Eur J Epidemiol (2017) 32:495–500
DOI 10.1007/s10654-017-0287-2

METHODS

The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening

Xabier García-Albéniz[1,2] · **John Hsu**[2,3] · **Miguel A. Hernán**[1,4,5]

Received: 3 June 2017 / Accepted: 17 July 2017 / Published online: 26 July 2017

Abstract Observational analyses for causal inference often rely on real world data collected for purposes other than research. A frequent goal of these observational analyses is to use the data to emulate a hypothetical randomized experiment, i.e., the target trial, that mimics the design features of a true experiment, including a clear definition of time zero with synchronization of treatment assignment and determination of eligibility. We review a recent observational analysis that explicitly emulated a target trial of screening colonoscopy using insurance claims from U.S. Medicare. We then compare this explicit emulation with alternative, simpler observational analyses that do not synchronize treatment assignment and eligibility determination at time zero and/or do not allow for repeated eligibility. This empirical comparison suggests that lack of an explicit emulation of the target trial leads to biased estimates, and shows that allowing for repeated eligibility increases the statistical efficiency of the estimates.

Keywords Colon cancer screening · Target trial · Causal inference

In theory, randomized controlled trials can validly address causal questions about the comparative effectiveness or safety of health-related interventions. In practice, the randomized trial that would answer a particular causal question—the target trial—may not be feasible, ethical, or timely. When this is the case, we can resort to observational analyses that use real world data to explicitly emulate the target trial of interest [1].

As an example of the above, consider the effect of screening colonoscopy on colorectal cancer (CRC) incidence [2]. Though screening colonoscopy is thought to reduce CRC incidence, no results from randomized trials will be available until the next decade [3–5]. In the meanwhile, evidence about the effectiveness of screening colonoscopy will necessarily rely on observational analyses of real world data [6]. We recently emulated a (hypothetical) target trial of screening colonoscopy using a large insurance claims database from the Medicare program in the United States. We estimated that screening colonoscopy lowers the eight-year risk of CRC risk by 0.63 percentage points in individuals aged 70–74 years [7].

To obtain this estimate of 0.63, we implemented an analytic approach that explicitly emulates a target trial of screening colonoscopy. Our approach was designed to reduce bias and to improve precision when using nonrandomized longitudinal data. However, our approach may appear unnecessarily complex when compared with more

Electronic supplementary material The online version of this article (doi:10.1007/s10654-017-0287-2) contains supplementary material, which is available to authorized users.

✉ Xabier García-Albéniz
xabi@post.harvard.edu

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

[2] Mongan Institute, Massachusetts General Hospital, Boston, MA, USA

[3] Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

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

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

frequently used observational analyses; whether the complexity is necessary remains an open question.

To address this question, here we compare the trial emulation approach with simpler approaches to estimate the effect of screening colonoscopy on the eight-year risk of CRC. We then discuss the relative advantages and disadvantages of our explicit emulation approach compared with these simpler approaches. First, we briefly review our approach.

Overview of the original observational analysis to emulate the target trial

We first specified the protocol of a target trial of colonoscopy (Table 1). To emulate this target trial, we identified Medicare beneficiaries who met the eligibility criteria and assigned them to either one of two comparison arms: (1) screening colonoscopy within the seven days after they met the eligibility criteria; or (2) no screening strategy. Follow-up ended at CRC diagnosis, death, disenrollment from Medicare (loss to follow-up), or December 31, 2012, whichever occurred earlier. We compared the eight-year risk of CRC under both strategies [7]. We emulated the target trial both assuming that randomization was conditional on the baseline covariates listed in Table 1 and that randomization was unconditional. In practical terms, this means that we conducted two analyses: one adjusted for the baseline covariates and another one unadjusted. Because adjustment for baseline variables did not materially affect the estimates, for simplicity we will only discuss unadjusted estimates here. Also for simplicity, we restrict our discussion to the analysis in the 70–74 years age group.

