**SPECIAL ISSUE ARTICLE**

WILEY

# Accounting for time-dependent treatment use when developing a prognostic model from observational data: A review of methods

**Romin Pajouheshnia[1,2]** (ID) | **Noah A. Schuster[1]** |
**Rolf H. H. Groenwold[3]** | **Frans H. Rutten[1]** |
**Karel G. M. Moons[1]** | **Linda M. Peelen[1]**

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[2]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

[3]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

**Correspondence**

Romin Pajouheshnia, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Huispost Str. 6.131, PO Box 85500, 3508 GA, Utrecht, The Netherlands; or Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, PO Box 80082, 3508 TB, Utrecht, The Netherlands.
Email: R.Pajouheshnia@uu.nl

**Funding information**

Netherlands Organisation for Scientific Research, Grant/Award Number: 917.16.430, 9120.8004, and 918.10.615

Failure to account for time-dependent treatment use when developing a prognostic model can result in biased future predictions. We reviewed currently available methods to account for treatment use when developing a prognostic model. First, we defined the estimands targeted by each method and examined their mechanisms of action with directed acyclic graphs (DAGs). Next, methods were implemented in data from 1,906 patients; 325 received selective β-blockers (SBBs) during follow-up. We demonstrated seven Cox regression modeling strategies: (a) ignoring SBB treatment; (b) excluding SBB users or (c) censoring them when treated; (d) inverse probability of treatment weighting after censoring (IPCW), including SBB treatment as (e) a binary or (f) a time-dependent covariate; and (g) marginal structural modeling (MSM). Using DAGs, we demonstrated IPCW and MSM have the best properties and target a similar estimand. In the case study, compared to (a), approaches (b) and (e) provided predictions that were 1% and 2% higher on average. Performance (c-statistic,

Brier score, calibration slope) varied minimally between approaches. Our review of methods confirmed that ignoring treatment is theoretically inferior, but differences between the prediction models obtained using different methods can be modest in practice. Future simulation studies and applications are needed to assess the value of applying IPCW or MSM to adjust for treatments in different treatment and disease settings.

**KEYWORDS**

Cox regression, inverse probability weights, marginal structural model, prediction, prognosis

# 1 | BACKGROUND

Prognostic models provide information that can be used to guide physicians and patients when making medical decisions. For example, if a patient has a high predicted probability of a poor health outcome, the physician may recommend preventative medication or referral for specialist treatment. For a prognostic model to be useful for guiding individual decisions regarding interventions, predictions provided by the model should ideally reflect what a person's risk of developing a certain outcome would be if there were to be no intervention (Liew, Doust, & Glasziou, 2011). This has been termed the "untreated risk" of an outcome (Groenwold et al., 2016; Pajouheshnia, Damen, Groenwold, Moons, & Peelen, 2017) and can be formally defined as the probability $\Pr(Y^{T=0} \mid P)$, where $Y^{T=0}$ is the outcome if individuals were to remain untreated and P is a vector of predictors (Sperrin et al., 2018).

Data used to develop prognostic models often come from observational studies or longitudinal medical databases (e.g., electronic health record data) in which the study sample is composed of both treated and untreated individuals. In particular, individuals who were not receiving treatment when entering the cohort may start treatment during follow-up—an issue termed treatment "drop-in" (Liew et al., 2011; Sperrin et al., 2018). Unfortunately, conventional approaches to develop a prognostic model using data from a partly treated study population will result in a model that does not provide predictions of untreated risk, instead predicting $\Pr(Y \mid P)$. Assuming the treatment is effective in reducing the risk of developing the outcome of interest, the observed risk of individuals in the development sample will, on average, be lower than the risk had they remained untreated. Therefore, risk predictions for future patients provided by the model will underestimate the true untreated risk of the outcome (Groenwold et al., 2016). This could have a number of negative consequences, including possible under-treatment in future patients. Given that a contemporary yet untreated population in many cases no longer exists, several researchers have suggested that the effects of treatment should be accounted for in the analysis when developing a prognostic model (Cheong-See et al., 2016; Groenwold et al., 2016; Liew et al., 2011; Liew, Doust, & Glasziou, 2012; Moons et al., 2015; Pajouheshnia et al., 2017; Sperrin et al., 2018). However, as of yet, there is no clear consensus over whether one analytical approach to deal with the effect of treatment is to be preferred over another.

Simulation studies have been conducted to investigate the impact of methods to account for the effects of treatments on the accuracy of prognostic models. Groenwold et al. (2016) showed that, for a point (time-invariant) treatment, including treatment as a covariate in a prediction model resulted in better predictive performance than ignoring treatment or excluding treated individuals from the development set. This was extended by Sperrin et al. (2018), who proposed the use of marginal structural models (MSMs) to model untreated risk. Through simulations, MSMs were shown to provide unbiased estimates of $Pr(Y^{T=0} | P)$, the target estimand of a prognostic model.

In this article, we review currently available methods that might be considered when trying to develop a prognostic model in a partly treated patient cohort, with specific focus on time-varying treatment use after the study baseline. We first outline the theoretical properties of different methods, using directed acyclic graphs (DAGs) to explain their underlying assumptions. Following this, we illustrate how the methods can be applied in practice in a case study and examine differences between the resulting models.

## 2 | METHODS TO ACCOUNT FOR TREATMENT USE

We consider six approaches that try to remove the effect of treatment drop-in in the development data. The first three approaches remove treatment from the data set in different ways. The next three approaches explicitly model treatment effects. All approaches are based on developing a Cox model with the same predictors as a standard model that ignores treatment use. In this section, we describe the effect of each approach on the associations between predictors in our model (P), treatment use (T), and mortality (Y), using causal diagrams. Figure 1a represents these relationships when treatment is ignored. For each of the approaches (or estimators), we consider whether and how they may provide unbiased estimates of our estimand of interest, namely, $Pr(Y^{T=0} | P)$.

