

BMJ Open Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models

Paulina Wieszczy ^{1,2}, Michal F Kaminski,^{1,2,3} Magnus Løberg,^{1,4} Marek Bugajski,^{2,3} Michael Bretthauer,^{1,4} Mette Kalager^{1,4}

To cite: Wieszczy P, Kaminski MF, Løberg M, *et al*. Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. *BMJ Open* 2021;**11**:e042158. doi:10.1136/bmjopen-2020-042158

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-042158>).

Received 26 June 2020
Revised 10 March 2021
Accepted 04 April 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Mette Kalager;
mkalager@hsph.harvard.edu

ABSTRACT

Objective To estimate overdiagnosis of colorectal cancer (CRC) for screening with sigmoidoscopy and faecal occult blood testing (FOBT).

Design Simulation study using data from randomised trials.

Setting Primary screening, UK, Norway

Participants 152 850 individuals from the Nottingham trial and 98 678 individuals from the Norwegian Colorectal Cancer Prevention (NORCCAP) trial.

Intervention CRC screening.

Outcome measure We estimated overdiagnosis using long-term data from two randomised trials: the Nottingham trial comparing FOBT screening every other year to no-screening, and the NORCCAP trial comparing once-only sigmoidoscopy screening to no-screening. To estimate the natural growth of adenomas to CRC, we used the following microsimulation models: (i) the Microsimulation Screening Analysis; (ii) the CRC Simulated Population model for Incidence and Natural history; (iii) the Simulation Model of Colorectal Cancer; (iv) a model derived by the German Cancer Research Center. We defined overdiagnosed cancers as the difference between the observed number of CRCs in the no-screening arm and the expected number of cancers in screening arm (sum of observed and prevented by adenoma removal). The amount of overdiagnosis is defined as the number of overdiagnosed cancers over the number of cancers observed in the no-screening arm.

Results Overdiagnosis estimates were highly dependent on model assumptions. For FOBT screening with 2354 cancers observed in control arm, four out of five models predicted overdiagnosis, range 2.0% (2400 cancers expected in screening) to 7.6% (2533 cancers expected in screening). For sigmoidoscopy screening with 452 cancers observed in control arm, all models predicted overdiagnosis, range 25.2% (566 cancers expected in screening) to 128.1% (1031 cancers expected in screening).

Conclusions The amount of overdiagnosis estimated based on the microsimulation models varied substantially. Microsimulation models may not give reliable estimates of the preventive effect of adenoma removal, and should be used with caution to inform guidelines.

Strengths and limitations of this study

- Data from two large randomised trials with long follow-up and detailed information of adenomas and cancers were used.
- Four different microsimulation models to estimate the natural growth of adenomas to colorectal cancer were applied.
- The amount of overdiagnosis was highly dependent on the transition rates in the different models of natural history.
- Microsimulation models may not give reliable estimates of the preventive effect of adenoma removal and should be used with caution.

INTRODUCTION

Overdiagnosis is diagnosis of a disease that would not have developed to cause symptoms or death in the patient's lifetime if not detected by screening.^{1 2} While largely disregarded 10 years ago, overdiagnosis is now recognised as a major harm of cancer screening.¹⁻³

A substantial amount of overdiagnosis has been observed for mammography screening for breast cancer, ultrasound screening for thyroid cancer, and prostate cancer by prostate-specific antigen.¹

Although colorectal cancer (CRC) screening is endorsed in the USA, Europe and in many other areas of the world, there is a striking paucity of studies on overdiagnosis. CRC screening entails different tests with varying characteristics, and most likely different amounts of overdiagnosis. The most commonly used tests are faecal occult blood testing (FOBT), sigmoidoscopy and colonoscopy.³ Faecal tests are mainly early detection screening tests as they primarily detect invasive cancers. Colonoscopy and sigmoidoscopy

are mainly preventive screening tests as they primarily detect and remove adenomas which are cancer precursors. Because all CRC screening tests entail colonoscopy (either for primary screening, or for individuals positive with FOBT or sigmoidoscopy), all entail a component of preventive screening.³

The few studies investigating overdiagnosis in CRC screening report overdiagnosis of CRC of 6%–9% after positive FOBT, and 1 per 1000 screening colonoscopies detect and remove an adenoma that would never progress to CRC. However, the studies have been hampered by the failure to recognise the cancer-preventive effect of screening to ascertain correct estimates of overdiagnosis, and a lack of proper comparison groups, which made reliance entirely on modelling necessary.^{4,5}

We took advantage of recent long-term follow-up data from large randomised trials of FOBT and sigmoidoscopy screening, with optimal comparative groups.^{6,7} The aim of the study was to estimate the risk of overdiagnosis of CRC based on the numbers of adenomas and cancers removed at FOBT and sigmoidoscopy screening.

METHODS

Concept of estimating overdiagnosis

Estimating overdiagnosis for CRC screening requires knowledge about the natural history of adenomas to CRC. The natural history of adenomas cannot be observed directly. However, microsimulation models for CRC screening which are widely used for guidelines and policy making, for example, by the US Preventive Services Task Force and the American Cancer Society⁸ have natural history assumptions as an important part of their models.^{9,10}

Study design

We estimated overdiagnosis of CRC by combining data from two sources:

1. We used data of observed adenomas and CRCs from two large randomised trials of FOBT screening⁶ and flexible sigmoidoscopy.⁷ Both trials entailed no-screening control arms.
2. For the natural history of adenomas to cancer, we applied estimates from the three most commonly used microsimulation models of CRC screening (a–c) and estimates derived from an observational setting (d).
 - a. The Microsimulation Screening Analysis (MISCAN).¹¹
 - b. The Colorectal Cancer Simulated Population model for Incidence and Natural history (CRC-SPIN).¹¹
 - c. The Simulation Model of Colorectal Cancer (SimCRC).¹¹
 - d. Model derived by the German Cancer Research Center (German).¹²

All microsimulation models (a–c) simulate the lifetime risk of developing adenoma and CRC for a large population of individuals. The models were developed independently and were calibrated to the same data regarding

adenoma prevalence, cancer incidence and stage distribution. The data were collected by Cancer Intervention and Surveillance Modeling Network. The model derived by the German Cancer Research Center (d) was developed using the German screening colonoscopy data and is based on the observed numbers of adenomas and cancers detected in the German population.

