Effects of Screening for Colorectal Cancer: Development and Validation of a Multistate Markov Model

Short title: Colorectal Cancer Markov Model

Authors:

Thomas Heisser[1,2] (Doctoral Researcher), Michael Hoffmeister[1] (Associate Professor of Epidemiology), Hermann Brenner[1,3,4] (Professor of Epidemiology)

Affiliations:

1 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

2 Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany.

3 Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.

4 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

Correspondence:

Thomas Heisser, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120 Heidelberg, Germany; E-mail: t.heisser@dkfz.de; phone: +49 (0)6221 42-1335

Word count:

Abstract: 250; Main text: 3,968

Keywords:

colorectal cancer; screening; modelling; validation

Acknowledgements:

The authors thank all individuals who participated in the KolosSal study, all participating gastroenterologists for the recruitment of study participants, and all staff involved in study coordination, documentation and data preparation.

Declarations

Funding:
Financial support for this study was provided in part by a grant by the German Federal Ministry of Education and Research (grant number 01GL1712). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. TH was supported by the Helmholtz International Graduate School for Cancer Research at the German Cancer Research Centre (DKFZ).

Conflict of interest:
The authors declare that they have no conflict of interest.

Ethics approval:
The KoloSsal study was approved by ethics committees of the University of Heidelberg (057/2005) and of the Medical Association of Saarland (54/05).

Consent to participate and for publication:
Informed consent was obtained from each participant.

Availability of data, code and material:
All analyses relevant to the study are included in the article or uploaded as supplementary information. The R code defining the core model is uploaded as supplementary information. All data requests regarding the KolosSal study should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

Authors' contributions:
HB and TH designed this study and developed the methodology. HB designed, initiated and led the KolosSal study. MH and HB contributed to the acquisition of data. TH conducted the statistical analyses and drafted the manuscript. All authors critically reviewed the manuscript, contributed to its revision, and approved the final version submitted. The researchers are independent from funders. All authors had full access to all of the data (including statistical reports and tables) used for the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Abstract

Simulation models are a powerful tool to overcome gaps of evidence needed to inform medical decision making. Here, we present development and application of a multistate Markov model to simulate effects of colorectal cancer (CRC) screening, along with a thorough assessment of the model's ability to reproduce real-life outcomes. Firstly, we provide a comprehensive documentation of the model development, structure and assumptions. Secondly, to assess the model's external validity, we compared model-derived cumulative incidence and prevalences of colorectal neoplasms to (1) results from KolosSal, a study in German screening colonoscopy participants, (2) registry-based estimates of CRC incidence in Germany, and (3) outcome patterns of randomized sigmoidoscopy screening studies. We found that (1) more than 90% of model-predicted neoplasm prevalences were within the 95% confidence intervals of the observed prevalences in the KolosSal study; (2) the 15-year cumulative CRC incidences estimated by simulations for the German population deviated by 0.0-0.2 percent units in men and 0.0-0.3 percent units in women when compared to corresponding registry-derived estimates; and (3) the time course of cumulative CRC incidence and mortality in the modelled intervention group and control group closely resembles the time course reported from sigmoidoscopy screening trials. Summarized, our model adequately predicted colorectal neoplasm prevalences and incidences in a German population for up to 25 years, with estimated patterns of the effect of screening colonoscopy resembling those seen in registry data and real-world studies. This suggests that the model represents a valid tool to assess the comparative effectiveness of strategies for CRC screening.

Introduction

Colorectal cancer (CRC) is one of few cancers for which effective screening options are established. While population-based CRC screening programs are already in place or being rolled out in numerous countries worldwide [1], health authorities remain concerned with identifying screening strategies which optimally balance health outcomes against resource use and test burden. This, however, is a challenging endeavor, as the lack of risk factors clearly elevating the CRC risk above the population average makes it difficult to identify high-risk subgroups for which screening would be particularly beneficial [2,3]. In addition, screening modalities vary largely in diagnostic performance, uptake, test burden, and costs. Combined, these factors allow for a multitude of possible screening strategies, only some of which have by now been assessed in randomized or observational studies, with inevitably limited scope of assessed strategies and follow-up periods [4–13].

Simulation models are a powerful tool to overcome these evidence gaps required to inform medical decision-making. However, as models are at least as imperfect as the real-world data used for their development, a model's success depends on the levels of trust decision-makers will have in its results [14]. To enhance a model's credibility, firstly, transparency regarding the model's structure, underlying parameters and assumptions is crucial. Secondly, a model should be subjected to a validation process by measuring modelled outcomes against observed real-world data the model ought to reproduce. Such external validation is regarded a particular robust form of validation and thus an indispensable step in model development [14].

In several previous analysis, we step-by-step developed a multistate Markov model to simulate effects of CRC screening in a hypothetical population [15–18]. The objective of the present study was, firstly, to provide a thorough and transparent documentation of the model development, structure and assumptions. Secondly, we sought to comprehensively address the model's ability to reproduce real-world outcomes by a three-way-validation approach: we compared model-derived outcomes to (1) results from a German colonoscopy screening study (KolosSal study), (2) estimates based on German registry data, and (3) outcome patterns reported from randomized controlled trials (RCTs) on the effects of screening sigmoidoscopy.

Methods

Model Documentation

An exhaustive documentation on the model's structure and data sources used for its development is given in **Supplementary Appendix 1a,** including overviews on all model parameters (**Appendix 1a, Tables 1-3**). Briefly, our multistate Markov model simulates the natural history of CRC in a hypothetical population for a predefined number of years. Screening can interfere with the natural history of CRC (**Figure 1**).

The analyses presented in this study are based on simulations for a hypothetical previously unscreened German population, in line with the used input parameters derived from Germany-centred data sources. The model can principally be used for simulating any population, provided updated or appropriately adjusted input parameters. The model was developed in the open-source statistical software R (version 3.6.3). The code defining the core model is available in **Supplementary Appendix 1b**, along with conceptual considerations.

Model Validation

To assess the model's ability to reproduce real-world outcomes, firstly, we used our model to simulate prevalences of adenomas and cancers detected at screening colonoscopy by age group and compared these to observed prevalences in the KolosSal study, a statewide cohort study in the German state of Saarland. The objective was to assess the agreement between the modelled and the observed prevalences.

Secondly, we simulated populations with levels of annual colonoscopy use representative for the German general population and compared the resulting cumulative incidences to those derived from German registry data before and after the introduction of a population-wide colonoscopy screening offer in the year 2002. The objective was to assess the agreement between the modelled and observed cumulative incidences over time.