### 2.1 | Excluding or censoring treatment users from the cohort

A seemingly straightforward way to remove the effect of treatment drop-in is to restrict the analysis to patients who do not receive the treatment during follow-up or to include follow-up of those patients until the moment they start treatment and censor them from that moment onward.

As illustrated in Figure 1b, restricting the analysis to individuals who did not receive a prescription (T = 0) will remove the effect of treatments on mortality from the cohort. However, as previously mentioned, in observational data, treatment is typically associated with predictors in the model (P). Restricting the analysis to the T = 0 subset has two problems: (a) the distribution of predictors and outcomes in the subset may not represent the target population for the model and (b) in the presence of an unobserved confounder (U) of the treatment–outcome association, T now serves as a collider, which is represented by the two incoming arrows. Conditioning on T will alter the relation between P and U, which may lead to biased estimation of the prognostic model (the association between P and Y; see Figure 1c; Hernán, Hernández-Díaz, & Robins, 2004). Alternatively, censoring individuals at the moment of treatment drop-in will retain a representative risk distribution in the data at baseline and retain more information on follow-up. However, as with restriction, if censoring is "informative," that is, associated with P and U, selection bias might affect the estimated model parameters, instead estimating Pr(Y | P, T = 0).
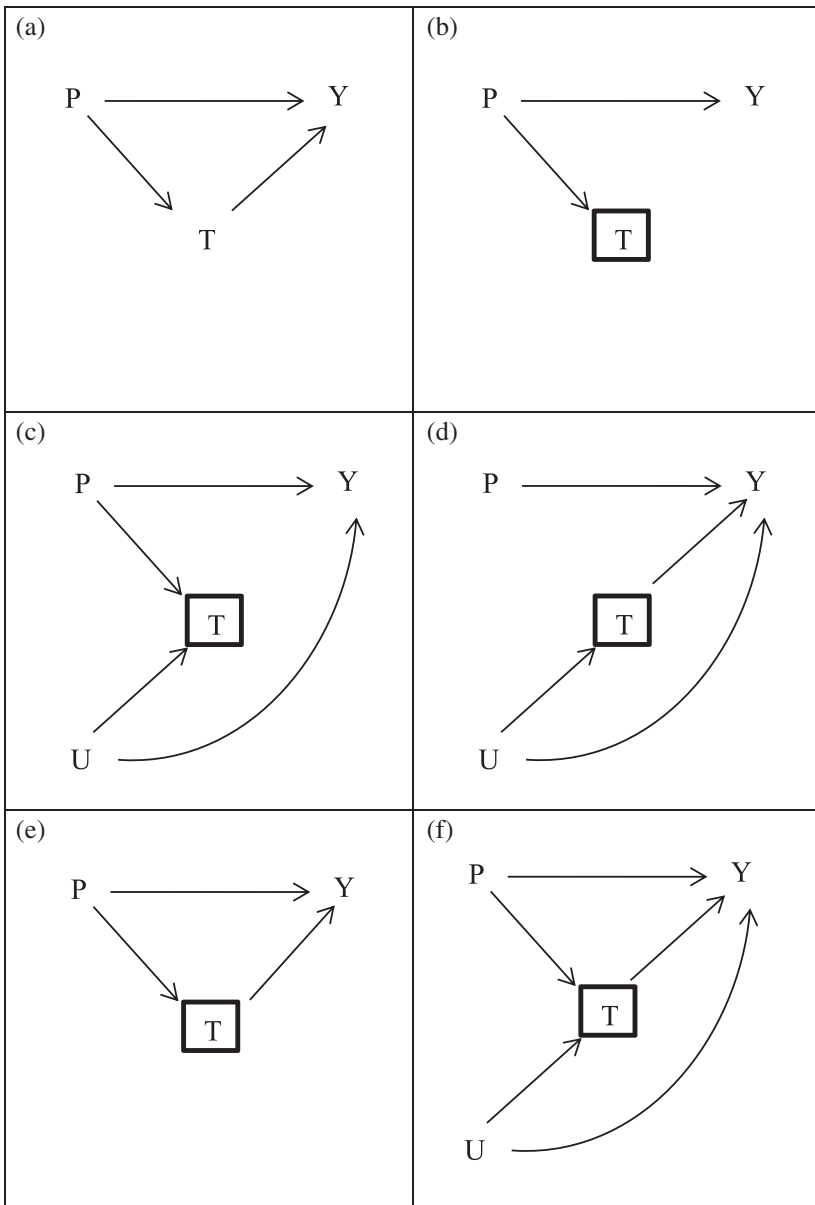
**FIGURE 1** The mechanisms underlying approaches to remove the effect of treatment drop-in on a prognostic model. For a model to provide predictions of the risk of mortality if an individual were not to receive treatment, the model parameters should be estimated such that the associations between P and Y are not influenced by the effect of T on Y. (a) The arrow from predictor to treatment (drop-in) represents a causal association. An indirect pathway from P to T to Y is present, as well as the direct path from P to Y. (b) Conditioning on treatment (represented by a black box) by restriction removes the path between T and Y. (c) In the presence of U, T becomes a collider, and restriction on T creates a backdoor path from P to Y, via T and U, inducing selection bias. (d) Censoring first conditions on T and removes the association between T and Y. The causal path from P to T is removed by inverse probability of treatment weighting. (e) Conditioning on T blocks the indirect path from P to Y via T. (f) In the presence of U, T becomes a collider, and restriction on T creates a backdoor path from P to Y, via T and U, inducing selection bias. For a marginal structural model, this may also create a backdoor path from P to U to Y (not shown). P = predictor(s); Y = outcome; T = treatment; U = unmeasured variable(s)

