Methods to Estimate the Comparative Effectiveness of Clinical Strategies that Administer the Same Intervention at Different Times

Citation: Huitfeldt, Anders, Mette Kalager, James M. Robins, Geir Hoff, and Miguel A. Hernán. "Methods to Estimate the Comparative Effectiveness of Clinical Strategies That Administer the Same Intervention at Different Times." Curr Epidemiol Rep 2, no. 3 (July 24, 2015): 149–161.

As Published: http://dx.doi.org/10.1007/s40471-015-0045-5

Publisher: Springer International Publishing

Persistent URL: http://hdl.handle.net/1721.1/104806

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Curr Epidemiol Rep (2015) 2:149–161
DOI 10.1007/s40471-015-0045-5

EPIDEMIOLOGIC METHODS (T LASH, SECTION EDITOR)

Methods to Estimate the Comparative Effectiveness of Clinical Strategies that Administer the Same Intervention at Different Times

Anders Huitfeldt[1,4] · **Mette Kalager**[1,4,5] · **James M. Robins**[1,2] · **Geir Hoff**[4,5,6] · **Miguel A. Hernán**[1,2,3]

Published online: 24 July 2015

Abstract Clinical guidelines that rely on observational data due to the absence of data from randomized trials benefit when the observational data or its analysis emulates trial data or its analysis. In this paper, we review a methodology for emulating trials that compare the effects of different timing strategies, that is, strategies that vary the frequency of delivery of a medical intervention or procedure. We review trial emulation for comparing (i) single applications of the procedure at different times, (ii) fixed schedules of application, and (iii) schedules adapted to the evolving clinical characteristics of the patients. For illustration, we describe an application in which we estimate the effect of surveillance colonoscopies in patients who had an adenoma detected during the Norwegian Colorectal Cancer Prevention (NORCCAP) trial.

Keywords Causal inference · Dynamic strategies · Inverse probability weighting · Colorectal cancer · Surveillance · Colonoscopy

This article is part of the Topical Collection on *Epidemiologic Methods*

✉ Anders Huitfeldt
ahuitfeldt@mail.harvard.edu

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

[2] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[3] Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA

[4] Institute of Health and Society, University of Oslo, Oslo, Norway

[5] Telemark Hospital, Skien, Norway

[6] Cancer Registry of Norway, Oslo, Norway

Introduction

Clinical decisions are increasingly reliant on guidelines, but clinical guidelines are only as good as the available evidence on the comparative effectiveness of interventions [1•]. Ideally, such evidence would come from randomized controlled trials. When a randomized trial is not available, it may be possible to emulate it using observational data [2•]. This approach requires appropriate confounding adjustment, avoidance of selection bias in the definition of the groups to be compared, and formulation of a research question that is relevant for decision makers.

Prior explicit attempts to emulate trials using observational data have studied, for example, postmenopausal hormone therapy [3], statins [4••], epoetin [5••], and antiretroviral therapy [6••]. Here, we review the emulation of trials to compare strategies that differ in the timing of the intervention of interest. As an example, we will consider post-polypectomy surveillance by colonoscopy. During this procedure, adenomas (benign tumors of the colon [7]) are detected and removed. Most adenomas will not develop into colorectal cancer, but most cancers arise from adenomas [8]. In patients with removed adenomas, surveillance colonoscopies are recommended to detect and remove future adenomas before they become malignant. The optimal interval between colonoscopies is not known. Current guidelines both in the USA [9] and the EU [10] are mostly based on expert opinion due to the scarcity of available evidence.

Besides reviewing a methodology to emulate trials for the comparison of strategies that administer the same intervention at different times, we also review a classification of these strategies. First, we consider point interventions to study the effectiveness of a single application of the treatment. Second, we consider sustained interventions to study the effectiveness of a fixed treatment schedule (e.g., colonoscopy at 3 years

after the initial procedure). Third, we consider sustained interventions to study the effectiveness of a personalized schedule of treatment (e.g., colonoscopy every year if the most recent procedure detected large adenomas, otherwise every 3 years). To fix ideas, we review the methodology in the context of its implementation to a cohort of Norwegian individuals. We start by describing this cohort.

Data

The Norwegian Colorectal Cancer Prevention (NORCCAP) screening study was a randomized clinical trial of once-only sigmoidoscopy screening versus no sigmoidoscopy, conducted in Oslo and Telemark counties in Norway between 1999 and 2001. Our analysis includes participants in the sigmoidoscopy arm in whom at least one adenoma was detected (n=2190). As part of the trial, endoscopies were conducted in these individuals until the bowel was free from adenomas. We excluded patients with history of serious gastrointestinal disease, known genetic predisposition to colorectal cancer, and cancer detected as a result of screening in NORCCAP.

In addition to the available data (age, sex, county, smoking, family history of colorectal cancer, and findings at NORCCAP colonoscopies), we conducted a manual chart review at all hospitals in Oslo and Telemark—guided by claims data from the governmental single-payer agency HELFO—to collect data on the date, findings (e.g., size and type of adenomas) and indication of all subsequent colonoscopies and sigmoidoscopies. Of the post-screening endoscopies, 64 % were for surveillance purposes (3 % sigmoidoscopies and 61 % colonoscopies); 30 % were clinically indicated because of symptoms (27 % colonoscopies, 3 % sigmoidoscopies); and 6 % were due to a recent incomplete endoscopy (4 % colonoscopies, 2 % sigmoidoscopies).

Our outcome of interest was incidence of colorectal cancer. For many surveillance interventions, the use of cancer incidence as an outcome is questionable because of potential lead time bias: [11] cancer cases will be detected earlier in patients with more intensive surveillance, which will make surveillance appear less beneficial. In this case, however, the use of the outcome cancer incidence is justified because most of the beneficial effect of surveillance colonoscopy seems to be due to removing adenomas before they become malignant [12], with only a small component of the effect due to earlier detection of prevalent cancer. Death from colorectal cancer could not be studied as an outcome because there were too few cases.

We refer to the date of the last NORCCAP colonoscopy as time of "first eligibility" for our analyses. For each individual, follow-up ends at colorectal cancer, death, sigmoidoscopy, emigration, or December 2011, whichever occurred first. Because we are trying to estimate the effects of post-baseline colonoscopies, which were not randomly assigned to the trial participants, ours is an analysis of observational data. The flow chart in Fig. 1 describes the enrollment of participants in our study. Table 1 displays the characteristics of the eligible individuals.