Thirdly, we modelled exemplary populations with and without baseline colonoscopy screening and assessed the CRC incidence and mortality over time. The objective was to assess the concordance of the resulting patterns when compared to the results of the screening sigmoidoscopy RCTs (the Italian Screening for COlon REctum (SCORE) trial [4], the Norwegian Colorectal Cancer Prevention (NORCCAP) trial [8,19], the UK Flexible Sigmoidoscopy Screening (UKFSS) trial [7], and the US Prostate, Lung, Colorectal and Ovarian (PLCO) screening trial [9]).

Data Sources

KolosSal Study

Details on the KolosSal study have been described previously [20–24]. Briefly, the KolosSal study recruited patients undergoing screening colonoscopy with the objective to monitor long-term CRC incidence and mortality. Residents of the federal state of Saarland aged 55 or older undergoing screening colonoscopy in a participating practice were eligible for enrolment. The study was approved by ethics committees of the University of Heidelberg and of the State Medical Association of Saarland. Informed consent was obtained from each patient.

This analysis is based on data of participants recruited between 2005 and 2013 in 33 gastroenterological practices in Saarland certified to perform screening colonoscopy. Patients were classified according to the most advanced of the following findings: CRC, advanced adenoma – defined as presence of at least 1 adenoma with at least 1 of the following features: >1cm in size, tubulovillous or villous components, high-grade dysplasia – other adenoma, hyperplastic or unspecified polyp, none of these [25,26].

German Registry Data

Age-group- and sex-specific incidence rates for the German general population were obtained from the interactive database of the German Centre for Cancer Registry Data [27], which provides combined and quality-checked data from epidemiological cancer registries in Germany. We extracted the CRC incidence rates for the years 2001, 2006, 2011, and 2016, i.e. one year before the introduction of the population-based colonoscopy screening offer in Germany in 2002, and then in steps of five years to allow for detecting effects of the successive uptake of the colonoscopy screening offer (**Supplementary Table 1**) [28].

Statistical Analysis

Part 1: Equivalence to KolosSal Study Prevalences

Of the KolosSal study population, we selected subjects aged 55-75 who had their first screening colonoscopy. We excluded patients with history of inflammatory bowel disease and patients where no information could be obtained on whether a previous colonoscopy had been conducted. We then assessed the detected prevalences of any neoplasm (ANN, any adenomas and cancers) and of any advanced neoplasm (ADN, advanced adenomas and cancer) at screening colonoscopy according to age at screening in steps of one year from age 55-75, stratified by sex, and calculated the 95% confidence intervals (CI).

We then modelled, for a hypothetical population of each 100,000 previously unscreened men and women, disease progression without screening from the states at ages 50-75. We assessed the prevalences of ANN and ADN which would have been detected if screening colonoscopy had been applied once at each individual age from ages 55-75.

Subsequently, we calculated the prevalence ratios (PRR) by dividing the modelled prevalence estimates by the prevalences observed in KolosSal, both age-specific and as meta-analyses combining all sex-specific PRRs, along with the respective CIs. Statistical equivalence between the modelled and observed prevalences was then tested by applying two one-sided t-tests (TOST) [29] to each of the sex-specific meta-analyses estimates. We chose a 20% margin of equivalence. A maximum deviation of 20% was considered acceptable as the model is not primarily intended to meticulously forecast absolute case numbers with or without screening but rather to assess the comparative performance of CRC screening strategies in terms of their contribution to the relative reduction of the CRC burden [14].

Part 2: Comparison to Registry-derived Cumulative Incidences

For each of the extracted years (2001, 2006, 2011, 2016) from the database of the German Centre for Cancer Registry Data, we estimated the 15-year cumulative CRC incidence for men and women of the German general population aged 50-64, 55-69 and 60-74. As the incidence rates as provided by the registry are computed for age intervals of five years, the 15-year cumulative rate is calculated as sum of five times the age-group-specific interval rate for the respective age-groups, assuming the age-specific rates remain constant over time. The cumulative risk in the absence of competing causes of death can then be approximated by the cumulative rate [30,31].

For comparison, we derived corresponding estimates for modelled populations of each 100,000 German men and women with levels of 2% and 6% annual colonoscopy use. These utilization levels were chosen as the estimated use of colonoscopy for any reason within 10 years in Germany was 24.3% and 63.9% for the periods 2000-2002 and 2008-2010, respectively, reflecting the increase triggered by the introduction of the colonoscopy screening offer in 2002 [28].

As the model starting prevalences were obtained from a population of first-time screening colonoscopy users [17], i.e. a selective population without or only negative prior fecal testing or (diagnostic) colonoscopy, we discounted adenoma and preclinical cancer prevalences at model start by 30%. This was considered reasonable in order to increase comparability to the German general population given that, already in the period 2000-2002, approximately 66%

of 55-59 year-olds had either fecal testing within two years or colonoscopy within ten years [28].

Part 3: Comparison to RCT-reported CRC Incidence and Mortality Patterns

Finally, we modelled each 100,000 50- and 60-year old men and women with and without screening colonoscopy at baseline with 16 years of follow-up, i.e. the approximate median follow-up length of by now reported results from screening-sigmoidoscopy RCTs [4,7–9]. We assumed 70% adherence to baseline colonoscopy in the screening group and 30% (additional) use of colonoscopy over the course of the follow-up period in both groups to account for colonoscopy utilization rates in the general population regardless of indication [32]. Between both groups, we compared the cumulative incidence and mortality after the end of follow-up and the annual incidence and mortality rate ratio over time. The thereof resulting patterns of the effect of screening colonoscopy were compared to the patterns reported in screening sigmoidoscopy RCTs.

Sensitivity Analyses

As the testing rule for statistical equivalence between modelled and observed outcomes in KolosSal is influenced by the sample size of the modelled population, we re-run the tests with reduced simulated populations of 1,000 (instead of 100,000) men and women, respectively, to account for this uncertainty. As well, for the comparison to registry-derived incidences, we modelled a population without discounting the starting prevalences. Finally, to explore the impact of uncertainty regarding the annual transition rates between states, all point estimates of the starting prevalences and transition rates were replaced by the lower and upper boundary of the 95%-CIs in all analyses.

Patient and Public Involvement

Patients and the public were neither involved in the design and conduct of this study, nor in writing or editing of this document. Research at the German Cancer Research Center (DKFZ) is generally informed by a Patient Advisory Committee.