## 2.2 | Reweighting using inverse probability of (censoring) treatment (IPCW) weights

To solve the issue of informative censoring, weights can be applied to patients so that the association between treatment and predictors is removed (Robins & Finkelstein, 2000). Given that treatment drop-in does not occur at a fixed point in time, time-dependent weights need to be derived by, for example, dividing patient follow-up into multiple periods and fitting logistic regression models in each of the periods, regressing the censoring events on predictors of the probability of being censored (i.e., treated), in order to derive time-varying censoring weights (Austin & Stuart, 2015; Cole & Hernán, 2008; Latimer et al., 2017). Stabilized weights for patient $i$ at time $j$ take the form

$$SW_{ij} = \prod_{k=0}^{j} \frac{KM}{f\{T_{ik} \mid P_{ik}\}}, \tag{1}$$

where $T(k)$ is treatment at time point $k$, P are predictors measured at baseline, and KM represents the Kaplan–Meier survival estimates produced by a null Cox model. In the presence of time-dependent predictor variables, this may yield biased estimates of $\Pr(Y^{T=0} \mid P)$. If information on time-dependent predictors $\overline{P}$ is available, this can be substituted for P, and baseline predictor values P can be used in the numerator to stabilize the weights.

The prognostic model is then developed, after censoring, using a weighted Cox model. As Figure 1d indicates, if the underlying assumptions of the weighting procedure are met, the association between P and T should be removed, on top of the removal of the association between T and Y by censoring, allowing for correct estimation of the associations between P and Y in a treatment-naïve pseudopopulation. This might not be the case, for example, if there is a very strong indication for treatment resulting in structural non-positivity (Cole & Hernán, 2008).

## 2.3 | Explicitly modeling treatment: time-dependent Cox model

Previous findings have suggested that the explicit modeling of treatment can correct for the effect of a point treatment on predictor–outcome associations (Groenwold et al., 2016). By conditioning the prognostic model on treatment use, the pathway from P to Y via T is blocked (see Figure 1e). A crude approach to model treatment drop-in is to include an "ever/never" indicator variable in the prognostic model. To make prognostic predictions for future patients using this model, we would set their value for the drop-in indicator to zero. However, drop-in occurs by definition after baseline, and conditioning on future treatment (and, thus, future survival) introduces immortal time bias into the estimated model coefficient for treatment drop-in (Lévesque, Hanley, Kezouh, & Suissa, 2010). As T is associated with P, this bias will affect the coefficients of the prognostic model and, hence, affect the predicted probabilities. To address this bias, treatment can be included as a time-dependent covariate in the prognostic model (Fisher & Lin, 1999). If information on treatment status throughout follow-up is available, a time-dependent Cox model can be constructed using start- and stop-time notation (Therneau & Grambsch, 2013). Unlike all the previous approaches, this approach directly models all changes in treatment, such as discontinuation of treatment over time, and treated patients will again contribute untreated person-time after discontinuation. Therefore, this approach makes maximal use of the available data. However, by explicitly modeling treatment, the targeted estimand becomes $\Pr(Y \mid P, T = 0)$, which, for Cox regression, may deviate from $\Pr(Y^{T=0} \mid P)$ due to non-collapsibility of the hazard ratio (Greenland, 1996). In addition, the prognosis of patients post-treatment might not represent the untreated

prognosis one intends to capture in a prognostic model, and if treatment during follow-up is dependent on prior values of predictors (as shown in Figure 1f), in the presence of residual confounding of the association between T and Y, conditioning on treatment opens a backdoor path from P to Y (via U) and may result in model coefficients that are still biased (Hernán et al., 2004).

## 2.4 | Explicitly modeling treatment: MSM

This approach formally derives a prognostic model within a counterfactual framework. As described elsewhere (Sperrin et al., 2018), to estimate the counterfactual, untreated risk $Pr(Y^{T=0} | P)$, we can first estimate stabilized inverse probability of treatment (IPTW) weights. As with the derivation of IPCW weights (as described in Section 2.2), we can calculate time-varying IPTW weights. As treatment users are not censored in this approach, weights are calculated for the full period of follow-up for both treated and untreated patients. A weighted prognostic model is derived, which includes the relevant predictor variables, as well as a term to model time-dependent treatment and terms for any interactions between treatment and other predictors in the model that may vary over time. As with explicitly modeling treatment, this approach aims to remove the association between treatment and predictors in the model and will incorporate patient follow-up after treatment discontinuation. Importantly, as with the time-dependent Cox method, this approach also remains sensitive to unmeasured confounding of the treatment–outcome association, as shown in Figure 1f, and requires correct specification of any treatment–predictor interactions. Censoring followed by IPCW does not require any modeling of the treatment–outcome relation, as treated follow-up time is completely removed from the data set.

## 3 | APPLIED EXAMPLE

We demonstrate the methods described in Section 2 in a case study in which we developed a (hypothetical) prognostic model to predict 5-year mortality risk for patients with chronic obstructive pulmonary disease (COPD) *if they were not to use cardio-selective β-blocking agents (SBBs)*. Due to a perceived association between SBB use and COPD exacerbations (Lipworth, Wedzicha, Devereux, Vestbo, & Dransfield, 2016), estimates of patient prognosis if SBBs were to be withheld could be used by physicians to weigh the potential benefits against the harms of prescribing SBBs for individual patients. However, in practice, some COPD patients do use SBBs, and these agents are known to improve survival (Du, Sun, Ding, Lu, & Chen, 2014; Etminan, Jafari, Carleton, & FitzGerald, 2012). All other ("background") treatments were considered to be a part of routine care and should not bias the estimation of untreated risk (Pajouheshnia et al., 2017). A summary of the case study is presented in Supplemental Figure 1.