Because eligible individuals could meet the eligibility criteria at multiple weeks during the follow-up period, we emulated 260 target trials (with each trial starting in a different week) and allowed individuals to contribute information to multiple trials depending on eligibility. On average, each individual was eligible for 49.7 of the 52 trials emulated each year for which she was eligible. We analyzed the data by pooling the 260 emulated trials. The statistical analysis accounted for the repeated use of individuals: bootstrapping was used to estimate the variance of the estimates. To decrease the computational demands imposed by the emulation of 260 target trials, we included only a random sample of 5% of individuals assigned to the no screening strategy each week. For comparability, we also used 5% of the eligible individuals in the simpler analyses described below.

Under this study design, all 685 CRC cases in the screening strategy were unique, but the 21,954 CRC cases in the no screening strategy corresponded to only 1086 unique cases. 12.8% of eligible individuals were censored due to loss to follow-up; adjustment for potential selection bias using inverse-probability weighting did not materially change the estimates. The CRC risk was greater at time zero in the screening strategy than in the no-screening strategy because previously undiagnosed cancers were diagnosed at baseline, but the curves crossed at about five years (Fig. 1). The eight-year risk difference was −0.63% (95% CI: −0.83, −0.43) (Table 2).

Three alternative, simpler observational analyses

We now describe three simpler analytic approaches that, unlike ours, include individuals in the analysis only once.

Treatment assigned at time zero/eligibility determined at time zero

A simpler emulation of the target trial would be identical to our analysis except that individuals would be used only once. That is, we assign individuals who receive a colonoscopy while meeting the eligibility criteria to the colonoscopy strategy (with time zero being the time of the colonoscopy), and individuals who did not received a colonoscopy at first eligibility to the no colonoscopy group (with time zero being their first eligible time during the study period).

This approach, which appropriately emulates the target trial, results in 685 unique cases in the screening strategy and 1086 unique cases in the no screening strategy. The risk curves are essentially identical to the ones obtained by using our original approach (Fig. 1). However, the 95% confidence intervals are wider than for our original approach, as expected. The eight-year risk difference is −0.67% (95% CI: −1.03, −0.28) (Table 2). Adjustment for baseline variables do not materially affect this estimate (Supplementary Figure).

Treatment assigned at time zero/eligibility determined after time zero

A variation of the previous approach is to redefine the no-colonoscopy group to ensure that no individuals in that group ever received a colonoscopy during the follow-up. That is, we assign individuals who receive a colonoscopy while meeting the eligibility criteria to the colonoscopy strategy (with time zero being the time of the colonoscopy), and individuals who did not receive a colonoscopy throughout the entire study period to the no-colonoscopy group (with time zero being their first eligible time during the study period).

Under this analytic approach, there are 685 cases in the screening group, but only 11 in the no-screening group.

Table 1 Specification and emulation of a target trial of screening colonoscopy using real world data from U.S. Medicare 1999–2012

Component	Target trial	Emulated trial using real world data
Aim	To estimate the effect of screening colonoscopy on the 8-year risk of CRC in U.S. individuals aged 70–74 years	Same
Eligibility	Persons without gastrointestinal symptoms aged 70–74 years with no history of CRC, and continuously enrolled in Medicare for 5 years with no adenoma, inflammatory bowel disease, colectomy, or CRC screening in that period, and who were regular users of preventive services (at least 2 of the following: influenza vaccine, preventive visit, breast or prostate screening, in the 2 years before enrollment)	Same, except CRC history is evaluated in the 5 years before enrollment
Treatment strategies	1. Screening colonoscopy at baseline 2. No screening for CRC at baseline Patients receive usual care after the intervention	Same
Treatment assignment	Patients are randomly assigned to either strategy	Patients are assigned to screening colonoscopy if they receive a screening colonoscopy in the 7 days following eligibility and to no screening otherwise. Randomization is emulated via adjustment for baseline covariates: sex, race, age, original reason for Medicare entitlement, use of preventive services, U.S. Census Bureau division, combined comorbidity score, calendar month, presence of each CCW condition (Alzheimer's disease, acute myocardial infarction, asthma, atrial fibrillation, cataract, chronic heart failure, chronic kidney disease, endometrial cancer, breast cancer, lung cancer, prostate cancer, chronic obstructive pulmonary disease, depression, diabetes, glaucoma, hip/pelvic fracture, hyperlipidemia, benign prostatic hyperplasia, hypertension, hypothyroidism, ischemic heart disease, osteoporosis, osteoarthritis, stroke)
Follow-up	Follow-up starts at treatment assignment and ends at CRC diagnosis, at death, at loss to follow-up, 8 years after baseline, or on 31 December 2012, whichever occurs first	Same
Outcome	CRC diagnosis within 8 years of baseline	Same
Causal contrast	Intention-to-treat effect, i.e., effect of being assigned to screening colonoscopy versus no screening at baseline. Per-protocol effect, i.e., effect of receiving screening colonoscopy versus no screening at baseline	Observational analog of per-protocol effect
Statistical analysis	Intention-to-treat analysis. Per-protocol analysis: comparison of 8-year CRC risk between groups receiving each treatment strategy with adjustment for baseline covariates (and post-baseline covariates when adjusting for loss to follow-up)	Same as per-protocol analysis