Results

Part 1: Equivalence to KolosSal Study Prevalences

Overall, 11,912 participants (5,916 men and 5,996 women) of the KolosSal study were included in the analysis (**Supplementary Table 2)**. The large majority (overall, 75 out of 80) of model-predicted ANN and ADN prevalences were within the 95% confidence intervals of the observed prevalences in the KolosSal study (**Figure 2**). **Supplementary Figure 1** shows the corresponding PRR. The sex-stratified meta-analyses of the PRR of ANN and ADN is shown in **Figure 3**. When tested for statistical equivalence at a 20% margin, all TOSTs yielded p-values <0.001.

Part 2: Comparison to Registry-Derived Cumulative Incidences

Table 1 shows the 15-year registry- vs model-derived cumulative CRC incidences. In men aged 50-64, 55-69, and 60-74, registry-derived cumulative incidence declined by approximately 0.3, 0.5, and 0.9 percent units, respectively, from 2001 to 2016. Corresponding changes in women were 0.3, 0.4, and 0.7 percent units, respectively. Model-derived cumulative incidences for men (women) declined by 0.2 (0.1), 0.4 (0.3) and 0.6 (0.5) percent units when comparing 2% vs 6% annual colonoscopy use over 15 years.

Overall, the model-derived cumulative incidence with 2% annual colonoscopy use compared best against the registry-derived cumulative incidence based on the year 2001 (deviation of 0.0-0.2 percent units in men and 0.1-0.3 percent units in women), while the model with 6% annual colonoscopy use was closest to the registry-derived estimate for the year 2011 (deviation of 0.0-0.1 percent units in men and 0.0-0.2 percent units in women). Deviations between registry- and model-based cumulative incidences tended to increase for older age groups.

Part 3: Comparison to RCT-Reported CRC Incidence and Mortality Patterns

The simulated cumulative incidence and mortality trends are shown in **Figure 4,** and the annual incidence and mortality rate ratios are provided in **Supplementary Figure 3**. **Table 2** shows the 15-year risk of developing and dying of CRC for control vs intervention groups, and the effect of screening colonoscopy at baseline.

Analogous to the overall patterns seen in the SCORE, NORCCAP, UKFSS and PLCO trials on sigmoidoscopy **(Supplementary Figure 2)**, the cumulative incidence in the colonoscopy screening group increased markedly in the year following the screening, followed by modest further increase until the end of follow-up. In the control group, the cumulative incidence

followed a steady growth trajectory. Both curves crossed approximately after 4-5 years, with the gap between curves widening up with longer follow-up duration, a pattern also observed in SCORE, UKFSS and PLCO (in NORCCAP, curves cross approximately after 6 years). While the cumulative mortality in both groups developed almost in parallel in the first years, the curves diverged more markedly with increasing follow-up duration, seemingly exceeding the trends seen in the sigmoidoscopy RCTs.

In the model, men and women benefitted at a similar magnitude from the effects of colonoscopy screening (**Supplementary Figures 4-7)**. This deviates from the results of the sigmoidoscopy RCTs: in SCORE, adjusted incidence rate ratios showed a trend towards a stronger effect in women [4]. Results from NORCCAP suggested that sigmoidoscopy screening reduced all-site CRC incidence and mortality in men but had little or no effect in women [8]. UKFSS and PLCO investigators detected a beneficial effect of screening in women, but to a lesser extent than men [7,9].

Sensitivity Analyses

Results of sensitivity analyses are reported in **Supplementary Appendix 3.** Testing for equivalence between modelled- vs KolosSal-derived prevalences was robust against using a reduced sample size in the model. Modelled cumulative incidences without discounting the model starting prevalences for prior use of stool test or diagnostic colonoscopy were notably elevated as compared to the registry-derived incidences. Overall, sensitivity analyses using adjusted starting prevalences and transition rates between states yielded similar findings as compared to the base case scenario.

Discussion

In this study, we firstly provided a comprehensive and transparent documentation of the development, structure and assumption of our multistate Markov model on the effects of CRC screening, including the core code used for its development. Secondly, we closely scrutinized the model's external validity. Overall, we found (1) a high level of agreement between model-derived and actually observed prevalences of colorectal neoplasms in a large screening colonoscopy study from Germany; (2) closely matching model- and registry-derived cumulative CRC incidences, absolute and in relative changes over time; and (3) similar trends in outcomes for a hypothetical colonoscopy screening RCT as compared to real-life sigmoidoscopy screening RCTs. Combined, our results provide comprehensive evidence on the model's ability to reproduce both the natural history of CRC as well as the effects of colonoscopy screening.

Validity of the Simulated Natural History

The parts 1 and 2 of the validation approach, based on the KolosSal study and the German cancer registries, illustrated the model's ability to predict colorectal neoplasm prevalences and cumulative cancer incidences for up to 25 years. Most remarkably, >90% of model-predicted prevalences were within the 95% CIs of the observed prevalences in the KolosSal study. The meta-analyses of prevalence rate ratios showed a numeric trend suggesting that the model might overestimate prevalences of advanced neoplasms in men. However, the sample size in KolosSal for age groups >70 years, where model predictions tended to yield higher than observed prevalences, was small (N < 200), increasing the associated uncertainty. Explorative testing with two one-sided t-tests still suggested statistical equivalence at a 20% equivalence margin.

In addition, a model with 2% annual colonoscopy use closely approximated the registry-derived cumulative CRC incidence for the year 2001, i.e. before colonoscopy screening was introduced. Model-derived incidences for 55-69 and 60-74-year-olds tended to slightly overestimate the registry-derived incidences. A likely explanation is that we assumed constant annual utilization patterns over time for the sake of simplicity, which does not consider that, in the period 2000-2002, older age groups were more likely to make use of colonoscopy than younger ones [28]. Furthermore, a certain degree of deviation was to be expected given the complexity of CRC epidemiology and CRC screening, and considering that the analyses were based on aggregated data.

Interestingly, the cumulative incidence for a model with 6% annual colonoscopy use matched best with the registry-derived estimate for the year 2011. This seems plausible given that the 6% utilization rate was derived for the period 2008-2010. More recent evidence from European EHIS data suggest that use of any colonoscopy within 10 years may have further increased [unpublished data], which may explain the even lower incidences for 2016.