## 3.1 | Case study data

A cohort was defined using electronic health record data from the Utrecht General Practitioners Network database, with patient entry and follow-up from July 1, 1996, to December 31, 2005 (Rutten et al., 2005; Rutten, Zuithoff, Hak, Grobbee, & Hoes, 2010). Prevalent and incident COPD cases entered the cohort on July 1, 1996, and on their date of diagnosis, respectively. In line with previous definitions in this database of an adult, new-user cohort, patients younger than 45 years or using β-blocker medication (ATC code C07) within 14 days prior to cohort entry were excluded.

In total, 1,906 patients were included; median follow-up was 6.7 years, and 559 (29.3%) patients died during follow-up, of which 372 died within 5 years.

Patients prescribed SBBs during follow-up according to their individual prescription records (Rutten et al., 2010) were classified as SBB "users" (yes/no). Prescription duration (days) was calculated as the total number of tablets prescribed divided by the daily tablet quantity prescribed. When multiple prescriptions overlapped in a treatment episode, the total number of overlapping days was added to the end date of the final prescription, to better capture the possibility that patients would still have medication available. In total, 325 (17.1%) patients began using SBBs after entering the cohort, and the median time on treatment was 10.3 months. Prescriptions varied considerably from patient to patient (see Supplemental Figure 2). In total, 68 SBB users (20.1%) died during follow-up, compared to 491 nonusers (31.1%), and SBB users more often had a history of cardiovascular comorbidities than non–SBB users (see Table 1). For all other treatments, only data on use at baseline were recorded. For the sake of illustration, missing data on the number of tablets prescribed (1.2%) and smoking status (33.3%) were singly imputed using

**TABLE 1** Baseline characteristics of 1,906 patients with a diagnosis of chronic obstructive pulmonary disease stratified by selective β-blocker use during follow-up

| Characteristics | All patients $n = 1,906$ | Selective β-blocker use $n = 325$ | No selective β-blocker use $n = 1,581$ | p value of difference between groups |
|---|---|---|---|---|
| Age at study entry* | 67.8 (10.9) | 67.1 (10.4) | 68.0 (11.0) | 0.20 |
| Male sex | 1005 (52.7) | 168 (51.7) | 837 (52.9) | 0.73 |
| Smoking† | | | | 0.05 |
| Current smokers | 708 (37.1) | 104 (32.0) | 604 (38.2) | |
| Former smokers | 934 (49.0) | 179 (55.1) | 755 (47.8) | |
| Never smokers | 264 (13.9) | 42 (12.9) | 222 (14.0) | |
| Hypertension | 716 (37.6) | 228 (70.2) | 488 (30.9) | < 0.005 |
| Diabetes | 312 (16.4) | 86 (26.5) | 226 (14.3) | < 0.005 |
| *History of cardiovascular diseases* | | | | |
| Angina pectoris | 302 (15.8) | 121 (37.2) | 181 (11.4) | < 0.005 |
| Myocardial infarction | 86 (4.5) | 34 (10.5) | 52 (3.3) | < 0.005 |
| Atrial fibrillation | 183 (9.6) | 51 (15.7) | 132 (8.3) | < 0.005 |
| Heart failure | 475 (24.9) | 99 (30.4) | 376 (23.8) | 0.01 |
| Stroke | 139 (7.3) | 35 (10.7) | 104 (6.6) | 0.01 |
| Peripheral arterial disease | 134 (7.0) | 40 (12.3) | 94 (5.9) | < 0.005 |
| *Drug use (ATC code) at baseline* | | | | |
| Drugs for obstructive airway diseases (R03) | 432 (22.7) | 61 (18.8) | 371 (23.5) | 0.08 |
| Corticosteroids for systemic use (H02) | 98 (5.1) | 15 (4.6) | 83 (5.2) | 0.74 |
| Lipid modifying agents (C10) | 79 (4.1) | 12 (3.7) | 67 (4.2) | 0.77 |
| Agents acting on the renin–angiotensin system (C09) | 159 (8.3) | 38 (11.7) | 121 (7.7) | 0.02 |
| Diuretics (C03) | 236 (12.4) | 49 (15.1) | 187 (11.8) | 0.13 |
| Antithrombotic agents (B01) | 135 (7.1) | 24 (7.4) | 111 (7.0) | 0.91 |

*Note.* Absolute numbers and percentages are reported. Characteristics of users and nonusers were compared using *t* tests (age) and chi-squared tests. (*) Age in years; mean (standard deviation) reported. (†) Missing values were singly imputed (33.3%).

predictive mean matching and polytomous logistic regression (*mice* package version 2.25; Buuren & Groothuis-Oudshoorn, 2010).

## 3.2 | Derivation of a hypothetical prognostic model

First, a prognostic model that completely ignored SBB drop-in was developed using Cox regression using the full follow-up time of all patients. Candidate predictors were selected based on the literature (Esteban et al., 2008; Singanayagam, Schembri, & Chalmers, 2013), the availability of measurements, and backward selection using likelihood ratio testing. The prediction model consisted of the estimated 5-year baseline hazard and 17 predictors: age, sex, smoking status, history of comorbidities (diabetes, hypertension, angina pectoris, myocardial infarction, atrial fibrillation, heart failure, stroke, and peripheral arterial disease), and baseline drug use (drugs for obstructive airway diseases, corticosteroids for systemic use, lipid modifying agents, agents acting on the renin–angiotensin system, diuretics, and antithrombotic agents). The proportional hazards assumption was assessed by visual inspection of Schoenfeld residuals, and no major violations were identified. As treatment use is typically ignored in most prognostic model development studies (Pajouheshnia et al., 2017), the Cox model ignoring treatment—now referred to as "Model 1"—served as a reference model.

