoscopy quality may be one of the relevant factors that policy makers may consider in selecting the most appropriate screening method for their particular setting. Research is needed to further investigate from what number of exams ADR can be

reliably estimated and used as a predictor of both screening outcomes in general, and comparative performance of alternative screening methods in particular.

To conclude, the relative cancer incidence and mortality reductions for annual FIT versus colonoscopy screening may differ based on colonoscopy quality, as measured by ADR. Although the predicted mortality benefits are similar for FIT and colonoscopy with average ADR levels, colonoscopy screening may result in fewer cancer deaths in settings with higher ADR levels, while annual FIT screening may result in fewer deaths in lower ADR settings.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST: None of the authors report conflicts of interest. Dr. Doubeni is a member of the US Preventive Services Task Force (USPSTF). This article does not necessarily represent the views and policies of the USPSTF.

GRANT SUPPORT: This publication was made possible by Grant Numbers U54 CA163262 (PROSPR), U01 CA152959 (CISNET), U01 CA151736, U24 CA171524, and P30 CA008748 from the National Cancer Institute (NCI). MISCAN-colon is part of the Cancer Intervention and Surveillance Modeling Network (CISNET). This work was supported by the NCI-funded consortium Population-Based Research Optimizing Screening through Personalized Regimens (PROSPR), the overall aim of which is to conduct multi-site, coordinated, transdisciplinary research to evaluate and improve cancer screening processes.

AUTHOR CONTRIBUTIONS: Reinier G.S. Meester Ph.D. made substantial contributions to the conception of the work, the acquisition, analysis and interpretation of the data, and drafting; Iris Lansdorp-Vogelaar Ph.D; Chyke A. Doubeni M.D. M.P.H., Marjolein van Ballegooijen M.D., Ph.D.[1]Douglas A. Corley M.D., Ph.D., Ann G. Zauber Ph.D. made substantial contributions to the conception of the work, the interpretation of the data, and critical revision for important intellectual content; Theodore R. Levin M.D., Christopher D. Jensen Ph.D., Virginia P. Quinn Ph.D. made substantial contributions to the critical revision for important intellectual content; Reinier G.S. Meester M.S., Iris Lansdorp-Vogelaar Ph.D had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data.

REFERENCES

1. Siegel R, Miller K, Fedewa S, et al. Colorectal Cancer Statistics, 2017. *CA Cancer J Clin* 2017; ***In Press***.
2. U. S. 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**(23): 2564-75.
3. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014; **370**(14): 1298-306.
4. Meester RG, Doubeni CA, Lansdorp-Vogelaar I, et al. Variation in Adenoma Detection Rate and the Lifetime Benefits and Cost of Colorectal Cancer Screening: A Microsimulation Model. *Jama* 2015; **313**(23): 2349-58.
5. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015; **64**(10): 1637-49.
6. Quintero E, Castells A, Bujanda L, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *N Engl J Med* 2012; **366**(8): 697-706.
7. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014; **370**(14): 1287-97.
8. Lee TJ, Rutter MD, Blanks RG, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. *Gut* 2012; **61**(7): 1050-7.
9. Zorzi M, Senore C, Da Re F, et al. Quality of colonoscopy in an organised colorectal cancer screening programme with immunochemical faecal occult blood test: the EQuIPE study (Evaluating Quality Indicators of the Performance of Endoscopy). *Gut* 2015; **64**(9): 1389-96.
10. Knudsen AB, Zauber AG, Rutter CM, et al. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force. *JAMA* 2016.
11. van Hees F, Zauber AG, van Veldhuizen H, et al. The value of models in informing resource allocation in colorectal cancer screening: the case of The Netherlands. *Gut* 2015; **64**(12): 1985-97.
12. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. *Cancer* 1975; **36**(6): 2251-70.
13. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, Zauber A, Habbema JD. A novel hypothesis on the sensitivity of the fecal occult blood test: Results of a joint analysis of 3 randomized controlled trials. *Cancer* 2009; **115**(11): 2410-9.
14. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010; **375**(9726): 1624-33.
15. Rutter CM, Knudsen AB, Marsh TL, et al. Validation of Models Used to Inform Colorectal Cancer Screening Guidelines: Accuracy and Implications. *Med Decis Making* 2016; **36**(5): 604-14.
16. Chiu HM, Chen SL, Yen AM, 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.
17. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ* 2014; **348**: g2467.
18. Zauber A, SJ Winawer, JD Waye et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *New England Journal of Medicine* 2012; **366**(8): 10.
19. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. *Ann Intern Med* 2009; **150**(12): 849-57, W152.
20. U. S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008; **149**(9): 627-37.

21. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012; **143**(3): 844-57.
22. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013; **369**(12): 1095-105.
23. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006; **101**(2): 343-50.
24. Rex DK, Johnson DA, Anderson JC, et al. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. *Am J Gastroenterol* 2009; **104**(3): 739-50.
25. Baxter NN, Sutradhar R, Forbes SS, Paszat LF, Saskin R, Rabeneck L. Analysis of administrative data finds endoscopist quality measures associated with postcolonoscopy colorectal cancer. *Gastroenterology* 2011; **140**(1): 65-72.
26. Cooper GS, Xu F, Barnholtz Sloan JS, Schluchter MD, Koroukian SM. Prevalence and predictors of interval colorectal cancers in medicare beneficiaries. *Cancer* 2012; **118**(12): 3044-52.
27. Kaminski MF, J. Regula, E. Butruk, et al. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *New England Journal of Medicine* 2010; **362**(19): 9.
28. Barclay RL, Vicari JJ, Doughty AS, Johanson JF, Greenlaw RL. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *N Engl J Med* 2006; **355**(24): 2533-41.
29. Adler A, Wegscheider K, Lieberman D, et al. Factors determining the quality of screening colonoscopy: a prospective study on adenoma detection rates, from 12,134 examinations (Berlin colonoscopy project 3, BECOP-3). *Gut* 2013; **62**(2): 236-41.
30. Holme O, Bretthauer M, Fretheim A, Odgaard-Jensen J, Hoff G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. *Cochrane Database Syst Rev* 2013; **9**: CD009259.
31. Doubeni C, Jensen C, Fedewa S, et al. Mailed fecal immunochemical test positivity and sensitivity vary with ambient temperature: a population-based study. *Journal of the American Board of Family Medicine* 2016; (In press).
32. Meester RG, Zauber AG, Doubeni CA, et al. Consequences of Increasing Time to Colonoscopy Examination Following Positive Result From Fecal Colorectal Cancer Screening Test. *Clin Gastroenterol Hepatol* 2016.
33. Winawer SJ, Fischer SE, Levin B. Evidence-Based, Reality-Driven Colorectal Cancer Screening Guidelines: The Critical Relationship of Adherence to Effectiveness. *JAMA* 2016; **315**(19): 2065-6.
34. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ* 2008; **86**(4): 317-9.
35. National Institute for Public Health and Environment. Bowel cancer screening programme,. 2014. http://www.rivm.nl/en/Topics/B/Bowel_cancer_screening_programme (accessed Sep 28 2016).
36. Kaminski MF, Anderson J, Valori R, et al. Leadership training to improve adenoma detection rate in screening colonoscopy: a randomised trial. *Gut* 2016; **65**(4): 616-24.
37. Wilschut JA, Hol L, Dekker E, et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology* 2011; **141**(5): 1648-55 e1.

Table 1 Base-case test-performance assumptions in MISCAN.

Performance characteristic, %	Colonoscopy, by quintile [a] (screening, diagnostic, surveillance)					FIT
	1	2	3	4	5	
Sensitivity per lesion [b]						
Adenomas ≤ 5 mm	14.7	29.6	41.0	66.2	98	0.0
Adenomas 6 - 9 mm	39.6	65.8	85	94.3	98	11.4 [2.5][e]
Adenomas ≥ 10 mm	88.0	92.2	95	96.8	98	15.9 [3.5][e]
Stage I-IV cancer	98	98	98	98	98	63/89 [3.5][e]
Specificity [c]	100	100	100	100	100	96.4 [1.25][e]
Completeness colonoscopy [d]	98	98	98	98	98	-

Abbreviations: FIT = faecal immunochemical test; ADR q*i* = adenoma detection rate quintile *i*.