By combining the natural history estimates from these models with the data of actual observed adenomas and cancers in the randomised trials, we estimated the number of expected cancers if adenomas had not been removed at screening, and compared this to the observed number of CRCs in the no-screening control arms of the two randomised trials.

If the observed and expected numbers of cancer in the screening arm are similar to the control arm, there is no overdiagnosis. If the numbers are higher, this would reflect overdiagnosis, or that the models, which have been validated and are used to inform guidelines, may be incorrect.

Data sources

Screening with FOBT

The Nottingham trial is a randomised trial comparing FOBT screening with no-screening.^{6,13} Between 1981 and 1991, 152 850 individuals age 45–74 years living in the Nottingham area in England were randomised in a ratio of 1:1 to either screening by guaiac-based FOBT every other year, or to no screening.¹³ Between three and six screening rounds were applied to the screening arm, depending on the date of study entry. Attendance to at least one FOBT screening round was 59.9% (44 838 individuals), attendance to all screening rounds was 38.2% (28 720 individuals).

For the current study, we used data from the most recent follow-up of the trial published in 2012 (median follow-up time 19.5 years) reporting on removed adenomas and on CRC incidence and mortality for the screening as compared with the control arm.⁶

Screening with sigmoidoscopy

The Norwegian Colorectal Cancer Prevention Trial (NORCCAP) is a randomised trial comparing once-only sigmoidoscopy screening to no screening.^{7,14} Between 1999 and 2001, all residents of Oslo and Telemark County in Norway aged 50–64 years were randomised to sigmoidoscopy screening with (10 388 individuals) or without (10 392 individuals) an immunochemical FOBT, or to no screening (79 430 individuals).¹⁴ In all analyses the number of observed cancers in the control arm is rescaled to fit the screening arm.

For the current study, we used individual patient data provided by the NORCCAP investigators, including sex, age, CRC diagnosis, death and immigration, and for the screening participants the date of adenoma removal. Median follow-up was 14.8 years.⁷ We used the actual age and sex distribution among individuals with adenomas,

consisting of 60% men and 40% women for adenomas removed at screening.

Model assumptions

Since we did not have access to individual data for adenoma patients for the FOBT trial, we assumed follow-up time of the entire 19-year study period for patients who had adenomas removed in the first screening round. For adenomas removed in subsequent rounds, we assumed uniform number of adenomas removed in each round, with corresponding shorter follow-up time. For detailed information on our assumptions, see online supplemental appendix 1.

Since we had access to individual data in the sigmoidoscopy trial, including date of adenoma removal, size and histology of the adenomas, and follow-up time, no assumptions were needed.

Modeling

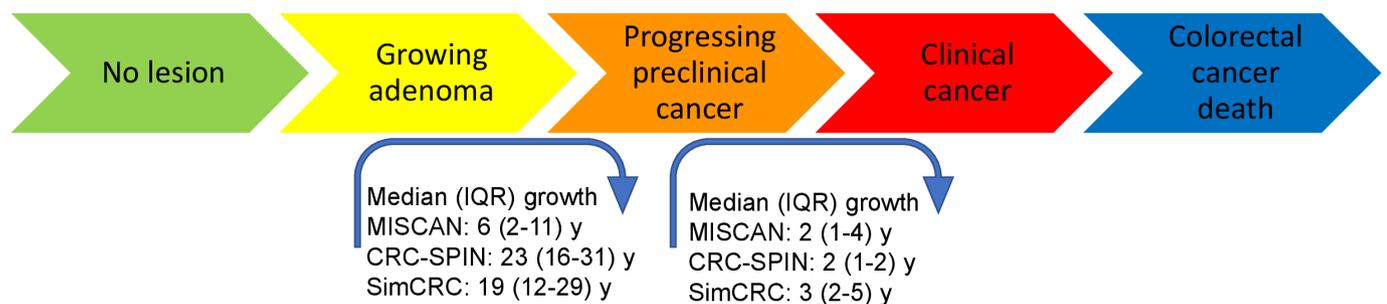
We used Markov models with discrete time (1-year intervals) to calculate the number of prevented cancers. Each year any adenoma patient could stay in the same state, move from the non-advanced/advanced adenoma state to preclinical cancer, or from preclinical cancer to cancer. Additionally, at each state patients could move to overall death. The state transition probabilities for overall death were obtained from population registries (see online supplemental appendix 1).

In order to estimate the transition probabilities from one state to another we used the four above-mentioned CRC screening models, including five scenarios.^{11 12}

Each of the models provides assumptions for the natural history of adenoma growth as presented in figure 1. The MISCAN, CRC-SPIN and SimCRC models simulate time (in years) from adenoma to preclinical cancer and from preclinical cancer to symptomatic cancer.¹¹ In our analyses using these three models, we assumed that all adenomas removed at screening were non-advanced (the most conservative approach). We did not adjust our analyses to multiple adenomas because as we focus on the time to the first cancer (ie, duration for one of the detected adenomas to develop to cancer). We used the median dwell time to estimate the annual transition between stages in the models. In the MISCAN model, non-progression of some adenomas is part of the natural history. We used probabilities of progressive adenomas reported by the authors, that is, 14% for individuals aged less than 65% and 14% to 96% (linearly increasing) for individuals aged 65–100 (25.7% progression at age 70, 49.1% progression at age 80, etc).¹⁵

The German model¹² estimates age-dependent and sex-dependent annual transition rates from non-advanced adenoma to advanced adenoma, from advanced adenoma to preclinical cancer, and from preclinical cancer to clinical cancer. In our estimates we used the highest (German, high) and lowest (German, low) point-estimates for annual transition rates. Since no information on adenoma histology was available for the FOBT trial, we defined all adenomas ≥ 10 mm as advanced, and all adenomas < 10 mm as non-advanced.

