# Misleading and avoidable: design-induced biases in observational studies evaluating cancer screening—the example of site-specific effectiveness of screening colonoscopy

**Authors**: Malte Braitmaier, Sarina Schwarz, Vanessa Didelez, Ulrike Haug

**Author information:**

Malte Braitmaier, Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany, ORCID: 0000-0001-7534-4068

Sarina Schwarz, Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany, ORCID: 0000-0002-7926-2032

Vanessa Didelez, Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany & Faculty of Mathematics and Computer Sciences, University of Bremen, Bremen, Germany, ORICD: 0000-0001-8587-7706

Ulrike Haug, Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany & Faculty of Human and Health Sciences, University of Bremen, Bremen, Germany, ORCID: 0000-0002-1886-2923

Corresponding author: Ulrike Haug, haug@leibniz-bips.de

**Abstract**

**Objective**: Observational studies evaluating the effectiveness of cancer screening are often biased due to an inadequate design where I) the assessment of eligibility, II) the assignment to screening vs. no screening and III) the start of follow-up are not aligned at time zero (baseline). Such flaws can entail misleading results but are avoidable by designing the study following the principle of target trial emulation (TTE). We aimed to illustrate this by addressing the research question whether screening colonoscopy is more effective in the distal vs. the proximal colon.

**Methods**: Based on a large German health care database (20% population coverage), we assessed the effect of screening colonoscopy in preventing distal and proximal CRC over 12 years of follow-up in 55–69-year-old persons at average CRC risk. We applied four different study designs and compared the results: cohort study with / without alignment at time zero, case control study with / without alignment at time zero.

**Results**: In both analyses with alignment at time zero, screening colonoscopy showed a similar effectiveness in reducing the incidence of distal and proximal CRC (cohort analysis: 32% (95% CI: 27% - 37%) vs. 28% (95% CI: 20% - 35%); case-control analysis: 27% vs. 33%). Both analyses without alignment at time zero suggested a difference in site-specific performance: Incidence reduction regarding distal and proximal CRC, respectively, was 65% (95% CI: 61% - 68%) vs. 37% (95% CI: 31% - 43%) in the cohort analysis and 77% (95% CI: 67% - 84%) vs. 46% (95% CI: 25% - 61%) in the case-control analysis.

**Conclusions**: Our study demonstrates that violations of basic design principles can substantially bias the results of observational studies on cancer screening. In our example, it falsely suggested a much stronger preventive effect of colonoscopy in the distal vs. the proximal colon. The difference disappeared when the same data were analyzed using a TTE approach, which is known to avoid such design-induced biases.

## Introduction

Randomized controlled trials (RCT) are the gold standard for evaluating the effectiveness of cancer screening. However, existing RCTs in this field do not answer all relevant research questions. For screening colonoscopy, for example, an RCT has recently been published (NordICC trial) demonstrating its effectiveness in reducing colorectal cancer (CRC) incidence overall [1], but it was not powered to compare the effectiveness in the distal vs. the proximal colon.

Complementary evidence from observational studies is therefore needed. Apart from potential confounding, there is a high risk of bias and thus of misleading results if such studies are inadequately designed. Indeed, several observational studies have reported a markedly stronger preventive effect of screening colonoscopy in the distal as compared to the proximal colon [2, 3, 4], while a cohort study designed following the principle of target trial emulation (TTE) showed a similar effectiveness of screening colonoscopy in the distal and the proximal colon [5]. We argued that the difference by site in the former studies was due to biases induced by non-alignment at “time zero”, i.e. at baseline. This means that I) the assessment of eligibility, II) the assignment to study arms and III) the start of follow-up were not aligned as they would be in an RCT and as it would be ensured in an observational study designed based on the principle of TTE [6]. Specifically, previous studies often defined exposure based on pre- or post-baseline information on colonoscopy. As we further argued, this lack of alignment in previous studies led to overestimating the effectiveness of screening colonoscopy. Due to the different age pattern of distal and proximal CRC, this bias affected distal CRC more than proximal CRC, i.e. the difference in effectiveness by site was an artefact.

To demonstrate this, we compared different study designs with and without alignment at time zero aiming to investigate the question of site-specific effectiveness of screening colonoscopy in reducing CRC incidence. For the two designs without alignment we used a cohort study design, where the assignment to study arms occurs *before* time zero (pre-baseline), and a nested case control study design, where the assignment to study arms occurs *after* time zero

(post-baseline). The current paper is part of a growing literature identifying violations of alignment at time zero as a potential source of major bias in observational studies [6, 7, 8].

## Methods

*Data source and study population*

We used the German Pharmacoepidemiological Research Database (GePaRD) which comprises claims data from four statutory health insurance providers in Germany and covers about 20% of the German population [9]. In GePaRD, information on utilization of screening colonoscopy, offered in Germany to persons aged 55 or older since 2002 (since 2019 also to men aged 50-54), is distinguishable from diagnostic colonoscopy. As previously described, the data source enables the valid identification of incident CRCs [10]. Furthermore, it contains appropriate information to apply in- and exclusion criteria and to adjust for confounding as relevant to the research question on the effectiveness of screening colonoscopy in reducing CRC incidence [5]. For the present study, we used data from 2004 to 2020.

Based on this data source, we applied four different study designs to address the research question, specifically a cohort and a case-control study design, each with and without alignment at time zero. The study designs without alignment at time zero were inspired by published examples [2, 11, 12, 13], and were partly complemented by sensitivity analyses. For each of these four studies, persons were selected from the same population. Specifically, the source population was a cohort of persons aged 55–69 at baseline, who were continuously insured for at least three years before baseline.

*Cohort study without alignment at time zero*

The cohort started in 2009 (baseline). Similar to a previous study [2], individuals were assigned to the screening colonoscopy arm if they had a screening colonoscopy any time before baseline, including the baseline quarter. Individuals were assigned to the control arm if they did not undergo screening colonoscopy any time before baseline, including the baseline quarter. In a sensitivity analysis, we considered both screening and diagnostic colonoscopies

for the assignment to the study arms, because some of the previous studies did not distinguish between these examinations. Eligibility criteria were checked at baseline and the outcome variable (incident CRC) was assessed beginning with baseline (start of follow-up). Persons were followed up until end of study period (end of 2020), end of continuous insurance coverage, death or CRC diagnosis, whichever occurred first. We also conducted sensitivity analyses starting the cohort in 2010 and 2011, respectively.

When using such a study design, the assessment of eligibility and the start of follow-up are aligned, but the assignment to the screening and the control arm is based on a period *before* time zero (pre-baseline). Specifically, individuals in the colonoscopy arm had the examination in the past (i.e. they were assigned to the screening arm based on past exposure) rather than *at* time zero.

