OPEN

Estimating the Effect of Targeted Screening Strategies
An Application to Colonoscopy and Colorectal Cancer

Duncan C. Thomas

Abstract: Screening behavior depends on previous screening history and family members' behaviors, which can act as both confounders and intermediate variables on a causal pathway from screening to disease risk. Conventional analyses that adjust for these variables can lead to incorrect inferences about the causal effect of screening if high-risk individuals are more likely to be screened. Analyzing the data in a manner that treats screening as randomized conditional on covariates allows causal parameters to be estimated; inverse probability weighting based on propensity of exposure scores is one such method considered here. I simulated family data under plausible models for the underlying disease process and for screening behavior to assess the performance of alternative methods of analysis and whether a targeted screening approach based on individuals' risk factors would lead to a greater reduction in cancer incidence in the population than a uniform screening policy. Simulation results indicate that there can be a substantial underestimation of the effect of screening on subsequent cancer risk when using conventional analysis approaches, which is avoided by using inverse probability weighting. A large case–control study of colonoscopy and colorectal cancer from Germany shows a strong protective effect of screening, but inverse probability weighting makes this effect even stronger. Targeted screening approaches based on either fixed risk factors or family history yield somewhat greater reductions in cancer incidence with fewer screens needed to prevent one cancer than population-wide approaches, but the differences may not be large enough to justify the additional effort required. See video abstract at, http://links.lww.com/EDE/B207.
(*Epidemiology* 2017;28: 470–478)

Because screening behavior depends on a number of time-dependent factors like previous screening history and possibly family members' behaviors, these factors can act as both confounders and intermediate variables on a causal pathway from screening to disease risk. Thus, conventional analyses that simply adjust for these variables can lead to incorrect inferences about the causal effect of screening.[1] For example, a positive result from a screening test (one's own or a family member's) might increase the frequency of screening in the future, conceivably making screening appear to be positively associated with subsequent cancer, even though the effect is actually beneficial.

To address this problem, Robins and collaborators[2–4] have suggested inverse probability weighting to analyze observational data as if it were a sequence of trials in which screening is applied at random, conditional on covariates, rather than self-selected. Essentially, each event and person time is weighted inversely by the probability of the observed screening history at that time (conditional on past history) to produce a "pseudo-population" in which screening assignment is unrelated to past measured history. The association between screening and subsequent cancer risk can then be analyzed directly in this pseudo-population without any confounding by measured risk factors. These complications do not arise in the evaluation of one-time only screening programs (e.g., Refs. 5 and 6), so they are not considered further. The targeted screening programs considered here are similar to those in various on-going studies of dynamic screening programs.[7–10]

These issues are illustrated by our previous use of a risk prediction model for colorectal cancer based on 27 variants from genome-wide association studies, family history, and

Editor's Note: A commentary on this article appears on p. 479.
Submitted 18 May 2016; accepted 28 March 2017.
From the Department of Preventive Medicine, University of Southern California, Los Angeles, CA.
Funded by National Institutes of Health Grants P30 CA014089, U19 CA148107, P30 ES 07048, R01 ES019876, and P01 CA196569. The methodologic work was supported in part by NIH Grants R01-ES19876 and P50-CA014098 (DCT). The DACHS study is supported by grants from the German Research Council (Deutsche Forschungsgemeinschaft, grants BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, and CH 117/1-1) and the German Federal Ministry of Education and Research (grants 01KH0404, 01ER0814, and 01ER1505A). The sponsors had no role in the design and conduct of the study.
The author reports no conflicts of interest.
Computer code for the simulation is provided in the Supplementary material (http://links.lww.com/EDE/B194). A version for the analysis of the real data analyses is available from the author on request. Request for access to the DACHS study data should be directed to Drs. Chang-Claude, Hoffmeister, and Brenner and will be based on established consortium data sharing policies. Updated recommendations for colonoscopy screening and new related research are also available in the supplementary material.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).
Correspondence: Duncan Thomas, Department of Preventive Medicine, University of Southern California, 2001 N. Soto Street, SSB-201F, Los Angeles, CA 90089. E-mail: dthomas@usc.edu.

ISSN: 1044-3983/17/2804-0470
DOI: 10.1097/EDE.0000000000000668

colonoscopy history to compare population-wide screening beginning at age 50 with a genetically targeted approach.[11] We computed the ages at start of screening that would yield the same risk as the population average and found that for men, this ranged from age 42 for those at the highest risk to 52 for those at lowest risk, with a similar differential for women about 5 years later. However, this analysis did not consider the dynamics of subsequent exams depending on the outcomes of prior ones.

The primary goal of this study was to compare methods for estimating the causal effect of screening for a precursor lesion on the subsequent risk of cancer incidence in observational epidemiology studies. I also sought to assess whether a targeted screening approach in which screening schedules depend on individuals' risk factors like a genetic risk index[11–13] would lead to a greater reduction in cancer incidence than a uniform population-wide one. The motivating example concerns the use of colonoscopy to screen for polyps (adenomas) as a precursor of colorectal cancer.[11,14,15]

Here, I illustrate the approaches first by simulation and then using data from the Darmkrebs: Chancen der Verhütung durch Screening (DACHS) study,[14,16] a large, on-going population-based case–control study with exquisitely detailed information on screening histories for colorectal adenomas (polyps), as well as various risk factors.

SIMULATION STUDY

Simulation Model

The conceptual model for the interplay between the underlying polyp/cancer disease process and the social dynamics of screening behavior is shown in Figure 1. The term "polyp" is used throughout to refer to an adenoma and "cancer" to a diagnosed colorectal carcinoma, whether clinically or screen detected. In simulating adenoma–carcinoma progression, any undiagnosed malignant clone is called a "tumor."