Model 1-3¹²



Model 4¹¹

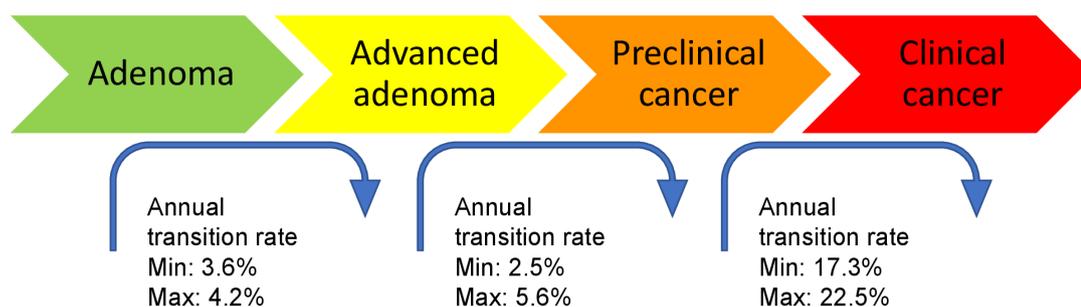


Figure 1 The assumptions of growth and time and rates from adenoma to cancer in the four different models. Model 1–3¹¹ is the Microsimulation Screening Analysis (MISCAN), the Colorectal Cancer Simulated Population model for Incidence and Natural history (CRC-SPIN) and the Simulation Model of Colorectal Cancer (SimCRC).¹¹ Model 4 is a model derived by the German Cancer Research Center (German).¹²

To obtain additional details on applied transition rates for adenomas and cancer, we contacted the MISCAN, CRC-SPIN and SimCRC investigators. We got responses from the MISCAN and CRC-SPIN modellers and following their advice, we performed the following sensitivity analyses. For the MISCAN model, the authors assume that the time from adenoma to cancer follows an exponential distribution with a mean of 140 years (meaning that most of the adenoma patients will not develop cancer within their lifespan). Thus, we drawn 1000 samples from an exponential distribution with mean of 140 years and calculate a median over the samples to estimate the annual transition between adenoma and cancer. For the CRC-SPIN we used annual transition rates for the following steps of the natural history: from adenoma <5 mm to adenoma 5–9 mm to adenoma ≥10 mm to preclinical cancer to cancer.¹⁶

Data analysis

Overdiagnosis calculations

We estimated overdiagnosis as the difference between the expected number of cancers in the screening arm if no adenomas had been removed at screening and the number of cancers observed in the control arm. In our *primary analysis*, we used the following formula to calculate the number of overdiagnosed cancers:

$$OD = CRC_{EXP}^{Src} - CRC_{OBS}^{Ctrl}$$

$$OD = (CRC_{OBS}^{Src} + CRC_{ADN}^{Src+}) - (CRC_{OBS}^{Ctrl}) \text{ (Primary analysis)}$$

where CRC_{EXP}^{Src} are the number of expected cancers among screening participants in the screening arm if no adenomas had been removed, and CRC_{OBS}^{Ctrl} is the number of observed cancers in the control arm. The number of expected cancers is the sum of the numbers of cancers observed (diagnosed) in the screening arm CRC_{OBS}^{Src} , and the numbers of cancers that would develop from adenomas if they were not removed at screening CRC_{ADN}^{Src+} (figure 2 and box 1).

In a *secondary analysis*, we included all adenomas, both in the screening and control arm of the FOBT trial. This was not done for the sigmoidoscopy trial since no information existed on adenoma removal in the non-attenders and in the control arm. The following formula was used to estimate overdiagnosis:

$$OD = \left(CRC_{OBS}^{Src} + CRC_{ADN}^{Src+} + CRC_{ADN}^{Src-} \right) - \left(CRC_{OBS}^{Ctrl} + CRC_{ADN}^{Ctrl} \right) \text{ (Secondary analysis)}$$

where CRC_{ADN}^{Src-} is the number of cancers expected if no adenomas had been removed among non-attenders and CRC_{ADN}^{Ctrl} are the number of cancers expected if no adenomas had been removed in the control arm.

We estimated the amount of overdiagnosis as number of overdiagnosed cancers divided by the number of observed cancers in the control arm. For transparency, we also calculated overdiagnosis using the observed and expected number of CRCs in the screening arm as an alternative denominator. In a non-randomised setting, without