*Cohort study with alignment at time zero*

As described previously [5], we emulated sequential trials for each calendar quarter from 2007 to 2011. The emulation of sequential target trials makes full use of the information from longitudinal data without violating principles of study design by using pre- or post-baseline information for the assignment to study arms. At the baseline quarter of each trial, eligibility was assessed and individuals with previous screening colonoscopy or CRC diagnosis were excluded. Individuals were then assigned to the screening arm if they underwent a screening colonoscopy *in the baseline quarter* of the respective trial and to the control arm otherwise. Individuals were followed up until end of study period (end of 2020, i.e. follow-up was longer than in our previous analysis), end of continuous insurance coverage, death or CRC diagnosis, whichever occurred first. This study design made sure that assessment of eligibility criteria, assignment to the screening and control arm, and start of follow-up were aligned at time zero as would be the case in an RCT.

*Case-control study without alignment at time zero*

We applied a case-control design frequently used in the published literature [11, 12, 13, 14, 15, 16]. Essentially, CRC cases are identified (date of diagnosis corresponds to index date) and matched with controls free of CRC at index date. Then screening colonoscopy use *ever before* or within a certain time period *before* the index date is assessed in cases and controls, i.e. colonoscopies leading to CRC diagnosis are not considered as exposure in this type of study. Here, we selected all individuals from the source population entering the cohort in 2009 with a CRC diagnosis in 2018-2020. For each case we matched up to five controls on age (+/- one year) and sex (sampling without replacement). The exposure variable was then defined as any screening colonoscopy between 2009 and the index date, i.e. exposure to colonoscopy use was assessed within 10-12 years *before* the index date. Colonoscopies conducted in the six months before CRC diagnosis were not considered in defining the exposure. As mentioned above, this approach corresponds to published case-control studies which ignore colonoscopies conducted as part of the diagnostic process leading to the current diagnosis [11, 12, 13, 14, 15, 16]. In general, it is a fundamental characteristic of traditional case-control studies to assess exposure before disease onset. In a sensitivity analysis, we considered both screening and diagnostic colonoscopies for the assignment to exposure groups. Again, we also conducted sensitivity analyses using the years 2010 and 2011 for cohort entry, i.e. the source population underlying this nested case-control study.

In the case-control design we used here (nested within a cohort), the assessment of eligibility and the start of follow-up were aligned, while the assignment to the screening and the control arm occurred *after* time zero (post-baseline) instead of *at* time zero. Note that in case-control studies not nested in a cohort, there typically are additional misalignments [11, 14]. Specifically, eligibility is assessed at index date and the start of follow-up is unclear.

*Case-control study with alignment at time zero*

Following the approach described by Dickerman et al. [17], a case-control study was nested within the original cohort of sequential emulated target trials, and colonoscopy use was assessed at baseline of each emulated trial. We included CRC patients with an incident CRC

diagnosis at any point during follow-up (until 2020) and then used risk set sampling to match up to five controls to each case. We sampled matched controls with replacement, i.e. the same control could be matched to more than one case. Matching variables were the same as above. The key difference to the case-control study without alignment is that exposure assignment was based on information available at the start of the emulated trial, i.e. at time zero, instead of information occurring after time zero. This approach has been shown to avoid self-inflicted biases in the same way as a prospective study using TTE [17].

*Data analysis*

For the cohort studies, we estimated cumulative incidence functions (CIF) via pooled logistic regressions, which were adjusted for baseline confounders via inverse probability of treatment weighting. Effects were estimated as adjusted relative risks (RR) at the end of follow-up based on these CIFs. As previously shown, adjustment yielded satisfactory covariate balance and a negative control analysis did not indicate any residual confounding [5]. Confidence intervals were estimated via person-level bootstrap. For the case-control studies, effects were estimated as adjusted odds ratios (ORs) obtained via conditional logistic regression. For the case-control analysis with alignment, no confidence intervals could be obtained due to computational limitations: The emulation of sequential trials with repeated cohort entry would require bootstrapping, where matching is repeated for every bootstrap sample, resulting in run times of several months.

## Results

*Cohort study without alignment at time zero*

We selected a random sample of 200,000 individuals in the control arm and 200,000 individuals in the screening colonoscopy arm. The adjusted relative risk after 12 years of follow-up was 0.35 for distal CRC and 0.63 for proximal CRC (Table 1). The adjusted cumulative incidence curves are given in Fig. **1**. As shown in Supplement 1, results were similar when the year 2010 or the year 2011 was used as baseline. In sensitivity analyses considering both

screening and diagnostic colonoscopies as exposure, the adjusted 12-year relative risk was 0.40 for distal CRC and 0.66 for proximal CRC (Supplement 2).

*Cohort study with alignment at time zero*

Overall, 192,054 persons were included in the screening colonoscopy arm. The 5% random sample (restriction due to computational limitations) of controls assigned to the no screening arm included 116,452 persons (1,241,071 non-unique). The adjusted relative risk after 12 years of follow-up was 0.68 for distal CRC and 0.72 for proximal CRC (Table 1). Figure 1 shows the adjusted cumulative incidence curves for distal and proximal CRC. The distribution of screen-detected and post-colonoscopy CRCs (i.e. non-screen-detected CRCs) by site is shown in Supplement 6.

*Case-control study without alignment at time zero*

Overall, 446 cases with distal CRC matched to 2,230 controls and 302 cases with proximal CRC matched to 1,510 controls were included. The adjusted ORs for distal and proximal CRC were 0.23 and 0.54, respectively (Table 2). When the year 2010 or the year 2011 was used to define the source population, the difference by site was similar (Supplement 1). The sensitivity analysis considering both screening and diagnostic colonoscopy as exposure yielded similar results; the adjusted ORs for distal and proximal CRC were 0.20 for distal CRC and 0.44 for proximal CRC, respectively (Supplement 2).

*Case-control study with alignment at time zero*

Overall, 8,382 cases with distal CRC matched to 40,925 controls and 4,463 cases with proximal CRC matched to 22,175 controls were included. The adjusted ORs for distal and proximal CRC were 0.73 and 0.67, respectively (Table 2).

## Discussion

To the best of our knowledge, our study is the first to systematically compare different study designs to assess the effectiveness of screening colonoscopy in reducing CRC incidence in

the distal vs. the proximal colon. Our cohort and case-control analyses with alignment at time zero showed no relevant difference in the effectiveness by site. Using study designs without alignment at time zero led to an overestimation of the effectiveness of screening colonoscopy overall. The overestimation affected distal CRCs considerably more than proximal CRCs , i.e. purely by design there appeared to be a difference in effectiveness by site. This finding held up in sensitivity analyses varying data years and the type of examinations considered for the exposure definition (only screening or also diagnostic colonoscopy). Our findings demonstrate that the difference in the effectiveness of colonoscopy by site reported by previous observational studies was due to bias introduced by inadequate study design.

