



## **Strong instrumental variables biased propensity scores in comparative effectiveness research: A case study in oncology**

Nicolas H Thurin, Jérémy Jové, Régis Lassalle, Magali Rouyer, Stéphanie Lamarque, Pauline Bosco-Levy, Corentin Segalas, Sebastian Schneeweiss, Patrick Blin, Cécile Droz-Perroteau

### **► To cite this version:**

Nicolas H Thurin, Jérémy Jové, Régis Lassalle, Magali Rouyer, Stéphanie Lamarque, et al.. Strong instrumental variables biased propensity scores in comparative effectiveness research: A case study in oncology. *Journal of Clinical Epidemiology*, 2023, 155, pp.31-38. <10.1016/j.jclinepi.2023.01.002>. <hal-03988528>

**HAL Id: hal-03988528**

**<https://hal.science/hal-03988528v1>**

Submitted on 14 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

## ORIGINAL ARTICLE

# Strong instrumental variables biased propensity scores in comparative effectiveness research: A case study in oncology

Nicolas H. Thurin<sup>a,\*</sup>, Jérémy Jové<sup>a</sup>, Régis Lassalle<sup>a</sup>, Magali Rouyer<sup>a</sup>, Stéphanie Lamarque<sup>a</sup>,  
Pauline Bosco-Levy<sup>a</sup>, Corentin Segalas<sup>b</sup>, Sebastian Schneeweiss<sup>c</sup>, Patrick Blin<sup>a</sup>,  
Cécile Droz-Perroteau<sup>a</sup>

<sup>a</sup>University of Bordeaux, INSERM CIC-P 1401, Bordeaux PharmacoEpi, Bordeaux, France

<sup>b</sup>University of Paris Cité, Centre of Epidemiology and Statistics (CRESS) INSERM, Paris, France

<sup>c</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Accepted 11 January 2023; Published online 16 January 2023

---

**Abstract**

**Background and Objectives:** Some medications require specific medical procedures in the weeks before their start. Such procedures may meet the definition of instrumental variables (IVs). We examined how they may influence treatment effect estimation in propensity score (PS)-adjusted comparative studies, and how to remedy.

**Study Design and Setting:** Different covariate assessment periods (CAPs) did and did not include the month preceding treatment start were used to compute PS in the French claims database (Système National des Données de Santé-SNDS), and 1:1 match patients with metastatic castration resistant prostate cancer initiating abiraterone acetate or docetaxel. The 36-month survival was assessed.

**Results:** Among 1, 213 docetaxel and 2, 442 abiraterone initiators, the PS distribution resulting from the CAP [-12; 0 months] distinctly separated populations ( $c = 0.93$ ; 273 matched pairs). The CAPs [-12; -1 months] identified 765 pairs ( $c = 0.81$ ). Strong docetaxel treatment predictors during the month before treatment start were implantable delivery systems (1% vs. 59%), which fulfilled IV conditions. The 36-month survival was not meaningfully different under the [-12; 0 months] CAP but differed by 10% points (38% vs. 28%) after excluding month -1.

**Conclusion:** In the setting of highly predictive pretreatment procedures, excluding the immediate pre-exposure time from the CAP will reduce the risk of including potential IVs in PS models and may reduce bias. © 2023 Elsevier Inc. All rights reserved.

**Keywords:** Propensity score; Instrumental variables; Matching; Bias; Prostate cancer; SNDS

---

Conflicts of interest: NH Thurin, J Jové, R Lassalle, M Rouyer, S Lamarque, P Blin, C Droz-Perroteau are researchers at Bordeaux PharmacoEpi, a research platform of the Bordeaux University and its subsidiary the ADERA, which performs financially supported studies for public and private partners in compliance with the ENCePP Code of Conduct. P Bosco-Levy is now employed by Horia, Bordeaux, France. C. Segalas has declared no conflicts of interest. Dr. Schneeweiss (ORCID# 0000-0003-2575-467X) is participating in investigator-initiated grants to the Brigham and Women's Hospital from Boehringer Ingelheim and UCB unrelated to the topic of this study. He is a consultant to Aetion Inc., a software manufacturer of which he owns equity. He is an advisor to Temedica GmbH, a patient-oriented data generation company. His interests were declared, reviewed, and approved by the Brigham and Women's Hospital in accordance with their institutional compliance policies.

