European Journal of Epidemiology
https://doi.org/10.1007/s10654-022-00929-7

ESSAY

Emulation of a target trial with sustained treatment strategies: an application to prostate cancer using both inverse probability weighting and the g-formula

Louisa H. Smith[1,2] · Xabier García-Albéniz[1,3] · June M. Chan[4,5] · Shoujun Zhao[5] · Janet E. Cowan[5] · Jeanette M. Broering[5,6] · Matthew R. Cooperberg[4,5] · Peter R. Carroll[5] · Miguel A. Hernán[1,7,8]

Received: 29 March 2022 / Accepted: 3 October 2022

Abstract

As with many chronic illnesses, recurrent prostate cancer generally requires sustained treatment to prolong survival. However, initiating treatment immediately after recurrence may negatively impact quality of life without any survival gains. Therefore, we consider sustained strategies for initiating treatment based on specific characteristics of prostate-specific antigen (PSA), which can indicate disease progression. We define the protocol for a target trial comparing treatment strategies based on PSA doubling time, in which androgen deprivation therapy is initiated only after doubling time decreases below a certain threshold. Such a treatment strategy means the timing of treatment initiation (if ever) is not known at baseline, and the target trial protocol must explicitly specify the frequency of PSA monitoring until the threshold is met, as well as the duration of treatment. We describe these and other components of a target trial that need to be specified in order for such a trial to be emulated in observational data. We then use the parametric g-formula and inverse-probability weighted dynamic marginal structural models to emulate our target trial in a cohort of prostate cancer patients from clinics across the United States.

Keywords Marginal structural models · g-formulza · Dynamic treatment strategies · Androgen deprivation therapy · Target trial

Introduction

When randomized trials are not feasible or timely, observational data can be used to emulate the randomized trial that, if conducted, would answer the question of interest – the target trial [1]. Observational emulations can result in effect estimates that match those from true randomized trials [2–5], but these comparisons typically benefit from the protocol of the target trial being explicitly specified [6]. In particular, the treatment strategies under comparison need to be unambiguously described, which may not be a simple task when the strategies are sustained over time.

As an example, consider the question of when to start treatment in people with previously treated prostate cancer who experience a rise in prostate-specific antigen (PSA) without overt metastasis or symptoms [7]. The specification of the treatment strategies includes not only the criteria for both treatment initiation (e.g., PSA greater than some value) and treatment discontinuation (e.g., side effects or a planned intermittent treatment strategy) [8, 9], but also the duration of the allowable period to start treatment after the

✉ Louisa H. Smith
l.smith@northeastern.edu

1 Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA

2 Roux Institute at Northeastern University, 100 Fore St, Portland, ME 04101, USA

3 RTI Health Solutions, Barcelona, Spain

4 Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, USA

5 Department of Urology, UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA

6 Department of Surgery, University of California, San Francisco, CA, USA

7 CAUSALab, Harvard TH Chan School of Public Health, Boston, MA, USA

8 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Published online: 26 October 2022

criteria are reached (the grace period), and the frequency of monitoring for those criteria.

Here we describe the components necessary to specify protocols for target trials involving sustained treatment strategies. As an illustration we specify and emulate in observational data a target trial of dynamic strategies for androgen deprivation therapy for recurrent prostate cancer. We illustrate the use of two methods to adjust for time-varying confounding: inverse probability weighting of dynamic marginal structural models [10, 11] and the parametric g-formula [12, 13].

The target trial

We previously emulated a target trial among people with prostate cancer and PSA-only relapse to compare immediate initiation of treatment vs. deferral of treatment [14]. Immediate initiation was defined as androgen deprivation therapy prescription or orchiectomy within three months of the PSA-based relapse, and deferral as a lack of treatment within two years of this relapse or until evidence of progression. The estimates from our emulation were compatible with those from two subsequent randomized trials [15, 16], which found small differences in all-cause mortality between the two treatment strategies. The 95% confidence intervals for both the observational and randomized effect estimates were very wide [14, 16].

One factor affecting prostate cancer prognosis after biochemical relapse is change in PSA over time [17]. However, the three previous studies [14–16] considered treatment initiation strategies that did not depend on evolving PSA levels. Assigning androgen deprivation therapy only to patients with worsening prognosis based on PSA kinetics might avoid or delay the costs and side effects of treatment [8] for others (who may not benefit). We therefore designed a target trial that assigns treatment only when rapid increases in PSA were observed. The key components of this protocol are summarized in Table 1.