As illustrated in Supplement 3 using directed acyclic graphs, the bias underlying studies using pre-baseline information on colonoscopy for the assignment to study arms can be expressed as a form of collider stratification bias [18, 19]. To give an intuitive explanation, let us revisit the study by Guo et al. [2, 5]: At baseline, patients were asked about past colonoscopy use and—based on this information—assigned as exposed or unexposed to colonoscopy. Persons reporting a prior CRC diagnosis at baseline were excluded [2]. Given that colonoscopy is one of the main tools by which CRC is diagnosed, this process removes individuals with previously diagnosed CRC from the exposed group, i.e. it enriches the exposed group with individuals who are known to be free of CRC. No such selection process takes place in the unexposed group. This leads to a lower prevalence of preclinical CRC at baseline in the exposed as compared to the unexposed group. As a consequence, this selection reduces the number of CRCs occurring during follow-up in the exposed group as compared to the unexposed group and thus leads to overestimation of the effect of screening on CRC incidence. As the vast majority of CRCs diagnosed at an age when persons are typically included into screening studies are in the distal colon [20] while proximal CRCs become more common at older age, this bias mainly affects results for distal CRC, i.e. as mentioned above there appeared to be a difference in effectiveness by site purely by design. We note that in addition to the initial exposure assignment, Guo et al. also used an updated exposure variable in a Cox model with

time-dependent covariates. However, this does not correct the initial selection issue at the start of follow-up.

The above argument applies to studies using pre-baseline information for the assignment to exposure groups. Many other studies used post-baseline information for the assignment to exposure groups, also inducing bias. We illustrated this by the case-control study without alignment at time zero: Whenever after baseline CRC is detected in a person at his or her first colonoscopy, as is the case for most screen-detected CRCs, this person is assigned to the unexposed group as there was no previous colonoscopy and the actual colonoscopy detecting the CRC is not considered as prior exposure. This enriches the unexposed group with CRCs and thus leads to overestimation of the effectiveness of screening. As the majority of screen-detected CRCs are in the distal colon, this bias predominantly affects CRCs in the distal colon and thus leads to an artificial difference in the effectiveness of colonoscopy by site (see also Supplement 4). In our case-control study design embedded in an emulated target trial with alignment at time zero, in which screen-detected CRCs are correctly assigned, no relevant difference in the effectiveness of colonoscopy by site was observed. Of note, misalignment due to post-baseline exposure assignment is typical of but not limited to case-control designs on cancer screening. It can also occur in inadequately designed cohort studies and is not overcome by using a time-varying exposure variable in a hazard model. This is explained in more detail in Supplement 5 based on the example of the study by Nishihara et al. [3]

In summary and more generally, both study designs without alignment at time zero have in common that there are mechanisms that lead to inappropriate consideration of screen-detected CRCs, i.e. in the screening arm there was no peak in CRC incidence immediately after baseline as it would be the case in an RCT. Of course, this overestimates the impact of screening on CRC incidence, particularly for distal CRC, as illustrated in Figure 1. The flawed approaches ignore the fact that a screening colonoscopy sometimes comes too late to prevent CRC. Following the publication of the NordICC study, there was a discussion whether it is appropriate to include persons with preclinical CRC, causing the peak at baseline, in a

prevention trial [21, 22]. However, from a public health perspective, it is important to also take into account CRCs that are not prevented by screening in order to avoid overestimating the effectiveness of CRC screening at the population level. Apart from this, it should be noted that studies without alignment at time zero do not provide a valid answer to the question regarding the size of the preventive effect of colonoscopy in persons free of CRC at baseline.

It should be noted that, although we focus our discussion on biases most relevant for site-specific effectiveness of screening colonoscopy, misalignment at time zero should also be avoided for many other reasons. Rasouli et al. [23] demonstrated that time related issues such as prevalent user bias or time-varying confounding are a threat to case-control designs not embedded in an emulated target trial. Also Dickerman et al. showed—based on case-control studies investigating the impact of statins on CRC risk—the biases inherent to traditional case-control studies and the potential of avoiding bias and wrong conclusions if the study is designed following the principle of TTE [17]. Similarly, there are many examples of biases other than those we discussed here that are inherent to cohort studies without alignment at time zero [8].

Our findings have several implications. First, regarding research on CRC screening, previous studies suggesting a lower effectiveness of colonoscopy in the proximal colon stimulated a search for reasons that may explain the occurrence of post-colonoscopy CRCs specifically in the proximal colon. It was suggested that one main reason relates to sessile serrated lesions as they are more difficult to detect and more often occur in the proximal colon [24]. While we do not question the important role of these lesions, our findings may encourage a broadening of the discussion of potential reasons leading to post-colonoscopy CRCs. Indeed, in our emulated target trial on screening colonoscopy, the proportion of post-colonoscopy CRCs located in the distal vs. the proximal colon was rather similar (Supplement 6). A one-sided focus on lesions that occur more frequently in the proximal colon therefore seems too narrow regarding the identification of lesions possibly leading to post-colonoscopy CRCs.

Our results also have implications beyond the specific research question of our study. Observational data are often used to evaluate the effectiveness of cancer screening. They

represent a valuable data source to complement RCT evidence in this field, as RCTs on cancer screening are scarce, were often conducted many years ago and are typically not powered to estimate, for example, subgroup-specific effects or differences by cancer subtype. However, our study illustrates that—in addition to appropriate control of confounding—it is of key importance to design these studies in a way to ensure alignment at time zero. This means that assessment of eligibility, assignment to the screening and control arm and start of follow-up must be aligned. Otherwise, there is a high risk of bias.

Specific strengths of our study include the systematic comparison of different study designs as well as the comprehensive sensitivity analyses. Given that all analyses were conducted using the same data source and referred to the same setting, there is no heterogeneity regarding, for example, the study variables or setting-related factors such as the uptake of surveillance colonoscopy or colonoscopy quality. This strengthens our conclusion that differing results of the analyses with and without alignment at time zero are exclusively due to the study design.

It should be noted that our findings apply to the population aged 55-69, covering the typical screening age range of CRC screening. Whether screening colonoscopy is equally effective in the distal and proximal colon in older age groups cannot be answered by our study, nor did we address the endpoint CRC mortality. These research questions were beyond our study's scope, as our primary objective was to illustrate the relevance of design-induced biases and the possibility to avoid them using TTE, exemplified by investigating site-specific effectiveness of screening colonoscopy in reducing CRC incidence.

In conclusion, our study demonstrates that violation of alignment at time zero can substantially bias the results of observational studies on cancer screening. In our example, it falsely suggested an almost doubled preventive effect of colonoscopy in the distal vs. the proximal colon. The difference disappeared when the same data were analyzed using a TTE approach, which is known to avoid design-induced biases.