Building on our earlier simulation,[17] I generated a cohort of 5,000 sibships of size 5. For each individual, two measured covariates and two unobserved frailties were generated as a multivariate normal deviate with correlations between the two components and among members of the same sibship. The first of these components influences the biologic behavior of the polyps, and the second influences the social behavior of the family's screening behavior. I also generated a random age at censoring for each individual, independent of risk factors.

The disease process was simulated first (eAppendix 1; http://links.lww.com/EDE/B194). A large number of times to polyp development were generated for each subject under the Armitage–Doll[18] multistage model at rates that depend on the observed and latent disease risk factors. Each polyp was then allowed to grow as the square root of time,[19] at a rate that also depend on risk factors, and each cell in each polyp could undergo malignant conversion as another multistage process. Once fully malignant, tumors were assumed to grow exponentially at a randomly chosen rate until they reached a randomly chosen size that would be clinically detectable. Age at cancer incidence is the first age at which any tumor reaches that clinically detectable size, if prior to the age of censoring.

Screening histories were generated in two stages, first for each individual's initial colonoscopy at an age that depends only on fixed covariates, then for each subsequent colonoscopy at intervals that depend on fixed covariates as well as the individual's and sibs' prior histories of screens and their outcomes (eAppendix 2; http://links.lww.com/EDE/B194). For each individual, a time to the first recommended colonoscopy was assigned along with a probability of actually complying, and after each subsequent exam, a time to the next one and compliance probability was also assigned. At each exam, each previously undetected polyp could be detected with probability given by a logistic function of its size. Age at diagnosis of cancer then becomes the earliest of the previously simulated times at which each tumor would be clinically diagnosed, if not previously screen detected as a polyp and if prior to the age at censoring. In addition, some malignant tumors might be detected at screening ("screen-detected cancers"). This process continues, generating the next scheduled colonoscopy within each family, based on the history up to that point in time, until each person has reached their preassigned ages at censoring or cancer diagnosis, whichever comes first.

Specific values for each of the parameters in the disease and screening models are given in eTable 1 (http://links.lww.com/EDE/B194). Only the observable data for each individual were used in the analysis, specifically the observed risk covariates, the screening times and whether polyps were detected at each, and the time to diagnosis of cancer or censoring, whichever came first. The computer code for the simulation, analysis and tabulation routines are provided in eAppendix 7 (http://links.lww.com/EDE/B194).

The parameters of the simulation model were chosen by trial and error to produce reasonable distributions of various summary statistics, such as the age distribution of polyps, mutations, and cancers, the ages at and outcomes of screening, and associations with subjects' observed risk factors and family histories.

Estimation Methods

The analysis of the relationship of cancer risk to past screening history was based on a standard case–control study design nested within the simulated cohort. (Although the full simulated cohort could have been analyzed, this was done for comparability with the DACHS study, which used a population-based case–control design. The resulting odds ratio parameters are consistent estimators of the corresponding hazard rates from a cohort study under the incidence density sample scheme described here.[20]) Each clinically detected incident cancer case was randomly matched with a single control from the cohort who had survived to the case's age at diagnosis (the "reference date" for the matched pair) still undiagnosed at that time. Other than sampling from each