## 3.3 | Application of methods

We applied each of the approaches to account for treatment drop-in described in Section 2: excluding treated patients (Model 2), censoring treated patients at the moment of treatment drop-in (Model 3), IPCW (Model 4), modeling drop-in as a binary covariate (Model 5), modeling treatment as a time-varying covariate (Model 6), and fitting an MSM with time-varying IPTW weights (Model 7). First, prognostic models were developed following each approach, and the resulting model coefficients and estimated 5-year baseline hazards were estimated. To derive the IPCW and IPTW weights, we divided the data into seven time periods of approximately 1 year and derived time-dependent stabilized weights for each time period using logistic regression (Austin & Stuart, 2015). While shorter periods of 3 months, for example, would have more accurately modeled treatment switches, the number of treated patients per 3-month period did not permit stable estimation of propensity scores. Next, to illustrate differences between the derived models, we used each of the models to calculate 5-year predicted mortality risks for patients in the development data, and we compared these predicted risks to the predicted risks produced by Model 1. Lastly, we assessed the apparent predictive performance (Harrell's c-index, Brier score, calibration slopes) in (a) the full development data set, with patients censored at the moment of SBB treatment, and (b) SBB nonusers only, to evaluate performance in untreated patients. All analyses were performed using the R statistical program (Version 3.2.2).

In addition, comparisons were repeated using a "highly treated" subset of the data (50% of the patients began using SBBs), to see whether the results of the modeling approaches differed more with more treatment use in the development sample.

## 3.4 | Case study findings

The regression coefficients differed between all seven models (see Table 2). The largest changes in regression coefficients compared to Model 1—developed by the conventional approach of ignoring treatment—were in Models 2 (SBB users excluded), 3 (SBB users censored), 4 (IPCW),

**TABLE 2** Baseline hazards and regression coefficients of all models

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|---|---|---|---|---|---|---|---|
| 5-year baseline hazard | 0.16 (0.01) | 0.18 (0.01) | 0.15 (0.01) | 0.13 (0.01) | 0.16 (0.01) | 0.15 (0.01) | 0.13 (0.01) |
| Age (years) | 0.08 (0.00) | 0.08 (0.01) | 0.08 (0.01) | 0.08 (0.01) | 0.08 (0.00) | 0.08 (0.00) | 0.08 (0.01) |
| Female sex | −0.43 (0.09) | −0.42 (0.10) | −0.47 (0.10) | −0.44 (0.10) | −0.41 (0.09) | −0.43 (0.09) | −0.44 (0.10) |
| Former smoker | −0.63 (0.09) | −0.56 (0.10) | −0.57 (0.10) | −0.58 (0.10) | −0.61 (0.09) | −0.64 (0.09) | −0.63 (0.11) |
| Never smoker | −0.81 (0.16) | −0.72 (0.17) | −0.72 (0.17) | −0.67 (0.17) | −0.80 (0.16) | −0.81 (0.16) | −0.71 (0.18) |
| Hypertension | −0.45 (0.10) | −0.43 (0.11) | −0.52 (0.11) | −0.48 (0.10) | −0.38 (0.10) | −0.47 (0.10) | −0.47 (0.10) |
| Diabetes | −0.04 (0.11) | 0.04 (0.12) | −0.08 (0.13) | −0.02 (0.13) | 0.02 (0.11) | −0.06 (0.11) | 0.00 (0.13) |
| Angina pectoris | −0.22 (0.12) | −0.08 (0.14) | −0.27 (0.14) | −0.21 (0.13) | −0.09 (0.12) | −0.26 (0.12) | −0.22 (0.14) |
| Myocardial infarction | 0.28 (0.17) | 0.33 (0.19) | 0.24 (0.19) | 0.24 (0.22) | 0.34 (0.17) | 0.28 (0.17) | 0.29 (0.20) |
| Atrial fibrillation | 0.07 (0.12) | 0.06 (0.13) | 0.02 (0.13) | 0.04 (0.13) | 0.09 (0.12) | 0.08 (0.12) | 0.12 (0.13) |
| Heart failure | 0.61 (0.09) | 0.66 (0.10) | 0.62 (0.10) | 0.64 (0.11) | 0.63 (0.09) | 0.61 (0.09) | 0.55 (0.10) |
| Stroke | 0.19 (0.13) | 0.16 (0.15) | 0.10 (0.15) | 0.07 (0.14) | 0.22 (0.13) | 0.17 (0.13) | 0.10 (0.14) |
| Peripheral arterial disease | 0.17 (0.15) | 0.13 (0.17) | 0.00 (0.17) | 0.08 (0.18) | 0.20 (0.15) | 0.18 (0.15) | 0.19 (0.17) |
| Baseline B01 | 0.02 (0.14) | 0.07 (0.16) | 0.11 (0.15) | 0.11 (0.16) | 0.03 (0.14) | 0.02 (0.14) | −0.12 (0.19) |
| Baseline C03 | 0.32 (0.12) | 0.36 (0.13) | 0.37 (0.13) | 0.36 (0.14) | 0.33 (0.12) | 0.32 (0.12) | 0.23 (0.15) |
| Baseline C09 | 0.20 (0.16) | 0.10 (0.18) | 0.10 (0.18) | 0.04 (0.18) | 0.18 (0.16) | 0.21 (0.16) | 0.23 (0.18) |
| Baseline C10 | 0.18 (0.24) | 0.01 (0.27) | 0.15 (0.27) | 0.15 (0.27) | 0.11 (0.24) | 0.19 (0.24) | 0.22 (0.25) |
| Baseline H03 | 0.76 (0.16) | 0.92 (0.16) | 0.78 (0.17) | 0.82 (0.20) | 0.80 (0.16) | 0.76 (0.16) | 0.81 (0.21) |
| Baseline R03 | −0.06 (0.11) | 0.03 (0.11) | 0.02 (0.12) | 0.01 (0.12) | −0.05 (0.11) | −0.06 (0.11) | −0.06 (0.13) |
| Selective β-blocker use | – | – | – | – | −0.56 (0.14) | 0.39 (0.19) | 0.06 (0.23) |