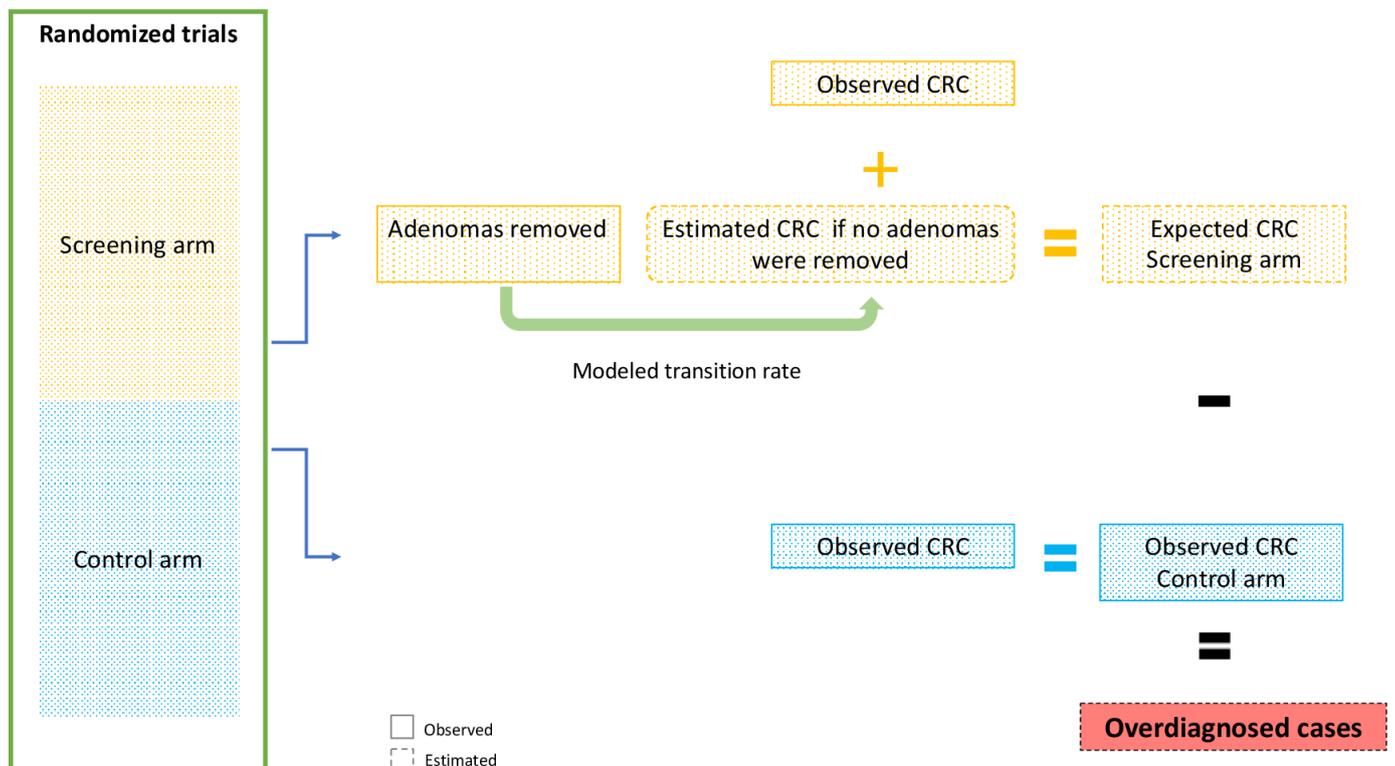


Figure 2 Illustrating estimates of overdiagnosis. CRC, colorectal cancer.

Box 1 Glossary of terms for overdiagnosis in preventive screening

Overdiagnosis: diagnosis of a disease that would not have developed to cause symptoms or death in the patient's lifetime if not detected by screening.

Number of overdiagnosed cancers: number of cancers expected in the screening arm subtracting number of cancers observed (diagnosed) in the control arm.

Number of cancers expected in the screening arm: the sum of the numbers of cancers observed (diagnosed) in the screening arm, and the numbers of cancers that would develop from precursor lesions if they were not removed at screening (estimated through microsimulation models).

Overdiagnosis: number of overdiagnosed cancers divided by the number of observed (diagnosed) cancers in the control arm.

a valid control group, number of cancers observed in the presence of screening may be the only number available and it is sometimes used as the denominator. However, for preventive screening methods (reducing cancer incidence) such as in colorectal screening, this may result in falsely high overdiagnosis estimates.¹⁷

RESULTS

Trial data

Table 1 shows the baseline characteristics of the randomised trials on FOBT and sigmoidoscopy screening, including detected cancers and adenomas, and cancer incidence in the follow-up period.

Table 1 Baseline characteristics of the randomised trials on FOBT and sigmoidoscopy screening, respectively^{6 7}

	Screening arm	Control arm
FOBT trial		
Number of patients included*	76 056	75 919
Person-years of follow-up	1 286 526	1 286 877
Attendance rate (to at least one screening round)	44 838 (59.9%)	NA
Positive test	2050 (2.7%)	NA
Number of individuals with at least one colonoscopy after positive test	1778 (2.3%)	NA
Number of cancers detected at screening	236	
CRC incidence	2279 (3.0%)	2354 (3.1%)
Total number of individuals with adenomas detected at screening		
Size <10 mm in diameter	159 (0.2%)	
Size ≥10 mm in diameter	615 (0.8%)	
Total number of individuals with adenomas detected (non-responders and interval adenomas)		
Size <10 mm in diameter	629 (0.8%)	601 (0.8%)
Size ≥10 mm in diameter	736 (1.0%)	883 (1.2%)
Sigmoidoscopy trial†		
Number of patients included	20 552	78 126
Person-years of follow-up	289 272	1 084 774
Attendance rate	12 955 (63.0%)	
Positive test	2639 (12.8%)	
Number of individuals with at least one colonoscopy after positive test	2520 (12.3%)	
Number of cancers detected at screening	41	
CRC incidence	393 (1.9%)	1751 (2.2%)
Total number of individuals with adenomas detected at screening‡		
Non-advanced adenoma	1628 (7.8%)	
Advanced adenoma	582 (2.8%)	

*875 subjects were excluded could not be traced by the Office for National Statistics or had emigrated and were excluded from the report.⁶

†Norwegian Colorectal Cancer Screening Trial randomised individuals aged 50–54 in 1:5.4 ratio and individuals 55–64 in 1:3 ratio in the screening and control arm, respectively.

‡Available only for screening participants. Size of adenoma refers to a patient's largest adenoma at an investigation. Advanced adenoma was defined as adenoma ≥10 mm in diameter or with villous histology or with high-grade dysplasia.

CRC, colorectal cancer; FOBT, faecal occult blood testing.

**Table 2** Transition time in years, and rates as annual percentages in different phases of the natural history of colorectal cancer for the different model scenarios^{11 12}

Name of the model	MISCAN	CRC-SPIN	SimCRC	German low	German high
<i>Phase I: from adenoma to preclinical cancer</i>					
Annual transition from non-advanced adenoma to advanced adenoma (%)				3.6	4.2
Annual transition from advanced adenoma to preclinical cancer (%)				2.5	5.6
Median dwell time (years)	6	23	19		
Annual transition from adenoma to preclinical cancer (%)*	10.9	3.0	3.6		
<i>Phase II: from preclinical cancer to cancer</i>					
Median of dwell time (years)	2	2	3		
Annual transition from preclinical cancer to cancer (%)*	29.3	29.3	20.6	17.3	22.5
<i>Total: from adenoma to cancer</i>					
Annual transition from adenoma to cancer (%)	0.72†	1.80	1.81	0.13‡	0.36‡

*Calculated based on reported median dwell time from adenoma to preclinical cancer and from preclinical cancer to cancer for discrete time Markov models with 1-year intervals.

†Assuming 14% progressive adenomas.

‡Assuming 100% non-advanced adenomas.

CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history; MISCAN, Microsimulation Screening Analysis; SimCRC, Simulation Model of Colorectal Cancer.