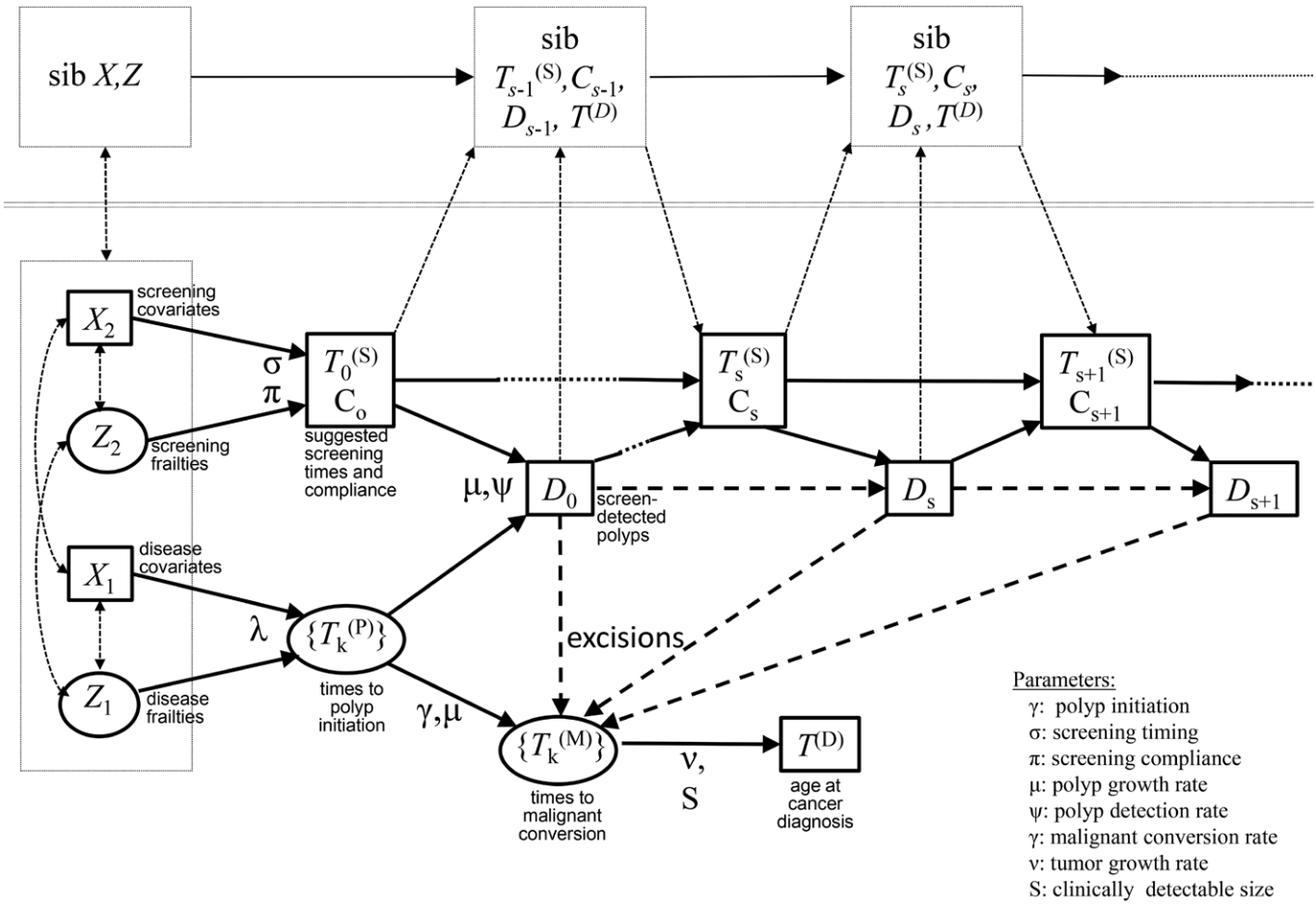

FIGURE 1. Conceptual model for polyp and cancer development and screening behavior. Squares represent measured variables and circles represent unobservable latent variables. The solid arrows show causal influences with rates given by the Greek letters, which themselves depend on the quantities at the heads of the corresponding arrows. Dashed lines represent structural events that affect at-risk status for the indicated events, for example, discovery of a polyp by screening leading to excision and hence removal from risk of subsequent rediscovery or malignant conversion. The first screening and its outcome is shown in the boxes labeled ($T_0^{(S)}$,C_0) and D_0, respectively. Thereafter, only two screening events at time s and $s + 1$ are shown (or $s - 1$ and s for sibs). Above the pair of dotted lines is a summary of data for other members of the family, with light dashed lines showing their influence on another family member, shown in greater detail.

case's Cox "risk sets," no other matching was done. Cases and controls were then compared in terms of the number of screens completed over various time intervals before the reference date, using conditional logistic regression.

For causal inference, the first step is the estimation of propensity scores, the probability that an individual is screened at each age as a function of their history. A logistic model was fitted to a dataset in which each individual is represented once for every year up to censoring or cancer diagnosis. The model was fitted twice, once with only baseline risk factors and once with the time-dependent covariates, and "stabilized inverse probability weights"[2–4] were computed as the ratio of these two sets of predicted propensity scores (for the exposed or one minus the scores for the unexposed) (eAppendix 3; http://links.lww.com/EDE/B194). These are used as weights in the conditional logistic regression model for disease. The inverse probability weighting essentially mimics an analysis in which screening history is unconfounded by any (measured) factors that influence screening behavior. Doing this in a time-dependent manner allows for such factors to include prior screening outcomes (for the same individual or family members).

Prediction of the Outcomes of Different Screening Regimens

Screening and cancer outcomes were computed based on the simulated disease process for the original cohort, imposing the following five counterfactual screening regimens:

- No screening.
- Untargeted: all individuals in the population are screened first at age 50, then at 5- or 10-year intervals, and thereafter depending on whether the previous screen was positive.

- Targeted: individuals' ages at first screen and intervals between screens depend on their fixed risk index value, as described below.
- Family history based: similar to the targeted schedule, except that individuals' timing depends instead on their evolving family history of cancer,
- Both: similar to the targeted schedule depending on both the fixed risk index and family history

and were compared with the results for the observed simulation. The timing of the first screens under the three targeted schemes was based on fitting logistic models for individuals' 10-year risk of cancer based on age and predictors (risk index and/or family history). The first target screen for each individual was then set to the age at which their predicted risk equals the 10-year risk at age 50 for the general population (eAppendix 4; http://links.lww.com/EDE/B194). Following a negative screen, the time to the next screen was set to the predicted time at which the cumulative risk would be the same as the 10-year risk at the age of the last screen. Following a positive screen, the same was done using the 5-year risk among screen-positive individuals.

I assumed these various screening policies affected only the history of screening events, not the underlying disease process, which proceeds exactly the same under all screening programs, the only difference being that discovered polyps are excised at discovery, thereby potentially altering the time at which the first tumor is clinically diagnosed.

For each screening schedule, I tabulated the following measures of performance up to the simulated age at censoring:

- The total number of screens performed, the proportion positive, and the average number of polyps detected per screen before any cancer diagnosis.
- The total number of clinically diagnosed cancers and the screen detected.
- The false-negative rate: the proportion of all undetected polyps that would produce a clinically detectable cancer within 10 years of screening.
- The false-positive rate: the proportion of all detected polyps that would not produce a clinically detectable cancer before censoring.

Simulation Results

Figure 2 compares the nested case–control study estimates of the effect of the number of previous screens on subsequent cancer risk over various intervals of time before the reference date (the "analysis window"), with or without adjustment for the fixed covariates and the number of positive screens. These effect estimates are expressed as odds ratios per screen conducted during the indicated window. When the inverse probability weights were used, the associations were considerably stronger than without them.