**Funding:** The authors received no financial support for the research, authorship, and publication of this article. However, the example used was drawn from the TherapeutiC strategy in *Metastatic castration-Resistant prostate cancer: target population and changes between 2012 and 2014 (CAMERRA) study*, which was funded by Janssen-Cilag, France and carried

out by the Bordeaux PharmacoEpi platform under the supervision of a Scientific Committee.

**Author contributions:** NH Thurin contributed to conceptualization, investigation, methodology, validation, roles/writing—original draft; J Jové contributed to data curation, formal analysis, methodology, software, visualization, roles/writing—original draft; R Lassalle contributed to methodology, validation, writing—review and editing; M Rouyer contributed to conceptualization, investigation, project administration, writing—review and editing; S Lamarque contributed to conceptualization, investigation, project administration, writing—review and editing; P Bosco-Levy contributed to methodology, writing—review and editing; C Segalas contributed to methodology, visualization, roles/writing—original draft; S Schneeweiss contributed to methodology, writing—review and editing; P Blin contributed to conceptualization, investigation, methodology, writing—review and editing; C Droz-Perroteau contributed to project administration, supervision, writing—review and editing.

\* Corresponding author. University of Bordeaux, INSERM CIC-P 1401, Bordeaux PharmacoEpi, 146 rue Léo Saignat – Case 41, 33076 Bordeaux cedex, France, Tel.: +33-557-57-92-09; fax: +33-557-57-47-40.

E-mail address: [nicolas.thurin@u-bordeaux.fr](mailto:nicolas.thurin@u-bordeaux.fr) (N.H. Thurin).

**What is new?****Key findings**

- Inclusion into propensity score of instrument-acting variable may have limited impact on bias increase, but may affect precision of the generated point estimates, jeopardizing potential comparison.
- Ideally, propensity score index date should be set at the time treatment is decided.

**What this adds to what was known?**

- Pre-treatment procedures can result in strong instruments, especially when comparing different modalities of treatment.
- In settings of highly predictive pre-treatment procedures, excluding the immediate pre-exposure time from the propensity covariate assessment period can reduce the risk of including potential instruments in propensity score models and may reduce bias.

**What is the implication and what would change known?**

- Presence of instruments should be carefully assessed in propensity score construction, and removed whenever possible.

**1. Introduction**

Nonexperimental and pharmacoepidemiological studies relying on longitudinal health care databases are considered as a valuable complementary source of evidence to randomized controlled trial (RCT) evidence. Measurement issues aside, inferring causality from nonexperimental studies requires adjustment for confounding, i.e., the potential distortion of the true relationship between the assessed drug and the outcome of interest that may result from imbalance in outcome predictors between the treatment groups [1–3]. Propensity scores (PSs) are frequently used to minimize confounding in the analysis of health care databases as they can adjust for many proxies of the patients' health status observable in secondary data [4]. A variety of ways to implement PS analyses in nonexperimental studies exists, including matching, weighting, stratification, and modeling [5–8]. A PS is defined as the probability for a patient to start a treatment according to its baseline characteristics. Under some assumptions [5], the PS can act as a balancing score that mimics randomization in RCTs conditional on the observable factors, resulting in comparative cohorts with similar distribution of baseline characteristics.

At the time of comparing the effect of two treatments, pharmacy dispensing records and by extension their

associated claims are considered reliable markers for drug exposure [9]. Consequently, in comparative cohort studies based on claims data, dates of initial dispensing are usually used as the beginning of exposure, as well as the end of the covariate assessment period (CAP) essential to the construction of a PS [6,10]. However, it is worth remembering that in its original definition, PS refers to the “probability of assignment to a particular treatment” [5], and so, to the time when the decision to treat was made. In some situations, this decision to treat may occur well before the actual treatment onset because of a specific pretreatment check (e.g., specific laboratory tests), or drug administration-related medical procedure (e.g., implantable chamber for chemotherapy). In such situations, treatment-specific services occurring in patient journeys before the treatment start but after the decision to treat was made may be strongly associated with the exposure (i.e., treatment proxies) but poorly associated with patient characteristics. Measurements of such pretreatment services therefore have properties similar to instruments (or instrumental variables—IVs) [11].