Briefly, the trial would include individuals diagnosed with clinically localized prostate cancer treated with curative intent who later had evidence of recurrence that was only apparent as a rise in PSA (approximately the same criteria – which depend on whether initial treatment included surgery and/or radiation – as in the previous studies). Patients would be assigned to a treatment initiation strategy that depends on the speed of PSA change. Each eligible individual would be followed from the assignment of the treatment strategy (time zero) until death (the outcome of interest), administrative end of follow-up (10 years after time zero), or loss to follow-up (2 years without contributing clinical data), whichever occurs first. The data from the target trial could be used to estimate both the intention-to-treat effect and the per-protocol effect [18]. The treatment strategies and statistical analysis plan are described in more detail in the next two sections.

Target trial: treatment strategies

The target trial would compare treatment initiation strategies that depend on the individual's rate of change in PSA levels. A common measure of PSA kinetics is the PSA doubling time (PSADT), that is, the estimated time over which PSA would double, given observed values [19]. Specifically, we calculate PSADT from consecutive measurements of PSA at time s and time t as $\frac{\{\log(2\times \mathrm{PSA}_s)-\log(\mathrm{PSA}_s)\}\times\{\mathrm{date}_t-\mathrm{date}_s\}}{\log(\mathrm{PSA}_t)-\log(\mathrm{PSA}_s)} = \log(2)\frac{\mathrm{date}_t-\mathrm{date}_s}{\log\left(\frac{\mathrm{PSA}_t}{\mathrm{PSA}_s}\right)}$ where the difference between measurement dates is in days. The trial would include treatment strategies of the form "Start androgen deprivation therapy the first time PSADT drops below x days," where the threshold x varies from 0 to 360 in increments of 10, resulting in 37 different treatment strategies. The lower the threshold x, the later treatment initiation is expected to occur. A threshold of 0 means that treatment would never be initiated. The treatment duration would be left to be decided by the physician and patient, but this target trial does not allow for intermittent treatment: once treatment is discontinued for longer than one month, it is not to be re-initiated. This description of the treatment strategies is, however, incomplete for the following four reasons:

First, because treatment may be clinically indicated in situations not defined solely by PSADT, the treatment strategies need to specify the situations under which treatment is indicated regardless of PSADT. For example, "Start treatment the first time PSADT drops below x days, *or if a patient shows other signs of progression based on imaging or severe symptoms*." This specification of the strategy makes clear that a threshold of 0 would not necessarily result in treatment never being initiated because treatment will be initiated if the disease progresses clinically, regardless of PSADT values.

Second, because immediate initiation of treatment may be unfeasible, we need to also specify the period during which treatment can be started (the grace period). For example, "Start treatment *within the three months following* the first time PSADT drops below x days or the time a patient shows other signs of progression based on imaging or severe symptoms." In practice, randomized trials rarely specify the duration of the grace period because it is understood that treatment will be initiated reasonably soon after randomization. However, specifying the duration of the grace period in the target trial is required to emulate it using observational data. Otherwise, it would not be possible to determine whether an

Table 1 Description of a target trial to identify the optimal androgen deprivation therapy timing with respect to prostate-specific antigen doubling time, and of its emulation in observational data

Components	Target trial	Emulation of trial using observational data
Eligibility criteria	1. Histologically confirmed adenocarcinoma of the prostate, clinically staged cT3aN0M0 or lower	1. Same
	2. PSA relapse after definitive radical treatment (prostatectomy and/or radiotherapy), as evidenced by one of the following: a) PSA rise above 0.2 ng/mL beyond post-treatment nadir if initial treatment was prostatectomy (with or without radiation); b) PSA that did not fall lower than 0.2 ng/mL if treated with prostatectomy with salvage radiation; c) Three successive PSA rises at least 30 days apart if initial treatment was only radiation	2. Same
	3. No symptomatic disease requiring therapy, or any evidence of metastatic disease	3. Same. Operationalized as lack of any a) Positive findings on pelvis MRI, abdomen CT, pelvis CT, or bone scan at any time in the past; b) Severe symptoms (fatigue, bone pain, anorexia, weight loss, abdominal pain) at the time of relapse; c) Progression noted in physician notes
	4. Naive to ADT treatment (no orchiectomy, and any previous ADT was more than 1 year in the past, and not for longer than 12 months)	4. Same. Operationalized as: a) Never had orchiectomy at any time in the past; b) No prescribed ADT within the past year, or for more than 12 months at any time
	5. Life expectancy at least 5 years	5. Same. Operationalized based on National Comprehensive Cancer Network guidelines principles of life expectancy estimation, using the Social Security Administration life tables
	6. PSA doubling time at relapse of 30 days or more	6. Same
Treatment strategies	Initiate ADT (LHRH-agonist, LHRH-antagonist anti-androgen, second-line hormonal medications, or orchiectomy) within 3 months after PSA doubling time drops below a prespecified value, from 0 to 360 days in increments of 10. Under all strategies, ADT will be started within 3 months after a patient experiences further disease progression. Continuation of ADT after initiation will be left at the physician's and patient's discretion. Once ADT has been discontinued, it will not be reinitiated. Under all strategies, individuals will continue to have PSA measured and symptoms assessed at the physician's and patient's discretion. Patients complete study-specific surveys at baseline and every six months thereafter	Same
Treatment assignment	Each individual is randomized a treatment strategy defined by a PSADT threshold	Each individual is assigned to all treatment strategies
Follow-up	Patients are followed from treatment assignment (time zero) until death, loss to follow-up (24 months without contact with the study team via returned surveys, or physician visits), or administrative end of follow-up	Same
Outcome	Death from any cause	Same
Causal contrast	Intention-to-treat effect and per-protocol effect	Observational analog of per-protocol effect
Analysis plan	Intention-to-treat analysis: Survival curves and 5- and 10-year mortality estimates within each treatment arm. Adjustment for potential selection bias due to loss to follow-up via IP weighting. Per-protocol analysis: same except that individuals are censored at non-adherence and uncensored individuals are assigned IP weighs that are a function of baseline and time-varying variables. Alternatively, the per-protocol analysis may be based on the g-formula	Observational analog of a per-protocol analysis: same as in target trial except that analyses are not conducted separately by assigned treatment strategy, and we created 36 individuals (clones) per eligible patient and assigned one to each strategy when using censoring plus IP weighting