Table 1 summarizes the predictions of potential outcomes under the various counterfactual screening programs for the analysis window 1−10 years prior, the interval chosen by the DACHS investigators in their previous publications. In general, any of the systematic approaches, whether targeted or not, required more screens than actually observed (mainly because subjects were assumed to be fully compliant). All four systematic screening schedules led to fewer clinical cases; the strategy based on using risk factors yielding the

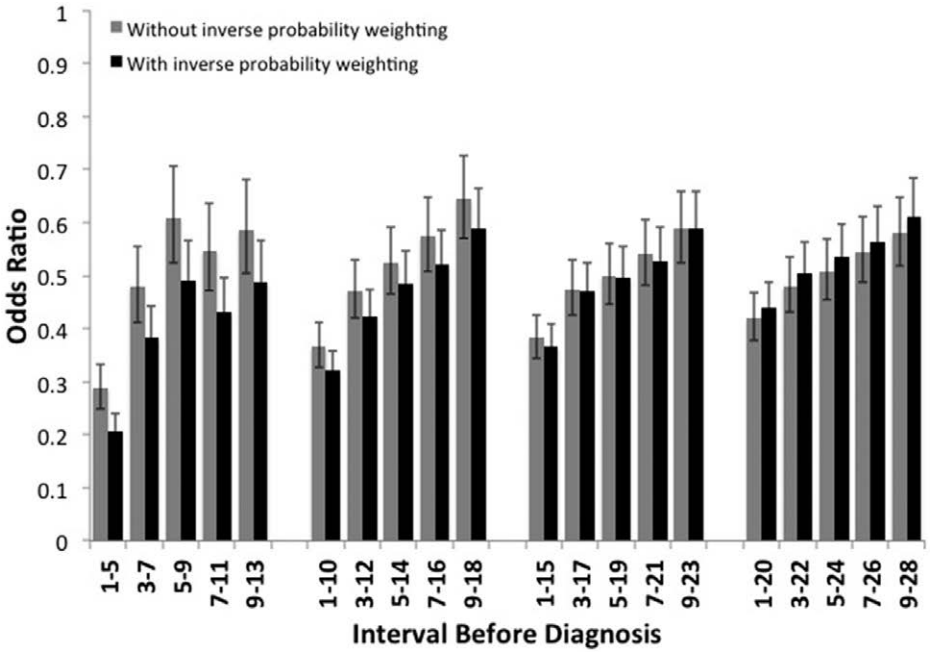

FIGURE 2. Odds ratio estimates per screen for the effect of the colonoscopies on colorectal cancer risk over various windows of time before the reference date (simulated data). Labels on each pair of bars refer to the interval of time before the reference date (the diagnosis of the case) over which the effect of screening is evaluated (the "analysis window"). Shaded bars are estimated without weighting, solid bars using inverse probability weighting.

TABLE 1. Projected Yields from Untargeted and Targeted Screening Regimens Among the 25,000 Simulated Individuals, Using an Analysis Window of 1–10 Years Before Reference Data (Simulated Data; Numbers in Parentheses Are 95% Confidence Intervals)

		Recommended Screening Schedule				
Performance Criterion	**No Screening**	**Observed Behavior**	**Population-wide Untargeted**	**Risk Factor Based**	**Family History Based**	**Risk Factor and Family History Based**
Total number of screens	0	27,996	61,041	75,432	60,209	73,055
Proportion positive	N.R.	23.6% (23.1, 24.1)	19.6% (19.1, 20.1)	16.6% (16.2, 17.1)	19.4% (18.9, 19.9)	16.6% (16.2, 17.1)
Number of clinically diagnosed cancers	2,559 (2,467, 2,654)	1,563 (1,490, 1,640)	906 (850, 966)	820 (767, 877)	915 (859, 975)	847 (793, 905)
Number of screen-detected cancers	0	7 (3, 15)	28 (19, 41)	24 (16, 36)	51 (39, 67)	30 (21, 43)
Cancers prevented	0 (ref)	1,003 (944, 1,066)	1,681 (1,605, 1,760)	1,763 (1,685, 1,844)	1,695 (1,619, 1,775)	1,742 (1,665, 1,823)
NNS	N.R.	27.9 (26.3, 29.7)	36.3 (34.7, 38.0)	42.8 (40.9, 44.8)	35.5 (33.9, 37.2)	41.9 (40.1, 43.9)
False-negative rate	N.R.	1.3% (1.1, 1.4)	1.3% (1.1, 1.5)	1.3% (1.2, 1.5)	1.3% (1.2, 1.5)	1.3% (1.2, 1.5)
False-positive rate	N.R.	75.7% (74.6, 76.8)	37.4% (36.0, 38.7)	50.0% (48.5, 51.5)	39.3% (38.0, 40.7)	49.1% (47.5, 50.6)

NNS indicates number needed to screen; N.R., not relevant.

largest reduction in cancers compared with the observed number. Likewise, the various systematic screening schedules yielded more screen-detected ("early") cancers than observed, although there were many fewer of these than the clinically diagnosed ("late") cancers since screening colonoscopies are aimed primarily at detecting polyps rather than cancers. Thus, defining the number "prevented" as the difference between a systematic program and observed behavior in the number of screen detected minus the clinically diagnosed cases, the number needed to screen (NNS)[21] to prevent one case is the lowest for the observed behavior because of the much lower number of screens performed, with relatively modest differences among the other programs. "False-negative" rates were low across all programs and "false-positive" rates lower than observed for all systematic programs.

In a simulation, it is possible to "observe" directly the potential outcomes of the same individual under alternative scenarios and cross-tabulate their outcomes. eTable 2 (http://links.lww.com/EDE/B194), does this for clinically diagnosed cancers and screen-detected cancers under pairwise comparisons of the observed screening history, population-wide, and fully targeted screening. For example, comparing clinically diagnosed cancers under targeted versus untargeted regimes yielded 1,956 cancers under both regimes, 1,668 under the untargeted regime that would not have occurred under the targeted one, and 1,432 under the targeted regime that would not have occurred under the untargeted one. There is thus a net benefit of 236 fewer cases per 100,000 under the targeted regime. Likewise, there is a net benefit of 8 (120−112) more screen-detected cancers under the targeted regime. As might be expected, the corresponding differences are much larger comparing the predicted outcomes under any of the systematic screening programs against the observed outcomes.