Three Hypothetical Randomized Trials

The design of any trial is determined by the causal question of interest, which in turn is determined by the population, the strategies being compared, and the outcome of interest to the decision makers [13]. For surveillance tests, the strategies are defined by the timing of the test. Some strategies involve a point intervention at baseline, whereas other strategies involve interventions that are sustained over time according to either a fixed schedule (e.g., do not perform a colonoscopy for 5 years after baseline, then perform a colonoscopy at the end of year 5) or a schedule that depends on each individual's time-evolving clinical characteristics (i.e., schedule the time of every colonoscopy according to the findings at the previous colonoscopy). We refer to sustained strategies with a fixed schedule as static and to those with a subject-specific schedule as dynamic.

Here, we review three types of hypothetical trials that compare static and dynamic strategies and therefore address different questions regarding the effectiveness of surveillance colonoscopy. In all trials, eligible individuals are followed until death, loss to follow-up (i.e., emigration out of Norway), sigmoidoscopy, occurrence of the outcome (here, diagnosis of colorectal cancer), or December 31, 2011, whichever occurred earlier. In all trials, individuals receive a colonoscopy whenever it is clinically indicated (e.g., due to symptoms) but a surveillance colonoscopy only according to the trial protocol. A graphical representation of each trial is shown in Fig. 2.

Trial type #1: point interventions assigned at a fixed time after first eligibility

Individuals who survived 36 months since first eligibility are randomized to either (1) immediate surveillance colonoscopy or (2) no surveillance colonoscopy. Additional eligibility criteria are no colorectal cancer, colonoscopy, or sigmoidoscopy during the 36 months before randomization. Individuals who reach age 70 or develop any invasive non-colorectal cancer before baseline also become ineligible (other comorbidities might be added to the exclusion criteria). For each individual, follow-up starts at the time of randomization, i.e., baseline is 36 months after first eligibility.

More generally, one can consider trials in which baseline is month z, where z ranges between 36 and 84. The effect estimates from these trials will only apply to survivors without

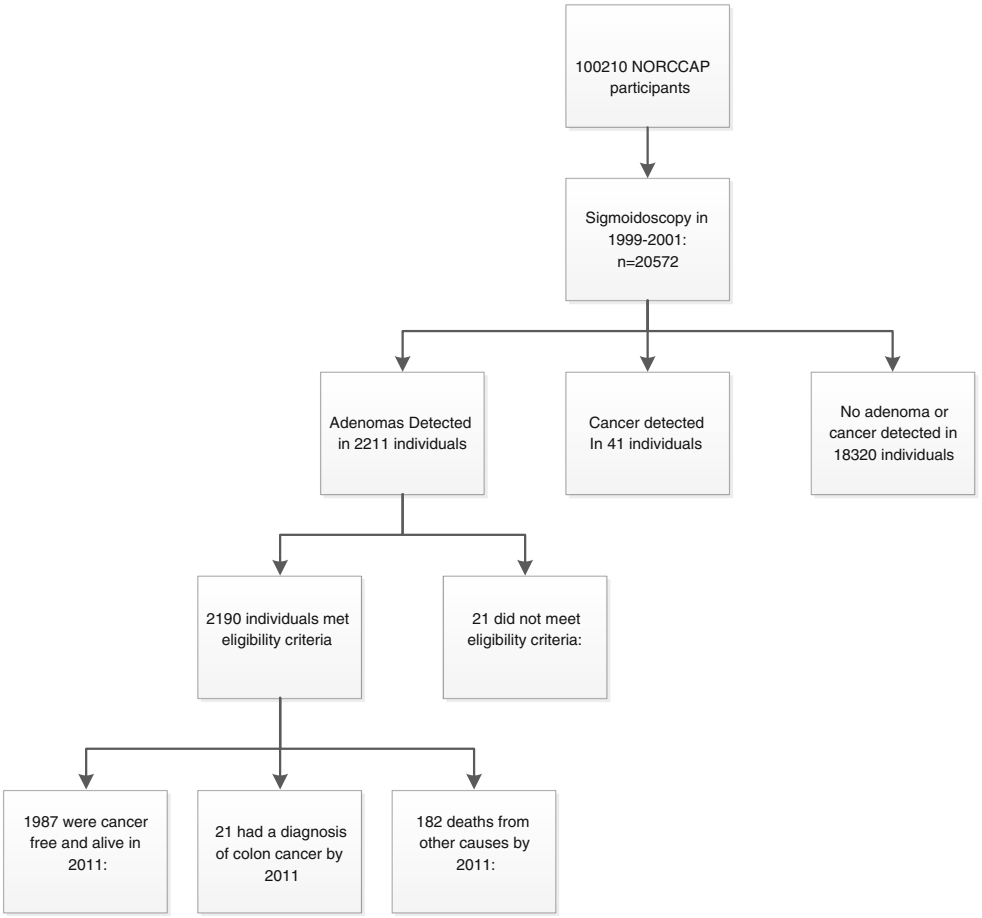

Fig. 1 Flowchart of selection of the 2190 eligible individuals from the intervention arm of the NORCCAP trial

Table 1 Characteristics of 2190 eligible individuals from the intervention arm of the NORCCAP trial

Number of men	1322 (60 %)
Average (SD) age at first eligibility, years	57.2 (3.8)
Median (IQR) duration of follow-up, months	134 (126–143)
Incident cases of colorectal cancer	21
Detected at surveillance colonoscopy	1
Deaths	187
From colorectal cancer	5
Number of colonoscopies during follow-up	819
Number of sigmoidoscopies	75
Number of people with at least one colonoscopy after first eligibility	577
Number of people whose first follow-up colonoscopy was for surveillance	395
Median (IQR) time to first follow-up colonoscopy, months	68 (51–91)
Number of colonoscopies per individual	
0	1613 (74 %)
1	389 (18 %)
2	140 (6 %)
3+	48 (2 %)

symptoms or cancer by z months after first eligibility. These trials will help determine the effect of undergoing a colonoscopy among the survivors, but it does not directly inform the decision of when to undergo the colonoscopy. The next trial does so.

Trial type #2: sustained static strategies assigned at first eligibility

Baseline is the time of first eligibility. Individuals are randomized to either (1) surveillance colonoscopy 36 months after baseline or (2) surveillance colonoscopy 84 months after baseline. Individuals in both arms who reach age 70 or develop malignancies other than colorectal cancer may have surveillance colonoscopies at any time as determined by their physician. More generally, one can consider additional arms in which 36 is replaced by any value of x between 36 and 84. We could also consider similar trials in which baseline is any month after first eligibility. For example, one could consider a trial in which individuals who have survived 36 months after first eligibility are randomized to either (1) immediate surveillance colonoscopy or (2) surveillance colonoscopy at month 84 after first eligibility (48 months after baseline at 36 months). We will only consider trials with baseline at first eligibility.

