Title

Dynamic Risk Prediction of Treatment Discontinuation Using Patient-Reported Outcomes Data in the Phase III NSABP B-35 Trial

Journal

Cancer Prevention Research, 16(11)

ISSN

1940-6207

Authors

Calsavara, Vinicius F

Henry, Norah L

Hays, Ron D

et al.

Publication Date

2023-11-01

DOI

10.1158/1940-6207.capr-23-0216

Peer reviewed

Dynamic Risk Prediction of Treatment Discontinuation Using Patient-Reported Outcomes Data in the Phase III NSABP B-35 Trial

Vinicius F. Calsavara[1], Norah L. Henry[2], Ron D. Hays[3], Sungjin Kim[1], Michael Luu[1], Márcio A. Diniz[1], Gillian Gresham[1], Reena S. Cecchini[4], Greg Yothers[4], Patricia A. Ganz[5], André Rogatko[1], and Mourad Tighiouart[1]

ABSTRACT

Predicting an individual's risk of treatment discontinuation is critical for the implementation of precision chemoprevention. We developed partly conditional survival models to predict discontinuation of tamoxifen or anastrozole using patient-reported outcome (PRO) data from postmenopausal women with ductal carcinoma *in situ* enrolled in the NSABP B-35 clinical trial. In a secondary analysis of the NSABP B-35 clinical trial PRO data, we proposed two models for treatment discontinuation within each treatment arm (anastrozole or tamoxifen treated patients) using partly conditional Cox-type models with time-dependent covariates. A 70/30 split of the sample was used for the training and validation datasets. The predictive performance of the models was evaluated using calibration and discrimination measures based on the Brier score and AUC from time-dependent ROC curves. The predictive models stratified high-risk versus low-risk early discontinuation at a 6-month horizon. For anastrozole-treated patients, predictive factors included baseline body mass index (BMI) and longitudinal patient-reported symptoms such as insomnia, joint pain, hot flashes, headaches, gynecologic symptoms, and vaginal discharge, all collected up to 12 months [Brier score, 0.039; AUC, 0.76; 95% confidence interval (CI), 0.57–0.95]. As for tamoxifen-treated patients, predictive factors included baseline BMI, and time-dependent covariates: cognitive problems, feelings of happiness, calmness, weight problems, and pain (Brier score, 0.032; AUC, 0.78; 95% CI, 0.65–0.91). A real-time calculator based on these models was developed in Shiny to create a web-based application with a future goal to aid healthcare professionals in decision-making.

Prevention Relevance: The dynamic prediction provided by partly conditional models offers valuable insights into the treatment discontinuation risks using patient-reported outcome data collected over time from clinical trial participants. This tool may benefit healthcare professionals in identifying patients at high risk of premature treatment discontinuation and support intervention to prevent potential discontinuation.

Introduction

In patients at high risk of developing breast cancer, endocrine therapy such as tamoxifen or an aromatase inhibitor reduces the risk of developing cancer by about 50% (1). However, many people stop taking the medication before the recommended 5-year duration, primarily because of bothersome side effects, thereby limiting the potential benefit of the treatment. Identifying the factors associated with treatment discontinuation is important to increase the possibilities for intervention and prevention of discontinuation.

In the phase III, randomized, double-blind, placebo-controlled NSABP B-35 clinical trial (2, 3), postmenopausal women with ductal carcinoma *in situ* (DCIS) treated with breast conserving therapy and whole-breast irradiation were randomized to either the aromatase inhibitor anastrozole or to tamoxifen for 5 years. Anastrozole was shown to significantly improve the breast cancer–free interval compared with tamoxifen, especially in women less than 60 years of age, although the absolute differences were small (3). Persistence rates were similar in the two study arms, with approximately 30% of the participants discontinuing treatment before the planned 5-year duration. Using data from this clinical trial, we developed dynamic risk prediction models for early discontinuation of each drug being given for chemoprevention.

[1]Cedars-Sinai Cancer Center, Cedars-Sinai Medical Center, Los Angeles, California. [2]University of Michigan Rogel Cancer Center, Ann Arbor, Michigan. [3]University of California Los Angeles, Department of Medicine, Los Angeles, California. [4]University of Pittsburgh, Pittsburgh, Pennsylvania. [5]University of California Los Angeles Jonsson Comprehensive Cancer Center, Los Angeles, California.

Corresponding Author: Vinicius F. Calsavara, Cedars-Sinai Cancer Center, Cedars-Sinai Medical Center, 700 N San Vicente Blvd, PDC Green, Los Angeles, CA 90069. E-mail: vinicius.calsavara@cshs.org

Cancer Prev Res 2023;XX:XX–XX

doi: 10.1158/1940-6207.CAPR-23-0216

Materials and Methods

Clinical trial data

From January 6, 2003, to June 15, 2006, the NSABP B-35, a phase III double blind randomized, placebo-controlled trial enrolled a total of 3,104 patients. Patient-reported outcomes (PRO) data were collected from the first 1,275 patients who were closed on December 28, 2004. Among the participants, 1,187 individuals had both baseline and at least one follow-up assessment were included in the analysis of the quality-of-life data (589 in the anastrozole and 598 in the tamoxifen group).

This study was conducted using de-identified data obtained from the NRG Oncology Statistical Data Management Center for the completed clinical trial whose primary results have been published. Use of these data was deemed exempt from the requirements for Institutional Review Board review and approval in accordance with federal regulations, 45 CFR 46.101(b). Informed consent was obtained from the participants in the original study. Additional details of the trial design have been reported elsewhere (2, 3).