[a] Adenoma detection rate (ADR) quintiles were derived from 57,588 colonoscopies performed by 136 gastroenterologists in Kaiser Permanente Northern California, a large integrated healthcare delivery system in the United States. The averages (and ranges) of ADR for quintiles 1 through 5 were 15.32% (7.35-19.05%), 21.27% (19.06-23.85%), 25.61% (23.86-28.40%), 30.89% (28.41-33.5.50%) and 38.66% (33.5.51-52.5.51%), respectively.[3]

[b] The adenoma sensitivity estimates for FIT (OC Sensor, cut-off >20 µg/g) were obtained by calibrating our model outcomes to the estimated per-person sensitivities from Imperiale et al.[7] The per-person sensitivity of FIT for adenomas, advanced adenomas, and cancer was 7.6, 23.8, 73.8, respectively. We assumed that faecal occult blood testing is more sensitive in cancers towards the end of the occult invasive period (close, time-wise, to becoming symptomatic): for preclinical cancers which will become symptomatic within the same stage, assumed test sensitivity was higher. This assumption showed good concordance with guaiac faecal occult blood test trial results.[13] Colonoscopy sensitivity estimates were derived elsewhere.[4]

[c] The probability of a false positive result was random in the base-case analysis, and independent of person or lesion. We assumed perfect specificity for colonoscopy including pathological examination of detected lesions.

[d] This is the proportion of colonoscopies visualising the maximum point of reach of the endoscope, i.e. the caecum.

[e] Standard deviation for the probabilistic sensitivity analysis is shown in brackets. A Beta distribution was assumed to reflect uncertainty. A difference of +/-5% point in assumed test sensitivity or of -/+2% in specificity corresponds roughly to a cutoff difference of + 20/-10 µg blood per g faeces.[37]

Table 2 Simulated benefit and burden of screening according to screening modality and ADR quintile[a].

Screening Program	ADR quintile	CRC Cases			CRC Deaths			LY gained		Colonoscopies for screening, follow-up and surveillance		Colonoscopy-related complications		
		mean	95% CI	RR	mean	95% CI	RR	mean	95% CI	mean	95% CI	mean	95% CI	RR
None	-	66.8	50.7-85.1		27.8	20.8-36.5		-	-	-	-	-	-	
FIT	1	49.7	37-65.2	*1.31*	10.1	7.3-13.6	*1.37*	200.9	149.5-263.7	2195	1478-2806	13.1	7-16.4	*0.96*
	2	41.8	31.1-54.7	*1.10*	8.2	5.9-11.2	*1.11*	222.4	164.5-290.4	2280	1533-2892	13.3	10.2-17.3	*0.98*
	3	36.9	27.0-49.5	*0.97*	7.1	5.1-9.9	*0.96*	235.3	174.3-306.9	2324	1563-2954	14.5	10.5-17.9	*1.07*
	4	32.5	23.6-44.7	*0.86*	6.1	4.3-8.8	*0.83*	246.3	183.0-322.1	2379	1594-3022	15.0	10.6-18.2	*1.11*
	5	28.8	20.9-40.8	*0.76*	5.4	3.7-7.8	*0.73*	254.6	187.6-331.4	2408	1612-3062	12.0	10.2-17.8	*0.89*
	1-5 (average)	37.9	27.9-50.7	-	7.4	5.3-10.2	-	231.9	172.5-302.3	2317	1557-2943	13.6	10.2-16.9	-
COL	1	48.1	36.1-62.2	*1.44*	11.8	8.6-15.8	*1.53*	183.1	137.7-241.1	3756	3505-4029	10.9	7.3-15.4	*0.78*
	2	38.8	28.9-49.9	*1.16*	9.0	6.5-12.0	*1.17*	212.5	158.2-277.1	4045	3729-4388	13.4	9.2-18.5	*0.96*
	3	32.9	24.4-42.9	*0.99*	7.5	5.4-9.9	*0.97*	230.0	171.2-301.6	4211	3849-4594	14.5	9.9-19.7	*1.03*
	4	26.4	19.7-34.7	*0.79*	5.8	4.2-7.9	*0.75*	248.9	186.1-329.7	4453	4021-4900	15.5	10.7-21.2	*1.11*
	5	20.6	15.4-27.1	*0.62*	4.4	3.2-5.9	*0.57*	265.0	196.9-350.4	4645	4151-5135	15.8	10.8-21.5	*1.13*
	1-5 (average)	33.4	24.8-42.8	-	7.7	5.6-10.2	-	227.9	170.2-300.4	4222	3856-4604	14.0	9.6-19.1	-

ADR = adenoma detection rate, RR = risk ratio, COL = colonoscopy, FIT = Faecal immunochemical test.

[a] Outcomes presented are per 1000 persons.

[b] Risk ratios compared to the average risk across quintiles.