310 **References**


311 1 Bretthauer M, Loberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K*, et al.* Effect of
312 colonoscopy screening on risks of colorectal cancer and related death. N Engl J Med 2022;**387**:1547-
313 56.
314 2 Guo F, Chen C, Holleczek B, Schottker B, Hoffmeister M, Brenner H. Strong reduction of
315 colorectal cancer incidence and mortality after screening colonoscopy: prospective cohort study from
316 Germany. Am J Gastroenterol 2021;**116**:967-75.
317 3 Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR*, et al.* Long-term colorectal-
318 cancer incidence and mortality after lower endoscopy. N Engl J Med 2013;**369**:1095-105.
319 4 Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening
320 colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of
321 randomised controlled trials and observational studies. BMJ 2014;**348**:g2467.
322 5 Braitmaier M, Schwarz S, Kollhorst B, Senore C, Didelez V, Haug U. Screening colonoscopy
323 similarly prevented distal and proximal colorectal cancer: a prospective study among 55-69-year-
324 olds. J Clin Epidemiol 2022;**149**:118-26.
325 6 Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using
326 real world evidence: an application to colorectal cancer screening. Eur J Epidemiol 2017;**32**:495-500.
327 7 Wakabayashi R, Hirano T, Laurent T, Kuwatsuru Y, Kuwatsuru R. Impact of "time zero" of
328 Follow-Up Settings in a Comparative Effectiveness Study Using Real-World Data with a Non-user
329 Comparator: Comparison of Six Different Settings. Drugs Real World Outcomes 2023;**10**:107-17.
330 8 Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents
331 immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol
332 2016;**79**:70-5.
333 9 Haug U, Schink T. German Pharmacoepidemiological Research Database (GePaRD). In:
334 Sturkenboom M, Schink T, eds. Databases for pharmacoepidemiolpogical research. Cham,
335 Switzerland: Springer, 2021:119-24.
336 10 Schwarz S, Hornschuch M, Pox C, Haug U. Colorectal cancer after screening colonoscopy: 10-
337 year incidence by site and detection rate at first repeat colonoscopy. Clin Transl Gastroenterol
338 2023;**14**:e00535.
339 11 Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of
340 colonoscopy and death from colorectal cancer. Ann Intern Med 2009;**150**:1-8.
341 12 Mulder SA, van Soest EM, Dieleman JP, van Rossum LG, Ouwendijk RJ, van Leerdam ME*, et al.*
342 Exposure to colorectal examinations before a colorectal cancer diagnosis: a case-control study. Eur J
343 Gastroenterol Hepatol 2010;**22**:437-43.
344 13 Doubeni CA, Weinmann S, Adams K, Kamineni A, Buist DS, Ash AS*, et al.* Screening
345 colonoscopy and risk for incident late-stage colorectal cancer diagnosis in average-risk adults: a
346 nested case-control study. Ann Intern Med 2013;**158**:312-20.
347 14 Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal
348 cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 2011;**154**:22-30.
349 15 Kahi CJ, Pohl H, Myers LJ, Mobarek D, Robertson DJ, Imperiale TF. Colonoscopy and colorectal
350 cancer mortality in the veterans affairs health care system: a case-control study. Ann Intern Med
351 2018;**168**:481-8.
352 16 Baxter NN, Warren JL, Barrett MJ, Stukel TA, Doria-Rose VP. Association between
353 colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and
354 colonoscopist specialty. J Clin Oncol 2012;**30**:2664-9.
355 17 Dickerman BA, Garcia-Albeniz X, Logan RW, Denaxas S, Hernan MA. Emulating a target trial in
356 case-control designs: an application to statins and colorectal cancer. Int J Epidemiol 2020;**49**:1637-
357 46.
358 18 Greenland S. Quantifying biases in causal models: classical confounding vs collider-
359 stratification bias. Epidemiology 2003;**14**:300-6.
360 19 Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias.
361 Epidemiology 2004;**15**:615-25.

362 20 Meza R, Jeon J, Renehan AG, Luebeck EG. Colorectal cancer incidence trends in the United
363 States and United kingdom: evidence of right- to left-sided biological gradients with implications for
364 screening. Cancer Res 2010;**70**:5419-29.
365 21 Song M, Bretthauer M. Interpreting epidemiologic studies of colonoscopy screening for
366 colorectal cancer prevention: understanding the mechanisms of action is key. Eur J Epidemiol
367 2023;**38**:929-31.
368 22 Brenner H, Heisser T, Cardoso R, Hoffmeister M. When gold standards are not so golden:
369 prevalence bias in randomized trials on endoscopic colorectal cancer screening. Eur J Epidemiol
370 2023;**38**:933-7.
371 23 Rasouli B, Chubak J, Floyd JS, Psaty BM, Nguyen M, Walker RL*, et al.* Combining high quality
372 data with rigorous methods: emulation of a target trial using electronic health records and a nested
373 case-control design. BMJ 2023;**383**:e072346.
374 24 Crockett SD, Nagtegaal ID. Terminology, molecular features, epidemiology, and management
375 of serrated colorectal neoplasia. Gastroenterology 2019;**157**:949-66 e4.

376

377

## Statements & Declarations


### *Funding*

BIPS intramural funding


### *Competing interests*

MB, SS, VD and UH are working at an independent, non-profit research institute, the Leibniz Institute for Prevention Research and Epidemiology – BIPS. Unrelated to this study, BIPS occasionally conducts studies financed by the pharmaceutical industry. These are post-authorization safety studies (PASS) requested by health authorities. The design and conduct of these studies as well as the interpretation and publication are not influenced by the pharmaceutical industry. The study presented was not funded by the pharmaceutical industry and was performed in line with the ENCePP Code of Conduct.

### *Author contributions*

MB: Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing – original draft, writing – review and editing; SS: Conceptualization, investigation, writing – review and editing; VD: Conceptualization, investigation, methodology, supervision, writing – review and editing; UH: Conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, writing – original draft, writing – review and editing.

### *Ethics approval & consent to participate*

In Germany, the utilisation of health insurance data for scientific research is regulated by the Code of Social Law. All involved health insurance providers as well as the German Federal Office for Social Security and the Senator for Health, Women and Consumer Protection in Bremen as their responsible authorities approved the use of GePaRD data for this study. Informed consent for studies based on claims data is required by law unless obtaining consent appears unacceptable and would bias results, which was the case in this study. According to the Ethics Committee of the University of Bremen studies based on GePaRD are exempt from institutional review board review.

### *Data availability*

As we are not the owners of the data we are not legally entitled to grant access to the data of the German Pharmacoepidemiological Research Database. In accordance with German data protection regulations, access to the data is granted only to employees of the Leibniz Institute for Prevention Research and Epidemiology – BIPS on the BIPS premises and in the context of approved research projects. Third parties may only access the data in cooperation with BIPS and after signing an agreement for guest researchers at BIPS.


### *Acknowledgements*

We thank Anja Gabbert and Inga Schaffer for the programming of analysis datasets.