Model data

The estimated annual transition rates from adenoma to cancer which are used in the five scenarios (four models) are displayed in [table 2](#). The highest annual transition rates from adenoma to cancer were found in the CRC-SPIN and the SimCRC models, and the lowest in the German model (German low) ([table 2](#)). The annual transition rate from adenoma to preclinical cancer ranged from 3.0% in the CRC-SPIN model to 10.9% for the progressive adenomas in the MISCAN model (not all adenomas are progressive in the MISCAN model). The transition rate from preclinical to clinical cancer ranged from 17.3% in the German model (German, low) to 29.3% in the MISCAN and CRC-SPIN models. In the MISCAN model, not all adenomas progress and the estimates are for the progressive adenomas. Accordingly, if one follows 100 adenoma patients for 10 years and assume median reported dwell times, between 1 (German, low) and 18 (CRC-SPIN and SimCRC) cancers would develop according to the different models. Annual transition rates for sensitivity analyses can be found in online supplemental table 1.

Overdiagnosis

For FOBT screening arm, the number of prevented cancers in the different models ranged from 72 to 254.

Four out of the five model scenarios predicted overdiagnosis ranging from 2.0% (2400 cancers expected in screening arm vs 2354 cancers observed in control arm) to 7.6% (2533 cancers expected in screening arm). The MISCAN model did not predict any overdiagnosis, [table 3](#). For sigmoidoscopy screening arm, the number of prevented cancers in the different models ranged from 173 to 638. All models predicted overdiagnosis, ranging from 25.2% (566 cancers expected in screening arm vs 452 observed in control arm) to 128.1% (1031 cancers expected in screening arm), [table 3](#). Similar results were obtained in the sensitivity analyses, see online supplemental tables 2 and 3. A comparison of cancers observed and cancers expected in screening and control arms can be found in [figure 3](#).

The overdiagnosis was highest when cancers observed in the screening arm were used as denominator (up to 7.9% for FOBT and 147.3% in sigmoidoscopy) and lowest when cancers expected in the screening arm was used as a denominator (up to 7.1% for FOBT and 56.2% for sigmoidoscopy), see online supplemental table 4.

In the secondary analysis for FOBT screening (when we included estimates of cancers prevented through adenoma removal outside of screening in the screening

Table 3 Number of observed and prevented CRCs and CRC overdiagnosis for the five different model scenarios for FOBT screening and sigmoidoscopy screening, respectively

	Screening arm				Control arm				Overdiagnosis (%)
	Prevented CRC*	Observed CRC	Expected CRC	Prevented CRC*	Observed CRC	Expected CRC	Overdiagnosed CRC*		
FOBT	MISCAN	72	2279	2351	90	2354	-3	-0.1	
	CRC-SPIN	200	2279	2479	138	2354	125	5.3	
	SimCRC	208	2279	2487	138	2354	133	5.6	
	German, low	121	2279	2400	58	2354	46	2.0	
	German, high	254	2279	2533	137	2354	179	7.6	
	Sigmoidoscopy†	207	393	600		452	148	32.7	
Sigmoidoscopy†	CRC-SPIN	610	393	1003		452	551	121.9	
	SimCRC	638	393	1031		452	579	128.1	
	German, low	173	393	566		452	114	25.2	
	German, high	393	393	786		452	334	73.9	

Overdiagnosis is calculated as the number of overdiagnosed CRC divided by the number of observed CRC in the control arm.

*Primary analysis included only adenomas in the screening attendees.

†Number of observed CRC in the control arm was standardised with the screening arm.

CRC, colorectal cancer; CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history; FOBT, faecal occult blood testing; MISCAN, Microsimulation Screening Analysis; SimCRC, Simulation Model of Colorectal Cancer.

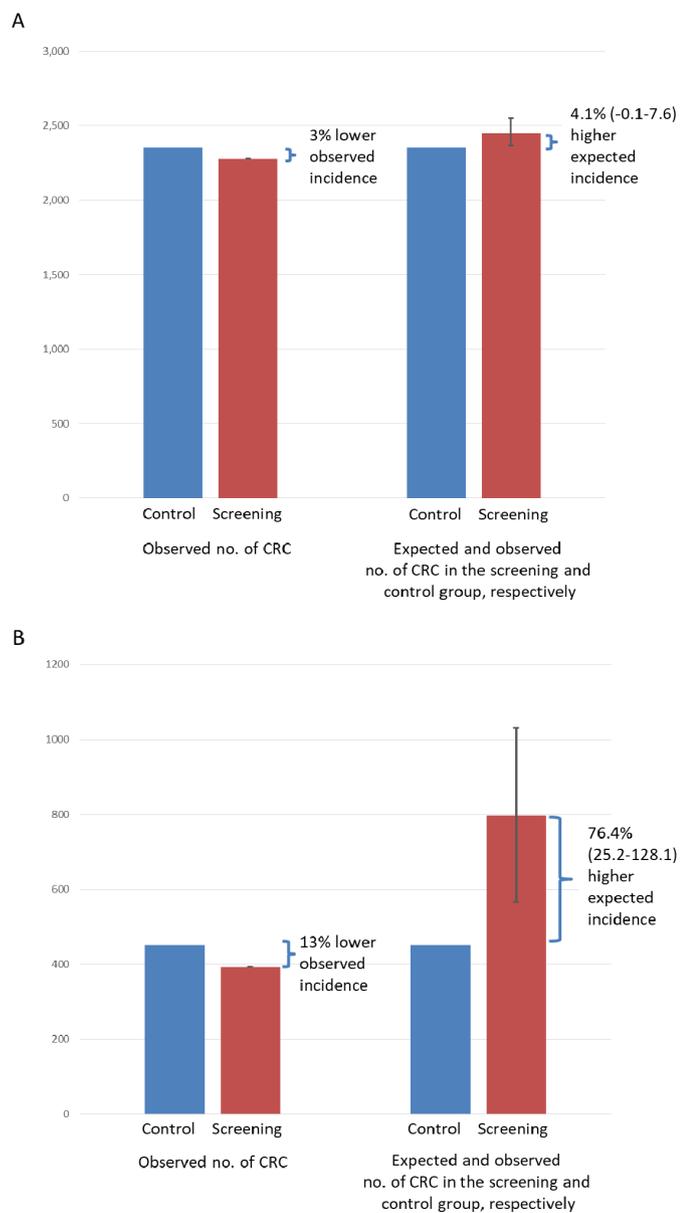


Figure 3 Numbers of observed and expected colorectal cancers (CRCs) in the screening arm and observed in the control arm for faecal occult blood testing (A) and sigmoidoscopy screening (B). Bar for expected number of cancers in screening arm is a mean over all microsimulation models. Error bar for expected number of cancers in screening arm is a range (min–max) over all microsimulation models.

arm and in the control arm), <1% overdiagnosis was estimated with the MISCAN model, whereas for the remaining four model scenarios, the estimated overdiagnosis ranged from 3.1% to 10.2%.