Validity of the Simulated Effects of Colonoscopy Screening

A striking finding was that the model was able to approximately reproduce the observed effect of the successive uptake of colonoscopy utilization in terms of relative changes in cumulative incidence patterns over time. The drop in 15-year cumulative incidences with varying levels of annual colonoscopy (0.2-0.6 percent units in men, 0.1-0.5 percent units in women) as predicted by the model was slightly more pronounced, but in its pattern overall comparable to the change over time calculated for registry-derived estimates before and after the introduction of colonoscopy screening in 2002 (0.1-0.4 percent units in both men and women). Sensitivity analyses not adjusting the starting prevalences for the observed high levels of prior use of screening tests yielded elevated absolute cumulative incidences, but similar levels of relative changes over time (men: 13-15%, women: 14-16%) as the base-case model with discounted starting prevalences (men: 12-14%, women: 14-19%). This still adequately compares to the drop seen in registry derived incidences from 2001 – 2011 (men: 6-13%, women: 9-13%), which reinforces our opinion that not absolute predictions, but relative changes, are to be considered the key outcome parameters of the model. Overall, the evidence from the registry-based validation part indicates that the modelled relative effect of colonoscopy, though inarguably dependent on appropriately specified input parameters, adequately reproduces the effects seen in the real-world.

Additional evidence supporting the external validity of the modelled effect of screening colonoscopy comes from the simulation of a hypothetical colonoscopy screening RCT, which yielded graphical outcome patterns comparable to those reported for the sigmoidoscopy screening RCTs. We chose the sigmoidoscopy-RCTs as basis for comparison since it appears reasonable to assume that long-term results from randomized studies comparing colonoscopy versus no screening, which are not to be expected before the late 2020s [33], will yield similar patterns. The seen initial increase in cumulative incidence in the intervention group may even be expected to be more pronounced, given that sigmoidoscopy is limited to the rectum and distal part (where approximately 2/3 of CRCs are found), while colonoscopy examines the entire colon.

Accordingly, the potential for relative reduction of the CRC incidence and mortality likely is even higher for colonoscopy as compared to sigmoidoscopy. A 2014 meta-analysis of observational studies on the effect of colonoscopy found an incidence reduction of 69% and a mortality reduction of 68% (**Table 3**) [6]. More recent studies found comparable estimates [10,11]. In the per-protocol analyses of our hypothetical RCT, we found an incidence reduction by screening colonoscopy of approximately 60-65%, with an associated mortality reduction of 75-80%. Results from the per-protocol analyses appear more appropriate for comparison with observational studies, as these deal with the impact of colonoscopies actually performed rather than the mere offer of screening colonoscopy [6], although existing RCTs did still not adjust for major utilization of endoscopic procedures in the follow-up period. Other simulation studies found similar-to-larger effects, reaching up to 88% CRC incidence reduction and 90% CRC mortality reduction by the CRC-SPIN model in a hypothetical US-American population [34]. Taken together, our model compares well to the results of the 2014 meta-analysis but may be more conservative than other simulation models, which seems appropriate given the large confidence intervals for the screening colonoscopy effect found in the 2014 meta-analysis.

Finally, our model predicted similar effects of screening colonoscopy in men and women. At this point, this seems reasonable as currently, to our knowledge, there is no evidence from colonoscopy studies suggesting otherwise. The evidence from sigmoidoscopy RCTs is ambiguous, indicating a possibly greater (SCORE), smaller (PLCO, UKFSS) or even no (NORCCAP) effect of screening in women. A pooled analysis suggested stronger incidence reduction in younger women as compared to men but no screening effect in older women [35]. Reasons why sigmoidoscopy screening should have limited or no effect in women are unclear with no plausible rationale at hand [8], and chance findings due to underpowered studies by limited sample size for sex-stratified analyses cannot be ruled out.

Strengths and Limitations

A number of simulation models on the effect of CRC screening have been proposed in the literature [13,36]. Several of them were built on a bouquet of in scope and set-up largely varying data sources, which augments the model-related uncertainty, negatively impacts a transparent view on the model, and complicates its reproducibility. In contrast, our motivation was to develop a model explicitly for the screening eligible population, derived from consistent, comprehensible and high-quality data sources strongly linked to the German CRC screening landscape, and reproducible with reasonable efforts.

Therefore, our model's strong reliance on parameters derived from very large, partly unique real world data sources (i.e. the German National Screening Colonoscopy Registry; population-based German cancer registries; and the German DACHS study, one of the world's largest population-based case-control studies, detailed in appendix 1) represents a key conceptual strength. However, the inherent downside of this reliance is that it complicates assessing the model's external validity, which evidently depends on the existence of data the model ought to reproduce. For instance, a microsimulation model calibrated to the Norwegian population was validated by replicating and predicting outcomes from the (Norwegian) NORCCAP RCT [37]. A major limitation of this study was that we could not provide a similar validation, as no comparably suitable large-scale trial at the highest level of evidence has been conducted in the German population.

As alternative, we chose the herein presented three-fold approach, which uses the currently best available evidence base to validate a simulation model developed from epidemiological data in Germany. This approach enabled us to scrutinize the model's natural history component (part 1 and 2) as well as the modelled effect of screening colonoscopy (part 2 and 3) at the same time. Colonoscopy (but not sigmoidoscopy) has been offered as primary screening modality in Germany since 2002, and inarguably constitutes the final pathway of any two-tier screening modality (including the since 1975 offered fecal stool tests) for the foreseeable future [13].

A further limitation of our validation approach was that the outcome data of screening colonoscopy participants collected in the KolosSal study are partly also documented in the National Screening Colonoscopy Registry, which was used to derive the transition rates between model states. However, the subsample of KolosSal participants from 2005-2012 (N = 11,912) was only a minute fraction (approximately 0.3%) of the Colonoscopy Registry data (N = 3.6 to 4.3 million), and excluding KolosSal participants recruited in 2005-2012 from the Colonoscopy Registry data which would have insured complete independence would not have changed the model parameters to any relevant extent.

Conclusions

Despite these limitations, this study provided comprehensive evidence illustrating our model's ability to predict colorectal neoplasm prevalences and incidences in a German population for up to 25 years, with estimated patterns of the effect of screening colonoscopy resembling those seen in registry data and real-world studies. This suggests that the model represents a

valid approach to assess the comparative effectiveness of strategies for colorectal cancer screening, and thus a powerful tool to inform medical decision making.

References

1. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015;64:1637–49.

2. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383:1490–502.

3. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394:1467–80.