Adjusting for preexposure patient characteristics that have properties of instruments may amplify any residual confounding [12–14], especially strong treatment predictors [15]. In practical settings, it is sometimes difficult to identify such IVs and if so then the impact on treatment effect estimates remains unclear. Although the theoretical process of variable selection in PS construction have been broadly discussed [13,16–18], few illustrative examples of results issued from designs based on an IV-biased PS have been published and contextualized. The objective of this article is to illustrate a situation of strong treatment predictors in oncology through a comparative cohort study conducted in the French Nationwide health care database (*Système National des Données de Santé*—SNDS) [19,20], to exemplify how including IVs in a PS can alter the composition of the comparator groups and impact effect estimates, and to introduce solutions to address this issue.

**2. Methods****2.1. General design**

As part of the project “TherapeutiC strAtegy in MEtastatic castration-Resistant pRostate cAnceR” (CAMERRA), a comparative cohort study was setup to assess the effectiveness of two first-line treatments used for metastatic castration-resistant prostate cancer (mCRPC) management in association with prednisone/prednisolone—abiraterone acetate and docetaxel [19,20]. Patients with mCRPC and initiating abiraterone acetate or docetaxel as first-line treatment in 2014 were identified and extracted from the SNDS with up to 5 years of history and 3 years of follow-up. Patients were then 1:1 matched based on high-dimensional PS (hdPS) and potentially other clinically pertinent variables, before

performing comparative analyses. The hdPS construction and hdPS assessment stages of CAMERRA were used throughout this article as example.

## 2.2. *Système National des Données de Santé (SNDS)*

The SNDS currently covers 98.8% of the French population. Health care resource consumption by each beneficiary can be tracked from birth (or immigration) to death (or emigration) using a unique national pseudonymized identifier [21,22]. All reimbursed resource consumptions in the community sector are documented as well as hospital discharge summaries. For each expenditure, dates, associated costs, prescriber, and caregiver information are provided. Though, neither medical indications nor results are recorded, the level of details of the captured information, enable an accurate characterization of patient health care journeys [23].

Access to SNDS data for the purposes of CAMERRA project was approved by the French national data protection agency (*Commission nationale de l'informatique et des libertés*), after positive opinion from the committee in charge of the ethical and scientific evaluation at the time of the study, the *Institut des Données de Santé*.

## 2.3. *Population*

Patients with prostate cancer aged  $\geq 40$  years covered by the general health insurance scheme were identified in 2014 based on specific prostate cancer drug dispensing or procedures, long-term registration and hospital stays for prostate cancer over a 5-year medical history period. A validated algorithm was then applied to detect castration resistance and metastasis management, and thus mCRPC status (positive predictive value = 97%) [19]. Patients with mCRPC initiating docetaxel or abiraterone acetate in 2014 were included in the cohort entering the day of treatment start (date of dispensing).

## 2.4. *Exposure*

Abiraterone acetate exposure was captured through outpatient pharmacy dispensing. Docetaxel was identified through hospital stays with “chemotherapy session for neoplasm” (International Classification of Diseases10 = Z51.1) as principal diagnosis and prostate cancer (C61) as related one, as long as any other identifiable chemotherapy was mentioned on the same date. This proxy was applied to patients with multiple cancers only when androgen deprivation therapy or surgical castration was recorded in their historic.

## 2.5. *Outcome*

Outcomes of interest were overall survival probability at 36 months following initiation of abiraterone acetate or docetaxel.

## 2.6. *Propensity score construction, assessment, and matching*

In this study, the hdPS represents the probability of being treated by abiraterone acetate vs. docetaxel as first-line treatment of mCRPC. Unlike classical PS for which the covariates are “manually” selected when they relate to the outcome even if they are unrelated to treatment assignment, the hdPS uses an automatized variable selection method to empirically rank and select covariates with the highest confounding potential to be included in the score [6,24]. This large number of considered covariates acts as a proxy for unmeasured confounders and so, may reduce residual confounding (e.g., tobacco consumption may not be captured but all diseases and cares resulting from heavy smoking are).