DISCUSSION

This study is the first to estimate overdiagnosis in colorectal cancer screening using randomised trials with valid comparison groups and sufficiently long follow-up. We found that overdiagnosis with FOBT screening is between 0% and 7.6% depending on adenoma transition

rates. The highest amount of overdiagnosis was observed for German model (German high), with 179 overdiagnosed cancers. For comparison only 236 cancers were detected at screening.

For sigmoidoscopy screening, our analyses yielded overdiagnosis between 25.2% and 128.1%. Two of the five models predicted overdiagnosis over 100%, which means that majority of cancers detected in the screening arm are overdiagnosed. The apparent difference in overdiagnosis between FOBT and sigmoidoscopy may be explained by different characteristics of adenomas detected at screening (less advanced in sigmoidoscopy and more advanced in FOBT). However, even if we use the lowest reported transition rates for adenomas removed in FOBT screening and highest reported for adenomas removed in sigmoidoscopy screening, the results are hard to believe clinically. Although, similar overdiagnosis (50%–70%) was estimated with MISCAN model for cervical cancer screening (which same as sigmoidoscopy is mostly preventive),¹⁸ there is no clinical justification for observed results.

Overdiagnosis was highly dependent on the model used, and thus the transition rates from adenoma to cancer. The choice of denominator or whether we included only adenomas detected at screening (primary analysis) or all adenomas (secondary analysis) did not change the results significantly. The CRC-SPIN and SimCRC models had the highest proportion of adenomas progressing to cancer within 10 years from adenoma detection (18%) for the German model (with the lowest annual transition rates, ie, German, low) it was 1.3%.

A reasonable explanation for the calculated overdiagnosis is that adenoma growth rates in at least some of the models are wrong. Our study may indicate that microsimulation models do not provide reliable estimates for the natural history of adenomas. As a consequence, because adenoma development is a central assumption in CRC screening microsimulation models, we are concerned that the model estimates for the effect of preventive CRC screening tests, such as sigmoidoscopy or colonoscopy screening, may be wrong. Since the estimates for flexible endoscopy screening were much more unlikely than for FOBT screening it may be that the models have specific problems with screening tests where multiple non-advanced adenomas are detected. It could be interpreted that either the proportion of non-advanced adenomas destined to progress to cancer or the time to progress to cancer is overestimated in these models.

Strengths of our study include the application of data from large randomised trials with long follow-up and detailed information of adenomas and cancers. Although some observational studies with long follow-up were also available^{19 20} we decided not to use them due to lack of appropriate control groups. Weaknesses include the inability to directly observe transition times of adenomas and cancers, which this study shares with all others in this area. Modelling studies of the natural history of CRC show a variation of mean transition time from adenoma

to preclinical cancer between 7.6 and 24.2 years.¹¹ The annual transition rate from advanced adenoma to CRC is estimated to be 2.5%–5.6% dependent on age and sex, and from adenoma to advanced adenoma from 3.6% to 4.2%.¹² Thus, differences in transition rates do impact on the results obtained. Our assumption on adenoma growth relied on the median transition rate from the microsimulation models¹¹ and the lowest and highest point estimate for annual transition rates from the German model.¹² The advantage of this assumption is that it is transparent and easy to understand. In addition to assumption on adenoma growth rate, the microsimulation models include several other assumptions such as participation rates, performance and quality of the screening intervention. We did not make assumptions on any of these variables, and some may considered this a limitation of our study. We were able to use the observed data from the published trials where these variables are observed in a real life scenario and do not have to be assumed. Since there is no randomised trials on faecal immunochemical test, used as benchmark test for CRC in the UK,²¹ we used data from the randomised trial of FOBT screening. We believe that our results can be generalisable for faecal immunochemical testing depending on the threshold for the test positivity which can be chosen with FIT, but not with FOBT screening.

Overdiagnosis in preventive screening is difficult to ascertain, because the early detection effect which may harbour the risk of overdiagnosis is counteracted by the preventive effect, which reduces cancer incidence. Disentangling the two effects is necessary to ascertain the net overdiagnosis. As all approaches used for CRC screening include colonoscopy which has a cancer prevention effect due to adenoma removal, it is invalid to assume that there is no overdiagnosis in CRC screening simply because the cumulative incidence of CRC is similar or lower in the screening arm compared with the control arm at the end of follow-up. In the present study, we achieved this by using information about removed adenomas in the randomised trials and by applying modelled transition rates from adenoma to cancer from the most commonly used microsimulation models. This approach should provide a good and valid approximation of the true amount of overdiagnosis in CRC screening, but as our results show, it did not result in plausible estimates for sigmoidoscopy and FOBT screening.

Microsimulation modelling has been used extensively to guide policy making about CRC screening. The models we used for this study are regarded as valid and trustworthy. Important guidelines which are followed by physicians around the world and their patients rely on these models for their recommendations for or against screening. The US Preventive Services Task Force is heavily relying on microsimulation modelling,⁸ and many other organisations do so increasingly.^{9,10} However, based on our results, the four presented microsimulation models may not give reliable estimates of preventive effect of screening and their results should be used with caution.

Author affiliations

¹Clinical Effectiveness Research Group, Institute of Health and Society, University of Oslo, Oslo, Norway

²Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, Warsaw, Poland

³Department of Cancer Prevention and Department of Oncological Gastroenterology, The Maria Skłodowska-Curie National Research Institute of Oncology, Warsaw, Poland

⁴Clinical Effectiveness Research Group, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway

Contributors PW: study design, data analyses, first draft of the paper. MFK: study design, data interpretation, drafting the paper. ML: study design, data interpretation, drafting the paper. MBugajski: study design, data interpretation, drafting the paper. MBretthauer: study design, data interpretation, drafting the paper. MK: study design, data analyses, data interpretation, drafting the paper.