Predictor variables

Baseline demographic and clinical characteristics collected were as follows: age at randomization (measured in years), race and ethnicity, and body mass index (BMI; kg/m^2). Patient-reported survey instruments administered at all timepoints included the Medical Outcomes Study (MOS)-Short Form 12 (SF-12; ref. 4), the SF-36 Vitality Scale (5), a shortened version of the Breast Cancer Prevention Trial symptom checklist (6–8), a 10-item version of the Center for Epidemiologic Studies Depression Scale (CES-D; refs. 9, 10), and the 4-item MOS Sexual Problems scale (11). Questionnaires were administered at baseline and every 6 months after treatment initiation. A complete list of the candidate predictor variables considered in the modeling procedures is provided in the Supplementary Table S1.

Outcome variable

The outcome was treatment discontinuation, defined as the time from the date of the first treatment to the date of treatment discontinuation. The reasons for discontinuation include: (i) side effects and toxicity; (ii) complications; (iii) withdrawal or refusal; (iv) alternative therapy; (v) closed site without reassignment; (vi) loss to follow-up; and (vii) other complicating diseases (2). Treatment completion and treatment discontinuation due to death or disease progression were right-censored (0.68%–1.2% for death and 4.75%–5.85% for breast cancer recurrence with anastrozole and tamoxifen treatment, respectively).

Missing data

To maximize precision and power, we imputed data for covariates post-baseline using the last observation carried forward (LOCF) method (12). This method is commonly used in longitudinal studies when the missing data are assumed to be missing at random. The number of patients with missing data at each timepoint by arm is provided in the Supplementary Tables S2 and S3. No missing data were observed for the non–time-varying baseline covariates (age, BMI, and race and ethnicity).

Partly conditional survival models

Partly conditional survival models are suitable for risk prediction of time-to-event outcomes with a limited number of longitudinal predictors. They provide a flexible framework for dynamic risk prediction by modeling future outcome conditional on remaining in treatment up to a landmark time (s), and information accrued by that time. The approach is based on the partly conditional models (13) and the novel two-stage partly conditional models (14) that focus on patients still at risk at the landmark time and relate the covariates' history up to time s ($s > 0$) to the residual survival time τ ($\tau > 0$). That is, they provide dynamic predictions in the τ time interval from s using the covariates information available up to time s.

On the basis of the partly conditional survival models, we were able to estimate the patient's risk of treatment discontinuation by time $\tau + s$ given that the patient has been on treatment up to time s. Associations between covariates (baseline characteristics and time-varying PROs) and treatment discontinuation (survival data) were modeled using a semiparametric Cox model (PC_{Cox} model; ref. 13) and the novel two-stage partly conditional models (14). For the latter, the smoothed curve of the trajectory of a single symptom over time is obtained by fitting a linear mixed effect model, and the estimated adverse event values are used for a new prediction based on the best linear unbiased predictor (BLUP) estimator, resulting in a partly conditional Cox BLUP model (PC_{Cox} BLUP model; Supplementary Methods S1). **Figure 1** illustrates the predicted risks of treatment discontinuation for a hypothetical individual over a time horizon τ based on the observed PROs (e.g., headaches, hot flashes, and joint pain) and their smoothed curves given a landmark time s. For this individual, data on headaches, hot flashes, and joint pain are available at times t_1, t_2, and $t_3 = s$. Using all data from the trial and the PC_{Cox} model, the estimated probability that this patient discontinues treatment after time s is given by the blue dotted line. This estimated probability is around 0.68 by time $s+\tau$. The risk of treatment discontinuation using the PC_{Cox} BLUP model is shown by the blue solid line.

Model performance

Calibration and discrimination were used to assess the predictive performance of the PC_{Cox} and PC_{Cox} BLUP models in estimating the conditional probability of remaining in the treatment. To quantify how well a dynamic prediction is calibrated in terms of prediction error (PE), we considered an extended Brier score version (Supplementary Methods S2) to correctly deal with longitudinal covariate measurements and a survival outcome (15). In addition, we estimated the time-dependent ROC curve and examined the area under the ROC (AUC; ref. 16) to measure the ability to discriminate between patients at high and low risk of a treatment discontinuation in the future.

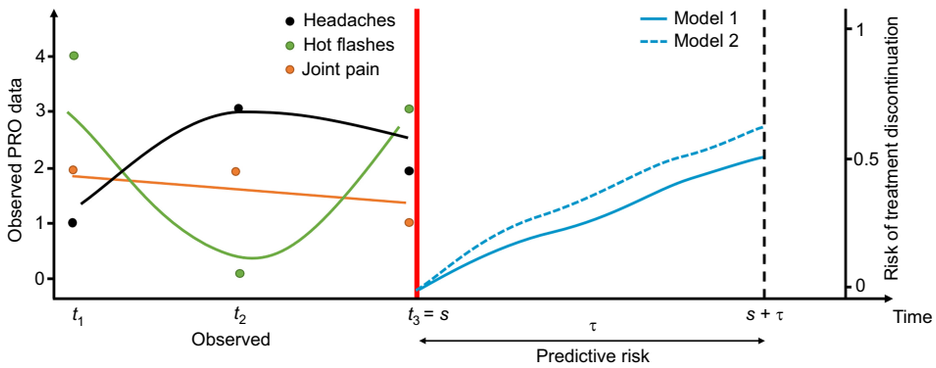

Figure 1.