### 2.6.1. *High-dimensional propensity score construction*

The hdPS algorithm from the Pharmacoepidemiology Toolbox, version 2.4.15 [25] was used to examine all covariates among different fixed patient characteristics (e.g., age and gender) and the five main SNDS data dimensions: long-term disease registration, hospital discharge diagnoses, drug dispensing, laboratory test, and medical and imaging procedure. From each dimension, the top 200 most prevalent codes were transformed into binary covariates according to three levels of within-patient frequency of occurrence. These binary covariates—3 000 in total—were then individually assessed for inclusion into the hdPS using an unconditional binary logistic regression model. The algorithm prioritized each of these variables by its potential to bias the exposure-outcome relationship using Bross’s bias formula, and includes the top 100 in the final model [24]. Upon recommendation of clinicians, age at treatment initiation, disease stage before treatment initiation, Charlson comorbidity index, and total health care costs were added to the 100 empirically selected covariates. The predicted probabilities for patients to initiate abiraterone acetate rather than docetaxel were then separately derived.

Key to this investigation is that we considered two CAPs to investigate the impact of the time window running from treatment decision to treatment start—i.e., the treatment implementation period—on the resulting hdPS distribution and treatment effect estimate (Fig. 1) [1]: from 12 months before the treatment dispensing date to treatment dispensing date [-12m; 0m] [2], from 12 months before treatment dispensing date to 1 month before treatment dispensing date [-12m; -1m]. Sensitivity analyses involving other lengths of CAP were performed: [-12m; -2m], [-24m; -1m].

### 2.6.2. *High-dimensional propensity score assessment and 1:1 matching*

Quality of the resulting hdPS models was assessed based on the overlap between the hdPS distribution of the two comparative groups. Prematching C-statistic was calculated

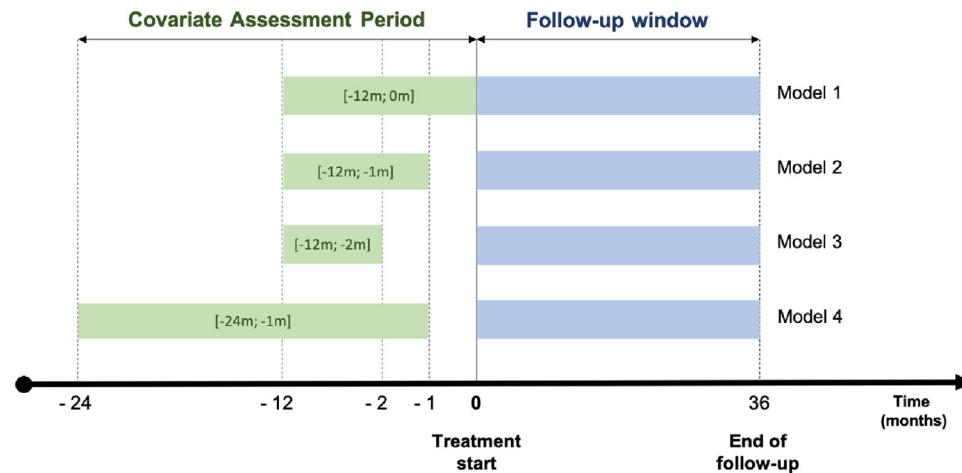


Fig. 1. Study design diagram representing the different covariate assessment periods used to compute high-dimensional propensity scores.

to inform on the ability of the model to predict treatment status (the closer to 1, the more discriminant). The distribution was trimmed, excluding values below the hdPS 2.5th percentile observed in the abiraterone acetate group and values above the hdPS 97.5th percentile observed in the docetaxel group. Treated patients from both groups were then 1:1 matched based on the  $\text{hdPS} \pm 0.01$ . The C-statistic of each hdPS model was then calculated in the matched population, i.e., postmatching. The closer to 0.5 is the C-statistic, the less discriminative is the model between treatment groups in the matched population [16]. The number of variables over the patient history with a standardized difference ( $\text{SDif}$ )  $\geq 10\%$  post matching was also calculated [26,27]. Association of these variables to the outcome of interest was assessed using a Cox proportional hazard regression model. Note that for similar post matching C-statistic, preference should be given to the settings with the lowest number of post matching  $\text{SDif} \geq 10\%$  associated to the outcome post matching, since they are the more likely to bias analyses [26,27].