individual who initiated treatment, say, 1 year after meeting initiation criteria has data compatible with the protocol of the target trial.

Third, because initiation of treatment during the grace period may follow many patterns (e.g., most people start treatment at the beginning of the grace period, or at the end of the grace period, or uniformly throughout the grace period), we also need to specify the expected rate of treatment initiation during the grace period. For example, "Start treatment *with equal probability* during any of the three months following the first time PSADT drops below x days, or if a patient shows other signs of progression based on imaging or severe symptoms." Again, actual randomized trials rarely specify the expected rate of treatment initiation during the grace period, but this information is required for the observational emulation, as we discuss below.

Finally, because the initiation of treatment depends on the frequency of measurement of PSA and other characteristics, we also need to specify the intensity of monitoring. For example, "Start treatment with equal probability during any of the three months following the first time PSADT drops below x days, or if a patient shows other signs of progression based on imaging or severe symptoms. *Participants must visit their physician for tests, imaging, and/or symptom assessment in addition to completing surveys at home not less than once every 2 years.*"

Target trial: intention-to-treat analysis

We refer to strategy x as the strategy in which treatment is initiated within 3 months after PSADT drops below x or disease progresses. To estimate the intention-to-treat effect, we compare the survival curves between individuals assigned to each strategy x. That is, we estimate $\Pr(Y_t = 1|X = x)$ where $t = 0, \ldots, 120$ months of follow-up and Y_t an indicator of death from any cause during or before month t.

We could estimate $\Pr(Y_t = 1|X = x)$ nonparametrically. However, with so many treatment arms, we may wish to obtain more precise estimates by making parametric assumptions, e.g., by fitting a pooled logistic regression model for the discrete-time hazard $\Pr(Y_t = 1|X = x, Y_{t-1} = 0)$ with a time-varying intercept, modeled as natural cubic spline terms, and a covariate for treatment strategy x, also modeled as cubic splines and product ("interaction") terms with time. The model's predicted values are then used to compute the survival curve for each strategy [11].

Additionally, if imbalances existed in baseline characteristics across groups, the model would include them as covariates. We would then standardize the estimated probabilities to the distribution of the covariates to estimate marginal survival curves. Finally, if necessary, inverse probability (IP) weighting would be used to adjust for selection bias from loss to follow-up [20].

Target trial: per-protocol analysis

To estimate the per-protocol effect, we would compare the survival curves under adherence to each of the strategies. Because adherence is not randomized, we would need appropriate adjustment for (possibly time-varying) confounders, that is, prognostic factors that are determinants of adherence to the assigned treatment strategy, or their proxies. Let L_t be the vector of measured covariates in month t, including an indicator of having a clinic visit in month t, an indicator of disease progression, an indicator of symptoms (bone pain, fatigue, weight loss, anorexia, and abdominal pain), and PSA. L_0 contains baseline covariates: D'Amico risk group, comorbidities, and age at diagnosis; time from diagnosis to relapse; and calendar year, PSA, and PSADT at relapse.

To adjust for these confounders, we can use IP weighting or the parametric g-formula. Both methods have been described elsewhere [11, 13]; we review them here. Under the assumption that losses to follow-up and non-adherence happen at random within levels of the confounders, and that all models (described below) are correctly specified, both methods consistently estimate the survival probabilities had everyone adhered to each strategy and stayed under follow-up throughout the duration of the study. We used the non-parametric bootstrap with 1000 samples to estimate 95% confidence intervals under each approach.