An illustration of an individual dynamic prediction for treatment discontinuation at a horizon time τ using the available longitudinal PRO data collected up to time s. The dots over time represent the patient-reported symptoms, while the solid lines between t_1 and t_3 depict the smoothed trajectories of PRO single-items obtained by fitting a linear mixed-effect model using the BLUP estimator. The blue dashed and solid lines indicate the dynamic predictions from the two predictive partly conditional (PC) survival models based on the observed PRO data and the smooth curves, respectively.

The dynamic risk predictions for every patient in our dataset were based on the PC_{Cox} and PC_{Cox} BLUP models using a 70/30 random split of the dataset into a training and validation sets, respectively. We fitted both models by including the logarithm of time, baseline characteristics, and longitudinal PRO data as the predictors. For each patient, smooth patient-reported symptom measurements over time were obtained by the BLUP using the restricted maximum likelihood estimates from the linear mixed model (LMM) in the training dataset. The LMM modeled each longitudinal PRO data using fixed effects and a random intercept and slope. The final predictive model was selected using the backward stepwise variables selection procedure based on the Akaike information criterion (17). The training dataset was used to build the model, and its predictive performance was checked on the validation set for the selected s and τ values. The predicted performance measures were calculated by considering clinically relevant predictions at horizon times of $\tau = 6$ and $\tau = 12$ months, conditioned on data available up to $s = 6$ and $s = 12$ months. The low-risk and high-risk groups were determined using the time-dependent ROC curve, where for each pair (s, τ), the risk threshold (c) was calculated using Youden's index (18). For the low- and high-risk groups classified by the risk threshold, we assigned a negative/positive label and calculated diagnostic measures such as sensitivity, specificity, and accuracy.

All analyses were conducted using R software [Research Resource Identifier (RRID): SCR_001905] version 4.0 (R: A Language and Environment for Statistical Computing, Vienna, Austria. 2020, R Development Core Team) with the package *partlyconditional* (https://github.com/mdbrown/partlyconditional). All hypotheses were two-tailed with a 5% significance level.

Data availability

The data that support the findings in this case study are available from NRG Oncology but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. However, the request for data can be made to NRG Oncology at https://www.nrgoncology.org/Resources/Ancillary-Projects-Data-Sharing-Application.

Results

Patient characteristics

Of the 3,104 participants randomly assigned to receive anastrozole or tamoxifen, a subsample (n = 1,223) were enrolled in the quality of life study. Data were available for 1,187 patients who received a treatment and completed both baseline and at least one follow-up questionnaire. Among these, 589 patients were treated with anastrozole (412 training dataset and 177 validation cohort) and 598 received tamoxifen treatment (418 training dataset and 180 validation cohort). Of the 1,187 patients available for analysis, 333 (28.1%) discontinued treatment within 5 years. Of these who discontinued treatment, 173 (29.4%) received anastrozole with 127 (30.8%) in the training dataset and 46 (26%) in the validation cohort. Meanwhile, 160 (26.8%) patients received tamoxifen when they discontinuated treatment, with 120 (28.7%) in the training dataset and 40 (22.2%) in the validation cohort. The overall rates of remaining on tamoxifen treatment at 12, 18, and 24 months were 90.6%, 86.9%, and 85%, respectively, versus 90%, 86.5%, and 84.1% for anastrozole, respectively. The rates of treatment continuation showed similarity between the two study groups, with roughly 30% of participants discontinuing therapy prior to the intended 5-year period (Supplementary Fig. S1). The rates of remaining on anastrozole/tamoxifen treatments both the training and validation datasets are provided in Supplementary Figs. S2 and S3. The baseline characteristics are presented in **Table 1**.

Individual prediction of treatment discontinuation

The multivariable models fit of PC_{Cox} and PC_{Cox} BLUP for each treatment arm in the training cohort are summarized

Q6 **Table 1.** Baseline characteristics stratified by drug and training/validation datasets.

	Anastrozole ($n = 589$)[a]				Tamoxifen ($n = 598$)[b]			
	Training dataset ($n = 412$)		Validation dataset ($n = 177$)		Training dataset ($n = 418$)		Validation dataset ($n = 180$)	
Variable	**Continued ($n = 285$)**	**Stopped Early ($n = 127$)**	**Continued ($n = 131$)**	**Stopped Early ($n = 46$)**	**Continued ($n = 298$)**	**Stopped Early ($n = 120$)**	**Continued ($n = 140$)**	**Stopped Early ($n = 40$)**
Age at random assignment (years), median (IQR)	60 (55–66)	60 (55.5–65.5)	60 (55–66)	59.5 (53–69)	60 (56–66)	61 (56–66.25)	60 (55–65)	60 (57–66)
Race, No. (%)								
Non-Hispanic White	239 (83.86%)	110 (86.61%)	109 (83.21%)	40 (86.96%)	256 (85.91%)	108 (90%)	115 (82.14%)	34 (85%)
Non-Hispanic Black	29 (10.18%)	9 (7.09%)	11 (8.4%)	3 (6.52%)	22 (7.38%)	8 (6.67%)	14 (10%)	3 (7.5%)
Non-Hispanic Others or Multiple Ethnicity	8 (2.81%)	3 (2.36%)	4 (3.05%)	1 (2.17%)	12 (4.03%)	2 (1.67%)	6 (4.29%)	2 (5%)
Hispanic	9 (3.16%)	5 (3.94%)	7 (5.34%)	2 (4.35%)	6 (2.01%)	2 (1.67%)	5 (3.57%)	1 (2.5%)
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (0.67%)	0 (0%)	0 (0%)	0 (0%)
BMI (kg/m^2), median (IQR)	28.97 (25.47–32.60)	28.75 (25.43–33.40)	28.57 (24.79–32.83)	28.50 (24.48–34.15)	28.21 (24.92 - 32.92)	28.45 (24.58–34.19)	28 (24.49–31.96)	26.91 (24.43–34.52)
BMI (kg/m^2), No. (%)								
Underweight (BMI < 18.5)	0 (0%)	1 (0.79%)	3 (2.29%)	1 (2.17%)	6 (2.01%)	1 (0.83%)	0 (0%)	1 (2.5%)
Normal weight (18.5 ≤ BMI < 25)	64 (22.46%)	26 (20.47%)	33 (25.19%)	11 (23.91%)	72 (24.16%)	33 (27.5%)	37 (26.43%)	11 (27.5%)
Overweight (25 ≤ BMI < 30)	98 (34.39%)	41 (32.28%)	39 (29.77%)	15 (32.61%)	101 (33.89%)	34 (28.33%)	50 (35.71%)	13 (32.5%)
Obesity (BMI ≥ 30)	123 (43.16%)	59 (46.46%)	56 (42.75%)	19 (41.30%)	119 (39.93%)	52 (43.33%)	53 (37.86%)	15 (37.5%)