APPLICATION TO THE DACHS STUDY

The DACHS study[14,16] is an ongoing population-based case–control study from Germany with detailed information on screening histories for colorectal adenomas, risk factors (sex, schooling, ever regular smoking, body mass index [5–14 years before the reference date], average metabolic equivalent of task, alcohol [average g/day], regular nonsteroidal anti-inflammatory drug use, hormone replacement therapy, and statins) and family history of colorectal cancer in first and second-degree relatives. No information was asked about screening for family members. Genotype data were not included in this analysis, as larger consortium analyses using some of these data are currently underway and others are still being generated. The current analysis includes 4,334 cases of colorectal cancer and 4,231 controls—about 3,000 additional subjects ascertained between 2011 and 2013 that were not included in previous publications. Details of the selection of data for use in this application are provided in eAppendix 5 (http://links.lww.com/EDE/B194). The original DACHS study was approved by the institutional review boards for the German Cancer Research Center (DKFZ) and participating sites; no additional data were gathered for this analysis, and no ethics review of this analysis was required.

Propensity Score Analyses

The simulation program was adapted to analyze the real data using the entire history of screening and family colorectal cancer data (eAppendix 3; http://links.lww.com/EDE/B194). The parameter estimates are provided in eTable 3 (http://links.lww.com/EDE/B194), and the resulting distributions of stabilized propensity scores for screening are shown in eFigure 1 (http://links.lww.com/EDE/B194). In addition to the effects of the fixed covariates, the probability of screening at any age

was strongly related to the total number of previous screens (lnOR = 0.425 ± 0.050 per screen), the history of positive screens (positively for the first and last and negatively for the total number of other positive screens), and the number of relatives with colorectal cancer.

I also fitted a logistic regression model to the screening outcomes (polyps removed at the first colonoscopy and the last three ones for which data were available). Only sex, smoking, family history, total number of previous screens (negative), and number of previous positive screens (positive) contributed appreciably to this model.

In addition to all the fixed covariates and family history at the reference date, the disease model included the number of colonoscopies over various intervals of time before the reference date as a covariate. As shown in Figure 3, the effect estimates were always negative, but somewhat stronger when using the inverse probability weights. For example, the number of screens 1−10 years previously yields an OR estimate of 0.50 per screen, (95% confidence interval = 0.45, 0.55) without weighting, compared with 0.43 (95% confidence interval = 0.38, 0.48) with weighting. Although this is a smaller effect than that from earlier analyses,[14,16] the units here are per number of screens in the interval rather than for a binary indicator for ever/never screened. Sensitivity analyses using the binary indicator yielded estimates closer to the previous publications, but the comparisons between analysis methods were virtually unchanged.

Evaluation of Outcomes Under Counterfactual Screening Regimens

In the simulations, the entire underlying state of the polyps/cancer process was known, so it was a simple matter to evaluate the effect of intervening on the screening process on the ultimate outcomes. Since the polyps and latent cancer processes are unobserved in the real data, the counterfactual analysis of alternative screening programs required use of simulation based on the fitted models, basically summing over all possible screening and polyp detection histories using the G-computation algorithm.[2] For each of the six screening programs described above, I simulated screening, polyp detection, and cancer histories for each subject 100 times and averaged the various outcome measures across these simulated replicates using a random effects model (eAppendix 6; http://links.lww.com/EDE/B194). The calculations were conducted in the same way as described above for the simulation, with minor differences explained in eAppendices (4–6; http://links.lww.com/EDE/B194).

The distribution of target ages at first screen, based on fixed covariates and/or family history, is shown in Figure 4, and the distributions of target intervals between screens for those with a negative and a positive previous screen are shown in Figure 5. As expected, these distributions are centered at age 50 (for the first screen) and at intervals of 5 and 10 years (for subsequent screening) but show very substantial variation around them.

The predicted outcomes under different screening regimens are summarized in Table 2. In a hypothetical cohort of 100,000 persons with a distribution of risk factors drawn from the case–control study, there would an expected reduction of cancer incidence from 4,413 cases with no screening to 4,225 under self-selected screening or 1,892 cases under uniform population-wide screening. This is further reduced to 1,241 cases by stratifying on family history alone, 756 by stratifying on the fixed covariates alone, or 208 by stratifying on both. Either population-wide or stratified screening required many more screens than the observed self-selected histories but yielded more polyps. Thus, the NNS was lowest (53) for the program that stratifies on both fixed covariates and family

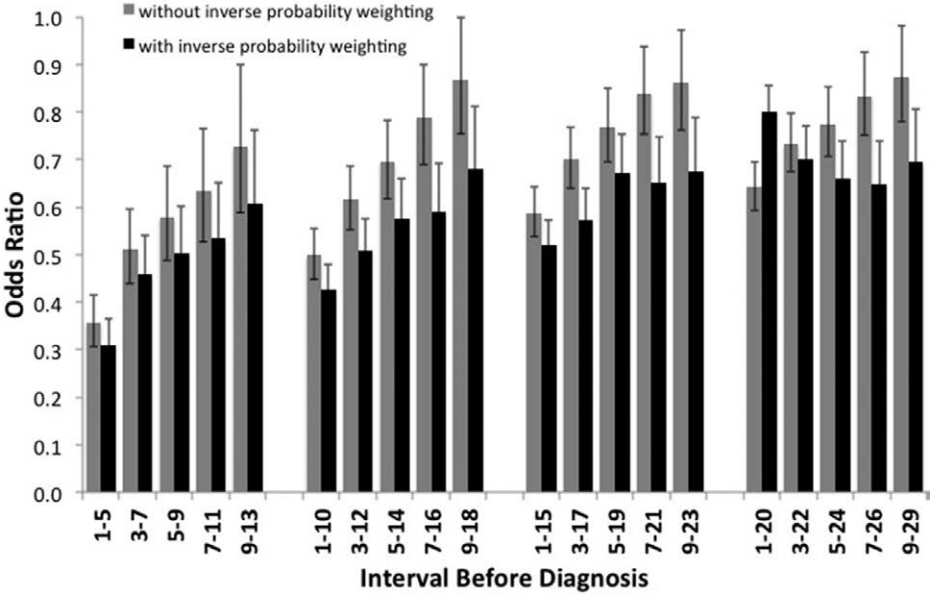

FIGURE 3. Odds ratio estimates per screen for the effect of the colonoscopies on colorectal cancer risk over various windows of time before the reference date (DACHS data). Labels on each pair of bars refer to the interval of time before the reference date (the diagnosis of the case) over which the effect of screening is evaluated (the "analysis window"). Shaded bars are estimated without weighting, solid bars using inverse probability weighting.