*Note.* Standard errors are presented alongside regression coefficients. Model 1 = treatment ignored; Model 2 = selective β-blocker (SBB) users excluded from the cohort; Model 3 = SBB users censored at the date of treatment; Model 4 = inverse probability of treatment weighting applied after censoring of treated patients; Model 5 = SBB drop-in included as a binary covariate in the model; Model 6 = SBB drop-in included as a time-varying covariate in the model; Model 7 = marginal structural model.

and 7 (MSM). Model 6 (time-dependent SBB) hardly deviated from Model 1. The regression coefficient for (binary) treatment use was negative (−0.56), representing a highly protective effect, but likely due to immortal time bias as described in Section 2.3. The coefficient for SBB use was positive (suggesting a harmful effect) when modeled directly as a time-dependent covariate (0.39) or in an MSM (0.06), which could be the result of residual confounding or misclassification of SBB exposure.

Risk predictions made by all models in the development data ranged across the whole spectrum of probabilities (0–1). As shown in Figure 2, Models 2 and 5 produced slightly higher risk predictions than Model 1 (treatment ignored). Models 3, 4, and 6 provided slightly lower risk predictions than Model 1. The means of the predicted risks in the full development cohort were 21.3% (Model 1), 22.7% (Model 2), 20.8% (Model 3), 20.8% (Model 4), 22.5% (Model 5), 21.1% (Model 6), and 21.6% (Model 7). When evaluated in full data, after censoring SBB users, the apparent predictive performance was consistent across the models: c-statistic = 0.79, Brier score = 0.09, and calibration slope ranged from 0.99 (Model 4) to 1.02 (Models 1, 5, and 6). Similar results were seen when the models were evaluated in the data with SBB users excluded.

Repeating the model development and evaluation in a highly treated subset of the development data intensified the differences between the approaches. As Figure 3 shows, the trends of higher predictions as compared to Model 1 (mean predicted risk: 18.6%) after excluding treatment (23.8%) and modeling binary treatment (23.5%) as well as lower predictions after censoring with (15.7%) or without (15.9%) IPCW were more prominent.
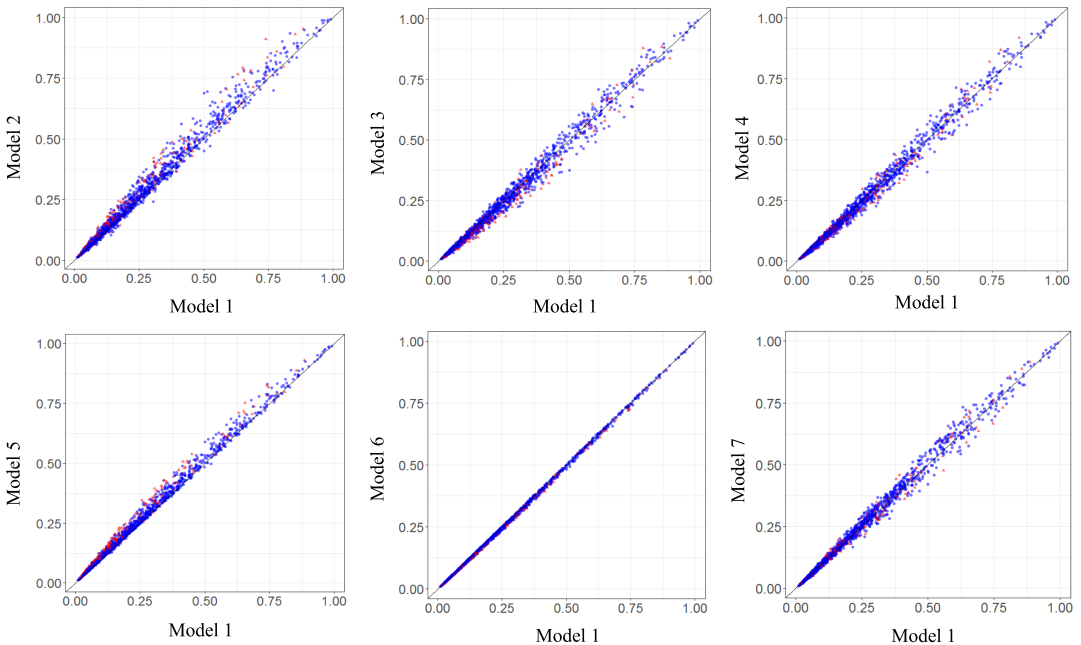