## Tables and Figures

*Table 1: Results of cohort study designs without and with alignment at time zero (adjusted for baseline covariates).*

| | ***Control group*** | ***Screening group*** | ***Adjusted relative risk*** | ***(95% CI)*** |
|---|---|---|---|---|
| Design without alignment at time zero | | | | |
| Number at risk | 200,000 | 200,000 | | |
| Number of CRC cases | | | | |
| Distal CRC | 2,472 | 829 | 0.35 | (0.32-0.39) |
| Proximal CRC | 1,290 | 823 | 0.63 | (0.57-0.69) |
| Design with alignment at time zero | | | | |
| Number at risk | 1,241,071 | 192,054 | | |
| Number of CRC cases | | | | |
| Distal CRC | 16,750 | 1,678 | 0.68 | (0.63-0.73) |
| Proximal CRC | 8,548 | 919 | 0.72 | (0.65-0.80) |

*Table 2: Results of case-control study designs without and with alignment at time zero.*

| ***Site*** | ***Case status*** | | ***Adjusted OR*** | |
|---|---|---|---|---|
| | ***Cases*** | ***Controls*** | | ***(95% CI)[§]*** |
| Design without alignment at time zero | | | | |
| Number of distal CRCs / controls | 446 | 2,230 | | |
| Thereof exposed | 36 | 653 | 0.23 | (0.16-0.33) |
| Number of proximal CRCs / controls | 302 | 1,510 | | |
| Thereof exposed | 54 | 434 | 0.54 | (0.39-0.75) |
| Design with alignment at time zero | | | | |
| Number of distal CRCs / controls | 8,382 | 40,925 | | |
| Thereof exposed | 799 | 5,695 | 0.73 | |
| Number of proximal CRCs / controls | 4,463 | 22,175 | | |
| Thereof exposed | 409 | 3,013 | 0.67 | |

[§:] Confidence intervals could not be obtained for the case-control analysis with alignment at time zero due to computational limitation (see methods section).
OR: Odds Ratio

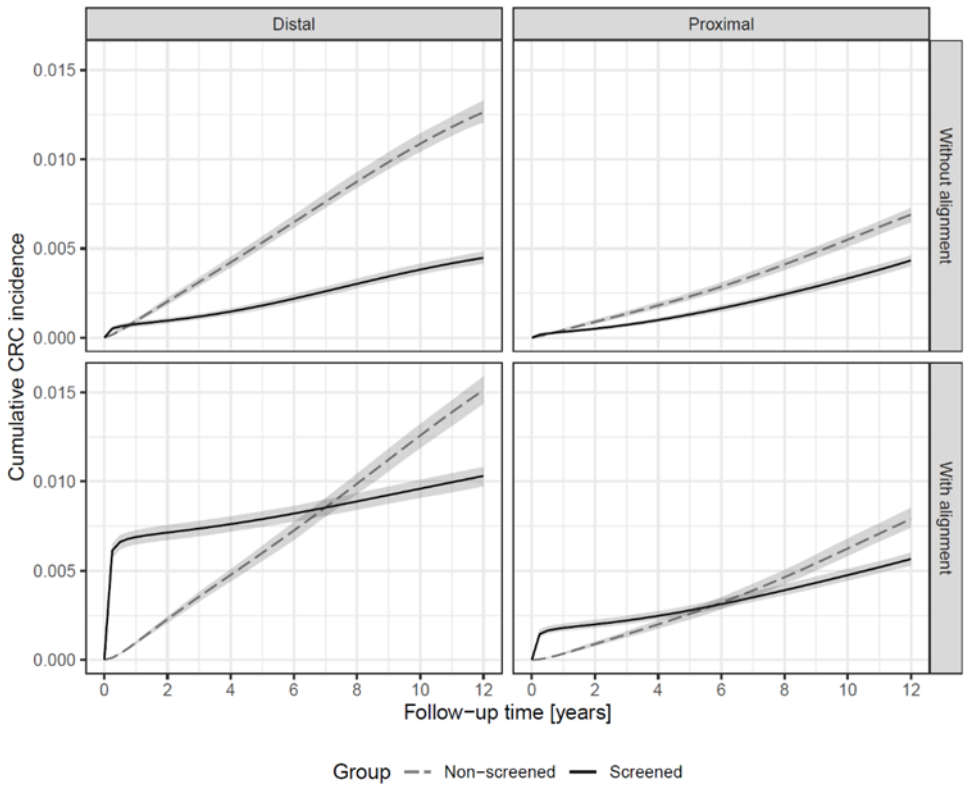


***Fig. 1**: Adjusted cumulative incidence functions for distal and proximal CRC from the cohort study design without alignment at time zero (top row) and the cohort study design with alignment at time zero (bottom row)*

# Supplement

## Supplement 1: Cohort and case-control study without alignment at time zero for different baseline years

As mentioned in the methods section, for the study designs without alignment at time zero, we selected individuals from the source population entering the cohort in 2009. In sensitivity analyses, we varied the baseline year, i.e. individuals entering the cohort in 2010 and 2011, respectively. The respective results are shown in Table S1 and Figure S1 for the cohort study and in Table S2 for the case-control study. For comparison, also the results of the base case analysis (baseline year 2009) are shown.

*Table S1: Results of cohort study designs without alignment at time zero for different baseline years.*

| ***Baseline year*** | | ***Control group*** | ***Screening group*** | ***Adjusted relative risk*** | ***(95% CI)*** |
|---|---|---|---|---|---|
| | Number at risk | 200,000 | 200,000 | | |
| | Number of CRC cases | | | | |
| 2009 | Distal, n | 2,472 | 829 | 0.35 | (0.32-0.39) |
| | Proximal, n | 1,290 | 823 | 0.63 | (0.57-0.69) |
| 2010 | Distal, n | 2,101 | 709 | 0.35 | (0.32-0.39) |
| | Proximal, n | 1,131 | 702 | 0.61 | (0.56-0.68) |
| 2011 | Distal, n | 2,048 | 642 | 0.33 | (0.30-0.37) |
| | Proximal, n | 1,056 | 640 | 0.54 | (0.49-0.59) |