IP weighting of a dynamic marginal structural model

We fit the same pooled logistic model as for the intention-to-treat analysis with two modifications.

First, individuals are censored if/when their observed treatment and covariate history is no longer consistent with their assigned strategy x. Censoring can occur for three reasons: the individual initiates treatment before their PSADT drops below x or before they experience disease progression, the individual does not initiate treatment within the 3-month grace period after meeting either of the criteria to start treatment, or the individual begins treatment again after having previously concluded it. During the grace period, no individual can be censored. Also, according to the protocol, individuals can stop at any time after treatment initiation, so no individuals can be censored while receiving treatment. However, the protocol does not allow for treatment re-initiation after

discontinuation, so individuals will be censored if they begin treatment after discontinuing it.

Second, to adjust for the potential selection bias introduced by censoring for non-adherence, at each month t we assign a time-varying IP weight to each individual. The denominator of the weight is the probability of remaining uncensored through month t, which is equal to the probability of remaining uncensored in months $k = 0, ..., t$. The probability of being uncensored is $f\left(A_k|\overline{L}_k, \overline{A}_{k-1}, \overline{Y}_{k-1} = 0\right)$, where A_t is an indicator of treatment during month t and $\underset{t}{A}$ the treatment history from time 0 through time t, during months with $A_{k-1} = 0$. Because our treatment strategies allow treatment discontinuation at any time, the denominator is 1 for months with $A_{k-1} = 1$. We can estimate $f\left(A_k|\overline{L}_k, A_{k-1} = 0, \overline{A}_{k-2}, \overline{Y}_{k-1} = 0\right)$ via a pooled logistic model, separately within treatment arms and separately among months with and without prior use of treatment during the study.

If the protocol of the target trial specifies strategies of the form: "Start treatment *with equal probability* during any of the three months following the [initiation threshold]," then we will ensure that the per-protocol effect is estimated under this initiation pattern by multiplying the weights during the grace period by an additional factor [11]. For a grace period of 3 months, that factor is $\frac{1}{4}$ for an initiator at eligibility, $\frac{1}{3}$ in the first month of the grace period, $\frac{1}{2}$ in the second, and 1 at the end of the grace period; for a non-initiator, the factors are $\frac{3}{4}, \frac{2}{3}, \frac{1}{2}$, and 1, respectively. This results in a pseudopopulation in which $\frac{1}{4}$ of those eligible initiate treatment immediately, $\frac{1}{3}$ of those remaining do so the next month, and so on, until all those who have not yet initiated do so at the conclusion of the grace period.

Finally, individuals are also censored at the end of any two-year period in which they did not visit a physician or complete a survey at least once. We can also estimate IP weights to adjust for potential selection bias due to this censoring [20].

Parametric g-formula

The g-formula can be viewed as a generalized form of standardization of the conditional hazard under each treatment strategy to the joint distribution of the time-varying covariates. To estimate each component of the g-formula, we can fit within each treatment arm a logistic model for $\Pr\left(Y_{t+1} = 1|\overline{L}_t, \overline{A}_t, Y_t = 0\right)$, and logistic or linear models for the conditional density of each of the time-varying covariates in the vector L_t. We also need to fit a logistic model for the conditional probability of discontinuation of treatment A_t because the protocol does not prescribe the probability of stopping treatment after initiating. In contrast, the probability of treatment initiation under each strategy is known: 0 before reaching the PSADT threshold, 1 within 3 months of reaching it ($\frac{1}{4}$ at eligibility, $\frac{1}{3}$ after one month of the grace period, $\frac{1}{2}$ after two, and 1 at the end of the grace period), and 0 again if treatment is discontinued.

Finally, we need to assign a monitoring strategy that aligns with our trial protocol. The probability of a clinic visit is estimated using a model in the observed data, but is set to 1 if there has been no visit in the last 23 months, guaranteeing monitoring at least every 2 years.

We then standardize the probability of the outcome under each strategy x by averaging over all treatment and covariate histories, using the modeled densities. The resulting integral can be approximated via Monte Carlo simulation.

Target trial: emulation

The Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) observational study of prostate cancer patients began enrollment in 1995. The study has been described in detail elsewhere [21, 22]. We used data through 2016, at which point over 14,000 biopsy-proven patients had been enrolled at over 40 U.S. clinics and followed prospectively. Physicians provided clinical data (diagnosis, start and stop dates of medications, outcomes, lab and imaging tests) and participants provided a follow-up survey approximately every six-12 months after a baseline questionnaire (quality of life, other health service use). Evidence of disease progression after relapse was based on clinical notes describing severe symptoms or metastases seen on imaging. We used consecutive PSA measurements to calculate PSADT. We set to the value of the 95th percentile of PSADT those values that were greater than the 95th percentile (because of very slow-growing PSA) [19] or were undefined because PSA was constant or decreased from one date to the next [23].