[a]In the training dataset, the 12-, 18-, and 24-month rates of remaining in treatment were 89.1%, 85.9% and 83.7%, respectively; In the validation dataset, the 12-, 18-, and 24-month rates of remaining in treatment were 92.1%, 88%, and 85.1%, respectively.

[b]In the training dataset the 12-, 18-, and 24-month rates of remaining in treatment were 89.9%, 85.8%, and 83.6%, respectively. In the validation dataset the 12-, 18-, and 24-month rates of remaining in treatment were 92.2%, 89.4%, and 88.2%, respectively.

Table 2. Estimated parameters of PC_{Cox} and two-stage PC_{Cox} BLUP models in the training dataset for Anastrozole as treatment.

		PC_{Cox} model			PC_{Cox} BLUP model		
Variable	**Category**	**Coefficient**	**robust SE**	***P* value**	**Coefficient**	**robust SE**	***P* value**
Log (BMI)	1-unit increment	0.682	0.522	0.192	0.854	0.519	0.100
Insomnia	1-unit increment	0.204	0.085	0.017	0.325	0.152	0.032
Joint pain	1-unit increment	−0.160	0.069	0.019	−0.396	0.128	0.002
Hot flashes	1-unit increment	0.154	0.079	0.051	0.277	0.126	0.028
Headaches	1-unit increment	−0.194	0.104	0.062	−0.259	0.212	0.220
Gynecologic symptoms	1-unit increment	0.673	0.244	0.006	1.570	0.539	0.004
Vaginal discharge	1-unit increment	−0.320	0.191	0.094	−0.927	0.464	0.046
Log (time)	1-unit increment	−0.335	0.131	0.011	−0.288	0.137	0.035

Abbreviation: SE denotes standard error.

in **Tables 2** and **3**. For the anastrozole-treated patients, BMI, insomnia, joint pain, hot flashes, headaches, gynecologic symptoms, and vaginal discharge were predictors of time to treatment discontinuation (**Table 2**), whereas for tamoxifen-treated patients the predictors were BMI, cognitive problems, joint pain, gynecologic symptoms, CESD-10, SF-12, weight problems, and pain with intercourse (**Table 3**). The results of the predictive models' performance in the validation cohort are summarized in **Table 4**. For the four sets of s and τ, the PEs ranged from 0.0391 to 0.0784 for the anastrozole arm and from 0.0242 to 0.0451 for the tamoxifen arm, regardless of the fitted model. As expected, the lowest PEs were observed when $\tau = 6$ months, regardless of the patient's treatment.

In both treatment arms, the best discrimination between those with and without treatment discontinuation was achieved for the PC_{Cox} BLUP model when predicting risk discontinuation within the next 6 months ($\tau = 6$) using patient characteristics and PRO history up to $s = 12$ months. For the anastrozole and tamoxifen arms, the model achieved an AUC = 0.76 [95% confidence interval (CI), 0.57–0.95] and AUC = 0.78 (95% CI, 0.65–0.91), respectively. For the pair, $(s = 12, \tau = 12)$, the PC_{Cox} BLUP achieved AUC = 0.69 (95% CI, 0.51–0.86) in anastrozole-treated patients and AUC = 0.73 (95% CI, 0.58–0.88) in tamoxifen-treated patients. PC_{Cox} model had the best discrimination ability for $(s = 6, \tau = 6)$ and $(s = 6, \tau = 12)$in the tamoxifen treatment with AUC = 0.74 (95% CI, 0.58–0.90) and AUC = 0.77 (95% CI, 0.67–0.88), respectively. On the other hand, the models were not useful in predicting early anastrozole-treatment discontinuation for the sets $(s = 6, \tau = 6)$ and $(s = 6, \tau = 12)$ because the 95% CIs for the AUC contain the value of 0.5. The estimated time-dependent ROC curves associated with the PC_{Cox} and PC_{Cox} BLUP models are shown in the Supplementary Figs. S4 and S5.