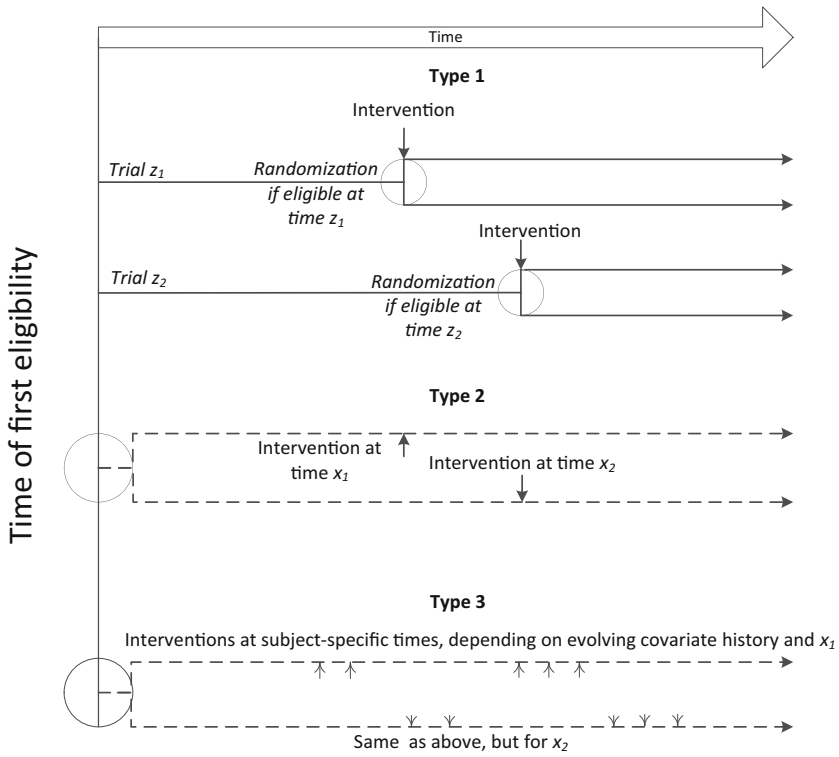

Fig. 2 The three trial types considered in this paper. *Circles* represent randomization, *dotted lines* represent periods when the strategy specifies all interventions (e.g., colonoscopy or no colonoscopy), *solid lines* represent periods when the strategy does not specify the intervention (e.g., anything goes, colonoscopy or no colonoscopy)

Both trial types #1 and #2 compare fixed surveillance schedules, but they address different questions. Trial #1 helps individuals who have survived z months after adenoma removal decide whether they should undergo a surveillance colonoscopy at that time. Trial #2 helps individuals who just had their adenomas removed decide how long they should wait before having a surveillance colonoscopy (if they plan to have only one surveillance colonoscopy). Neither trial type considers strategies that assign different surveillance schedules to different individuals (i.e., dynamic strategies). The next trial type does so.

Trial type #3: sustained dynamic strategies assigned at first eligibility

Individuals at first eligibility are randomized to either (1) receive surveillance colonoscopies according to the following rules:

- First surveillance colonoscopy at 36 months if the adenomas detected at baseline sigmoidoscopy were low risk (1 or 2 small adenomas without villous features) and 12 months earlier (at month 24) otherwise.
- Follow-up surveillance colonoscopy 36 months after the previous colonoscopy (surveillance or clinical) if low-risk adenomas were detected, 12 months earlier (24 months after the previous colonoscopy) if high-risk adenomas (more than two, or large, or containing villous features) were detected, and 12 months later (48 months) if no adenomas were detected.

or (2) surveillance colonoscopies according to similar rules, but where 36 months is replaced by 84 months. During the follow-up, individuals in both arms of the trial may also receive a colonoscopy whenever it is clinically indicated due to symptoms. Individuals who reach age 70 or develop malignancies other than colorectal cancer after baseline may have surveillance colonoscopies at any time as determined by their physician. For each individual, follow-up starts at the time of randomization, i.e., baseline is the time of first eligibility.

More generally, one can consider additional arms in which 36 is replaced by x with x ranging from 36 to 84, or trials in which the time until the next surveillance colonoscopy is obtained by adding or subtracting y (rather than 12) months.

Emulating the Design of the Hypothetical Trials

In this section, we review how to emulate the design of each of the above hypothetical trials by setting up a database with the same structure as that of the trial. In the next section, we review how to mimic the analysis of the hypothetical trials.

Trial type #1: point intervention assigned at a fixed time after first eligibility

We emulated 49 "trials," one starting at each month z between months 36 and 84 after first eligibility. For the "trial" starting in month z, we identified the individuals who met the eligibility criteria at baseline, i.e., all individuals with adenomas

detected and removed at first eligibility who were alive and had not yet had a post-screening colonoscopy/sigmoidoscopy or been diagnosed with colorectal cancer by z months of follow-up. For each trial, individuals were classified into the colonoscopy arm if they received a colonoscopy during month z and into the control arm otherwise.

We identified 2028 eligible individuals. On average, each participated in 45 trials, of which at most 1 was in the colonoscopy arm. The number of eligible individuals who received a colonoscopy at baseline ranged between 0 (in several trials) and 16 (in trial z=61). See Appendix Table 3 for details. Unfortunately, all trials had zero cancers among the exposed, which means the data from NORCCAP cannot be used for a meaningful emulation of trial type #1.

Trial type #1 has the advantage of being easy to emulate and analyze when sufficient observational data are available. This approach has been used in observational studies to estimate the observational analog of the intention-to-treat effect of statin therapy [4••] and postmenopausal hormone therapy [3]. Here, we will not consider this trial type further.

Trial type #2: sustained static strategies assigned at first eligibility

We emulated a randomized trial with 49 arms, in which the participants were assigned at first eligibility to colonoscopy at a randomly assigned time ranging from month 36 to 84 after first eligibility. Classifying the 2190 eligible individuals into a single arm is not possible because, at baseline, each individual's data are consistent with all 49 arms. To overcome this problem we created an expanded dataset with 49 clones of each individual, and assigned each of them to a different arm [14]. The 2190 eligible subjects contributed 107,309 clones to this trial. See Appendix Table 4 for details.

The clones in the expanded dataset were censored at the time their data deviated from the strategy to which they were assigned. For example, in arm 84, 12.9 % of participants were censored for having a surveillance colonoscopy too early (before month 84), 73.5 % of participants were censored for failing to have a surveillance colonoscopy in time (in month 84), and 0.5 % were censored for having a sigmoidoscopy. Those who received a colonoscopy for clinical reasons or developed malignancies other than colorectal cancer were subsequently considered "immune" from censoring.