Sensitivity analyses were performed forcing different variables in the hdPS, as well as trying different matching strategies using hdPS and other fixed variables based on medical expert recommendations.

### 2.7. Description of health care utilization during the month before treatment start

In order to identify which treatment-specific medical services were performed during the interval between treatment decision and treatment start, we listed the prevalence of select services and computed the prevalence differences, including drug dispensing (Anatomical Therapeutic Chemical - Classification System level 2 and 5), medical and imaging procedure (*Classification commune des actes médicaux* codes), and laboratory tests ordered (*Nomenclature des actes de biologie médicale* codes) during the

month (30 days) before treatment start and PS matching [28].

## 3. Results

### 3.1. Population

A total of 386,127 prevalent prostate cancer cases were identified in the French population covered by the general health insurance scheme in 2014, including 12,951 mCRPC. Among them, 1,213 initiated docetaxel and 2,442 abiraterone acetate, with a similar distribution of initiation according to calendar months (chi-Square test  $P$ -value = 0.52).

### 3.2. High-dimensional propensity score assessment

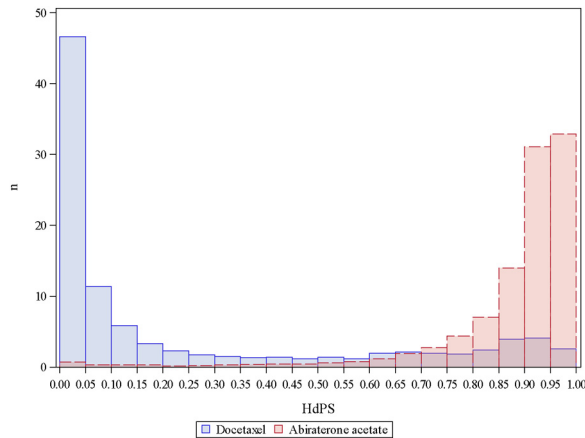
The two different CAPs—whose only difference is the ending point—were used to compute the hdPS and to execute the 1:1 matching on  $\text{hdPS} \pm 0.01$ .

The CAP [-12m; 0m] that includes the month before treatment initiation gave a poor overlap of hdPS distribution, clearly indicating two dissimilar populations (Fig. 2) with a prematching C-statistic of 0.931. Matching on hdPS led to 273 potential patient pairs with a postmatching C-statistic of 0.713.

Overlap between the two groups was clearly improved when the month before treatment initiation was excluded from the CAP [-12m; -1m]. Prematching C-statistic was 0.813. The number of potential matched-patient pairs was almost three times higher, attaining 765, and the post matching C-statistic was improved, getting to 0.614. Fifty-two variables remained with a  $\text{SDif} \geq 10\%$  post matching, including 14 variables statistically associated with the outcome. The resulting overlap after trimming and matching is plotted on Figure 3.

Using the same hdPS model and matching strategy, sensitivity analyses showed that the exclusion of the





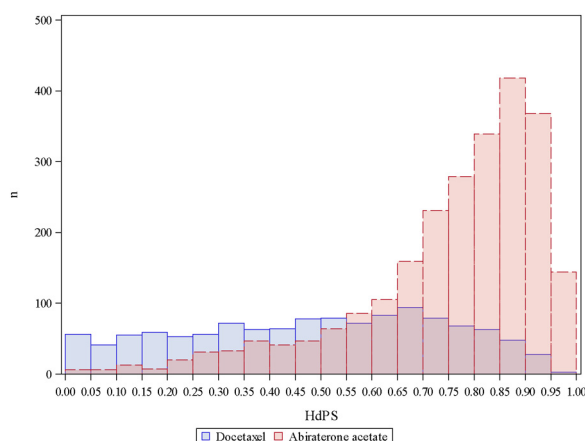
**Fig. 2.** Histogram describing the overlap between the hdPS distributions of both docetaxel and abiraterone acetate groups for the CAP [-12m; 0m] where 0 denotes the treatment dispensing date. hdPS, high-dimensional propensity score; CAP, covariate assessment period.