**FIGURE 2**  Five-year mortality risk predictions provided by the seven models for patients in the development data set. Predictions using Models 2–7 are plotted against predictions provided by Model 1 (treatment ignored). Blue points represent predictions for selective β-blocker (SBB) nonusers in the cohort; red points represent predictions for SBB users
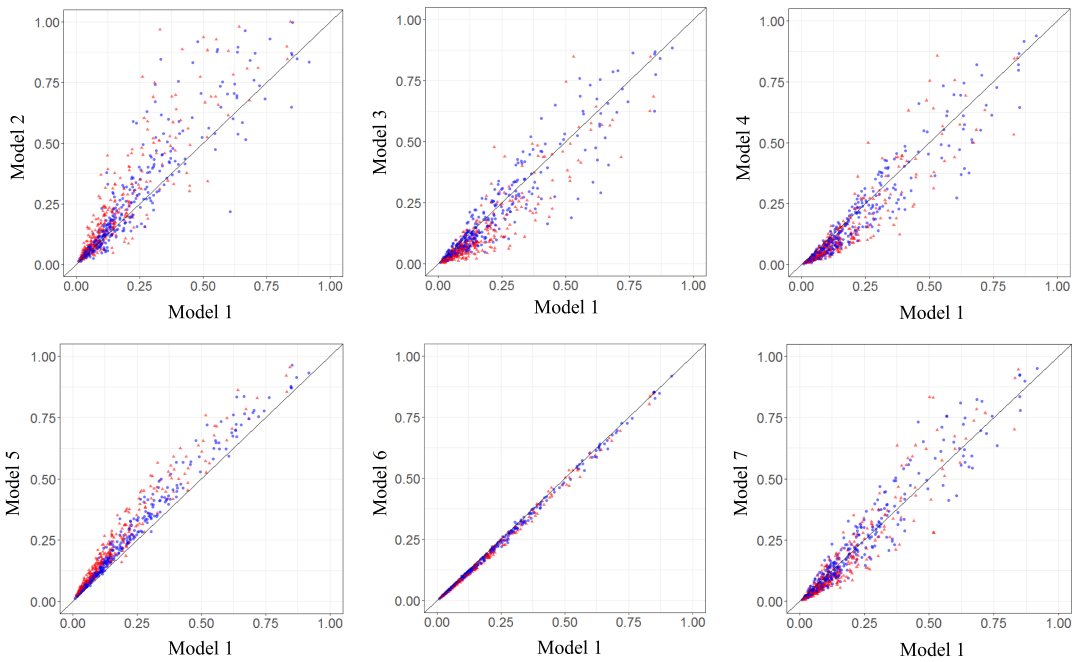
**FIGURE 3** Five-year mortality risk predictions provided by the seven models for patients in a "highly treated" subset of the development data (50% treatment drop-in). Predictions using Models 2–7 are plotted against predictions provided by Model 1 (treatment ignored). Blue points represent predictions for selective β-blocker (SBB) nonusers in the cohort; red points represent predictions for SBB users

## 4 | DISCUSSION

This study reviews seven approaches to account for treatment use when developing a prognostic model using observational data. Previous work has focused on methods to account for point treatments, that is, treatment use that does not change over time (Groenwold et al., 2016; Pajouheshnia, Peelen, Moons, Reitsma, & Groenwold, 2017). The applied case study that we present shows the use of treatments can vary greatly over time and across individuals (see Supplemental Figure 2). The seven models, derived by ignoring treatment or following one of six approaches, varied in their coefficients and estimated slightly different risks for individuals in the development data set. We expected the use of effective treatments in the development set to result in a model that underestimates untreated risks and, therefore, that approaches to correct for this would result in models that yielded higher predicted risks on average. However, in our case study, the impact was minimal.

There are a number of explanations for the limited and, where present, unexpected differences between approaches that we observed in the case study. First, it may be that the effect of treatment in the development cohort was too small to impact the performance of our prognostic model. The abundant literature supporting the protective effect of SBBs on mortality for patients with COPD (suggesting a risk ratio of approximately 0.7) suggests there was indeed a true effect of treatment in the patients in our data set who did receive treatment. However, it may be that the 17.1% drop-in rate was insufficient to affect predictions. Additionally, most patients were prescribed SBBs for less than 1 year, which may have limited the impact on their overall risk profile during the study. With the exception of modeling SBB use with an MSM or as a time-dependent

covariate, the approaches did not account for the possibility that individuals may have stopped treatment. We also assumed that treatment prescriptions directly represent treatment use, which may have overestimated the total amount of actual treatment use. Indeed, when the analysis set was sampled such that the SBB drop-in rate was 50%, the effects of removing or modeling treatment on risk predictions provided by the models became more apparent. The difference between drug prescription or dispensing and actual use is an inherent challenge when using routinely collected data to model treatment use. In addition, the limited effects of the more advanced IPCW and time-dependent covariate approaches may also be due to residual confounding. While this study has demonstrated that it is feasible to account for the time-dependent use of a pharmacological treatment in practice using prescription information, due to a lack of available information on disease severity and comorbidity *during* follow-up, we were not able to incorporate this important factor in our analyses. Unobserved confounding of predictor–treatment associations can bias IPCW and MSM estimates, as discussed elsewhere (Cole & Hernán, 2008), and, as indicated in Figure 1, may result in a failure to remove the association between predictors and treatment. In such a case, these methods would not correctly estimate $Pr(Y^{T=0} \mid P)$. This highlights the challenge of implementing more complex approaches when using observational data with limited available data.

Our structured approach to considering the potential bias caused by treatments falls in line with the ideas recently described by Sperrin et al. (2018), who recommended using MSMs to address treatment use for prognostic model development. In our review of methods, we include an alternative estimator, namely, censoring followed by IPCW. Although all of the approaches considered in this paper attempt to remove the influence of treatment on predictor–outcome associations in a prognostic model, subtle differences between them may affect their suitability. We have demonstrated that these differences can be largely explained in terms of the target estimands of the different approaches and the potential mechanisms by which they can introduce selection bias. As previously discussed (Robins, Hernan, & Brumback, 2000; Sperrin et al., 2018), an advantage of the MSM approach is that, unlike simply modeling treatment as a time-dependent covariate, it can account for the association between prior treatment use and time-dependent confounders. Furthermore, as with the time-dependent Cox approach, this method utilizes all available data by also modeling periods after treatment discontinuation. However, correct specification of the MSM to capture differences between the pre- and post-treatment prognoses of patients may be challenging, and censoring followed by IPCW could be considered a safer, more conservative approach.