Trial type #3: sustained dynamic strategies assigned at first eligibility

We emulated a trial with 49 arms, one for each value of x in the dynamic strategies defined above. The 2190 individuals were classified into the arm that was consistent with their observed data. Like in the previous trial, individuals cannot be assigned to a single arm at baseline, so we created an expanded dataset with 49 clones of each individual and assigned each of them to a different arm. The clones were censored at the time they deviated from the strategy to which they were assigned. For example, in arm 84, 11.3 % of participants were censored for having a surveillance colonoscopy too early, 79.7 % of participants for failing to have a surveillance colonoscopy in time, and 1.3 % for having a sigmoidoscopy. The 2190 eligible subjects contributed 107,309 clones to this trial. See Appendix Table 5 for details.

Emulating the Design of Hypothetical Trials with a Grace Period

So far, we have implicitly assumed that it is possible to administer a colonoscopy at a precisely specified time point, e.g., month 36. However, in many clinical settings, this may not be feasible. We may therefore be more interested in emulating trials with a grace period, that is, a window of m months during which the patient may undergo colonoscopy. For example, in trial type #2, patients would be assigned to interventions of the form "surveillance colonoscopy between x and $x+m$ months after baseline." Trials with a grace period more accurately reflect clinical practice in which administrative delays and patient availability may prevent an immediate intervention.

Strategies with a grace period are emulated using "clones" as described above, but with different criteria for censoring. Suppose we use a grace period of m=6 months. An individual who received a surveillance colonoscopy in month 40 now has data consistent with arm 36 because subjects assigned to this arm are allowed to have a colonoscopy at any time between months 36 and 42. Therefore, his clones assigned to arms 36 to 40 will not be censored whereas his clones assigned to arm 41 will be censored because he received a surveillance colonoscopy before the assigned time.

The addition of a grace period requires us to specify the distribution of the interventions during the grace period. For example, we might ask whether most colonoscopies are performed during the first 2 months of the grace period or whether they are more equally distributed during the grace period. In our application, we will specify a uniform distribution of colonoscopies during the grace period [14].

In both trials #2 and #3 with a 6-month grace period, each of the 2190 eligible individuals in the original dataset contributed 49 clones, for a total of 107,310 clones to the expanded dataset. In trial #2, the average censoring time ranged between 41.9 months for x=36 to 89.1 months for x=84. In arm 84, 12.9 % of participants were censored for having a surveillance colonoscopy too early (before month 84), 71.5 % of participants were censored at month 90 for failing to have a surveillance colonoscopy in time, 0.1 % were censored after month 90 for having a second surveillance colonoscopy, and 0.6 % were censored for having a sigmoidoscopy. Across the 49

arms, there were 381 incident cases of colorectal cancer in the clones, which occurred in 12 unique individuals.

In trial #3, the average censoring time ranged from 34.2 months for x=36 to 78.1 months and for x=84. For arm 84, 11.3 % of participants were censored for having a surveillance colonoscopy too early, 77.6 % for failing to have a surveillance colonoscopy in time, and 1.4 % for having a sigmoidoscopy. In total, there were 254 incident cases of colorectal cancer in 13 unique individuals. See Appendix Tables 4 and 5 for details.

Emulating the Analysis of the Hypothetical Trials

After reviewing how to create observational databases with the same structure as hypothetical randomized trials, we review how to use those databases to estimate the cumulative incidence curves (or their complement, the survival curves) that would have been observed under each strategy if all individuals had fully adhered to their original arm assignment. In a slight abuse of notation, we index the strategies by the variable x, which was defined in the previous sections. For example, in trial #2, x=78 corresponds to the strategy "surveillance colonoscopy between 78 and 78+6 months after baseline."

In a true randomized trial with many arms x, we could estimate these curves nonparametrically (Kaplan-Meier curves) or parametrically by fitting a pooled logistic model of the form $logit\Pr(Y_{t+1}=0|Y_t=D_t=0,x)=\alpha_{0,t}+\alpha_1 f(x)+\alpha_2 f(x)\times t$, where t denotes time (in months), Y_t is an indicator of colorectal cancer by t, D_t is an indicator of death by t, $\alpha_{0,t}$ is a time-varying intercept (estimated, for example, via restricted cubic splines for time with knots at 30, 60, 90, and 120 months), $f(x)$ is a function of x (for example, a second degree polynomial), and $f(x)\times t$ is a product term to allow the hazard ratio to vary during the follow-up. For example, for the first 36 months of follow-up, the hazard is known to be identical under all strategies, but it may change after that if colonoscopy has a non-null effect on colorectal cancer incidence.

We would then calculate the predicted values for each value of x and compute their product in order to estimate the survival curves. Pointwise 95 % confidence intervals for the curves can be obtained via a non-parametric bootstrap. In our emulated trials, however, the above logistic model needs to be adjusted by both baseline and post-baseline (time-varying) confounders. The procedure then needs to be modified as we now describe.

Adjustment for Covariates

In both trials # 2 and #3, we need to adjust for covariates that jointly predict surveillance colonoscopy A_t (and therefore censoring) and subsequent outcome. Some of these variables are fixed at the baseline of each trial; others vary during the follow-up. Let L_0 represent the vector of baseline covariates, which include age at baseline, sex, family history of colorectal cancer, history of smoking, and findings at NORCCAP colonoscopies (number of adenomas, size, histology, and presence of villous elements). Let L_t represent the vector of time-varying covariates, which include an indicator for incident non-colorectal malignancies, and a vector of the findings from the most recent colonoscopy (number of adenomas, size of largest adenoma, histological grade, and presence of villous elements).

To adjust for L_0, one could fit the pooled logistic model $logit\Pr(Y_{t+1}=0|Y_t=0,x,L_0)=\alpha_{0,t}+\alpha_1 f(x)+\alpha_2 f(x)\times t+\alpha_3 L_0$ to the expanded dataset of each trial separately. To obtain the survival curves under each strategy x, one would then calculate the predicted values for each value of x, standardized them by L_0, and compute their product. However, the time-varying covariates L_t cannot be added to the logistic model because these variables may be affected by prior treatment [10, 11] (a colonoscopy may change the findings at future colonoscopies, for example by removing adenomas; see Appendix). We therefore need to use IP weighting to adjust for L_t.

The subject-specific, time-varying IP weights are $W_t=\prod_{j=0}^{t}\frac{1}{f\left(A_j|\overline{A}_{j-1},\overline{L}_j,Y_j=D_j=0\right)}$. Informally, the denominator of the weights is each subject's conditional probability of having, at each time t, his or her own surveillance colonoscopy history. We use overbars to denote history, i.e., $\overline{L}_t=(L_0,L_1,L_2,\ldots,L_t)$.