We used these data to emulate the eligibility criteria, treatment strategies, outcome, and follow-up of the target trial as summarized in Table 1.

Per-protocol analysis via IP weighting of a dynamic marginal structural model

We carried out the analysis described for the target trial (specifications for each model are shown in Table A1 in the Online Resource). However, since treatment strategies are not explicitly assigned at baseline in observational studies, some modifications had to be made. First, we estimated

the probability of treatment initiation among all eligible individuals, instead of separately within treatment arms. Second, we allowed for each individual to be part of the analysis for *each* strategy by copying the dataset for each value of x considered, and then censoring participants separately within each dataset when their observed data were not consistent with that strategy [10, 11].

We truncated the total weights at the 99th percentile to avoid near positivity violations, but the use of nontruncated weights resulted in similar estimates. All analyses were conducted in R version 3.4.3 [24].

Per-protocol analysis via the parametric g-formula

The estimation procedures for the observational emulation were the same as for the target trial. We used the R package gfoRmula [25].

Sensitivity analyses

To explore the effects of our choice of target trial protocol, we repeated the two analyses using a different distribution for treatment initiation during the grace period. We specified that the rate of treatment initiation during the grace period would be the same that would have been observed in the absence of an intervention until the end of the grace period, at which point treatment would be initiated if it had not been previously. For the IP weighting approach, this meant that the factors in the weights for treatment during the grace period were equal to 1. For the g-formula, we additionally fit a model for treatment distribution and during the grace period drew treatment values with probabilities estimated from that model.

In addition, we investigated whether assigning treatment initiation thresholds based on another function of PSA would better target those in need of treatment. Specifically, we repeated our original analyses but assigned treatment when *average* PSADT since relapse reached cutoffs between 0 and 1800, in increments of 150.

Finally, we conducted an unadjusted analysis by fitting an unweighted pooled logistic regression model for mortality in the censored and concatenated datasets, using only the terms for time and treatment strategy.

Results

After applying the eligibility criteria, we found 1,229 eligible individuals (Fig. 1). Their baseline characteristics are shown in Table 2. About 60% underwent radical prostatectomy as their original treatment, and 47% were assigned to a medium clinical risk group at that time. The median time to biochemical recurrence after diagnosis was 3.3 years.

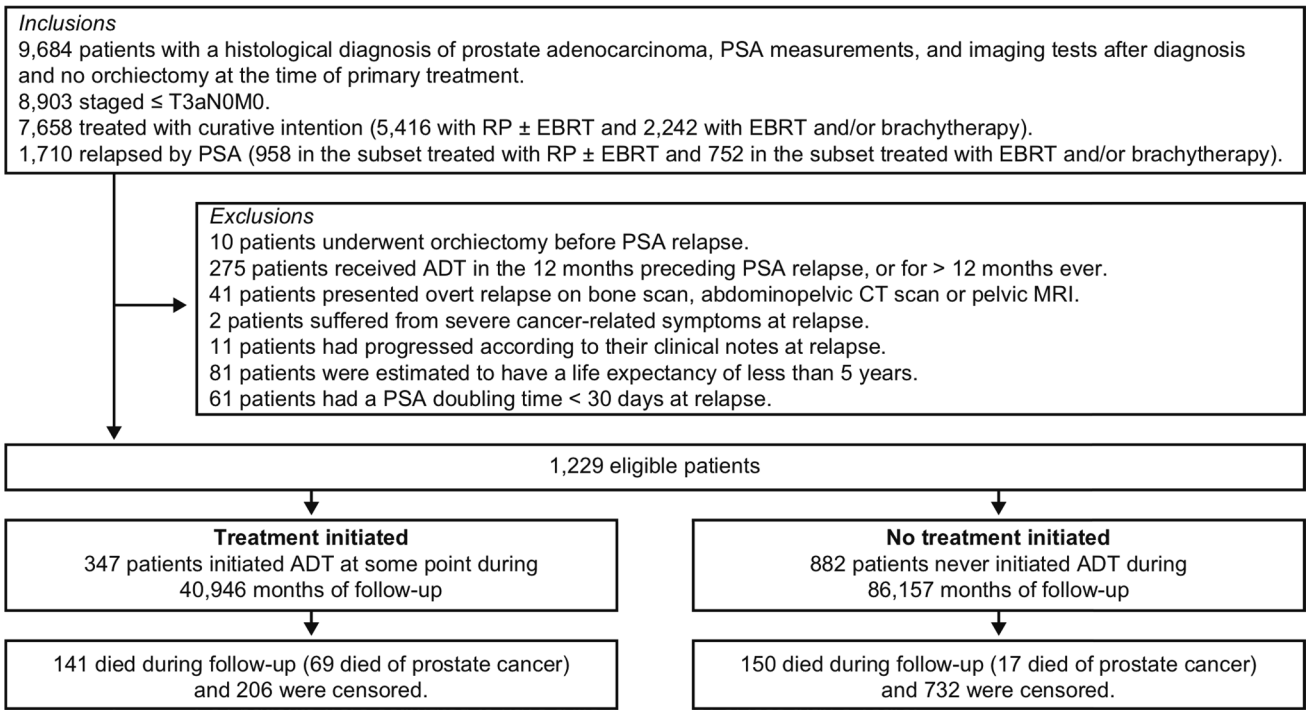

Fig. 1 Flowchart of patient selection from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) database through 2016 into the present study