4. Segnan N, Armaroli P, Bonelli L, Risio M, Sciallero S, Zappa M, et al. Once-Only Sigmoidoscopy in Colorectal Cancer Screening: Follow-up Findings of the Italian Randomized Controlled Trial—SCORE. J Natl Cancer Inst. 2011;103:1310–22.

5. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus Fecal Immunochemical Testing in Colorectal-Cancer Screening. N Engl J Med. 2012;366:697–706.

6. 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.

7. Atkin W, Wooldrage K, Parkin DM, Kralj-Hans I, MacRae E, Shah U, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. Lancet. 2017;389:1299–311.

8. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Long-Term Effectiveness of Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality in Women and Men: A Randomized Trial. Ann Intern Med. 2018;168:775–82.

9. Miller EA, Pinsky PF, Schoen RE, Prorok PC, Church TR. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: long-term follow-up of the randomised US PLCO cancer screening trial. Lancet Gastroenterol Hepatol. 2019;4:101–10.

10. Kahi CJ, Pohl H, Myers LJ, Mobarek D, Robertson DJ, Imperiale TF. Colonoscopy and Colorectal Cancer Mortality in the Veterans Affairs Health Care System: A Case-Control Study. Ann Intern Med. 2018;168:481--488.

11. Doubeni CA, Corley DA, Quinn VP, Jensen CD, Zauber AG, Goodman M, et al. Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. Gut. 2018;67:291–8.

12. Grobbee EJ, van der Vlugt M, van Vuuren AJ, Stroobants AK, Mallant-Hent RC, Lansdorp-Vogelaar I, et al. Diagnostic Yield of One-Time Colonoscopy vs One-Time Flexible Sigmoidoscopy vs Multiple Rounds of Mailed Fecal Immunohistochemical Tests in Colorectal Cancer Screening. Clin Gastroenterol Hepatol. 2020;18:667–75.

13. Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for Colorectal Cancer Screening. Gastroenterology. 2020;158:418–32.

14. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Value Health. 2012;15:843–50.

15. Brenner H, Altenhofen L, Stock C, Hoffmeister M. Prevention, early detection, and overdiagnosis of colorectal cancer within 10 years of screening colonoscopy in Germany. Clin Gastroenterol Hepatol. 2015;13:717–23.

16. Brenner H, Altenhofen L, Stock C, Hoffmeister M. Expected long-term impact of the German screening colonoscopy programme on colorectal cancer prevention: Analyses based on 4,407,971 screening colonoscopies. Eur J Cancer. 2015;51:1346–53.

17. Brenner H, Kretschmann J, Stock C, Hoffmeister M. Expected long-term impact of screening endoscopy on colorectal cancer incidence: A modelling study. Oncotarget. 2016;7:48168–79.

18. Chen C, Stock C, Hoffmeister M, Brenner H. How long does it take until the effects of endoscopic screening on colorectal cancer mortality are fully disclosed?: a Markov model study. Int J Cancer. 2018;143:2718–24.

19. Holme Ø, Løberg M, Kalager M, Bretthauer M, Hernán MA, Aas E, et al. Effect of Flexible Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality: A Randomized Clinical Trial. JAMA. 2014;312:606–15.

20. Brenner H, Haug U, Arndt V, Stegmaier C, Altenhofen L, Hoffmeister M. Low risk of colorectal cancer and advanced adenomas more than 10 years after negative colonoscopy. Gastroenterology. 2010;138:870–6.

21. Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, Haug U. Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. J Natl Cancer Inst. 2010;102:89–95.

22. Tao S, Hoffmeister M, Brenner H. Development and Validation of a Scoring System to Identify Individuals at High Risk for Advanced Colorectal Neoplasms Who Should Undergo Colonoscopy Screening. Clin Gastroenterol Hepatol. 2014;12:478–85.

23. Hoffmeister M, Holleczek B, Stock C, Zwink N, Stolz T, Stegmaier C, et al. Utilization and determinants of follow-up colonoscopies within 6 years after screening colonoscopy: Prospective cohort study. Int J Cancer. 2019;144:402–10.

24. Peng L, Balavarca Y, Weigl K, Hoffmeister M, Brenner H. Head-to-Head Comparison of the Performance of 17 Risk Models for Predicting Presence of Advanced Neoplasms in Colorectal Cancer Screening. Am J Gastroenterol. 2019;114:1520–30.

25. Arditi C, Gonvers JJ, Burnand B, Minoli G, Oertli D, Lacaine F, et al. Appropriateness of colonoscopy in Europe (EPAGE II) – Surveillance after polypectomy and after resection of colorectal cancer. Endoscopy. 2009;41:209–17.

26. Gimeno García AZ, González Y, Quintero E, Nicolás-Pérez D, Adrián Z, Romero R, et al. Clinical validation of the European Panel on the Appropriateness of Gastrointestinal Endoscopy (EPAGE) II criteria in an open-access unit: a prospective study. Endoscopy. 2011;44:32–7.

27. Cancer Statistics for Germany. Interactive Database of the German Centre for Cancer Registry Data (ZfKD) [Internet]. 15 June 2018. Berlin: German Centre for Cancer Registry Data / Robert-Koch-Institut (RKI).; 2019 [cited 2019 Nov 20]. Available from: www.krebsdaten.de

28. Guo F, Chen C, Schöttker B, Holleczek B, Hoffmeister M, Brenner H. Changes in colorectal cancer screening use after introduction of alternative screening offer in Germany: Prospective cohort study. Int J Cancer. 2020;146:2423–32.

29. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. J Gen Intern Med. 2011;26:192–6.

30. Brenner H, Hoffmeister M, Arndt V, Haug U. Gender differences in colorectal cancer: implications for age at initiation of screening. Br J Cancer. 2007;96:828–31.

31. Bray F FJ. Age standardization. In: Bray F CM Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J, editor. Cancer Incid Five Cont Vol XI [Internet]. Lyon: International Agency for Research on Cancer.; 2017 [cited 2020 Mar 3]. Available from: ci5.iarc.fr

32. Starker A, Buttmann-Schweiger N, Krause L, Barnes B, Kraywinkel K, Holmberg C. Cancer screening in Germany: availability and participation (Krebsfrüherkennungsuntersuchungen in Deutschland: Angebot und Inanspruchnahme). Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz. 2018;61:1491–9.

33. Robertson DJ, Kaminski MF, Bretthauer M. Effectiveness, training and quality assurance of colonoscopy screening for colorectal cancer. Gut. 2015;64:982.