This approach is biased because in the real world most CRCs are eventually diagnosed via a (screening or diagnostic) colonoscopy so, by excluding those who received a colonoscopy after time zero, individuals in the no-screening strategy group have little opportunity to have a CRC diagnosed. In fact, this approach excludes approximately 99% of the cases in the no-screening group. As a result of artificially lowering the CRC risk in the no-screening strategy, screening looks harmful at all times during the follow-up (Fig. 1) and the eight-year risk difference is 1.7% (95% CI: 1.4, 2.1). Adjustment for baseline variables do not materially affect this estimate (Supplementary Figure).

Treatment assigned before time zero/eligibility determined at time zero

Previously published analyses [8] were based on yet another analytic approach: selecting an arbitrary time zero to ascertain eligibility, e.g., January 1, 2004, and then assigning all eligible individuals to the colonoscopy strategy if they received a colonoscopy in the previous five years, or to the no-screening strategy otherwise.

This approach is subject to selection bias [9] because colonoscopies performed before the assessment of eligibility may affect eligibility. In particular, a colonoscopy that detects CRC or precursor lesions in the previous five years will result

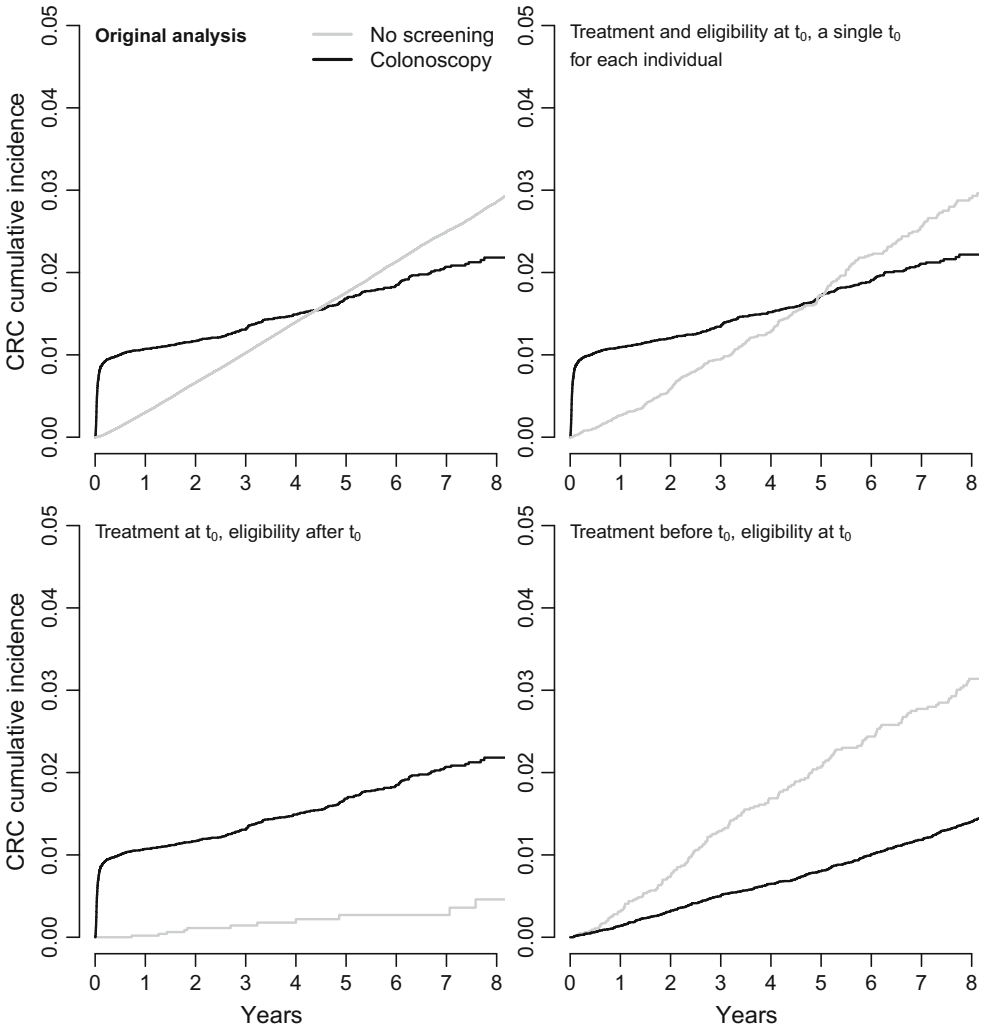

Fig. 1 Cumulative incidence curves under four observational analyses, Medicare 1999–2012. We use t_0 to denote time zero of follow-up

Table 2 Estimates of the effect of colonoscopy screening on 8-year risk of colorectal cancer in individuals aged 70–74 years old under four observational analyses, U.S. Medicare 1999–2012

Treatment assigned	Eligibility determined	Individuals used multiple times	Strategy	N	CRC cases	CRC risk, % (95% CI)	Difference, % (95% CI)
At time zero	At time zero	Yes	No screening	1,762,816	21,954	2.8 (2.7, 2.8)	Ref.
			Screening	46,872	685	2.1 (2.0, 2.3)	−0.63 (−0.83, −0.43)
At time zero	At time zero	No	No screening	72,249	1086	2.8 (2.5, 3.2)	Ref.
			Screening	46,872	685	2.2 (2.0, 2.4)	−0.67 (−1.03, −0.28)
At time zero	After time zero	No	No screening	6241	11	0.4 (0.2, 0.7)	Ref.