While this study provides insight into how various methods to account for treatment drop-in work and highlights the assumptions they require, we do not provide empirical evidence to support one method over another. Ideally, to compare different modeling approaches empirically, the performance of models derived following each approach would need to be compared in a treatment-naïve validation data set. Such data were not available for our case study, due to the widespread use of SBBs, and thus, we could only evaluate model performance in an untreated subset of the development cohort. Future studies using real data may need to consider employing advanced methods to account for treatment use in a validation data set (Pajouheshnia et al., 2017). Alternatively, future research could evaluate model performance in a data set with simulated untreated outcomes. In addition, in this study, we selected the parameters in our prognostic model based on clinical knowledge and a backward selection procedure. It is not yet established whether shrinkage methods for model selection could influence the effect of treatment on a prediction model, or vice versa, thus requiring further investigation. However, our applied case study does demonstrate that the choice of approach affects the risk predictions made by a developed

prognostic model. Whether these approaches are necessary for a given study will depend on factors such as the strength of the treatment effect, the duration of treatment, the number of individuals treated, and the strength of any associations between predictors in the model and treatment use; all of which can be investigated in additional case studies and simulation studies.

In conclusion, the available methods that could be used to account for treatment drop-in when developing a prognostic model vary in terms of how they account for causal pathways in the predictor–treatment–outcome association and in terms of their target estimands. Censoring followed by IPCW or MSMs should be considered in future model development studies.

## ORCID

*Romin Pajouheshnia* https://orcid.org/0000-0002-4208-3583

## REFERENCES

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679.

Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 1–68.

Cheong-See, F., Allotey, J., Marlin, N., Mol, B. W., Schuit, E., Ter Riet, G., … Thangaratinam, S. (2016). Prediction models in obstetrics: Understanding the treatment paradox and potential solutions to the threat it poses. *BJOG: An International Journal of Obstetrics & Gynaecology*, *123*(7), 1060–1064.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664.

Du, Q., Sun, Y., Ding, N., Lu, L., & Chen, Y. (2014). Beta-blockers reduced the risk of mortality and exacerbation in patients with COPD: A meta-analysis of observational studies. *PLOS ONE*, *9*(11), e113048.

Esteban, C., Quintana, J. M., Aburto, M., Moraza, J., Egurrola, M., España, P. P., … Capelastegui, A. (2008). Predictors of mortality in patients with stable COPD. *Journal of General Internal Medicine*, *23*(11), 1829–1834.

Etminan, M., Jafari, S., Carleton, B., & FitzGerald, J. M. (2012). Beta-blocker use and COPD mortality: A systematic review and meta-analysis. *BMC Pulmonary Medicine*, *12*, Article No. 48.

Fisher, L. D., & Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, *20*(1), 145–157.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*, *7*, 498–501.

Groenwold, R. H. H., Moons, K. G., Pajouheshnia, R., Altman, D. G., Collins, G. S., Debray, T. P. A., … Peelen, L. M. (2016). Explicit inclusion of treatment in prognostic modelling was recommended in observational and randomised settings. *Journal of Clinical Epidemiology*, *78*, 90–100.

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, *15*(5), 615–625.

Latimer, N. R., Abrams, K. R., Lambert, P. C., Crowther, M. J., Wailoo, A. J., Morden, J. P., … Campbell, M. J. (2017). Adjusting for treatment switching in randomised controlled trials—A simulation study and a simplified two-stage method. *Statistical Methods in Medical Research*, *26*(2), 724–751.

Lévesque, L. E., Hanley, J. A., Kezouh, A., & Suissa, S. (2010). Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ*, *340*, b5087.

Liew, S. M., Doust, J., & Glasziou, P. (2011). Cardiovascular risk scores do not account for the effect of treatment: A review. *Heart*, *97*(9), 689–697.

Liew, S. M., Doust, J., & Glasziou, P. (2012). Systematic review did not consider problem of treatment effects. *BMJ: British Medical Journal*, *345*.

Lipworth, B., Wedzicha, J., Devereux, G., Vestbo, J., Dransfield, M. T. (2016). Beta-blockers in COPD: Time for reappraisal. *European Respiratory Journal*, *48*(3), 880–888.

Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., Collins, G. S. (2015). Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1–W73.

Pajouheshnia, R., Damen, J. A. A. G., Groenwold, R. H. H., Moons, K. G. M., & Peelen, L. M. (2017). Treatment use in prognostic model research: A systematic review of cardiovascular prognostic studies. *Diagnostic and Prognostic Research*, *1*(1), Article No. 15.

Pajouheshnia, R., Peelen, L. M., Moons, K. G. M., Reitsma, J. B., & Groenwold, R. H. H. (2017). Accounting for treatment use when validating a prognostic model: A simulation study. *BMC Medical Research Methodology*, *17*, Article No. 103.

Robins, J. M., & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, *56*(3), 779–788.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

Rutten, F. H., Moons, K. G. M., Cramer, M. J., Grobbee, D. E., Zuithoff, N. P., Lammers, J. W., & Hoes, A. W. (2005). Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: Cross sectional diagnostic study. *BMJ*, *331*(7529), 1379.

Rutten, F. H., Zuithoff, N. P., Hak, E., Grobbee, D. E., & Hoes, A. W. (2010). Beta-blockers may reduce mortality and risk of exacerbations in patients with chronic obstructive pulmonary disease. *Archives of Internal Medicine*, *170*(10), 880–887.

Singanayagam, A., Schembri, S., & Chalmers, J. D. (2013). Predictors of mortality in hospitalized adults with acute exacerbation of chronic obstructive pulmonary disease. A systematic review and meta-analysis. *Annals of the American Thoracic Society*, *10*(2), 81–89.

Sperrin, M., Martin, G. P., Pate, A., Van Staa, T., Peek, N., & Buchan, I. (2018). Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Statistics in Medicine*, *37*(28), 4142–4154.

Therneau, T. M., & Grambsch, P. M. (2013). *Modeling survival data: Extending the Cox model* (pp. 69–75). New York, NY: Springer Science & Business Media.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.