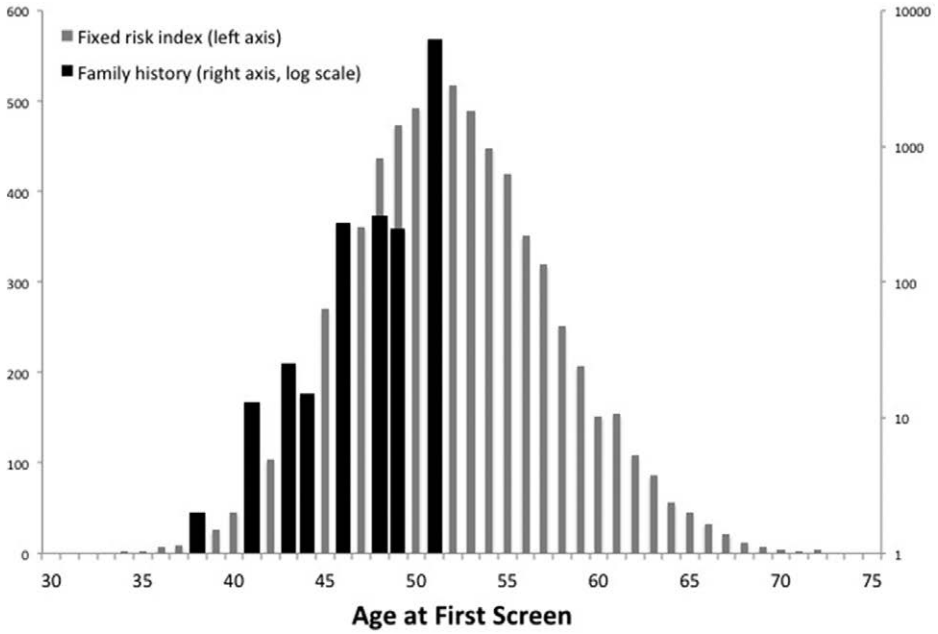

FIGURE 4. Distribution of target ages at first exposure (based only on fixed risk factors, family history, or both (DACHS data): family history-targeted distribution (solid bars) is plotted on the right axis on the log scale to better visualize the small numbers of individuals with positive family histories; the fixed risk index distribution is plotted on the left axis on the natural scale.

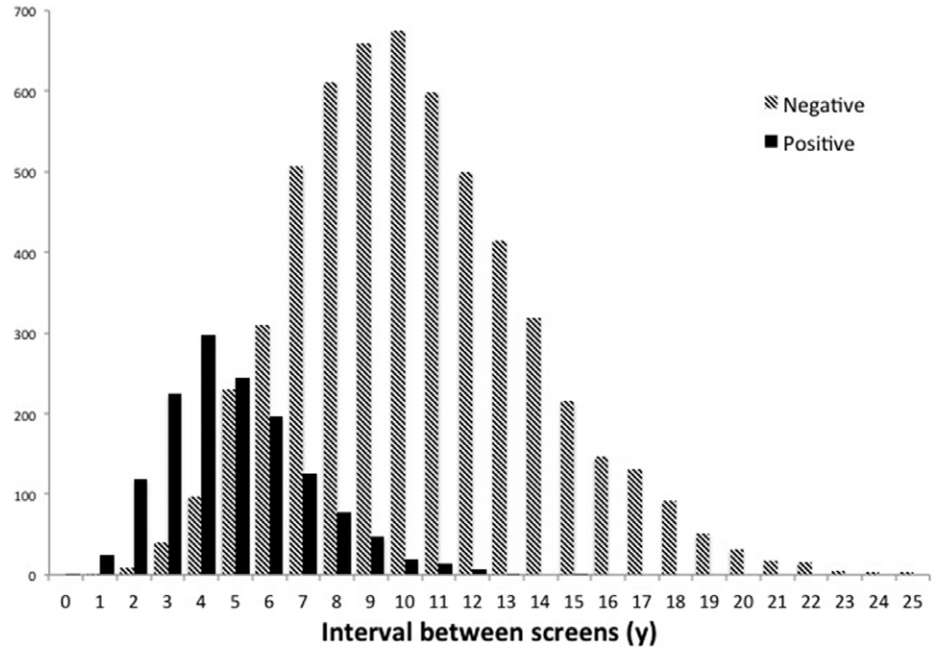

FIGURE 5. Distribution of target intervals between screens based on fixed risk factors and family history: shaded bars, following a negative screen (same as population 10-year risk); solid bars, following a positive screen (same as population 5-year risk)

history or on fixed covariates alone (58), compared with 80 for screening on the basis of family history or 105 for a population-wide screening program.

All the parameter estimates in the screening and polyp detection models and the predicted outcomes of different screening programs take the case–control sampling into account[22] (eAppendix 5; http://links.lww.com/EDE/B194). Because cases are overrepresented in the case–control sample relative to the population (and hence higher risk covariate values are as well), failure to take this into account would have led to overestimates of predicted cancers.