Funding This work was supported by Fulbright scholarship (Wieszczyni P) and research grants from Norwegian Research Council and Norwegian Cancer Society (Bretthauer M, Kalager M, Loberg M). Award/Grant number is not applicable to any of the fundings.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethics approval was not required for this study. The authors affirm that the manuscript is an honest, accurate and transparent account of the study. The Nottingham trial was approved by the Nottingham Family Practitioner Committee and BMA Medical Ethics committee. The NORCCAP trial was approved by the Ethics Committee of South East Norway and the Norwegian Data Protection Authority.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data and computing code available on request to the corresponding author.

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 iD

Paulina Wieszczyni <http://orcid.org/0000-0001-7930-1305>

REFERENCES

- Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015;350:g7773.
- Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605–13.
- Kalager M, Wieszczyni P, Lansdorp-Vogelaar I, et al. Overdiagnosis in colorectal cancer screening: time to acknowledge a blind spot. *Gastroenterology* 2018;155:592–5.
- Brenner H, Altenhofen L, Stock C, et al. Prevention, early detection, and overdiagnosis of colorectal cancer within 10 years of screening colonoscopy in Germany. *Clin Gastroenterol Hepatol* 2015;13:717–23.
- Luo D, Cambon AC, Wu D. Evaluating the long-term effect of FOBT in colorectal cancer screening. *Cancer Epidemiol* 2012;36:e54–60.
- 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.



- 7 Holme Øyvind, Løberg M, Kalager M, *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;168:775–82.
- 8 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.
- 9 Peterse EFP, Meester RGS, Siegel RL, *et al.* The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: Microsimulation analysis I to inform the American cancer Society colorectal cancer screening guideline. *Cancer* 2018;124:2964–73.
- 10 Wolf AMD, Fontham ETH, Church TR, *et al.* Colorectal cancer screening for average-risk adults: 2018 guideline update from the American cancer Society. *CA Cancer J Clin* 2018;68:250–81.
- 11 Kuntz KM, Lansdorf-Vogelaar I, Rutter CM, *et al.* A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Med Decis Making* 2011;31:530–9.
- 12 Brenner H, Altenhofen L, Stock C, *et al.* Natural history of colorectal adenomas: birth cohort analysis among 3.6 million participants of screening colonoscopy. *Cancer Epidemiol Biomarkers Prev* 2013;22:1043–51.
- 13 Hardcastle JD, Chamberlain JO, Robinson MH, *et al.* Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;348:1472–7.
- 14 Holme Øyvind, Løberg M, Kalager M, *et al.* Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA* 2014;312:606–15.
- 15 Loeve F, Boer R, Zauber AG, *et al.* National Polyp study data: evidence for regression of adenomas. *Int J Cancer* 2004;111:633–9.
- 16 Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiol Biomarkers Prev* 2010;19:1992–2002.
- 17 Ripping TM, Verbeek AL, Broeders MJ. Overdiagnosis in cancer screening: the need for a standardized denominator. *J Med Screen* 2016;23:111–3.
- 18 van Luijt PA, Rozemeijer K, Naber SK, *et al.* The role of pre-invasive disease in overdiagnosis: a microsimulation study comparing mass screening for breast cancer and cervical cancer. *J Med Screen* 2016;23:210–6.
- 19 Atkin W, Wooldrage K, Parkin DM, *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;389:1299–311.
- 20 Wieszczy P, Kaminski MF, Franczyk R, *et al.* Colorectal cancer incidence and mortality after removal of adenomas during screening Colonoscopies. *Gastroenterology* 2020;158:875–83.
- 21 NHS bowel cancer screening (BCSP) programme. Public health England bowel cancer screening: programme overview [11/30/2020], 2015. Available: <https://www.gov.uk/guidance/bowel-cancer-screening-programme-overview>

Supplementary Appendix

Content

1. Assumption of distribution of adenomas in FOBT trial

1.1. Screening group

1.1.1. First round

1.1.2. Subsequent screening rounds

1.1.3. Non-participants

1.2 Control group

2. Population

Supplementary Table 1. Annual transition rates from adenoma to colorectal cancer (sensitivity analysis).

Supplementary Table 2. Number of observed and prevented cancers (sensitivity analysis).

Supplementary Table 3. Number of overdiagnosed cancers and amount of overdiagnosis (in percent) in the different models (sensitivity analysis).

Supplementary Table 4. Number of overdiagnosed cancers and amount overdiagnosis (in percent) in the different models.

1. Assumption of distribution of adenomas in FOBT trial

1.1 Screening group

1.1.1. First round

In the FOBT trial report (6) the adenomas removed in patients were reported in the following groups: at first round of screening; at subsequent screening; in individuals who were randomized to screening but did not participate (non-participants); adenomas detected in the interval between two screening rounds either within 2 years or after 2 years; adenomas detected during follow-up; in the control group. To take account of length of follow-up time in our analyses, we assumed that all adenomas detected in the first screening round were followed for 19 years.

1.1.2. Subsequent screening rounds

Since biennial screening by FOBT was offered, we assumed that maximum follow-up time for subsequent screening is 17 years (first 2 years we assigned as first round) and subsequent screening took 8 years (total enrollment period was 10 years). We also assumed that constant number of adenomas was removed each year. Thus that out of adenomas removed at subsequent screening, 1/8 of adenomas had 17 years of follow-up, 1/8 of adenomas had 16 years of follow-up, ect. Detailed assumptions on the follow-up time for the rest of the sub-groups are described below.

1.1.3. Non-participants

We assumed that more adenomas occurred later in the follow-up, as the risk of adenomas increases with age. We used exponential distribution with $\lambda=0.12$ to calculate probability (p) of detecting the adenoma $p_t = \lambda \exp(-\lambda \cdot (19 - t))$, where t is time $t=1,2,\dots,19$. The parameter

λ was chosen, so that approximately 33% of adenomas were detected during the first 11 years of follow-up and 66% of adenomas were detected in the last 8 years of follow-up.

1.2 Control group

We assumed that the detection rate of adenomas in the control group was higher in the late period of follow-up, i.e. that 33% of adenomas were detected during first 11 years of follow-up and 66% of adenomas were detected during last 8 years of follow-up. Exponential distribution with $\lambda=0.12$ was used.