34. Knudsen AB, Zauber AG, Rutter CM, Naber SK, Doria-Rose VP, Pabiniak C, et al. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force. JAMA. 2016;315:2595–609.

35. Holme Ø, Schoen RE, Senore C, Segnan N, Hoff G, Løberg M, et al. Effectiveness of flexible sigmoidoscopy screening in men and women and different age groups: pooled analysis of randomised trials. BMJ. 2017;356:i6673.

36. Silva-Illanes N, Espinoza M. Critical Analysis of Markov Models Used for the Economic Evaluation of Colorectal Cancer Screening: A Systematic Review. Value Health. 2018;21:858–73.

37. Buskermolen M, Gini A, Naber SK, Toes-Zoutendijk E, de Koning HJ, Lansdorp-Vogelaar I. Modeling in Colorectal Cancer Screening: Assessing External and Predictive Validity of MISCAN-Colon Microsimulation Model Using NORCCAP Trial Results. Med Decis Making. 2018;38:917–29.

Overview of tables and figures

Table 1. Registry*- vs. model**-derived cumulative incidence of colorectal cancer after 15 years from baseline for ages 50-64, 55-69, and 60-74

Table 2. Risk of developing and dying from CRC, and the effect of screening colonoscopy at baseline for modelled control vs intervention groups* with different starting ages after 15 years of follow-up

Table 3. Comparison of effects of screening colonoscopy seen in the model to those reported in a meta-analysis of observational studies

Figure 1. Schematic illustration of the Markov Model

Figure 2. Prevalences and 95% confidence bands for any neoplasms (any adenoma or cancer) and advanced neoplasms (advanced adenomas or cancer) for ages 55-75 in the KolosSal study and as estimated by our model

Figure 3. Testing for equivalence between model and real study outcomes by applying two one-sided t-tests (TOST)

Figure 4. Per-protocol analysis of cumulative incidence and mortality in simulated cohorts of each 100,000 men and women in an unscreened control vs an intervention group with baseline colonoscopy screening

Table 1. Registry*- vs. model**-derived cumulative incidence of colorectal cancer after 15 years from baseline for ages 50-64, 55-69, and 60-74

	Men						Women					
	CRC Risk (%)						CRC Risk (%)					
	German Registry				Model		German Registry				Model	
	year				annual colonoscopy use		year				annual colonoscopy use	
Age at baseline	**2001**	**2006**	**2011**	**2016**	**2%**	**6%**	**2001**	**2006**	**2011**	**2016**	**2%**	**6%**
50	1.7	1.7	1.6	1.4	1.7	1.5	1.1	1.1	1.0	0.8	1.0	0.9
55	2.7	2.7	2.5	2.2	2.8	2.4	1.6	1.6	1.4	1.2	1.7	1.4
60	4.0	4.0	3.6	3.1	4.2	3.6	2.4	2.3	2.0	1.7	2.7	2.2

**Registry-derived cumulative incidence was estimated based on age-group and sex-specific incidence rates extracted one year before the introduction of the screening colonoscopy offer in Germany in 2002 and then in steps of 5-years.*

*** Model-derived cumulative incidence assumes approximate levels of annual colonoscopy use before (2%) and after (6%) the introduction of the colonoscopy screening offer.*

CRC, colorectal cancer.

Table 2. Risk of developing and dying from CRC, and the effect of screening colonoscopy at baseline for modelled control vs intervention groups* with different starting ages after 15 years of follow-up. Shown are results analysed according to the per protocol (PP, only those adherent) and intention to treat (ITT, all randomized) principle

Overall

		Risk of Developing CRC		Risk of Dying from CRC		Effect of Baseline Colonoscopy *(Numbers given per 1,000 colonoscopies)*			
Age at Baseline	**Analysis Strategy**	**Control**	**Intervention**	**Control**	**Intervention**	**Incidence Reduction**	**No of Cases Prevented**	**Mortality Reduction**	**No of Deaths Prevented**
50	PP	1.9%	0.7%	0.8%	0.4%	61.3%	12	75.0%	6
	ITT	1.9%	1.1%	0.8%	0.8%	42.9%	8	52.5%	4
60	PP	4.2%	1.5%	1.8%	0.8%	63.6%	27	79.0%	14
	ITT	4.2%	2.4%	1.8%	1.6%	44.5%	19	55.3%	10

Men

		Risk of Developing CRC		Risk of Dying from CRC		Effect of Baseline Colonoscopy *(Numbers given per 1,000 colonoscopies)*			
Age at Baseline	**Analysis Strategy**	**Control**	**Intervention**	**Control**	**Intervention**	**Incidence Reduction**	**No of Cases Prevented**	**Mortality Reduction**	**No of Deaths Prevented**
50	PP	2.4%	0.9%	1.1%	0.3%	61.0%	14	74.6%	8
	ITT	2.4%	1.3%	1.1%	0.5%	42.7%	10	52.2%	6
60	PP	5.2%	2.0%	2.3%	0.5%	62.2%	32	78.4%	18
	ITT	5.2%	2.9%	2.3%	1.0%	43.5%	23	54.9%	13

Women

		Risk of Developing CRC		Risk of Dying from CRC		Effect of Baseline Colonoscopy *(Numbers given per 1,000 colonoscopies)*			
Age at Baseline	**Analysis Strategy**	**Control**	**Intervention**	**Control**	**Intervention**	**Incidence Reduction**	**No of Cases Prevented**	**Mortality Reduction**	**No of Deaths Prevented**
50	PP	1.4%	0.5%	0.6%	0.1%	61.7%	9	75.8%	4
	ITT	1.4%	0.8%	0.6%	0.3%	43.2%	6	53.1%	3
60	PP	3.3%	1.1%	1.3%	0.3%	65.7%	22	80.1%	10
	ITT	3.3%	1.8%	1.3%	0.6%	46.0%	15	56.1%	7