DISCUSSION

The simulation results indicate that there can be a substantial bias in the estimated effect of screening on subsequent cancer incidence when using conventional analysis approaches and treating screening histories and their determinants as time-dependent covariates. Under some scenarios, screening might even appear detrimental despite its truly protective effect. However, in the real data application, the difference was much smaller and no apparently detrimental effect was seen even using the conventional analysis. I conclude that evaluation of dynamic screening strategies needs

TABLE 2. Summary of Predicted Outcomes of Different Screening Programs Using Estimates of Screening Benefits Over the 1–10 Years Before Reference Date Window (DACHS Data)

Screening Program	Mean Total Number of Screens per Person	Positive Screens		Cancers per 100,000 Persons	NNS to Prevent One Cancer
		Mean per Person	Mean per Screen		
No screening	0.00	0.000	(N.R.)	4,413 (3,989, 4,883)	Ref
Observed behavior	0.20 (0.19, 0.21)	0.044 (0.041, 0.048)	0.218 (0.203, 0.234)	4,225 (3,810, 4,685)	108 (103, 114)
Population-wide untargeted	2.64 (2.54, 2.74)	0.519 (0.502, 0.536)	0.196 (0.190, 0.203)	1,892 (1,621, 2,208)	105 (99, 112)
Risk factor based	2.13 (2.06, 2.20)	0.350 (0.337, 0.362)	0.164 (0.158, 0.170)	756 (592, 965)	58 (54, 63)
Family history based	2.53 (2.39, 2.66)	0.416 (0.389, 0.442)	0.164 (0.154, 0.175)	1,241 (1,026, 1,502)	80 (75, 85)
Risk factor and family history based	2.22 (2.19, 2.25)	0.394 (0.381, 0.408)	0.177 (0.172, 0.184)	208 (131, 331)	53 (49, 58)

NNS indicates number needed to screen; N.R., not relevant.

to use some form of causal inference techniques, such as the inverse probability weighting method discussed here. The second goal of the article was to show how counterfactual modeling approaches could be used to compare alternative dynamic strategies, such as population-wide or personalized ones. The results in Tables 1 and 2 suggest that there could be a greater benefit from strategies based on individual's environmental and/or genetic risk factors, although the magnitude of the differences may not be large enough to justify their greater complexity.

There are many assumptions involved in both the screening and the disease process parts of the simulation models. Several other colorectal cancer simulation models have previously been published[23–28] and used for evaluation of screening programs (e.g., Greuter et al.[29] using the ASCCA[24] model or Zauber[30] using the MISCAN[31] and SimCRC[32] models). Although the specific predictions of the model may differ, both in terms of cancer risks and the effects of different screening methods, I expect that for comparing different methods of analysis, any realistic model would produce similar results.

The cost–benefit tradeoffs implicit in the NNS[21] involve more than just its effect on cancer incidence. Although reducing incidence is a worthy goal in itself, an evaluation of mortality must also consider whether early detection (of polyps or cancers) really improves survival. This issue is complicated by the well-known problems in the screening literature[33,34] of lead-time and length bias. In simulations, one could address this by adding a survival outcome but that would require a number of additional modeling assumptions, including the effects of risk factors and family history on survival, the benefit from early detection, the effects of stage at diagnosis, treatment and their indications, and so on. For real data, one would need to collect survival outcomes, along with information on stage and environmental[35,36] and genetic[37,38] indications for and interactions with treatment. Ultimately, only a randomized controlled trial could really answer the question of the effect of screening on mortality.

Another element that belongs in any cost–benefit assessment is false negatives and false positives, which also have real if sometimes unquantifiable costs associated with them. In our simulations, I defined false positives as polyps detected that would not, if left intact, have produced a cancer within 10 years, and a false negative as a potentially detectable polyp missed by screening that would have produced a cancer within the person's lifetime. Such outcomes are not directly observable in real data and can only be estimated probabilistically by simulating the G-computation formula.

Other causal inference techniques were used by Valeri et al.[39] in an analysis of the effect of intervening on stage at diagnosis distributions on racial disparities in survival. Only one article[1] has used propensity score methods to address screening, using data from the Norwegian Colorectal Cancer Prevention (NORCCAP) trial. Their analysis was similar to ours but aimed only at comparing the average causal effects of static versus dynamic screening programs (based on evolving characteristics), not personalized versus population-wide strategies.

For the DACHS study analysis, no data were available on genetics, although this is currently being collected and data from some of the participants were included in the consortium article[11] described earlier. Although there are currently no genes known to have a major effect on colorectal cancer risk, genome-wide association studies have identified 27 loci that in the aggregate are associated with a roughly two-fold range in risk across the quartiles of the distribution of the polygenic risk index, and this will doubtless increase as more loci are discovered. Ultimately, it is likely that a genetic risk score might produce a greater advantage for targeted screening than using family history, although using both types together should be even better.

Colonoscopy is currently considered the most effective means of detection of adenomas, although fetal occult blood testing is widely used in some settings.[40] Although fecal immunochemical tests or DNA-based testing procedures[41] also have been shown to be effective for the detection of colorectal cancer, their sensitivity is much lower for the detection of adenomas[5,42] but will also likely improve over time. Although emphasizing the substantial net benefit of colorectal cancer screening, the US Preventive Services Task Force[43] recently concluded that presently available evidence does not provide a clear basis for choosing among the alternative tests.

ACKNOWLEDGMENTS

The author thanks Drs. Hermann Brenner, Michael Hoffmeister, and Jenny Chang-Claude for providing the DACHS data for the application section and for many helpful suggestions. The author also thanks Viola Walter for data management and consultation in the DACHS study. The author also thanks Graham Byrnes, Martyn Plummer, and Paolo Vineis for discussions over early drafts of the manuscript. Three peer reviewers had many very helpful suggestions that have been incorporated into the final version.