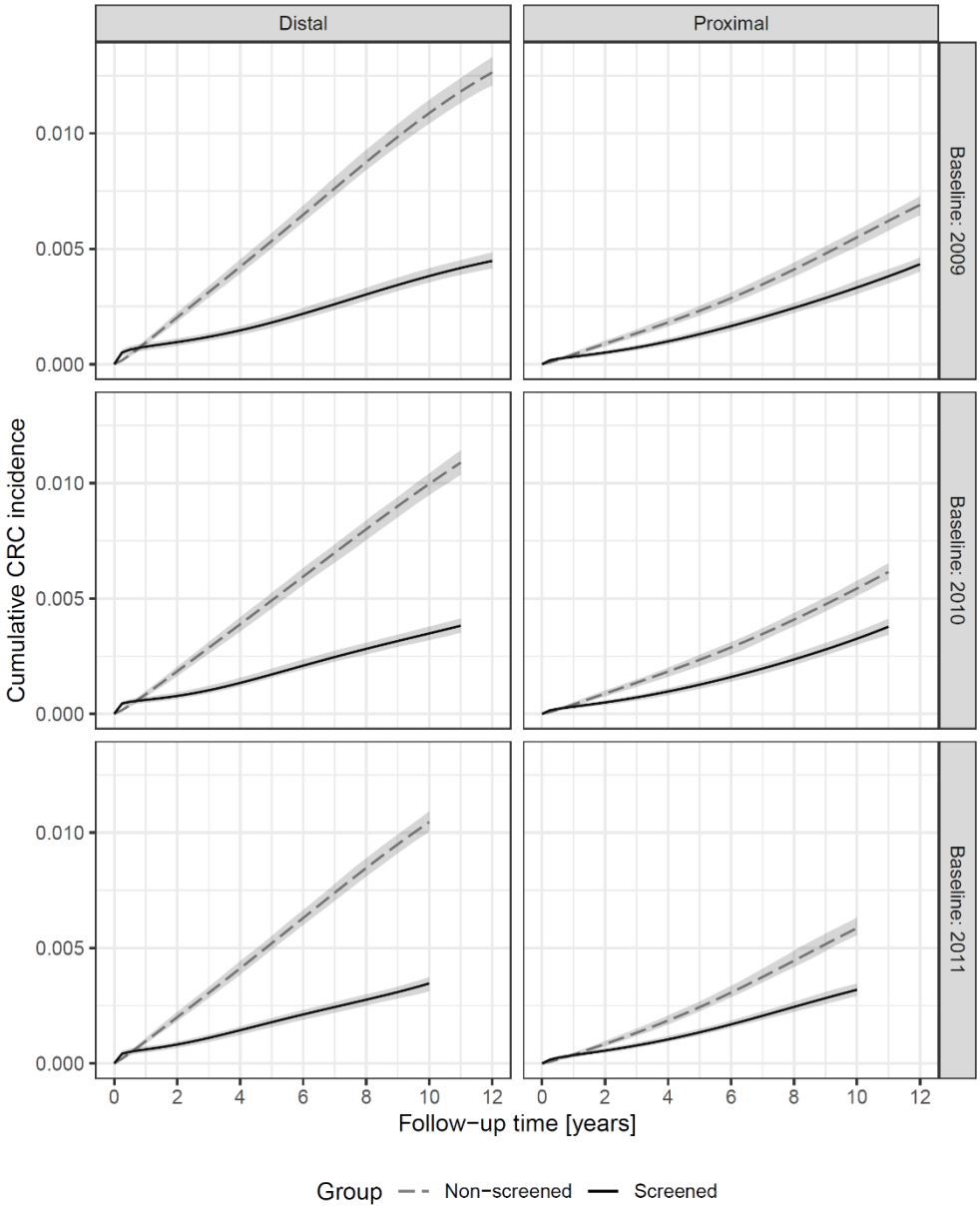


*Figure S1: Adjusted cumulative incidence functions for distal and proximal CRC from the cohort study design without alignment at time zero for different baseline years.*

*Table S2: Results of case-control designs without alignment at time zero for different baseline years.*

| ***Site*** | ***Case status*** | | ***Adjusted OR*** | |
|---|---|---|---|---|
| | ***Cases*** | ***Controls*** | | ***(95% CI)*** |
| Baseline year 2009 | | | | |
| Number of distal CRCs / controls | 446 | 2,230 | | |
| Thereof exposed | 36 | 653 | 0.23 | (0.16-0.33) |
| Number of proximal CRCs / controls | 302 | 1,510 | | |
| Thereof exposed | 54 | 434 | 0.54 | (0.39-0.75) |
| Baseline year 2010 | | | | |
| Number of distal CRCs / controls | 430 | 2,150 | | |
| Thereof exposed | 29 | 607 | 0.19 | (0.13-0.29) |
| Number of proximal CRCs / controls | 264 | 1,320 | | |
| Thereof exposed | 38 | 345 | 0.46 | (0.32-0.68) |
| Baseline year 2011 | | | | |
| Number of distal CRCs / controls | 408 | 2,040 | | |
| Thereof exposed | 24 | 500 | 0.20 | (0.13-0.31) |
| Number of proximal CRCs / controls | 254 | 1,270 | | |
| Thereof exposed | 31 | 298 | 0.47 | (0.31-0.71) |

## Supplement 2: Cohort and case-control study without alignment at time zero: considering both screening and diagnostic colonoscopy for the assignment to exposure groups

As mentioned in the methods section regarding the study designs without alignment at time zero, only screening colonoscopies were considered for the assignment to exposure groups in the base case analysis. In a sensitivity analysis, we considered both screening and diagnostic colonoscopies for the exposure assignment. The respective results are shown in Table S4 and Figure S2 for the cohort study design and in Table S5 for the case-control study.

*Table S4: Results of cohort study designs without alignment at time zero: sensitivity analyses considering both screening and diagnostic colonoscopy for the assignment to exposure groups. For comparison, also the results of the base case analysis are shown.*

| | ***Control group*** | ***Screening group*** | ***Adjusted relative risk*** | ***(95% CI)*** |
|---|---|---|---|---|
| Base case analysis | | | | |
| Number at risk | 200,000 | 200,000 | | |
| Number of CRC cases | | | | |
| Distal, n | 2,472 | 829 | 0.35 | (0.32-0.39) |
| Proximal, n | 1,290 | 823 | 0.63 | (0.57-0.69) |
| Sensitivity analysis | | | | |
| Number at risk | 200,000 | 200,000 | | |
| Number of CRC cases | | | | |
| Distal, n | 2,657 | 982 | 0.40 | (0.37-0.44) |
| Proximal, n | 1,348 | 893 | 0.66 | (0.60-0.72) |