[c] Colonoscopy without polypectomy was not associated with a higher risk of complications. The risk of complications for polypectomy increased exponentially with age. Complications include serious gastrointestinal events such perforation and gastrointestinal bleeding requiring blood transfusions; other gastrointestinal events such as paralytic ileus, nausea, vomiting and dehydration, and abdominal pain; and cardiovascular events such as myocardial infarction or angina, arrhythmias, congestive heart failure, cardiac or respiratory arrest, or syncope, hypotension, or shock. The fatal perforation rate was derived from estimates of the incidence of perforation and case fatality for perforation.

John Wiley & Sons, Inc.

Table 3 Sensitivity analysis results: % increase in simulated incidence and mortality per 5% point lower ADR. [a]

Scenario	Colonoscopy		FIT	
	Incidence	Mortality	Incidence	Mortality
1. Base case	12.3	13.3	9.0	9.9
2. Alternative FIT performance				
a. Lower FIT sensitivity [b]	-	-	8.0	8.7
b. Higher FIT sensitivity	-	-	9.7	10.8
c. Lower FIT specificity [c]	-	-	10.2	11.3
d. Higher FIT specificity	-	-	7.0	7.6
e. Correlated FIT results [d]	-	-	6.6	7.9
3. Alternative colonoscopy performance (variation in ADR attributed to -)[e]				
a. More variation in adenoma ≤5mm	9.7	11.3	7.1	7.7
b. Less variation in adenoma ≤5mm	12.6	14.1	11.3	12.6
c. Variation in caecal intubation	11.4	13.5	9.3	11.1
4. Biennial FIT screening[f]	-	-	6.1	6.1

FIT = Faecal immunochemical test, ADR = adenoma detection rate

[a] Mean simulated outcome differences per 5% decrease in ADR were derived by linear regression and presented relative to the model outcomes for ADR quintile 1 ($5 \times \beta_{ols}/outcome_{q1}$). The actual ADR-outcome relationship was slightly convex (rather than perfectly linear), particularly for FIT screening outcomes: the outcome impact of ADRs changes was somewhat larger at lower baseline ADR levels than at higher levels (see **Figure 3**).

[b] We assumed 5% point lower/higher sensitivity of FIT for adenomas, and 10-15% point lower/higher values for cancer.

[c] We assumed 2.5% point lower/higher specificity of FIT.

[d] It was assumed that patients with a false negative result were likelier than average to test negative again in subsequent screening rounds, Similarly, false-positive results were assumed to occur predominantly in specific patients with an inclination for faecal blood loss from causes other than adenoma. Positivity decreases as soon as these predisposed patients become ineligible for screening (i.e. after follow-up colonoscopy receipt).

[e] With more emphasis on small adenomas, all variation in ADRs was attributed to sensitivity of colonoscopy for adenomas smaller than 5 mm, which varied from 5.4%, lowest, to 98%, highest quintile. With less emphasis, all ADR variation was attributed equally to sensitivity for small, medium, and large adenomas, which varied from 26.0% to 98%. With variation in caecal intubation rates, ADR variation was attributed to both colonoscopy sensitivity and reach, the latter varying from 75% to 98% from lower to higher ADR quintiles,

[f] Biennial FIT screening resulted in higher CRC incidence and mortality risks than annual FIT screening of, on average, 45.1 and 9.9 per 1000 patients, respectively.

Figure 1 Simulated colorectal cancer incidence (a) and mortality (b) with FIT versus colonoscopy screening, according to ADR quintile. [a]

Abbreviations: ADRqi = Adenoma Detection Rate quintile i (i = 1,…,5).

[a] Colonoscopy screening outcomes were previously published.[4] Whiskers represent 95% probability intervals from multivariate probabilistic sensitivity analysis. See **Figure 3** for a plot of outcomes against actual, average ADR levels per quintile.

Figure 2 Simulated relative risks of colorectal cancer incidence (a) and mortality (b) for FIT versus colonoscopy screening. [a]

Abbreviations: ADRqi = Adenoma Detection Rate quintile i (i = 1,…,5).

[a] Whiskers represent 95% probability intervals from multivariate probabilistic sensitivity analysis. The variable width of probability intervals from the probabilistic sensitivity analysis was due to interaction of colonoscopy and FIT performance: in the model, lower ADRs decreased the outcome effect of FIT's variable false positive rates and the associated colonoscopy receipt, and higher ADRs increased the effect.