			Screening	46,872	685	2.1 (1.9, 2.3)	1.7 (1.4, 2.1)
Before time zero	At time zero	No	No screening	6507	178	3.1 (2.7, 3.6)	Ref.
			Screening	37,844	492	1.4 (1.3, 1.5)	−1.7 (−2.2, −1.3)

in the individual being excluded from the analysis. As a result of removing cases from the colonoscopy group, screening looks implausibly beneficial at all times during the follow-up (Fig. 1) and the eight-year risk difference is −1.7% (95% CI: −2.2, −1.3) (Table 2). Adjustment for baseline variables do not materially affect this estimate (Supplementary Figure).

Discussion

Our explicit emulation of a target trial yielded effect estimates that are consistent with the expected effect of screening colonoscopy: an initially increased risk due to the diagnosis of subclinical cancers at baseline followed by a decreased risk due to the elimination of cancer precursors during colonoscopy. A simplified variation of our original approach that does not allow for repeated eligibility resulted in very similar point estimates, but wider 95% confidence intervals.

Our target trial emulation respected a basic principle of trial design: both treatment assignment and the determination of eligibility occur simultaneously at time zero [10]. In contrast, observational analyses that violated this basic principle yielded implausible estimates of the effect of screening colonoscopy. The exclusion of post-baseline colonoscopies from the no screening group (similar to commonly used, naïve per-protocol analyses in randomized trials) led to a questionable estimate of continuous harm throughout the entire follow-up. The classification into treatment groups before the determination of eligibility led to a logically impossible estimate of benefit that missed the early increased in risk due to the detection of undiagnosed cancers via the screening colonoscopy. The later approach is similar to the one that created confusion about the effect of postmenopausal hormone therapy in observational studies [1, 11]. Note that adjustment for multiple covariates could not correct the bias introduced by the incorrect design of these two observational analyses.