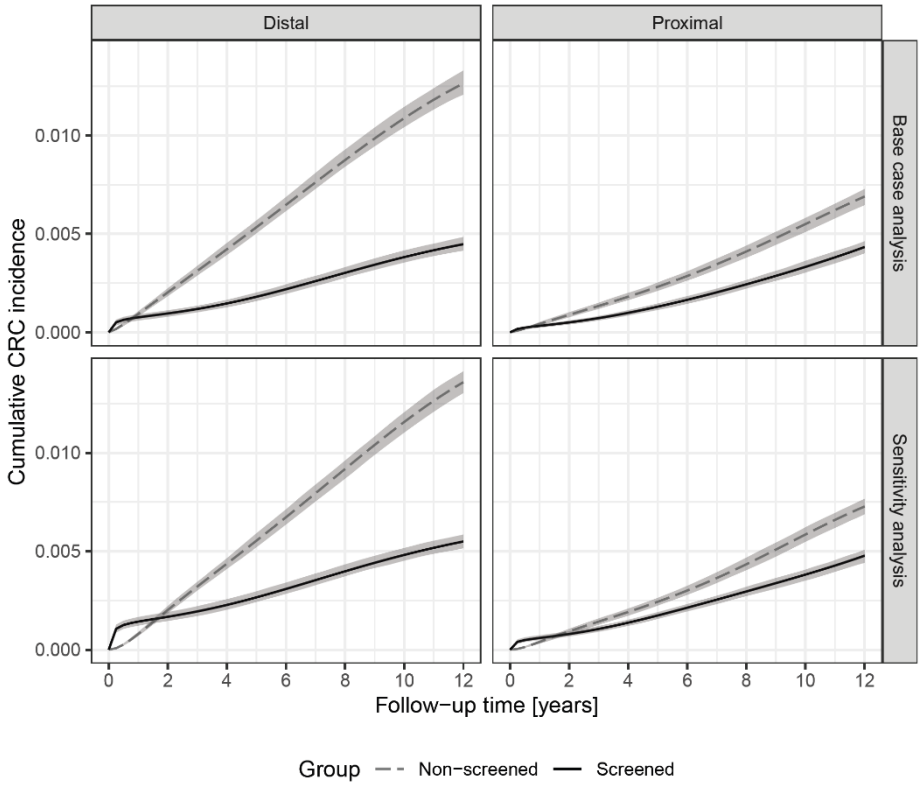


*Figure S2: Adjusted cumulative incidence functions for distal and proximal CRC from the cohort study design without alignment at time zero: sensitivity analysis considering both screening and diagnostic colonoscopy for the assignment to exposure groups. For comparison, also the cumulative incidence functions of the base case analysis are shown.*

*Table S5: Results of case-control designs without alignment at time zero: sensitivity analyses considering both screening and diagnostic colonoscopy for the assignment to exposure groups. For comparison, also the results of the base case analysis are shown.*

| ***Site*** | ***Case status*** | | ***Adjusted OR*** | |
|---|---|---|---|---|
| | ***Cases*** | ***Controls*** | | ***(95% CI)*** |
| Base case analysis | | | | |
| Number of distal CRCs / controls | 446 | 2,230 | | |
| Thereof exposed | 36 | 653 | 0.23 | (0.16-0.33) |
| Number of proximal CRCs / controls | 302 | 1,510 | | |
| Thereof exposed | 54 | 434 | 0.54 | (0.39-0.75) |
| Sensitivity analysis | | | | |
| Number of distal CRCs / controls | 446 | 2,230 | | |
| Thereof exposed | 60 | 1,020 | 0.20 | (0.15-0.26) |
| Number of proximal CRCs / controls | 302 | 1,510 | | |
| Thereof exposed | 82 | 685 | 0.44 | (0.33-0.58) |

## Supplement 3: Structural explanation of the bias inherent to study designs using pre-baseline information for the assignment to exposure groups

For simplicity, we divide time into three periods $t \in \{-1, 0, 1\}$ with $t = -1$ being the pre-baseline period, $t = 0$ the baseline and $t = 1$ the post-baseline or follow-up period. Let $E_t \in \{0, 1\}$ described a person's exposure to screening colonoscopy at time $t$. Let $P_t \in \{0, 1\}$ indicate the presence of colorectal precursors at time $t$ and $C_t \in \{0, 1\}$ the onset of preclinical CRC by time $t$. Let $Y_t \in \{0, 1\}$ indicate a diagnosis of colorectal cancer by time $t$. Finally, let $S = 1$ denote selection into the study cohort.

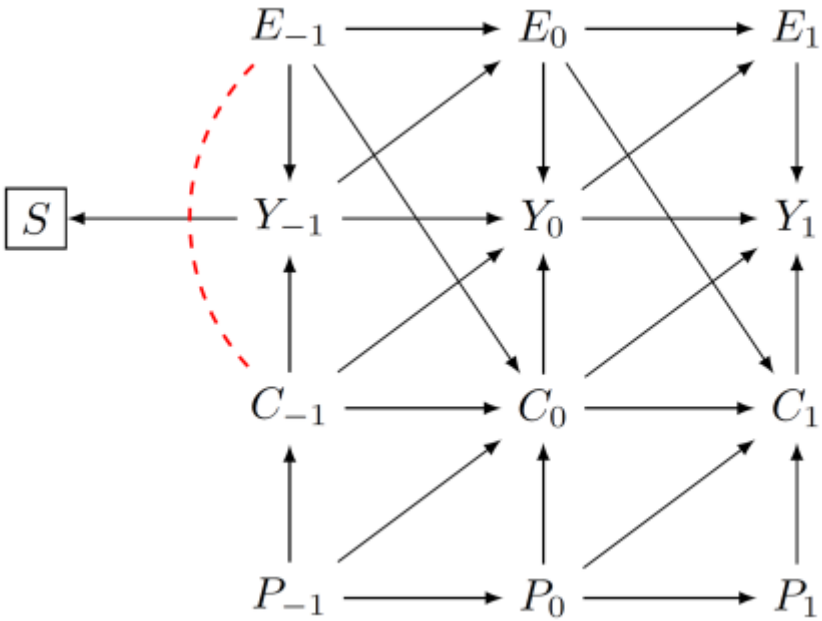


*Figure S3: DAG of bias resulting from violation of alignment at time zero in the form of exposure assessment based on pre-baseline information.*

As shown in Figure S3, at time point $t$ the causal mechanism that leads to a diagnosis of CRC is as follows: Precursors $P_t$ lead to the development of CRC $C_t$, which in turn progress to the outcome of interest, CRC diagnosis $Y_t$. At the same time, exposure to screening colonoscopy $E_t$ leads to CRC diagnosis $Y_t$ at the same time point, if the disease is present. Furthermore, exposure at time $t$ prevents disease onset at the later time $t + 1$ by removing precursor stages present at time $t$.

Importantly, the variable $Y_t$ is a collider variable on the path $P_t \rightarrow C_t \rightarrow Y_t \leftarrow E_t$. When cohort selection $S$ is based on this collider, a non-causal association is introduced between $E_t$ and $C_t$. If the cohort selection process excludes individuals with CRC diagnosis before baseline ($Y_{-1}$) while including individuals with past exposure $E_{-1}$ in the exposed group of the analysis dataset, the unexposed group appears to have a higher CRC incidence. Individuals who were screened in the past and had prevalent CRC received a diagnosis and were filtered out of the study cohort. Individuals who were screened in the past and did not have prevalent CRC are included in the exposed group. No such selection takes place in the unexposed group, where individuals must not have had any screening colonoscopy before baseline. Therefore, there is a non-causal association between exposure before baseline and prevalent, undiagnosed CRC before baseline. This non-causal association means that there are now open backdoor paths from

exposure before baseline to the study outcome at later time points. The resulting bias, therefore, can be expressed as a form of collider stratification bias.