2. Population

Our analyses were based on a gender distribution of 60% males and 40% females (13). We based our models on the UK population from 1981 to 2009. All models were adjusted for mortality rates based on life tables (available at www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/pastandprojecteddatafromtheperiodandcohortlifetables/2012baseduk19812062, access on Mar 5, 2018). On average, individuals with adenomas were 60 years old at the beginning of the study period and 78 years old at the end of follow-up. Since the individuals were recruited between 1981 and 1991, we calculated an average mortality in those years for the 1st year of follow-up, an average for years 1982 to 1992 for the 2nd year of follow-up, etc.

Supplementary Table 1. Annual transition rates from adenoma to colorectal cancer (sensitivity analysis).

Name of the model	MISCAN	MISCAN	CRC-SPIN
Model description	Exponential distribution of dwell time with mean of 140**	Exponential distribution of dwell time with mean of 140***	Using estimates from (16)
Annual transition from adenoma to cancer (%)*	0.69 (0.26)	0.71	0.07

This table is a continuation of Table 2 in the main paper. The rates presented here comes from the data provided by the models' authors after contacting them. *Number of cancers that would be predicted to develop from 100 patients with at least one adenoma followed for 10 years. **We drawn 100 samples (corresponds to 100 patients) from exponential distribution with mean of 140 (time from adenoma to cancer) and count how many of them had a value ≤ 10 (corresponds to the 10 years of follow-up). Such experiment was repeated 1,000 times and mean (standard deviation) was calculated over the observations. ***We drawn a sample of 100,000 from exponential distribution with mean of 140 years and calculated a median over the observations. Annual transition rate was calculated using discrete time Markov model with one year interval.

Supplementary Table 2. Number of observed and prevented cancers (sensitivity analysis).

		Screening group			Control group		Primary analysis	Secondary analysis
		Prevented CRC among attenders	Prevented CRC among non-attenders and by interval adenomas	Observed CRC	Prevented CRC	Observed CRC	Overdiagnosed cancers*	Overdiagnosed cancers**
FOBT	MISCAN ^a	71 (8)	83 (9)	2,279	60 (7)	2,354	-4	19
	MISCAN ^b	70	83	2,279	61	2,354	-5	17
	CRC-SPIN ^c	92	73	2,279	50	2,354	17	40
Sigmoidoscopy [§]	MISCAN ^a	214 (14)		393	452		155	
	MISCAN ^b	212		393	452		153	
	CRC-SPIN ^c	135		393	452		76	

*Primary analysis (overdiagnosis include only adenomas in the screening arm). **Secondary analysis (overdiagnosis include all adenomas in the screening and control arms; only FOBT). [§]Number of observed CRC in control group was recalculated to adjust for unbalanced sample size of screening and control group. ^aExponential distribution of dwell time with mean of 140 (mean and standard deviation over 1,000 replications). ^bExponential distribution of dwell time with mean of 140 (using median time from adenoma to cancer). ^cUsing estimates from (16).

Supplementary Table 3. Number of overdiagnosed cancers and the amount of overdiagnosis (in percent) in the different models (sensitivity analysis).

		Primary analysis				Secondary analysis				
		Number of overdiagnosed cancers	Rate of overdiagnosis			Number of overdiagnosed cancers	Overdiagnosis			
			% of CRC observed in the control group	% of CRC observed in the screening group	% of CRC expected in the screening group		% of CRC observed in the control group	% of CRC expected in the control group	% of CRC observed in the screening group	% of CRC expected in the screening group
FOBT	MISCAN ^a	-4	-0.2	-0.2	-0.2	19	0.8	0.8	0.8	0.8
	MISCAN ^b	-5	-0.2	-0.2	-0.2	17	0.7	0.7	0.7	0.7
	CRC-SPIN ^c	17	0.7	0.7	0.7	40	1.7	1.7	1.8	1.7
Sigmoidoscopy [§]	MISCAN ^a	155	34.3	39.4	25.5					
	MISCAN ^b	153	33.8	38.9	25.3					
	CRC-SPIN ^c	76	16.8	19.3	14.4					

*Primary analysis (overdiagnosis include only adenomas in the screening arm). **Secondary analysis (overdiagnosis include all adenomas in the screening and control arms; only FOBT). [§]Number of observed CRC in control group was recalculated to adjust for unbalanced sample size of screening and control group. ^aExponential distribution of dwell time with mean of 140 (mean over 1,000 replications). ^bExponential distribution of dwell time with mean of 140 (using median time from adenoma to cancer). ^cUsing estimates from (16).

Supplementary Table 4. Number of overdiagnosed cancers and the amount of overdiagnosis (in percent) in the different models

		Primary analysis				Secondary analysis							
		Number of overdiagnosed cancers	Overdiagnosis			Prevented CRC among non-attenders and by interval adenomas	Prevented CRC in control group	Number of overdiagnosed cancers	Overdiagnosis				
			% of CRC observed in the control group	% of CRC observed in the screening group	% of CRC expected in the screening group				% of CRC observed in the control group	% of CRC expected in the control group	% of CRC observed in the screening group	% of CRC expected in the screening group	
FOBT	MISCAN	-3	-0.1	-0.1	-0.1	113	90	20	0.8	0.8	0.9	0.9	
	CRC-SPIN	125	5.3	5.5	5.0	215	138	202	8.6	8.1	8.9	8.1	
	SimCRC	133	5.6	5.8	5.3	217	138	212	9.0	8.5	9.3	8.5	
	German, low	46	2	2	1.9	86	58	74	3.1	3.1	3.2	3.1	
	German, high	179	7.6	7.9	7.1	198	137	240	10.2	9.6	10.5	9.5	

Sigmoidoscopy [§]	MISCAN	148	32.7	37.7	24.7
	CRC-SPIN	551	121.9	140.2	54.9
	SimCRC	579	128.1	147.3	56.2
	German, low	114	25.2	29	20.1
	German, high	334	73.9	85	42.5

*Primary analysis (overdiagnosis include only adenomas in the screening arm). **Secondary analysis (overdiagnosis include all adenomas in the screening and control arms; only FOBT). [§]Number of observed CRC in control group was recalculated to adjust for unbalanced sample size of screening and control group.

Abbreviations: CRC: colorectal cancer