REFERENCES

1. Huitfeldt A, Kalager M, Robins JM, Hoff G, Hernán MA. Methods to estimate the comparative effectiveness of clinical strategies that administer the same intervention at different times. *Curr Epidemiol Rep*. 2015;2:149–161.
2. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
3. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19:766–779.
4. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–570.
5. Quintero E, Castells A, Bujanda L, et al.; COLONPREV Study Investigators. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *N Engl J Med*. 2012;366:697–706.
6. Bretthauer M, Kaminski MF, Løberg M, et al.; Nordic-European Initiative on Colorectal Cancer (NordICC) Study Group. Population-based colonoscopy screening for colorectal cancer: a randomized clinical trial. *JAMA Intern Med*. 2016;176:894–902.
7. Jover R, Bretthauer M, Dekker E, et al. Rationale and design of the European Polyp Surveillance (EPoS) trials. *Endoscopy*. 2016;48:571–578.
8. Cairns SR, Scholefield JH, Steele RJ, et al.; British Society of Gastroenterology; Association of Coloproctology for Great Britain and Ireland. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). *Gut*. 2010;59:666–689.
9. Atkin WS, Valori R, Kuipers EJ, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition—Colonoscopic surveillance following adenoma removal. *Endoscopy*. 2012;44(suppl 3):SE151–SE163.
10. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR; United States Multi-Society Task Force on Colorectal Cancer. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2012;143:844–857.
11. Hsu L, Jeon J, Brenner H, et al.; Colorectal Transdisciplinary (CORECT) Study; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology*. 2015;148:1330–1339.e14.
12. Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer*. 2011;104:1656–1663.
13. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med*. 2008;358:2796–2803.
14. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology*. 2014;146:709–717.
15. Good NM, Suresh K, Young GP, Lockett TJ, Macrae FA, Taylor JM. A prediction model for colon cancer surveillance data. *Stat Med*. 2015;34:2662–2675.
16. Brenner H, Chang-Claude J, Jansen L, Seiler CM, Hoffmeister M. Colorectal cancers occurring after colonoscopy with polyp detection: sites of polyps and sites of cancers. *Int J Cancer*. 2013;133:1672–1679.
17. Schmit SL, Figueiredo JC, Cortessis VK, Thomas DC. The influence of screening for precancerous lesions on family-based genetic association tests: an example of colorectal polyps and cancer. *Am J Epidemiol*. 2015;182:714–722.
18. Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer*. 1957;11:161–169.
19. Dowty JG, Byrnes GB, Gertig DM. The time-evolution of DCIS size distributions with applications to breast cancer growth and progression. *Math Med Biol*. 2014;31:353–364.
20. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol*. 1982;116:547–553.
21. Vineis P, Satolli R. From figures to values: the implicit ethical judgements in our measures of health. *Public Health Ethics*. 2012;5:22–28.
22. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172:1339–1348.
23. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiol Biomarkers Prev*. 2010;19:1992–2002.
24. Greuter MJ, Xu XM, Lew JB, et al. Modeling the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA). *Risk Anal*. 2014;34:889–910.
25. Neilson AR, Whynes DK. Cost-effectiveness of screening for colorectal cancer: a simulation model. *IMA J Math Appl Med Biol*. 1995;12:355–367.
26. Eddy DM, Nugent FW, Eddy JF, et al. Screening for colorectal cancer in a high-risk population. Results of a mathematical model. *Gastroenterology*. 1987;92:682–692.
27. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Comput Biomed Res*. 1999;32:13–33.
28. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2008;149:659–669.
29. Greuter MJ, Demirel E, Lew JB, et al. Long-term impact of the Dutch colorectal cancer screening program on cancer incidence and mortality-model-based exploration of the serrated pathway. *Cancer Epidemiol Biomarkers Prev*. 2016;25:135–144.
30. Zauber AG. The impact of screening on colorectal cancer mortality and incidence: has it really made a difference? *Dig Dis Sci*. 2015;60:681–691.
31. Loeve F, Brown ML, Boer R, van Ballegooijen M, van Oortmarssen GJ, Habbema JD. Endoscopic colorectal cancer screening: a cost-saving analysis. *J Natl Cancer Inst*. 2000;92:557–563.
32. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA*. 2000;284:1954–1961.
33. Cole P, Morrison AS. Basic issues in population screening for cancer. *J Natl Cancer Inst*. 1980;64:1263–1272.
34. Day NE. The assessment of lead time and length bias in the evaluation of screening programmes. *Maturitas*. 1985;7:51–58.
35. Walter V, Jansen L, Hoffmeister M, Ulrich A, Chang-Claude J, Brenner H. Smoking and survival of colorectal cancer patients: population-based study from Germany. *Int J Cancer*. 2015;137:1433–1445.
36. Carr PR, Jansen L, Walter V, et al. Associations of red and processed meat with survival after colorectal cancer and differences according to timing of dietary assessment. *Am J Clin Nutr*. 2016;103:192–200.
37. Phipps AI, Passarelli MN, Chan AT, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis*. 2016;37:87–95.
38. Kap EJ, Seibold P, Scherer D, et al. SNPs in transporter and metabolizing genes as predictive markers for oxaliplatin treatment in colorectal cancer patients. *Int J Cancer*. 2016;138:2993–3001.
39. Valeri L, Chen JT, Garcia-Albeniz X, Krieger N, VanderWeele TJ, Coull BA. The role of stage at diagnosis in colorectal cancer black-white survival disparities: a counterfactual causal inference approach. *Cancer Epidemiol Biomarkers Prev*. 2016;25:83–89.
40. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med*. 2013;369:1106–1114.
41. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370:1287–1297.
42. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med*. 2014;160:171.
43. U.S._Preventive_Services_Task_Force, Bibbins-Domingo K, Grossman DC, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA*. 2016;315:2564–2575.