Table 2 Baseline characteristics of the analytic sample of prostate cancer patients with biochemical recurrence in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE), 1995–2016 (n = 1229)

	N (%)
Age at diagnosis (%)	
40–49	22 (1.8)
50–59	246 (20.0)
60–69	577 (46.9)
70–79	367 (29.9)
80–89	17 (1.4)
Clinical risk group at diagnosis (%)	
Low	397 (32.3)
Medium	576 (46.9)
High	256 (20.8)
Comorbidities (%)	
0 or 1	512 (41.7)
More than 1	504 (41.0)
Missing	213 (17.3)
Original treatment (%)	
Radical prostatectomy	733 (59.6)
Radiotherapy	496 (40.4)
Relapse date (%)	
1980s	2 (0.2)
1990s	406 (33.0)
2000s	741 (60.3)
2010s	80 (6.5)
Years to relapse (median [IQR])	3.3 [2.1, 5.3]
PSA at relapse (median [IQR])	0.5 [0.3, 1.4]
PSADT at relapse (days) (median [IQR])	255.9 [124.1, 615.9]

Of the eligible individuals, 347 actually received androgen deprivation therapy of any kind at some point during follow-up, and 291 died from any cause. Because many individuals never initiated treatment, there were fewer person-months in the data that were consistent with treatment strategies defined by higher PSADT thresholds. For the treatment strategy defined by a threshold of 0, there were 64,247 person-months and 145 deaths. For the treatment strategy defined by a threshold of 360, there were 25,205 person-months and 64 deaths (Table A2 in the Online Resource).

The estimated survival was similar under all strategies (Fig. 2), though estimates were imprecise. Risk differences for 10-year mortality comparing the highest threshold (360) with the lowest (0) (i.e., earliest vs. latest initiation) were 0.02 (-0.31, 0.44) when estimated via IP weighting and -0.02 (-0.05, 0.04) when estimated via the parametric g-formula (Table A3 in the Online Resource). Results were similar when we varied the target trial protocol and the PSADT truncation level (Table A3 in the Online Resource), and the treatment thresholds (Table A4 in the Online Resource). When not adjusting for measured confounders, the risk was lower for a threshold of 0 than for earlier treatment (Fig. 2).

Discussion

We used observational data to emulate a target trial of several strategies for the initiation of treatment in individuals with asymptomatic biochemical recurrence of prostate cancer. We showed that estimating the per-protocol effect requires an unambiguous description of the treatment strategies, including the specification of instructions for treatment changes, the grace period for initiation, and the patterns of initiation during the grace period.

Our study does not add any conclusive evidence to determine the optimal treatment initiation in individuals with asymptomatic biochemical recurrence of prostate cancer. Though we estimated that initiating androgen deprivation therapy on the basis of PSA doubling time has little impact on all-cause mortality, very wide 95% confidence intervals imply that our data are equally compatible with harm, benefit, or no effect of early initiation of androgen deprivation therapy on survival. As expected, confidence intervals from the parametric g-formula approach were narrower, reflecting the additional parametric assumptions compared with the IP weighting approach.

Ideally, risk of deadly metastatic disease should be balanced against the threat to quality of life that hormonal treatment poses. US guidelines refrain from recommending a standard treatment in this situation due to uncertainty about timing after early signs of biochemical recurrence [26]. One reason for hesitation in assigning all relapsing patients to immediate therapy is the prolonged timeline of cancer spread in most patients. On average, clinical metastasis becomes apparent 7–8 years after biochemical recurrence; due to the age of the affected population, this is beyond or around many patients' expected lifespan even without cancer [7]. One trial investigating delayed treatment found that 41% of individuals in the delayed therapy arm died without progressing to a need for androgen deprivation therapy [15]. Furthermore, the therapy leads to a number of short and long-term side effects, including weight gain and loss of muscle mass, osteoporosis and anemia, sexual dysfunction, and dementia [27] which can reduce quality of life with possibly little benefit.