Figure 3 The shape of the simulated relationship between incidence (a) and mortality (b) and ADR level [a]

[a] Colonoscopy screening outcomes were previously published.[4] Whiskers represent 95% probability intervals from multivariate probabilistic sensitivity analysis (FIT in colour, colonoscopy in black). Data points on the x-axis represent ADR quintile averages.

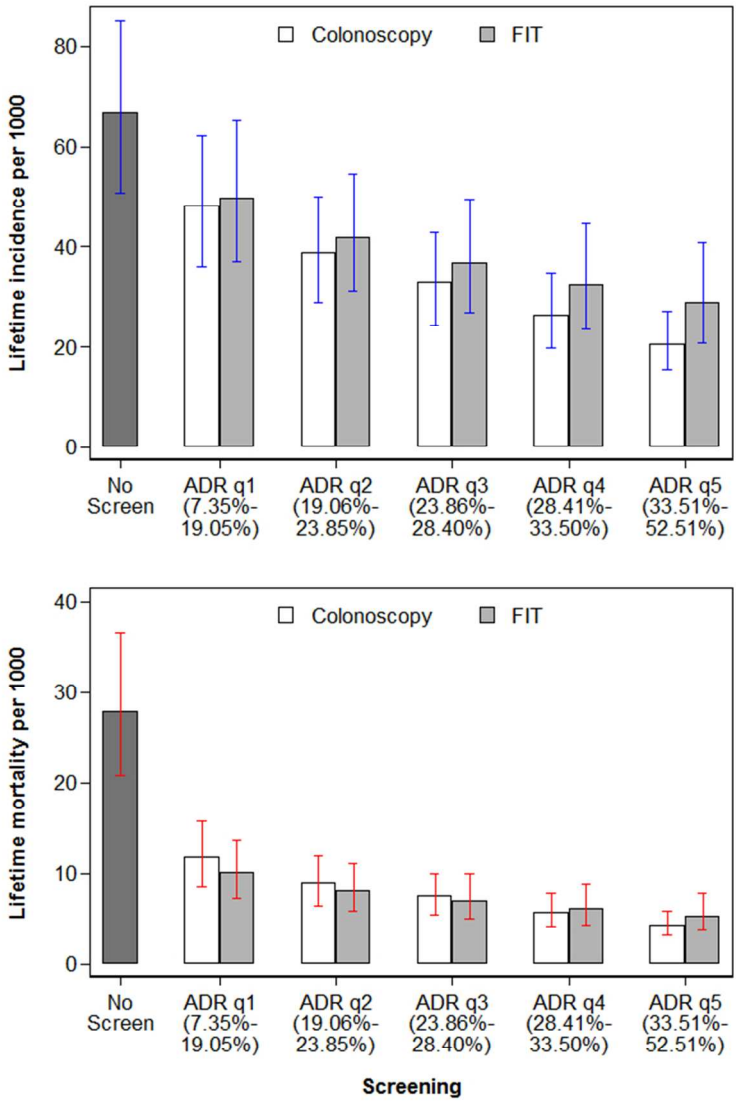


Figure1

227x331mm (72 x 72 DPI)