These results suggest that the increased complexity and computational demands of our observational analysis were justified. An explicit emulation of a target trial was able to reproduce the non-monotonic effect estimates and, when combined with multiple eligibility entry points for each individual (who can therefore be considered exposed and unexposed at different times), also reduced the variance of the estimates compared with simpler approaches.

An open question is how much the implementation of repeated eligibility reduces the variance of the estimates. It turns out that, in our setting, we can empirically answer this question because we used only 5% of the eligible individuals. After conducting the correct analysis with single eligibility using an increasing proportion of eligible individuals, we found that only after including about 50% of the eligible individuals in the no screening group, the variance of the estimates was close to that of the estimates in the original analysis. That is, allowing the individuals to contribute to multiple emulated trials was the equivalent of increasing the sample size of the unexposed group approximately tenfold.

Of course, the effect estimate of our emulation approach might still be biased because of confounding (if individuals assigned to the screening and no screening groups differ with respect to their risk factors for CRC) or selection bias (if, for example, the loss to follow-up was differential across groups). However, these biases seem unlikely to explain our results because adjustment for multiple demographic, clinical, and health care utilization covariates (which are either confounders or proxies for confounders) did not materially affect the estimates.

The validity of our approach is also supported by the consistency between the findings from emulated trials and of actual randomized trials for two other types of screening, for which randomized trial data exist. First, the colonoscopy effect estimates in our emulated trial were consistent with the estimates from several randomized trials of sigmoidoscopy screening (a less invasive procedure than colonoscopy). Second, the effect estimates from an emulated trial of fecal occult blood test (FOBT) effects were consistent with those of FOBT randomized trials [12, 13].

In summary, in this setting with a point intervention, we empirically demonstrated that the lack of synchronization of treatment assignment and eligibility determination at time zero leads to biased estimates, and also that allowing for repeated eligibility increases the statistical efficiency of the estimates. The publication of the findings of several screening colonoscopy randomized trials in the next decade will provide the ultimate test of our approach for studying the comparative effectiveness and safety of health interventions.

Acknowledgements This work was supported by the National Institutes of Health grants K99-CA207730, P01-CA134294, R01-CA164023, R01-HS023128.

Compliance with ethical standards

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

References

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758–64.
2. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, et al. Screening for colorectal cancer: US preventive services task force recommendation statement. JAMA. 2016;315:2564–75.
3. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366:697–706.
4. Kaminski MF, Bretthauer M, Zauber AG, Kuipers EJ, Adami H-O, van Ballegooijen M, et al. The NordICC study: rationale

and design of a randomized trial on colonoscopy screening for colorectal cancer. Endoscopy. 2012;44:695–702.
5. Colonoscopy Versus Fecal Immunochemical Test in Reducing Mortality From Colorectal Cancer (CONFIRM) [Internet]. [cited 2015 Jun 3]. https://clinicaltrials.gov/ct2/show/NCT01239082.
6. Smith RA, Andrews K, Brooks D, DeSantis CE, Fedewa SA, Lortet-Tieulent J, et al. Cancer screening in the United States, 2016: a review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2016;66:96–114.
7. García-Albéniz X, Hsu J, Bretthauer M, Hernán MA. Effectiveness of screening colonoscopy to prevent colorectal cancer among medicare beneficiaries aged 70 to 79 years. Ann Intern Med. 2017;166:18.
8. Stock D, Paszat LF, Rabeneck L. Colorectal cancer mortality reduction is associated with having at least 1 colonoscopy within the previous 10 years among a population-wide cohort of screening age. Gastrointest Endosc. 2016;84:133–41.
9. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615–25.
10. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016;79:70–5.
11. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19:766–79.
12. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut. 2011;61:1036–40.
13. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota colon cancer control study. N Engl J Med. 1993;328:1365–71.