The risk threshold to classify patients into either the low- or high-risk groups was different between the two models (Supplementary Table S4). Of note, when considering the PC_{Cox} BLUP model for tamoxifen-treated patients and the following pairs: i) $(s = 12, \tau = 6)$; ii) $(s = 12, \tau = 12)$, the sensitivity and specificity at the risk threshold (c) values were: i) 100% (95% CI, 56.55%–100%) and 60.14% (95% CI, 51.95%–67.8%) at $c = 0.0324$, and ii) 100% (95% CI, 60.97%–100%) and 48.59% (95% CI, 50.52%–56.74%) at $c = 0.0483$, respectively. For anastrozole-treated patients, the risk threshold

Table 3. Estimated parameters of PC_{Cox} and two-stage PC_{Cox} BLUP models in the training dataset for Tamoxifen as treatment.

		PC_{Cox} model			PC_{Cox} BLUP model		
Variable	**Category**	**Coefficient**	**robust SE**	***P* value**	**Coefficient**	**robust SE**	***P* value**
BMI (kg/m^2)	Normal	1 (Reference)			1 (Reference)		
	Obesity	−0.220	0.287	0.445	−0.324	0.311	0.298
	Overweight	−0.723	0.310	0.020	−0.840	0.322	0.009
	Underweight	−2.174	1.124	0.053	−2.200	1.125	0.051
Cognitive problems	1-unit increment	0.331	0.107	0.002	0.528	0.156	0.001
Joint pain	1-unit increment	0.150	0.074	0.043	0.285	0.138	0.039
Gynecologic symptoms	1-unit increment	0.246	0.154	0.11	0.378	0.325	0.245
CESD-10: happiness item	1-unit increment	0.137	0.069	0.048	0.350	0.239	0.143
SF-12: calm and peaceful item	1-unit increment	−0.145	0.085	0.088	−0.395	0.203	0.052
Weight problems	1-unit increment	0.158	0.079	0.045	0.297	0.167	0.075
Lack of sexual interest	1-unit increment	−0.189	0.094	0.045	−0.322	0.154	0.037
Pain with intercourse	1-unit increment	0.126	0.084	0.134	0.248	0.13	0.056
Log (time)	1-unit increment	−0.629	0.126	<0.001	−0.777	0.135	<0.001

Abbreviation: SE denotes standard error.

Table 4. PE and area under the time-dependent ROC curve (AUC) based on a partly conditional Cox model (PC_{Cox} model) and partly conditional Cox BLUP model (PC_{Cox} BLUP model) by treatment.

Anastrozole				PC_{Cox} model		PC_{Cox} BLUP model	
s	τ	Events[a]	Patients at risk	PE	AUC (95% CI)	PE	AUC (95% CI)
6	6	7	166	0.0405	0.60 (0.35–0.85)	0.0405	0.58 (0.33–0.83)
	12	14	166	0.0783	0.45 (0.28–0.63)	0.0784	0.59 (0.42–0.76)
12	6	6	156	0.0391	0.71 (0.53–0.90)	0.0391	0.76 (0.57–0.95)
	12	11	156	0.0693	0.64 (0.45–0.82)	0.0695	0.69 (0.51–0.86)
Tamoxifen							
6	6	4	143	0.0242	0.74 (0.58–0.90)	0.0243	0.68 (0.51–0.86)
	12	8	143	0.0449	0.77 (0.67–0.88)	0.0451	0.73 (0.62–0.84)
12	6	5	138	0.0317	0.73 (0.58–0.87)	0.0315	0.78 (0.65–0.91)
	12	6	138	0.0339	0.69 (0.54–0.83)	0.0331	0.73 (0.58–0.88)

[a]Events represent the number of treatment discontinuation events that occurred between s and $s + \tau$.

values did not yield reasonable estimates of the sensitivity and specificity.

We illustrated our method for predicting tamoxifen-treatment discontinuation using the BMI collected at enrollment and the repeated measures of eight patient-reported symptoms listed in (**Table 3**). We predicted early treatment discontinuation at a horizon time of $\tau = 12$ months using information collected up to $s = 12$ months. **Fig. 2A** shows the