The factors in the denominator of the weights were set to 1 in months following age 70, a non-surveillance colonoscopy, or the diagnosis of malignancies other than colorectal cancer because the individual has a probability 1 of remaining uncensored during those months. The factors in the denominator were also set to 1 during the first 9 months after a colonoscopy is received, because no surveillance colonoscopies were performed during this period (only colonoscopies due to symptoms or to incompleteness of the preceding colonoscopy). In previous applications of IP weighting for strategies with grace periods, the investigators were interested only in strategies that were not sustained beyond the initial decision to treat [14]. Therefore, the contributions to the weights were set to 1 for all time periods after treatment was first received.

For all other months, we estimate the denominator by fitting a logistic model for the conditional probability of receiving a colonoscopy to the original, unexpanded study population. We fit the model

$$logit\Pr\left(A_t=1|\overline{A}_{t-1},\overline{L}_t\right)=\beta_{0,t}+\beta_1 g\left(\overline{A}_{t-1}\right)P_t+\beta_2 L_0+\beta_3 L_t P_t$$

where $\beta_{0,t}$ is a time-varying intercept estimated via restricted cubic splines with knots at 30, 60, 90, and 120 months, $g\left(\overline{A}_{t-1}\right)$ is the time since the most recent colonoscopy, and covariate history $\overline{L}_t$ is summarized via the time-varying covariates L_t and the baseline variables L_0, which include age (restricted cubic splines with knots at 50, 55, 60, and 65 years); sex; family history of colorectal cancer (yes/no); history of smoking (yes/no); findings at the NORCCAP colonoscopies (indicators for three or more adenomas, adenoma greater than 10 mm, adenoma with villous component); and histological grade (1 if high grade dysplasia, 0 otherwise). The variables $g\left(\overline{A}_{t-1}\right)$ and L_t are entered to the model only in a product ("interaction") term with P_t, an indicator for prior colonoscopy (1 if the individual had a colonoscopy before t, 0 otherwise), such that the terms are zero in individuals who have not had a previous surveillance colonoscopy.

Because the IP weights already adjusted for the baseline covariates L_0, we did not include them as covariates in the outcome model. That is, we fit the weighted pooled logistic model $logit\Pr(Y_{t+1}=0|Y_t=0,x)=\alpha_{0,t}+\alpha_1 f(x)+\alpha_2 f(x)\times t$. To check the robustness of our estimates to different choices of functional form for time and x, we explored different parameterizations of the outcome model, including a quadratic functional form for time, cubic terms for x, and additional interaction terms between $f(x)$ and time.

Grace Period

Because our strategies of interest include grace periods, the above-mentioned IP weights W_t need to be modified [14]. Specifically, the numerator of the factors corresponding to months included in the grace period need to change to ensure that surveillance colonoscopies will be uniformly distributed during the grace period. For trial #2, the numerator of factors corresponding to month j of the grace period is replaced by $\frac{1}{m+1-j}$ with j=0, 1, … 5 when A_t=1, and replaced by $\frac{m-j}{m+1-j}$ when A_t=0. For trial #3, where there can be multiple surveillance colonoscopies, we use the same approach during all grace periods.

Estimates from NORCCAP Data

Table 2 shows the 5- and 10-year risks of colorectal cancer for arms 36 and 84 in trials #2 and #3. For both static and dynamic strategies, earlier surveillance colonoscopy resulted in a lower risk. The estimated survival curves for selected arms of trials #2 and #3 are shown in Fig. 3. As expected, the survival curves are essentially identical over the first 3 years, as the strategies are the same during this time period. Results were similar in sensitivity analyses using different functional forms for $f(x)$ and time.

Note that had the dataset included no cancer diagnoses after surveillance colonoscopy, the conclusion that delaying colonoscopy increases risk would be foregone. In our dataset, only one individual who has a surveillance colonoscopy between months 36 and 84 subsequently developed colorectal cancer, and he was censored before getting cancer under most clinically relevant strategies. Any changes to the strategies that led to him not being censored would result in substantial changes to the estimates. Therefore, our analysis needs to be replicated in a larger dataset.

Conclusions

After a medical procedure or medication has been shown to be effective, the next question is usually how often it should be administered. In this paper, we reviewed an approach that, when applied to a sufficiently large and rich dataset, helps decide among various timing strategies. Specifically, we outlined the design and analysis of hypothetical randomized trials to compare different strategies, and provided a methodology for emulating these trials using observational data.

Table 2 Estimated risk of colorectal cancer at 5 and 10 years under selected surveillance strategies, intervention arm of the NORCCAP trial

	Risk, % (95 % CI) x=36	Risk, % (95 % CI) x=84	Risk difference, % (comparing x=36 with x=84) (95 % CI)	Risk ratio (comparing x=36 with x=84) (95 % CI)
Static strategies				
At 5 years	0.15 (0.03–0.37)	0.30 (0.08–0.59)	−0.15 (−0.31–0.00)	0.47 (0.06–0.87)
At 10 years	0.31 (0.05– 0.69)	0.63 (0.27–1.14)	−0.32 (−0.67–0.01)	0.49 (0.10–1.01)
Dynamic strategies				
At 5 years	0.12 (0.00–0.36)	0.25 (0.01–0.50)	−0.13 (−0.30–0.01)	0.49 (0.03–1.18)
At 10 years	0.30 (0.05–0.90)	0.44 (0.17–0.76)	−0.14 (−0.46–0.03)	0.67 (0.10–1.76)