When estimating the effect of treatment initiation strategies via IP weighting, the pattern of initiation during the grace period and the instructions for treatment continuation after initiation may be left unspecified. The effect estimates will then correspond to the patterns

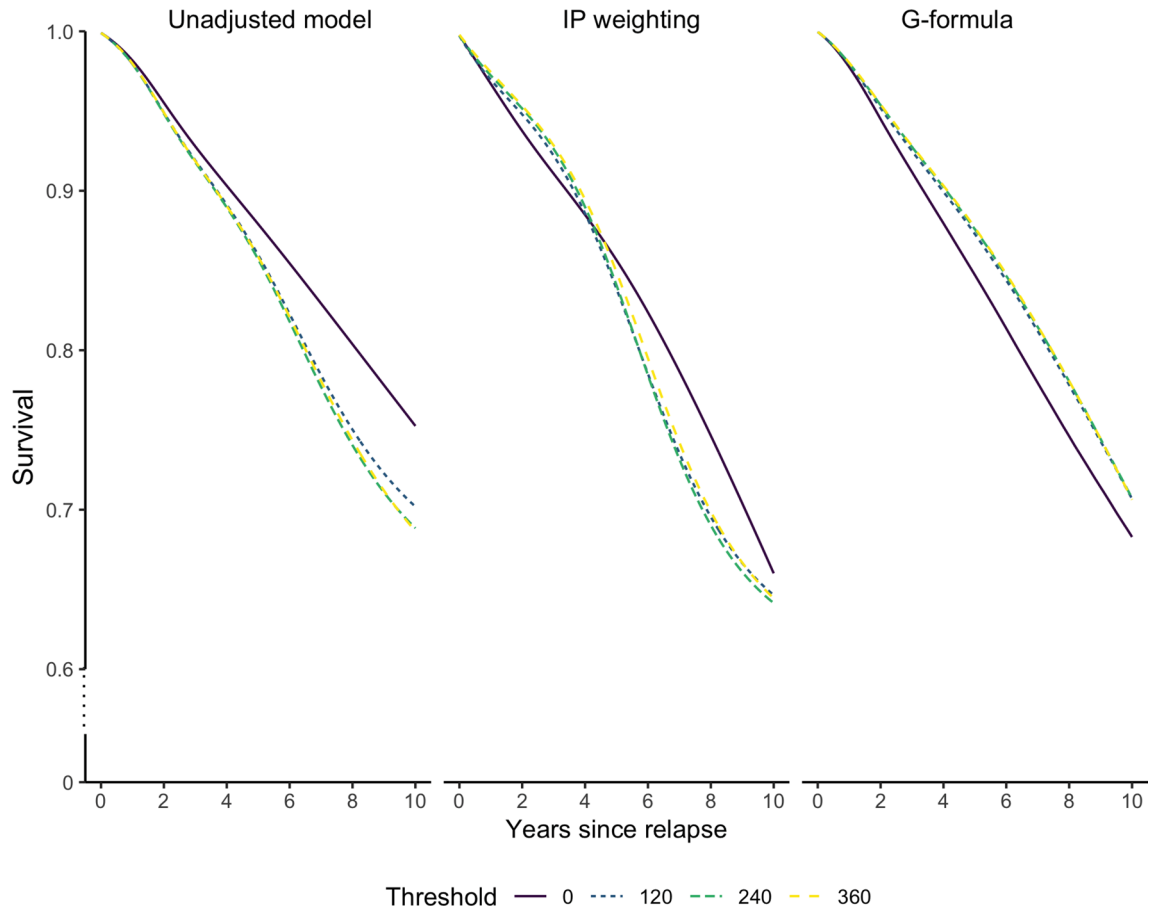

Fig. 2 Survival curves estimated via various methods, comparing several treatment strategies defined by PSA doubling time thresholds. Thresholds range from 0 (purple, least treatment) to 360 (yellow, most treatment)

of initiation during the grace period and of treatment continuation in the study population from which the data were collected. For example, if most individuals had intermittent treatment with androgen deprivation therapy after initiation (something that has been suggested to result in similar survival but better quality of life than continuous therapy [28]), then the IP weighted effect estimate would be interpreted as the effect of different treatment initiation strategies in a setting in which most individuals are treated intermittently thereafter. In contrast, when using the parametric g-formula, one needs to explicitly specify those treatment patterns. In our analysis, we specified continuous therapy after initiation when using the g-formula, a treatment pattern that roughly corresponds to what happened in the real world.

Estimating counterfactual quantities under exactly the same treatment strategy with both the IP weighting and g-formula approaches allows for direct comparison of different model specifications to estimate the same estimand. Comparable results are reassuring and suggest that model misspecification is not a serious problem. However, both methods rely on the same measured confounders and thus both sets of estimates would be biased if treatment initiation, or loss to follow-up, depended on factors not recorded in the database or recorded imperfectly. Note that adjustment for confounding was necessary because those treated at higher thresholds were at greater risk of mortality in unadjusted analyses but not after adjustment for the measured covariates.