2 months before the index date of the CAP [-12m; -2m] led to a slightly higher number of matched patient pairs (865) and a post matching C-statistic of 0.578 with 39 SDif  $\geq 10\%$ , including 12 associated with the outcome.

The lowest post matching C-statistic (0.603) and number of variables with a SDif  $\geq 10\%$  statistically associated with the outcome [5], leading to 716 patient pairs, was reached using a hdPS model excluding total health care costs and forcing cancer management specific variables, a CAP running from 2 years to 1 month prior to index date [-24m; -1m], and matching on hdPS  $\pm 0.01$ , previous stage of prostate cancer and prostate cancer diagnosis date  $\pm 1$  year (for more details see Appendix A).

### 3.3. Survival probabilities

Overall survival probabilities corresponding to the different CAPs are provided in Table 1. Overall survival probabilities appear as significantly different between



**Fig. 3.** Histogram describing the overlap between the hdPS distributions of both docetaxel and abiraterone acetate groups for the CAP [-12m; -1m]. hdPS, high-dimensional propensity score; CAP, covariate assessment period.

abiraterone acetate and docetaxel for all CAPs except [-12m; 0m]. The full final results of this comparative study have been published elsewhere [20].

### 3.4. Description of health care utilization during the month before treatment start

Important discrepancies were observed between abiraterone acetate vs. docetaxel groups in terms of drug dispensing, performed laboratory tests and procedures during the month prior to index date (see Appendix B). More patients received drugs for functional gastrointestinal disorders (7.6% vs. 29.2%) and antiemetics (1.2% vs. 32.0%), such as aprepitant, in the docetaxel group than in the abiraterone acetate group. Similar trends were also noticed for anesthetics (2.5% vs. 32.4%), antiseptics (4.8% vs. 41.2%), systemic corticosteroids (10.9% vs. 56.7%), and sodium bicarbonate (0.4% vs. 19.4%). Much more placements of subcutaneous implantable delivery system were observed in the docetaxel group than in the abiraterone acetate group (0.7% vs. 58.5%), as well as chest X-ray (5.1% vs. 34.3%). Laboratory tests such as blood and platelets count, prothrombin time, and activated cephalin time were also more present in docetaxel group (respectively 59.5% vs. 85.6%, 9.0% vs. 31.4%, and 10.6% vs. 37.0%).

## 4. Discussion

In this comparative cohort study, two different treatments against prostate cancer were compared using a PS matching strategy. The CAP running over the year before treatment initiation led to a poor overlap of the populations to compare. The exclusion of the month prior to treatment initiation drastically improved the overlap of the PS distribution between the two groups.

The limited overlap between populations as observed on Figure 2 should not stay unnoticed when performing comparative analysis in practice, as either external validity can be seriously compromised (i.e., there are two dissimilar populations and the results will only be relevant to very reduced proportion of selected patients) [29], or the PS model is biased. As such, it is an important (nonspecific) diagnostic and screening tool. The nature of the issue can be elucidated by a close examination of the variables used to construct the score and their clinical significance in the light of the clinical recommendations and drug indications in force at the time of the study. In 2014, abiraterone acetate and enzalutamide were both indicated as first-line treatment of mCRPC. Moreover, the review of the health care utilization during the month before treatment start highlighted that many encounters were directly related with the subcutaneous implantable delivery system required for the chemotherapy administration (anesthetic and antiseptic dispensing, hemostasis panel to prepare hospital visit, imaging procedure, and anesthesia during the intervention),