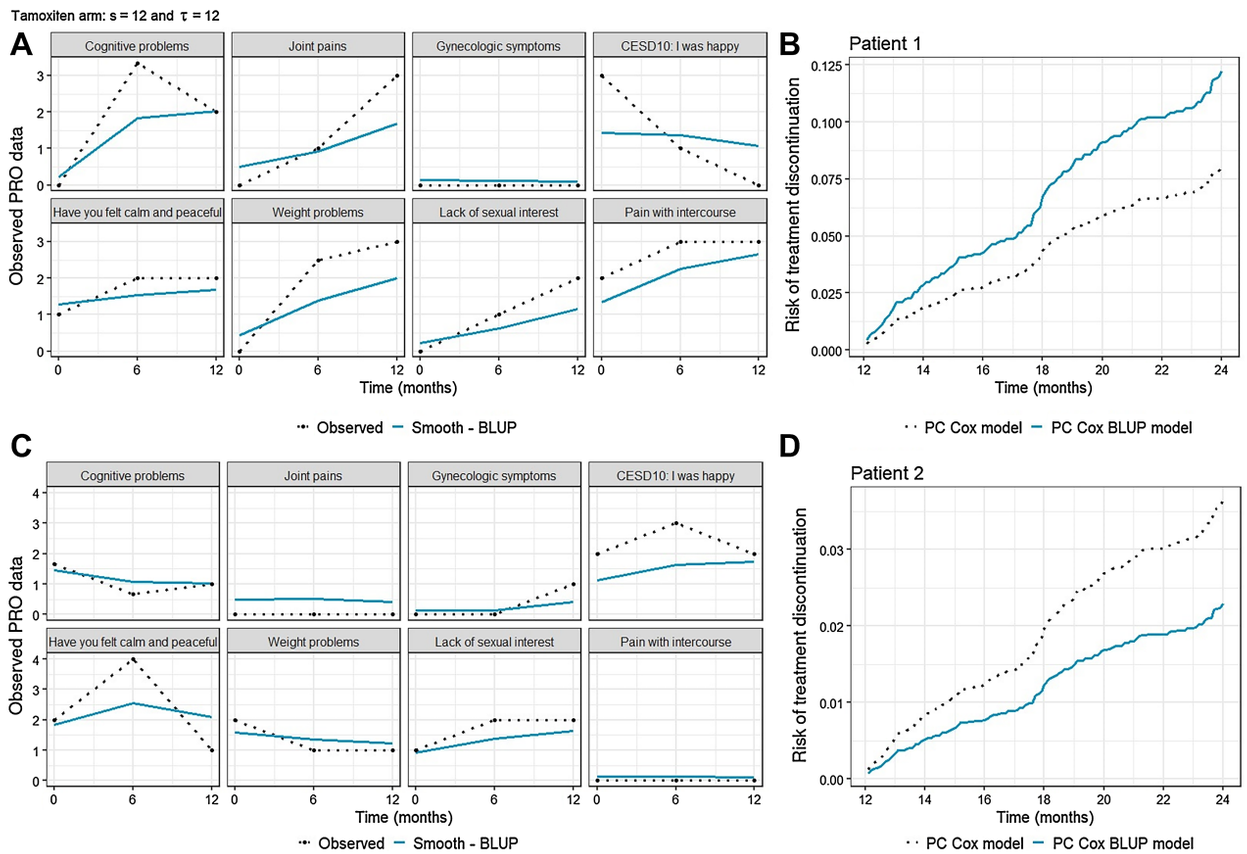

Figure 2.
Individual risk predictions for overweight patients in the validation cohort obtained from the PC_{Cox} and PC_{Cox} BLUP models at a horizon time of 12 to 24 months ($\tau = 12$), based on observed PRO data collected up to $s = 12$ months and their corresponding smooth trajectories using the BLUP estimator. **A,** Observed PRO data (dotted lines) over three timepoints and their corresponding smooth curves (blue solid lines) from Patient 1, who had early treatment discontinued at 14 months. **B,** Individual risk predictions for Patient 1. **C,** Observed PRO data (dotted lines) and their smooth trajectories (blue solid lines) from Patient 2, who completed treatment at 60 months. **D,** Individual risk predictions for Patient 2.

actual eight PROs of patient 1 at three time points, baseline, 6 months, and 12 months (shown by the black dotted line). This patient is also overweight at baseline ($25 \leq \text{BMI} < 30$). In the clinical data set, this patient discontinued treatment at 14 months, but this information was not used in the model. **Fig. 2B** shows the estimated probability of treatment discontinuation for patient 1 any time after 12 months but before 24 months. The risk of treatment discontinuation using the PC_{Cox} model by 24 months is 7.5%. Similar interpretation is made using the PC_{Cox} BLUP model. **Fig. 2C** shows similar information for patient 2 who is also overweight. However, patient 2 completed treatment by 60 months. As expected, the risk of treatment discontinuation using the PC_{Cox} model by 24 months is lower relative to patient 1, around 3.5%, see **Fig. 2D**.

In a prediction window length of 12 to 24 months, a higher estimated risk of treatment discontinuation was observed in patient 1, regardless of the predictive partly conditional survival models. According to the risk threshold values defined in the two sets ($s = 12,\ \tau = 6; c = 0.0324$) and ($s = 12,\ \tau = 12; c = 0.0483$), patient 1 was classified into the high-risk group, whereas patient 2 was assigned to the low-risk group. Overall, higher patient-reported symptom scores over time were observed for patients who discontinued treatment (see the time-varying PROs of the eight predictors on the top left of **Fig. 2**). A similar pattern was also observed for the dynamic prediction in a window length of 6 to 18 months based on available information collected up to 6 months (Supplementary Fig. S6).

Web-based treatment discontinuation predictive tool

We developed an online tool to facilitate the application of our predictive partly conditional survival models. Users can input the values of baseline characteristics and longitudinal predictors, and the tool produces the conditional probability of treatment discontinuation at a specific time horizon conditioned on a given landmark time (https://cshsbiostats.shinyapps.io/risk_anastrozole/ and https://cshsbiostats.shinyapps.io/risk_tamoxifen/). Additional details of the predictive tools are provided in the Supplementary Fig. S7.

Discussion

Patient-reported symptoms are commonly collected from patients at baseline and over the course of a clinical trial as indicators of toxicity, and are associated with a shorter time to treatment discontinuation (2). Premature discontinuation of treatment can impact the assessment of treatment efficacy. Therefore, identifying the predictors of early discontinuation is important for both routine clinical care and for the conduct of clinical trials.

We used partly conditional survival models (PC_{Cox} and PC_{Cox} BLUP models) based on trajectories of patient-reported symptoms and time to treatment discontinuation in postmenopausal women with DCIS treated with breast-conserving therapy. We also used partly conditional models to obtain dynamic risk predictions at the patient level, providing practical and useful information to support individualized decisions for the patient's treatment.