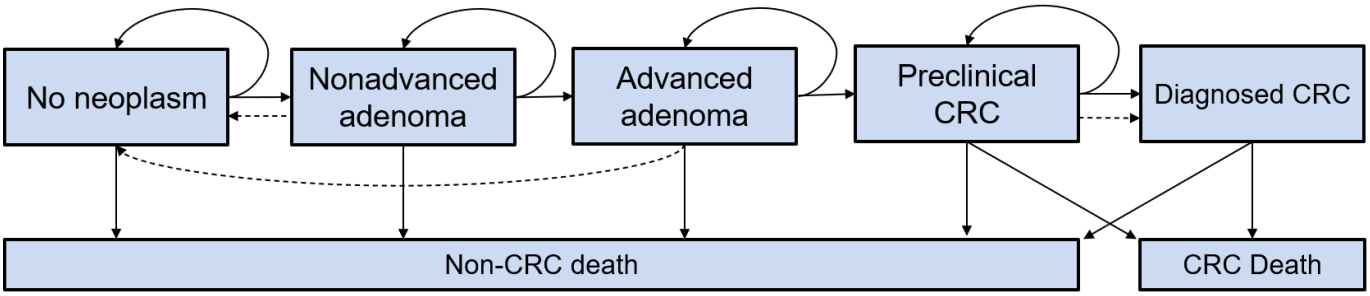
**100,000 men and women in both groups. Adherence of 70% to baseline colonoscopy screening in the intervention group and 30% (additional) use of colonoscopy for any reason over time in both groups. CRC, colorectal cancer, PP, per protocol, ITT, intention to treat.*

Table 3. Comparison of effects of screening colonoscopy seen in the model to those reported in a meta-analysis of observational studies

	Effect of screening colonoscopy at baseline (per protocol)		**Meta-Analysis Brenner 2014**[6]	
Age	**Incidence Reduction (%)**	**Mortality Reduction (%)**	**Incidence Reduction (%, 95% CI)**	**Mortality Reduction (%, 95% CI)**
50	61	75	69 (23 - 88)	68 (57 - 77)
60	64	79		

CI, confidence interval.

Figure 1. Schematic illustration of the Markov Model



Solid lines represent the progression of colorectal disease through the adenoma-carcinoma sequence in the absence of screening; dashed lines show the movement between states because of the detection and removal of adenomas and the detection of asymptomatic CRC at screening.

CRC, colorectal cancer.

Figure 2. Prevalences and 95% confidence bands for any neoplasms (any adenoma or cancer) and advanced neoplasms (advanced adenomas or cancer) for ages 55-75 in the KolosSal study and as estimated by our model. Left side: men. Right side: women**.**

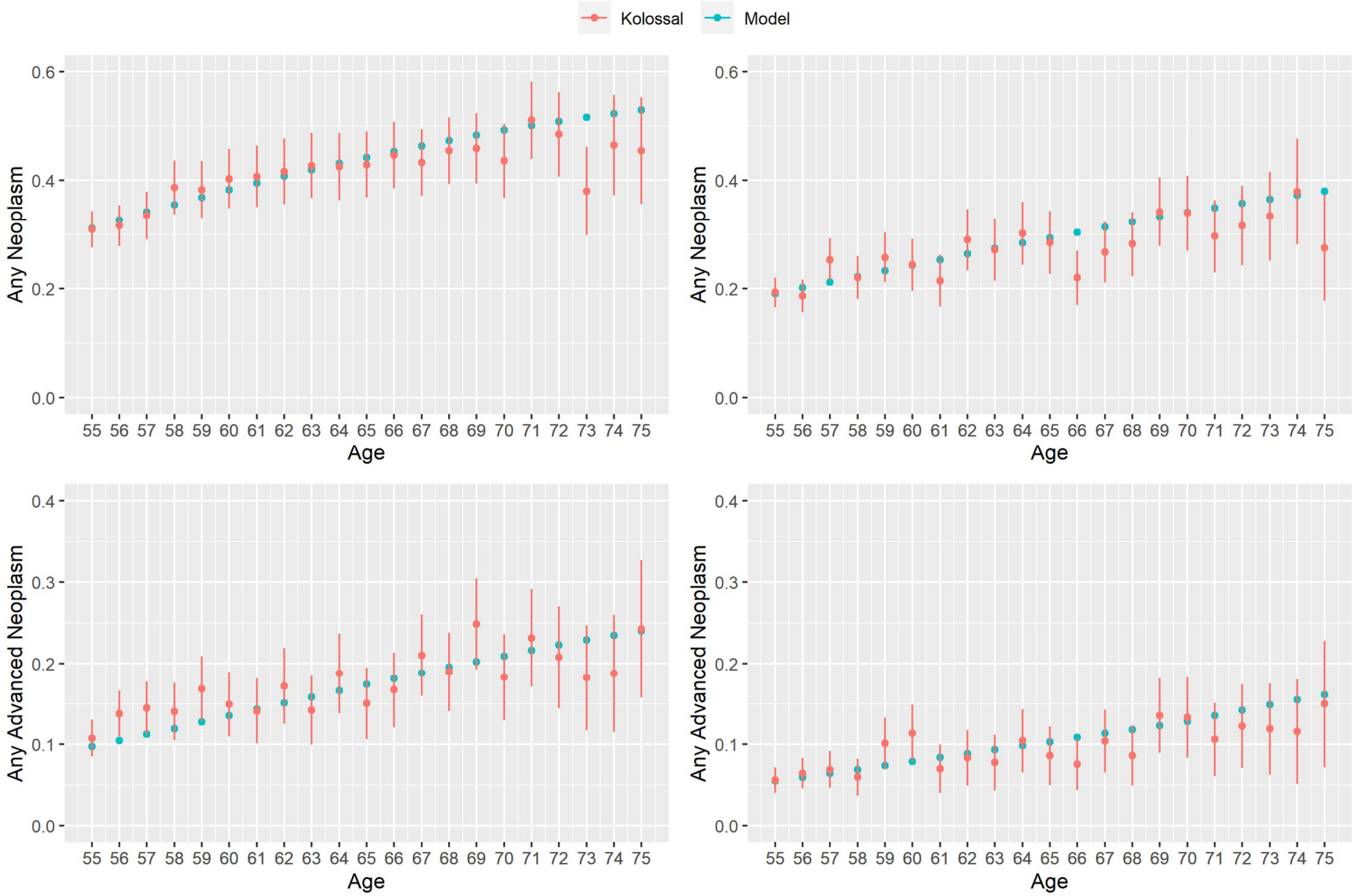
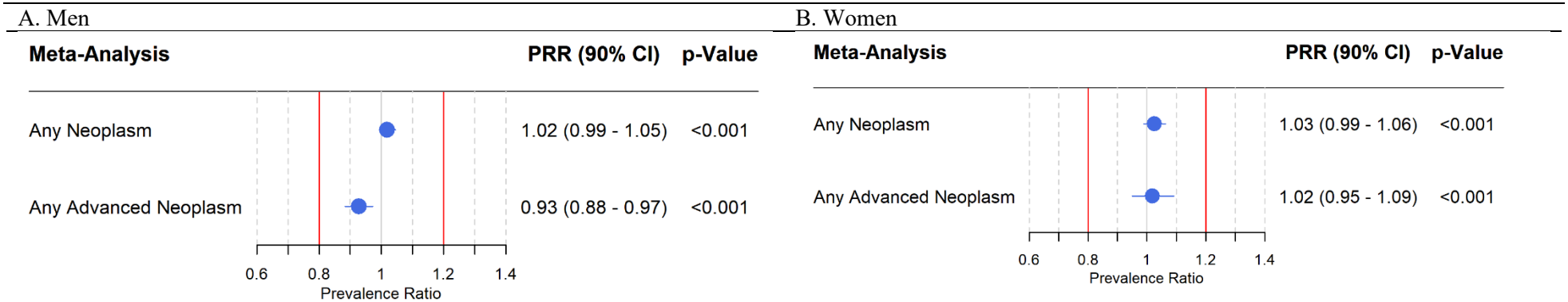