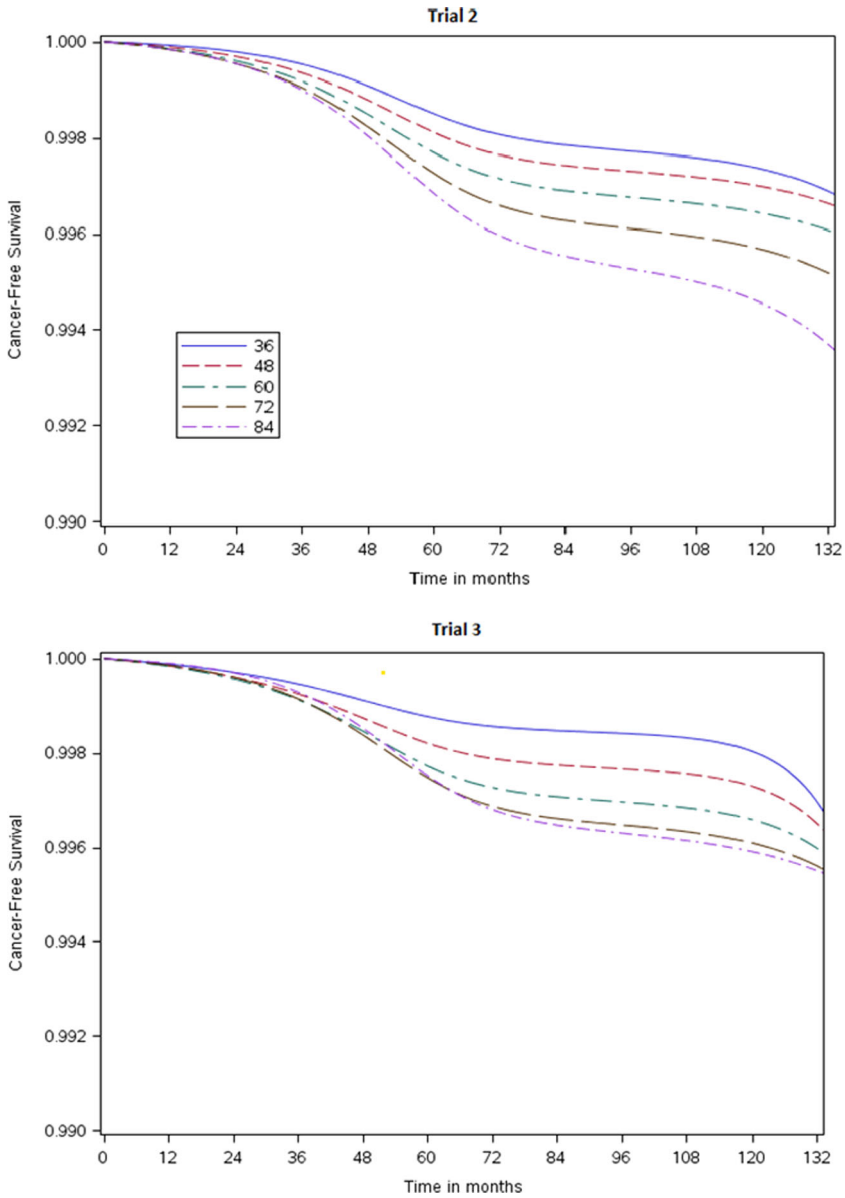

Fig. 3 Estimated survival curves for trials #2 and #3, intervention arm of the NORCCAP trial

As a motivating example, we compared the effectiveness of different strategies for scheduling surveillance colonoscopies in patients with adenomas, a clinical question for which the available evidence is sparse [9, 15–20]. Our analysis suggests that more frequent surveillance colonoscopies leads to a greater reduction in colorectal cancer risk; as expected, the analysis also suggests that dynamic strategies are more effective than static strategies. However, our analysis is more an example of implementation than an attempt at providing definite answers to the clinical question because the sample size of our study was small.

The application of the methods outlined in this review allowed us to specify a research question that is directly relevant to decision makers interested in timing questions. Though these methods allow adjustment for both baseline and time-varying covariates, the possibility of unmeasured confounding remains as in any observational study.

Compliance with Ethics Guidelines

Conflict of Interest Anders Huitfeldt declares that he has no conflict of interest.

Mette Kalager declares that she has no conflict of interest.

James M. Robins declares that he has no conflict of interest.

Geir Hoff declares that he has no conflict of interest.

Miguel A. Hernán declares that he has no conflict of interest.

Human and Animal Rights and Informed Consent All studies by A Huitfeldt, M Kalager, JM Robins, G Hoff, and MA Hernàn involving human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

Funding NIH grant P01 CA134294

Appendix. The Need for Inverse Probability Weights

To see why inverse probability weighting is required to adjust for previous findings at colonoscopy, consider the directed acyclic graph in Appendix Fig. 4. A_t is an indicator for colonoscopy at time t, L_t is an indicator for the (possibly unknown to the investigator) presence of adenomas at time t, L_t* is an indicator for the presence of *known* adenomas at time t. Adenomas only become known through colonoscopy: If $A_t=1$ then L_{t+1}*=L_t; otherwise, L_{t+1}*=L_t*. U represents the common causes of adenomas and colorectal cancer, such as genetics. Y is an indicator of colorectal cancer by the end of follow-up.

According to this causal diagram, L_{t+1}* is a confounder for the effect of A_{t+1} on Y. Knowledge of adenomas at time $t+1$ predicts colonoscopy at time $t+1$, and is also a marker for actual adenomas L_t, which cause cancer at time $k>t$. However, confounding adjustment via conditioning on the collider L_{t+1}* would open the biasing path $A_t \rightarrow L_{t+1}* \leftarrow L_t \rightarrow Y$. Note that, to avoid clutter, we chose not to include the direct arrow from A_{t-1} (not shown on graph) to L_t, which would only increase the number of biasing paths.

Another possible problem is that conditioning on L_{t+1}* may partially block the effect of A_t through the path $A_t \rightarrow L_{t+1}* \rightarrow Y$. The arrow $L_{t+1}* \rightarrow Y$ exists because the detection of polyps necessarily leads to polypectomy, which affects the risk of cancer at later times.

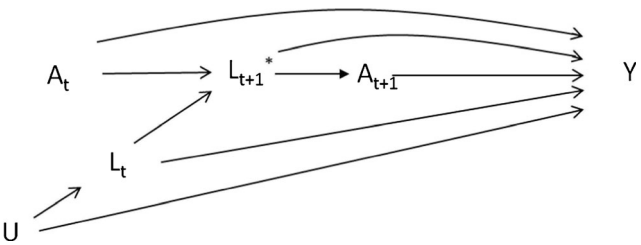

Fig. 4 Causal directed acyclic graph to represent the effect of A_t (colonoscopy at time t followed by polypectomy if necessary) on colorectal cancer Y. L_t is an indicator for the presence of adenomas and L_t* is an indicator for the presence of *known* adenomas at time t

Table 3 Emulation of trial type #1 (point interventions assigned at a fixed time after first eligibility) using the control arm of the NORCCAP randomized trial