Importantly, the strength of the bias will depend on the prevalence of $C_{-1}$. If, conceptually, the prevalence of CRC before baseline were to approach zero, no such selection would take place. In the age group under study here, the prevalence of proximal CRC before baseline will be much lower than the prevalence of distal CRC before baseline, which means that this bias will impact the effect estimate for distal CRC more severely.

## Supplement 4: Illustration of the mechanism underlying the misallocation of screen-detected CRCs in case-control studies without alignment at time zero

Figure S4 illustrates the mechanism that underlies the misallocation of screen-detected CRCs in case-control studies without alignment at time zero, resulting in an overestimate of the effectiveness of screening colonoscopy. First, let us imagine a hypothetical RCT investigating the effectiveness of screening colonoscopy on CRC incidence. At baseline, screening-naïve persons are randomly assigned to either the screening or the control arm. Analysing this data as a case-control study without alignment at time zero would mean that for CRCs occurring in both arms, it is assessed whether they had a colonoscopy *before* CRC diagnosis. Given that screen-detected CRCs did not have a colonoscopy *before* CRC diagnosis, they are assigned (post-baseline, i.e. after randomization) to the control arm and are thereby classified as unexposed. This overestimates the effectiveness of screening given that CRCs accumulate in the control group (unexposed group). Given that screen-detected CRCs are more frequent in the distal colorectum, the resulting bias affects distal CRC more severely than proximal CRC.

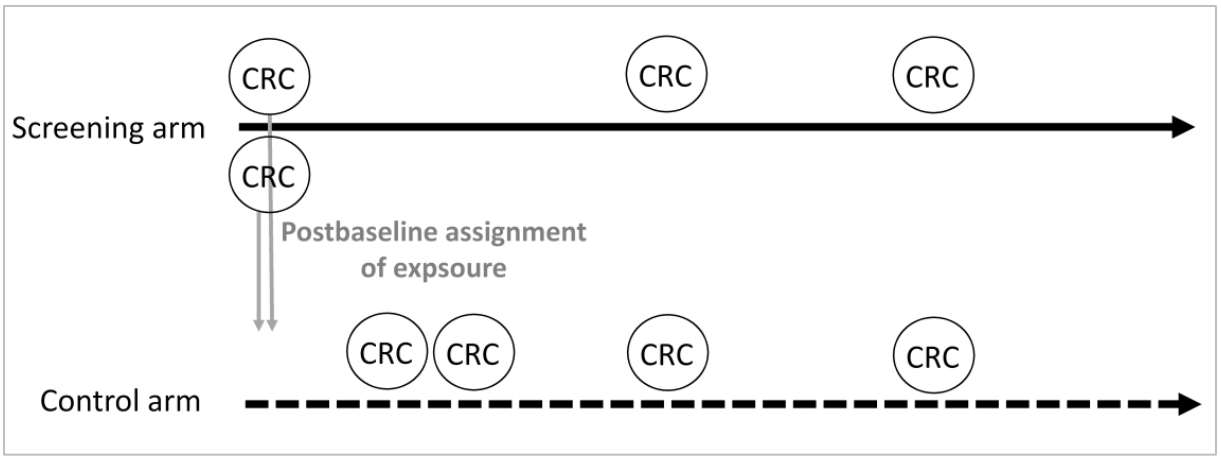


*Figure S4: Illustration of the mechanism of misallocation of screen-detected CRCs in case-control studies without alignment at time zero*

Of note, in published case-control studies investigating the effectiveness of screening colonoscopy based on primary data, selection bias in the control arm (higher prevalence of screening colonoscopy as compared to the general population) can—as an additional mechanism—also contribute to overestimating the effectiveness of screening colonoscopy, but it is not expected that this bias leads to a difference in the effectiveness by site.

In our case-control study without alignment at time zero, there was a second mechanism leading to overestimating the effectiveness of screening colonoscopy due a compromise we had to make because of the left truncation of our data. Specifically, we had to select CRC cases diagnosed in 2018-2020 from those entering the cohort in 2009 (see methods section) in order to be able to assess exposure in the 10 years prior to CRC diagnosis. CRCs diagnosed between 2009 and 2017 in the context of screening, which are more often in the distal than in the proximal colon, were not included in the final set of cases, i.e. distal CRCs exposed to screening colonoscopy were underrepresented in the final set of cases. We conducted additional analyses to disentangle the effect of both mechanism (data not shown), which did not change our conclusion, i.e. that the mechanism described in Figure S4 (also) leads to an artificial difference in the effectiveness of colonoscopy by site.

## Supplement 5: Bias due to post-baseline information for exposure assignment in a cohort study

In the cohort study by Nishihara et al. the assessment of eligibility criteria (e.g. no prior cancer except for nonmelanoma skin cancer, no prior endoscopy) as well as the start of follow-up was in 1988 (baseline). As part of a questionnaire administered every 2 years, participants were then asked whether they had undergone either sigmoidoscopy or colonoscopy and, if so, the reason for the investigation and whether there was a diagnosis of colorectal polyps. This means that the assignment to exposure groups used information after the assessment of eligibility and the start of follow-up, and it was updated every two years, i.e. post-baseline information was used to determine exposure. The outcome was the incidence of colorectal cancer, which was compared between participants without a lower endoscopy (control group), participants with a polypectomy, participants with a negative sigmoidoscopy and participants with a negative colonoscopy.

The mechanism described for the case-control study without alignment at time zero also applies to this design. In each two-year time interval CRCs detected in persons who had their first colonoscopy during this two-year time interval are—per definition—assigned to the unexposed group as they had no colonoscopy prior to CRC diagnosis. This overestimates the effectiveness of screening because CRCs are filtered to the unexposed group. As the majority of screen-detected CRCs are in the distal colon, this bias predominantly affects CRCs in the distal colon and thus leads to an artificial difference in the effectiveness of colonoscopy by site.

## Supplement 6: Post-colonoscopy CRC diagnoses

For the cohort analysis with alignment at time zero, we quantified the occurrence of post-colonoscopy CRC diagnoses occurring in the screening arm and assessed their site distribution. CRC diagnoses with a screening colonoscopy in the same calendar quarter or in the 180 days before CRC diagnosis were considered screen-detected and were not counted as post-colonoscopy CRC. The frequencies and percentages are given in the below Table:

| ***Site*** | ***N*** | ***%*** |
|---|---|---|
| Distal CRC | 541 | 39.3 |
| Proximal CRC | 633 | 46.0 |
| Both/unknown | 203 | 14.7 |
| Total | 1377 | |