The analytical framework applied in this study provides insights into the probability of early treatment discontinuation in the future, using a patient's baseline characteristics and longitudinal assessments obtained early in the treatment course. In addition, in the presence of large within-patient variability in the longitudinal measurements, predictions based on the two-stage PC_{Cox} BLUP model provides a more robust approach once the patient-reported data are smoothed prior to estimation, which can improve the prediction model's performance.

The predictive models were trained separately for each drug and were internally validated. The patient's risk models included BMI, insomnia, joint pain, hot flashes, headaches, gynecologic symptoms, and vaginal discharge for the anastrozole-treated patients and BMI, cognitive problems, joint pain, gynecologic symptoms, CESD-10: happiness item, SF-12: calm/peaceful item, weight problems and pain with intercourse for the tamoxifen-treated patients. The PC_{Cox} BLUP model showed good calibration and discriminative ability for both drugs to predict treatment discontinuation at horizon times $\tau = 6$ and 12 months using information collected up to $s = 12$ months. In the tamoxifen group, the PC_{Cox} model achieved higher AUC values than the PC_{Cox} BLUP model in predicting premature treament discontinuation in the timeframes of 6 and 12 months using the trajectory history up to $s = 6$ months. In the anastrozole group, the both models displayed poor performance in accurately predicting premature treatment discontinuation at 6 and 12 months, using available information up to $s = 6$ months.

Our study has several strengths associated with the use of partly conditional models. Predictive models were developed using novel statistical approaches to identify the important predictors of outcome. Obtaining dynamic predictions at the patient level allowed us to identify critical timepoints that could alert healthcare providers and guide treatment. The predictive performance of our models achieved satisfactory results for calibration and discrimination measures in the validation cohort to predict premature treatment discontinuation in both arms, except when information was available for up to 6 months for patients receiving anastrozole. The highest AUC values were obtained for the timeframe of 6 months using accumulated information up to $s = 12$ months. Furthermore, we developed an online tool for clinicians to facilitate practical application of our predictive models.

There are also some caveats related to the interpretation of the study results. Firstly, we evaluated the predictive models' performance using the area under the time-dependent ROC curve, a measure that is insensitive to detecting small differences in discriminative ability between the two models (19, 20). The premature treatment

discontinuation rates were low in both study arms at the timepoints used in the analysis (12-, 18-, and 24-month dropout rates were approximately 10%, 13%, and 15%, respectively). Approximately 30% of the participants discontinued treatment before the intended 5-year duration, and the persistence rates over time showed similarity between the two study arms. The predictive models were trained and validated using B-35 clinical trial participants, and may reflect a more motivated patient group compared with the general population. When PRO data were missing, the LOCF method was applied. This approach has been criticized in the statistical literature (21). In addition, missing baseline data associated with PRO data were not imputed, reducing the amount of available information from baseline covariates, which can reduce the predictive models' performance.

In conclusion, our study identified important patient-reported symptoms and baseline factors that can be used to predict early treatment discontinuation using two models suitable for dynamic risk prediction that incorporate longitudinal PRO data. The incorporation of these well-performing survival models into an online tool is of potential benefit for healthcare professionals to identify patients at high risk of premature treatment discontinuation and intervention to prevent potential discontinuation. Future research should externally validate partly conditional models and test the feasibility and acceptability of the Shiny web-based prediction tool.

Authors' Disclosures

N.L. Henry reports grants from NIH during the conduct of the study; personal fees from Up-To-Date, AstraZeneca; and personal fees from Myovant Sciences outside the submitted work. R.D. Hays reports grants from NIH during the conduct of the study. S. Kim reports grants from NIH/NCI during the conduct of the study. R.S. Cecchini reports grants from NCI during the conduct of the study. P.A. Ganz reports grants from NCI during the conduct of the study. A. Rogatko reports grants from NCI during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

V.F. Calsavara: Conceptualization, software, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. **N.L. Henry:** Conceptualization, writing–review and editing, interpreted the results. **R.D. Hays:** Conceptualization, writing–review and editing, interpreted the results. **S. Kim:** Data curation, writing–review and editing, interpreted the results. **M. Luu:** Software, writing–review and editing, interpreted the results. **M.A. Diniz:** Conceptualization, writing–review and editing, interpreted the results. **G. Gresham:** Writing–review and editing, interpreted the results. **R.S. Cecchini:** Writing–review and editing, interpreted the results. **G. Yothers:** Writing–review and editing, he has acquired the data and interpreted the results. **P.A. Ganz:** Methodology, writing–review and editing, interpreted the results. **A. Rogatko:** Conceptualization, methodology, interpreted the results. **M. Tighiouart:** Conceptualization, supervision, methodology, writing–review and editing, interpreted the results.

Acknowledgments

This work is supported in part by the NCI of the NIH (1U01CA232859–01; to V.F. Calsavara, N.L. Henry, R.D. Hays, S. Kim, M. Luu, M.A. Diniz, G. Gresham, R.S. Cecchini, G. Yothers, P.A. Ganz, A. Rogatko, M. Tighiouart); the NIH National Center for Advancing Translational Sciences UCLA CTSI (UL1 TR001881–01; to V.F. Calsavara, M.A. Diniz, M. Tighiouart, A. Rogatko); and NCI grants P01CA233452–02, U10CA180868, UG1CA189867, and U10CA180822. This program is supported by funding provided through the Cancer Moonshot.

The authors acknowledge members of the U01 Scientific Advisory Committee (Lari Wenzel, PhD; Claire Snyder, PhD; Michael Brundage, MD; and Elisa Long, PhD).

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevres.aacrjournals.org/).