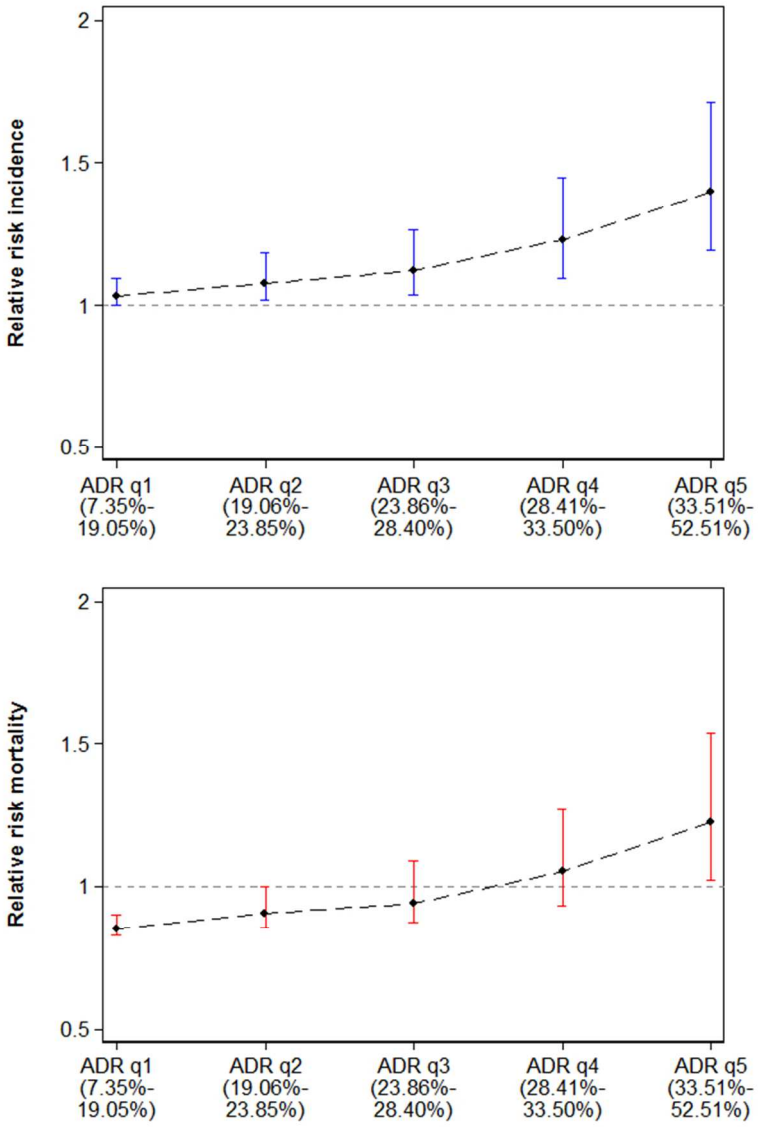
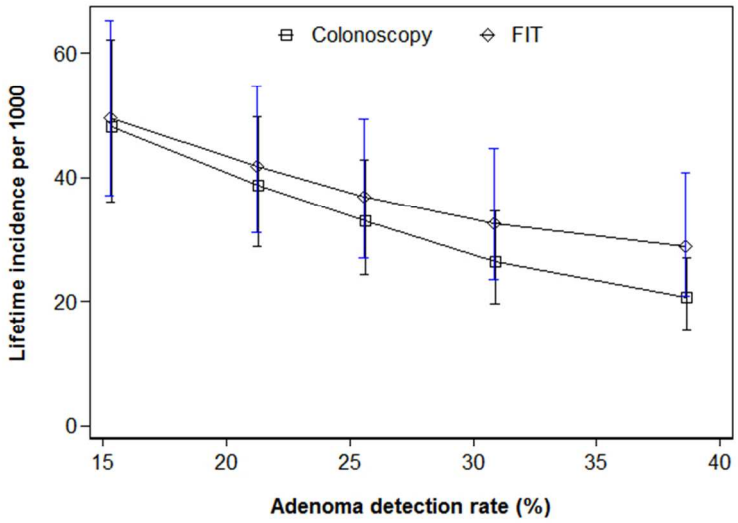


Figure2

227x331mm (72 x 72 DPI)



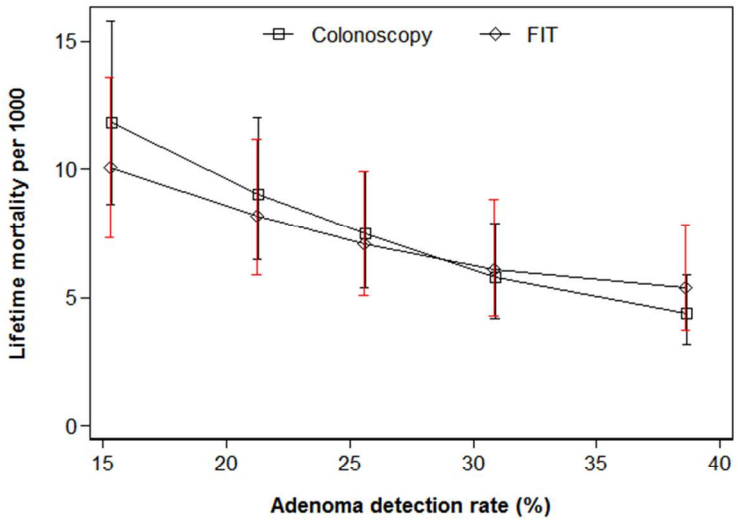


Figure3

227x331mm (72 x 72 DPI)