In summary, estimating per-protocol effects from observational data requires a detailed specification of the protocol of the target trial with a special emphasis on the specification of the treatment strategies under comparison. The use of IP weighting, but not the parametric g-formula, allows to omit an explicit specification of some elements of the treatment strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10654-022-00929-7.

Funding The authors declare that no funds, grants, or other support directly supported the preparation of this manuscript.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study was approved by the Harvard Longwood Campus Insitutional Review Board (IRB13-2919).

Consent to participate Informed consent was obtained from all CaPSURE participants.

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. 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;19:766–79.
3. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ (Online). 2009;338:395–9.
4. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. Am J Epidemiol. 2012;175:250–62.
5. Petito LC, García-Albéniz X, Logan RW, et al. Estimates of overall survival in patients with cancer receiving different treatment regimens. JAMA Netw Open. 2020;3:1–13.
6. Lodi S, Phillips A, Lundgren J, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. Am J Epidemiol. 2019;188:1569–77.
7. Artibani W, Porcaro AB, De Marco V, et al. Management of biochemical recurrence after primary curative treatment for prostate cancer: a review. Urol Int. 2018;100:251–62.
8. Higano CS. Intermittent versus continuous androgen deprivation therapy. J Natl Compr Canc Netw. 2014;12:727–33.
9. Nguyen PL, Alibhai SMH, Basaria S, et al. Adverse effects of androgen deprivation therapy and strategies to mitigate them. Eur Urol. 2015;67:825–36.
10. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content. Int J Biostat. 2010;6:44398–9.
11. Cain LE, Robins JM, Lanoy E, et al. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. Int J Biostat. 2010;6:1–42.
12. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chronic Dis. 1987;40(Suppl 2):139S-161S.
13. Young JG, Cain LE, Robins JM, et al. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci. 2011;3:119–43.
14. Garcia-Albeniz X, Chan JM, Paciorek A, et al. Immediate versus deferred initiation of androgen deprivation therapy in prostate cancer patients with PSA-only relapse. An observational follow-up study. Eur J Cancer. 2015;51:817–24.
15. Duchesne GM, Woo HH, Bassett JK, et al. Timing of androgen-deprivation therapy in patients with prostate cancer with a rising PSA (TROG 03.06 and VCOG PR 01–03 [TOAD]): A randomised, multicentre, non-blinded, phase 3 trial. Lancet Oncol. 2016;17:727–37.
16. Loblaw A, Bassett J, D'Este C, et al. Timing of androgen deprivation therapy for prostate cancer patients after radiation: Planned combined analysis of two randomized phase 3 trials. J Clin Oncol. 2018;36:5018–5018.
17. Klayton TL, Ruth K, Buyyounouski MK, et al. Prostate-specific antigen doubling time predicts the development of distant metastases for patients who fail 3-dimensional conformal radiotherapy or intensity modulated radiation therapy using the Phoenix definition. Pract Radiat Oncol. 2011;1:235–42.
18. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med. 2017;377:1391–8.
19. Pound CR, Partin AW, Eisenberger MA, et al. Natural history of progression after PSA. J Am Med Assoc. 1999;281:1591–7.
20. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56:779–88.
21. Lubeck DP, Litwin MS, Henning JM, et al. The CaPSURE database: a methodology for clinical practice and research in prostate cancer. Urology. 1996;48:773–7.
22. Cooperberg MR, Broering JM, Litwin MS, et al. The contemporary management of prostate cancer in the United States: lessons from the cancer of the prostate strategic urologic research Endeavor (CaPSURE), a national disease registry. J Urol. 2004;171:1393–401.
23. Studer UE, Collette L, Whelan P, et al. Using PSA to guide timing of androgen deprivation in patients with T0–4 N0–2 M0 prostate cancer not suitable for local curative treatment (EORTC 30891). Eur Urol. 2008;53:941–9.
24. R Core Team. R 3.4.3. 2017.
25. McGrath S, Lin V, Zhang Z, et al. gfoRmula: An R package for estimating the effects of sustained treatment strategies via the parametric g-formula. Patterns. 2020;1: 100008.
26. NCCN Clinical Practice Guidelines in Oncology. Prostate Cancer. National Comprehensive Cancer Network; 2018.
27. Lonergan PE, Washington SL, Cowan, JE, et al. Androgen deprivation therapy and the risk of dementia after treatment for prostate cancer. J Urol. 2021; epub ahead of print.
28. Botrel TEA, Clark O, dos Reis RB, et al. Intermittent versus continuous androgen deprivation for locally advanced, recurrent or metastatic prostate cancer: a systematic review and meta-analysis. BMC Urol. 2014;14:9.