"Trial"	Baseline month	Eligible individuals	Incident cancers	Exposed individuals	Incident cancers among the exposed
1	36	2028	14	1	0
2	37	2025	14	2	0
3	38	2020	14	4	0
4	39	2013	14	3	0
5	40	2009	14	3	0
6	41	2004	14	1	0
7	42	2001	13	3	0
8	43	1995	13	3	0
9	44	1988	13	6	0
10	45	1980	13	1	0
11	46	1975	12	0	0
12	47	1971	12	1	0
13	48	1963	12	3	0
14	49	1956	12	7	0
15	50	1948	12	1	0
16	51	1945	12	2	0
17	52	1935	12	3	0
18	53	1932	12	4	0
19	54	1923	12	5	0
20	55	1916	12	3	0
21	56	1907	12	2	0
22	57	1903	12	6	0
23	58	1892	10	10	0
24	59	1876	10	4	0
25	60	1870	10	7	0
26	61	1856	10	16	0
27	62	1836	10	13	0
28	63	1818	9	10	0
29	64	1805	9	5	0
30	65	1794	9	10	0
31	66	1779	9	6	0
32	67	1768	9	8	0
33	68	1758	9	6	0
34	69	1747	9	8	0
35	70	1738	9	10	0
36	71	1725	9	3	0
37	72	1720	9	4	0
38	73	1708	9	4	0
39	74	1700	9	5	0
40	75	1690	9	4	0
41	76	1679	8	5	0
42	77	1669	8	7	0
43	78	1655	7	4	0
44	79	1647	7	4	0
45	80	1638	7	2	0
46	81	1633	7	6	0
47	82	1620	7	0	0
48	83	1617	7	4	0
49	84	1609	6	0	0
Pooled analysis		88,156	497	228	0

Table 4 Emulation of trial #2 (sustained strategies to determine timing of first surveillance colonoscopy, with and without grace periods)

	Without grace period							With grace period					
x	Participants	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time of censoring	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time of censoring
36	2189	82	2066	29	9	3	36.0	100	2039	35	11	4	42.0
37	2190	84	2064	30	9	3	37.0	104	2036	35	11	4	43.0
38	2190	88	2058	32	9	3	38.0	111	2026	38	11	4	44.0
39	2190	93	2053	32	9	3	39.0	113	2024	38	11	4	45.0
40	2190	95	2050	32	10	3	40.0	114	2021	39	11	5	46.0
41	2190	96	2046	34	11	3	41.0	117	2016	41	11	5	47.0
42	2190	100	2040	35	11	4	42.0	123	2010	41	11	5	48.0
43	2190	104	2036	35	11	4	43.0	132	2000	42	11	5	49.0
44	2190	111	2026	38	11	4	44.0	133	1999	42	11	5	49.9
45	2190	113	2024	38	11	4	45.0	136	1994	44	11	5	50.9
46	2190	114	2021	39	11	5	46.0	144	1983	46	12	5	51.9
47	2190	117	2016	41	11	5	47.0	148	1979	46	12	5	52.9
48	2190	123	2010	41	11	5	48.0	154	1972	47	12	5	53.9
49	2190	132	2000	42	11	5	49.0	158	1967	48	12	5	54.9
50	2190	133	1999	42	11	5	49.9	164	1961	48	12	5	55.9
51	2190	136	1994	44	11	5	50.9	170	1953	50	12	5	56.9
52	2190	144	1983	46	12	5	51.9	180	1940	51	12	7	57.9
53	2190	148	1979	46	12	5	52.9	187	1932	52	12	7	58.9
54	2190	154	1972	47	12	5	53.9	195	1923	53	12	7	59.9
55	2190	158	1967	48	12	5	54.9	216	1900	55	12	7	60.9
56	2190	164	1961	48	12	5	55.9	229	1883	58	13	7	61.9
57	2190	170	1953	50	12	5	56.9	241	1868	60	13	8	62.9
58	2190	180	1940	51	12	7	57.9	246	1862	61	13	8	63.9
59	2190	187	1932	52	12	7	58.9	258	1849	62	13	8	64.9
60	2190	195	1923	53	12	7	59.9	267	1839	63	13	8	65.9
61	2190	216	1900	55	12	7	60.9	276	1829	64	13	8	66.9
62	2190	229	1883	58	13	7	61.9	283	1821	65	13	8	67.9
63	2190	241	1868	60	13	8	62.9	292	1810	66	14	8	68.9
64	2190	246	1862	61	13	8	63.9	302	1796	70	14	8	69.9
65	2190	258	1849	62	13	8	64.9	307	1790	71	14	8	70.9
66	2190	267	1839	63	13	8	65.9	312	1785	71	14	8	71.9
67	2190	276	1829	64	13	8	66.9	321	1774	73	14	8	72.9

Table 4 (continued)

	Without grace period							With grace period					
x	Participants	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time of censoring	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time of censoring
68	2190	283	1821	65	13	8	67.9	329	1763	76	14	8	73.9
69	2190	292	1810	66	14	8	68.9	335	1757	76	14	8	74.9
70	2190	302	1796	70	14	8	69.9	341	1747	79	14	9	75.9
71	2190	307	1790	71	14	8	70.9	349	1736	82	14	9	76.9
72	2190	312	1785	71	14	8	71.9	355	1726	84	15	10	77.9
73	2190	321	1774	73	14	8	72.9	360	1718	87	15	10	78.9
74	2190	329	1763	76	14	8	73.9	363	1709	91	16	11	79.9
75	2190	335	1757	76	14	8	74.9	370	1700	92	17	11	80.9
76	2190	341	1747	79	14	9	75.9	372	1698	92	17	11	81.9
77	2190	349	1736	82	14	9	76.9	377	1693	92	17	11	82.9
78	2190	355	1726	84	15	10	77.9	380	1689	92	17	12	83.9
79	2190	360	1718	87	15	10	78.9	385	1683	93	17	12	84.9
80	2190	363	1709	91	16	11	79.9	389	1678	94	17	12	85.9
81	2190	370	1700	92	17	11	80.9	399	1668	94	17	12	86.9
82	2190	372	1698	92	17	11	81.9	403	1661	97	17	12	87.9
83	2190	377	1693	92	17	11	82.9	408	1653	100	17	12	88.9
84	2190	380	1689	92	17	12	83.9	412	1649	100	17	12	89.9

Table 5 Emulation of trial #3 (sustained strategies where the timing of colonoscopies is a function of evolving subject-specific characteristics)