Received May 25, 2023; revised August 22, 2023; accepted September 22, 2023; published first September 26, 2023.

References

1. Nelson HD, Fu R, Zakher B, Pappas M, McDonagh M. Medication use for the risk reduction of primary breast cancer in women: updated evidence report and systematic review for the US preventive services task force. JAMA 2019;322:868–86.
2. Henry NL, Kim S, Hays RD, Diniz MA, Luu M, Cecchini RS, et al. Toxicity index, patient-reported outcomes, and early discontinuation of endocrine therapy for breast cancer risk reduction in NRG Oncology/NSABP B-35. J Clin Oncol 2021;39:3800–12.
3. Margolese RG, Cecchini RS, Julian TB, Ganz PA, Costantino JP, Vallow LA, et al. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma *in situ* undergoing lumpectomy plus radiotherapy (NSABP B-35): a randomized, double-blind, phase III clinical trial. Lancet 2016;387:849–56.
4. Ware J Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.
5. Ware JE. SF-36 health survey. Manual and interpretation guide. The Health Institute 1993;6:1–6:22.
6. Day R, Ganz PA, Costantino JP, Cronin WM, Wickerham DL, Fisher B. Health-related quality of life and tamoxifen in breast cancer prevention: a report from the national surgical adjuvant breast and bowel project P-1 study. J Clin Oncol 1999;17:2659.
7. Cella D, Land SR, Chang C-H, Day R, Costantino JP, Wolmark N, et al. Symptom measurement in the Breast Cancer Prevention Trial (BCPT) (P-1): psychometric properties of a new measure of symptoms for midlife women. Breast Cancer Res Treat 2008; 109:515–26.
8. Stanton AL, Bernaards CA, Ganz PA. The BCPT symptom scales: a measure of physical symptoms for women diagnosed with or at risk for breast cancer. J Natl Cancer Inst 2005;97:448–56.
9. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D. Am J Prev Med 1994;10:77–84.
10. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1: 385–401.

11. Ganz PA, Cecchini RS, Julian TB, Margolese RG, Costantino JP, Vallow LA, et al. Patient-reported outcomes with anastrozole versus tamoxifen for postmenopausal patients with ductal carcinoma *in situ* treated with lumpectomy plus radiotherapy (NSABP B-35): a randomized, double-blind, phase III clinical trial. Lancet 2016;387:857–65.
12. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.
13. Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. Biometrics 2005;61:379–91.
14. Maziarz M, Heagerty P, Cai T, Zheng Y. On longitudinal prediction with time-to-event outcome: comparison of modeling options. Biometrics 2017;73:83–93.
15. Schoop R, Graf E, Schumacher M. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. Biometrics 2008;64:603–10.
16. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337–44.
17. Yamashita T, Yamashita K, Kamimura R. A stepwise AIC method for variable selection in linear regression. Commun Stat - Theory Methods 2007;36:2395–403.
18. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.
19. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
20. Sweeting MJ, Barrett JK, Thompson SG, Wood AM. The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study. Stat Med 2017;36:4514–28.
21. Lachin JM. Fallacies of last observation carried forward analyses. Clinical Trials 2016;13:161–8.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Q1: Page: 1: Author: Per journal style, genes, alleles, loci, and oncogenes are italicized; proteins are roman. Please check throughout to see that the words are styled correctly. AACR journals have developed explicit instructions about reporting results from experiments involving the use of animal models as well as the use of approved gene and protein nomenclature at their first mention in the manuscript. Please review the instructions at http://aacrjournals.org/content/authors/editorial-policies#genenomen to ensure that your article is in compliance. If your article is not in compliance, please make the appropriate changes in your proof.

Q2: Page: 1: Author: Please verify the drug names and their dosages used in the article.

Q3: Page: 1: Author: Please verify the affiliations and their corresponding author links.

Q4: Page: 1: Author: Please verify the corresponding author details.

Q5: Page: 3: Author: Please confirm quality/labeling of all images included within this article. Figure labels should be legible at 100% zoom of the PDF file to ensure readability in print. Please flag any figures and/or label font sizes that should be adjusted.

Q6: Page: 4: Author: Please verify the layout of Tables for correctness.

Q7: Page: 8: Author: The Authors' Disclosures statement that appears in the proof incorporates the information from forms completed and signed off on by each individual author. No factual changes can be made to disclosure information at the proof stage. However, typographical errors or misspelling of author names should be noted on the proof and will be corrected before publication. Please note if any such errors need to be corrected. Is the disclosure statement correct?

Q8: Page: 8: Author: The contribution(s) of each author are listed in the proof under the heading "Authors' Contributions." These contributions are derived from forms completed and signed off on by each individual author. If you make changes to these contributions, you must inform the affected author(s).

AU: Below is a summary of the name segmentation for the authors according to our records. The First Name and the Surname data will be provided to PubMed when the article is indexed for searching. Please check each name carefully and verify that the First Name and Surname are correct. If a name is not segmented correctly, please write the correct First Name and Surname

on this page and return it with your proofs. If no changes are made to this list, we will assume that the names are segmented correctly, and the names will be indexed as is by PubMed and other indexing services.

First Name	**Surname**
Vinicius F.	Calsavara
Norah L.	Henry
Ron D.	Hays
Sungjin	Kim
Michael	Luu
Márcio A.	Diniz
Gillian	Gresham
Reena S.	Cecchini
Greg	Yothers
Patricia A.	Ganz
André	Rogatko
Mourad	Tighiouart