Figure 3. Testing for equivalence between model and real study outcomes by applying two one-sided t-tests (TOST). Shown are the meta-analyses estimates and 90% confidence intervals* for the summarized prevalence ratios at ages 55-75 for any neoplasm and any advanced neoplasm. Only the higher of the two p-values for the TOST are shown, values <0.05 suggest statistical equivalence at a 20% margin.



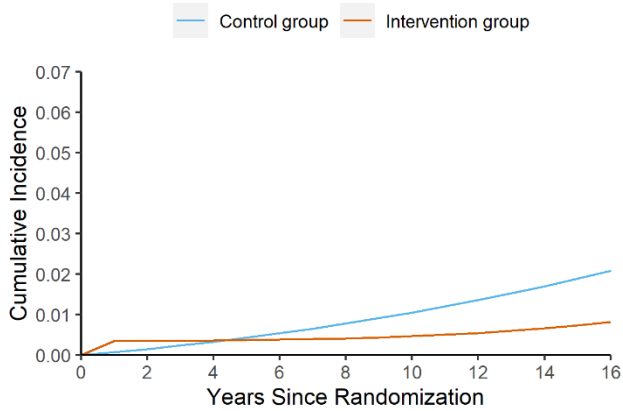
**Using a 90% confidence interval yields a 0.05 significance level for testing equivalence (as two one-sided tests without adjustment for multiple testing are performed))*

PRR, Prevalence Ratio

Figure 4. Per-protocol analysis of cumulative incidence and mortality in simulated cohorts of each 100,000 men and women in an unscreened control vs an intervention group with baseline colonoscopy screening. Shown are different starting ages at start of the simulation. All cohorts had 16 years of follow-up.

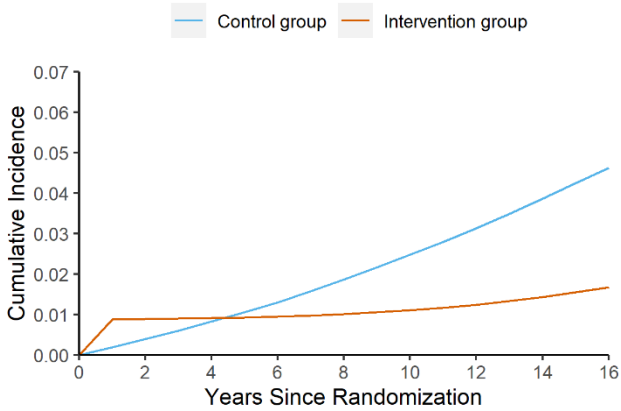
Cumulative Incidence

A. At age 50



Control At risk, n	200,000	199,861	199,536	199,149	198,698	198,184	197,607	196,953	196,233
Intervention At risk, n	140,000	139,521	139,509	139,489	139,454	139,395	139,303	139,165	138,980

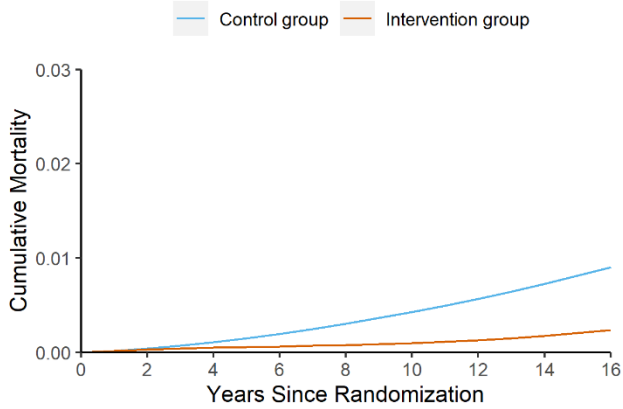
B. At age 60



Control At risk, n	200,000	199,625	198,799	197,895	196,848	195,662	194,426	193,026	191,506
Intervention At risk, n	140,000	138,775	138,744	138,701	138,632	138,522	138,369	138,141	137,833

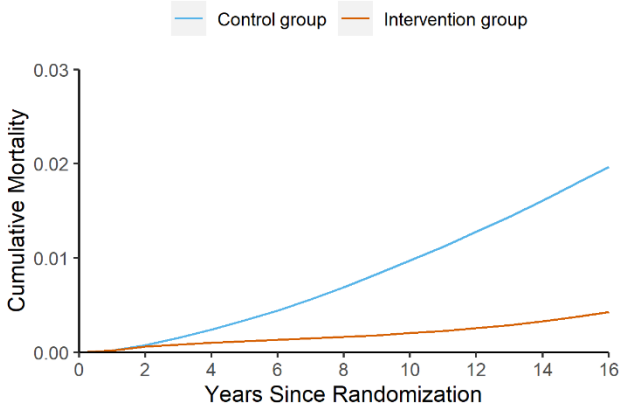
Cumulative Mortality

C. At age 50



Control At risk, n	200,000	199,968	199,857	199,700	199,503	199,271	199,008	198,708	198,375
Intervention At risk, n	140,000	139,978	139,944	139,923	139,903	139,878	139,843	139,790	139,716

D. At age 60



Control At risk, n	200,000	199,960	199,698	199,325	198,875	198,344	197,758	197,119	196,423
Intervention At risk, n	140,000	139,972	139,886	139,837	139,795	139,746	139,683	139,596	139,475