		Without grace period						With grace period					
x	Participants	Reached end of follow-up	Censored	Deaths	Emigration	Incident cancers	Average time to censoring	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time to censoring
36	2189	20	2144	18	6	1	28.1	31	2122	23	10	3	34.2
37	2190	20	2142	20	7	1	29.1	31	2120	26	10	3	35.2
38	2190	20	2142	20	7	1	30.1	33	2116	28	10	3	36.2
39	2190	22	2140	20	7	1	31.1	31	2119	28	9	3	37.2
40	2190	22	2139	21	7	1	32.0	31	2118	29	9	3	38.0
41	2190	22	2139	21	7	1	33.0	33	2117	29	8	3	39.0
42	2190	24	2132	23	8	3	34.0	34	2116	29	8	3	40.0
43	2190	24	2129	26	8	3	35.0	33	2117	29	8	3	40.9
44	2190	26	2125	28	8	3	35.9	35	2113	31	8	3	41.9
45	2190	26	2125	28	8	3	36.9	34	2114	31	8	3	42.8
46	2190	27	2123	29	8	3	37.9	38	2109	32	8	3	43.8
47	2190	29	2121	29	8	3	38.8	42	2102	34	9	3	44.8
48	2190	29	2121	29	8	3	39.8	42	2101	35	9	3	45.7
49	2190	30	2120	29	8	3	40.8	41	2102	35	9	3	46.8
50	2190	30	2118	31	8	3	41.8	41	2101	36	9	3	47.8
51	2190	30	2118	31	8	3	42.7	49	2091	38	9	3	48.7
52	2190	32	2115	32	8	3	43.7	49	2087	40	9	5	49.6
53	2190	34	2110	34	9	3	44.6	49	2086	41	9	5	50.5
54	2190	34	2109	35	9	3	45.6	52	2082	42	9	5	51.5
55	2190	34	2109	35	9	3	46.6	54	2078	44	9	5	52.4
56	2190	34	2108	36	9	3	47.5	64	2067	45	9	5	53.4
57	2190	39	2101	38	9	3	48.4	63	2066	47	9	5	54.3
58	2190	40	2096	40	9	5	49.4	65	2061	49	10	5	55.2
59	2190	40	2095	41	9	5	50.3	67	2059	49	10	5	56.3
60	2190	42	2092	42	9	5	51.3	67	2059	49	10	5	57.4
61	2190	44	2088	44	9	5	52.2	70	2055	50	10	5	58.3
62	2190	46	2085	45	9	5	53.2	65	2060	50	10	5	59.2
63	2190	46	2083	47	9	5	54.1	65	2060	50	10	5	60.1
64	2190	47	2079	49	10	5	55.1	69	2055	50	10	6	61.0
65	2190	49	2077	49	10	5	56.0	71	2052	51	10	6	62.0
66	2190	49	2077	49	10	5	56.9	76	2047	51	10	6	62.8
67	2190	49	2076	50	10	5	57.8	79	2042	53	10	6	63.7
68	2190	50	2075	50	10	5	58.8	89	2026	58	11	6	64.6
69	2190	50	2075	50	10	5	59.7	95	2019	58	11	7	65.6
70	2190	50	2074	50	10	6	60.6	98	2013	61	11	7	66.6
71	2190	50	2073	51	10	6	61.5	99	2010	63	11	7	67.3
72	2190	53	2070	51	10	6	62.4	102	2006	63	12	7	68.2
73	2190	55	2067	52	10	6	63.3	107	1998	66	12	7	69.0
74	2190	61	2055	57	11	6	64.2	105	1997	68	13	7	69.8
75	2190	63	2052	57	11	7	65.1	105	1995	69	14	7	70.5
76	2190	65	2048	59	11	7	66.0	106	1992	71	14	7	71.3
77	2190	66	2046	60	11	7	66.9	108	1989	72	14	7	72.3
78	2190	68	2043	60	12	7	67.8	107	1990	72	14	7	73.1
79	2190	70	2037	64	12	7	68.6	103	1993	73	14	7	73.9
80	2190	71	2033	66	13	7	69.5	100	1995	74	14	7	74.8

Table 5 (continued)

		Without grace period						With grace period					
x	Participants	Reached end of follow-up	Censored	Deaths	Emigration	Incident cancers	Average time to censoring	Reached end of follow-up	Censored	Deaths	Emigrations	Incident cancers	Average time to censoring
81	2190	74	2028	67	14	7	70.4	104	1991	74	14	7	75.6
82	2190	76	2023	70	14	7	71.2	105	1988	75	14	8	76.5
83	2190	76	2021	72	14	7	72.1	103	1986	79	14	8	77.2
84	2190	76	2021	72	14	7	73.0	107	1978	82	14	9	78.1

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1.• Institute of Medicine. Ethical and scientific issues in studying the safety of approved drugs. Washington, DC: The National Academies Press; 2012. **This document uses the concept of observational studies as attempts to emulate randomized trials for drug safety**.

2.• Hernán MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. Epidemiology (Cambridge, Mass). 2011;22(3):290–1. **This editorial discusses the need to frame research questions as hypothetical randomized trials in order to make them directly relevant for decision making**.

3. Hernan 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(6):766–79.

4.•• Danaei G, Rodriguez LA, Cantero OF, Logan R, Hernan MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat Methods Med Res. 2013;22(1):70–96. **An explicit attempt to emulate a trial with static strategies**.

5.•• Zhang Y, Thamer M, Kaufman J, Cotter D, Hernán MA. Comparative effectiveness of two anemia management strategies for complex elderly dialysis patients. Med Care. 2014;52 Suppl 3: S132–9. **An explicit attempt to emulate a trial with dynamic strategies**.

6.•• Cain LE, Logan R, Robins JM, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. Ann Intern Med. 2011;154(8):509–15. **This paper develops the theory of inverse probability weighted estimators for dynamic strategies, and also discusses the necessity of grace periods**.

7. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer. 1982;49(4):819–25.

8. Winawer SJ, Fletcher RH, Miller L, et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology. 1997;112(2):594–642.

9. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US multi-society task force on colorectal cancer. Gastroenterology. 2012;143(3):844–57.

10. 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:SE151–63.

11. Prorok PC, Connor RJ, Baker SG. Statistical considerations in cancer screening programs. Urol Clin North Am. 1990;17(4):699–708.

12. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med. 1993;329(27):1977–81.

13. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine—how to practice and teach EBM. New York: Churchill Livingstone; 1997.

14. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernan MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. Int J Biostat. 2010;6(2):Article 18.

15. Jørgensen OD, Kronborg O, Fenger C. A randomized surveillance study of patients with pedunculated and small sessile tubular and tubulovillous adenomas. The Funen Adenoma Follow-up Study. Scand J Gastroenterol. 1995;30(7):686–92.

16. Kronborg O, Jørgensen OD, Fenger C, Rasmussen M. Three randomized long-term surveillance trials in patients with sporadic colorectal adenomas. Scand J Gastroenterol. 2006;41(6):737–43.

17. Lund JN, Scholefield JH, Grainge MJ, et al. Risks, costs, and compliance limit colorectal adenoma surveillance: lessons from a randomised trial. Gut. 2001;49(1):91–6.

18. Lieberman DA, Weiss DG, Harford WV, et al. Five-year colon surveillance after screening colonoscopy. Gastroenterology. 2007;133(4):1077–85.

19. Winawer SJ, Zauber AG, Gerdes H, et al. Risk of colorectal cancer in the families of patients with adenomatous polyps. National Polyp Study Workgroup. N Engl J Med. 1996;334(2):82–7.

20. Von Karsa L, Segnan N, Patnick J. European guidelines for quality assurance in colorectal cancer screening and diagnosis. 2010.