**Table 1.** Overall survival probabilities at 36 months according to the different covariate assessment periods

	Model 1 <sup>a</sup>	Model 2 <sup>a</sup>	Model 3 <sup>a</sup>	Model 4 <sup>b</sup>
	CAP = [-12m; 0m]	CAP = [-12m; -1m]	CAP = [-12m; -2m]	CAP = [-24m; -1m]
■	273 patient-pairs	765 patients-pairs	865 patients-pairs	716 patients-pairs
Abiraterone acetate				
Survival probability % [95% CI]	31.5 [26.7; 37.2]	37.7 [34.5; 41.1]	37.3 [34.3; 40.5]	34.6 [31.5; 38.1]
Docetaxel				
Survival probability % [95% CI]	27.8 [23.1; 33.3]	27.9 [25.0; 31.1]	26.9 [24.3; 29.9]	27.9 [25.0; 31.2]
P-value <sup>c</sup>	0.2944	< 0.0001	< 0.0001	0.0027

CAP, covariate assessment period; CI, confidence interval; hdPS, high-dimensional propensity score; mCRPC, metastatic castration-resistant prostate cancer.

<sup>a</sup> Forced covariates = age, disease stage before mCRPC status, Charlson comorbidity index, costs; Matching variable(s) = hdPS ( $\pm 0.01$ ).

<sup>b</sup> Forced covariates = age, disease stage before mCRPC status, Charlson comorbidity index,  $\geq 1$  denosumab dispensing,  $\geq 1$  antineoplastic agent dispensing,  $\geq 1$  urethrovessical fibroscopy; Matching variable(s) = hdPS ( $\pm 0.01$ ), disease stage before mCRPC status, year of prostate cancer diagnosis ( $\pm 1$  year).

<sup>c</sup> Log-rank test.

or with the management of potential well-known chemotherapy side effects (setrons, aprepitant, metoclopramide, or domperidone to relieve nausea and vomiting, and sodium bicarbonate to treat mouth sores). Corticosteroid therapy is supposed to be associated with both docetaxel and abiraterone acetate, but future docetaxel patients took benefits from the pharmacy visit prior to treatment initiation to receive their corticoids, whereas future abiraterone acetate patients must receive their corticoids along with abiraterone acetate dispensing, which explains the observed imbalance. The covariates listed above are strongly conditioned by the treatment itself, but poorly associated with the outcome, except through the treatment. They qualify as IVs in important characteristics. The inclusion in the PS model of covariates acting as IVs prior to the index date (here, the month before treatment initiation), directly impacted the estimated propensity for a patient to initiate a treatment. It is worth noting that in the present case, the high prematching C-statistic value for [-12m; 0m] (0.931) probably reflects the presence of these variables in PS model, and in no way an indicator of its quality [16].

Survival probabilities resulting from the populations based on CAP [-12m; 0m] and [-12m; -1m] were slightly dissimilar but had overlapping confidence intervals. These results are consistent with Myers et al article, highlighting that though estimate which are conditional on perfect IV or near-IV may have larger bias and variance than the unconditional estimates, this increase of bias was generally small, if not negligible, when residual confounding was small [15]. Here, it can be assumed that whatever the CAP used, the hdPS implementation achieved a correct control of measured and unmeasured confounding, diminishing strength of the IV association with exposure, and so diminishing its impact on the final estimate. However, the fact remains that by limiting the number of patient-pairs eligible for comparison, the inclusion of instruments in the PS may sensibly alter the precision of the generated estimate, which can impact results interpretation. In the present

example, [-12m; 0m] CAP led to a nonsignificant difference between abiraterone acetate and docetaxel overall survival probabilities (31.5% [26.7; 37.2] vs. 27.8% [23.1; 33.3],  $P < 0.29$ ), whereas the same comparison appeared as significantly different with [-12m; -1m] CAP (37.7% [34.5; 41.1] vs. 27.9% [25.0; 31.1],  $P < 0.01$ ), probably by excluding instruments.

As our empirical example describes, the inclusion of strong IVs into PS may have serious repercussion in comparative effectiveness research. Since this induced bias results from the characteristics of the included variables [15], such repercussion could be observed in all approaches relying on PS, whether automated or not, in secondary or primary collected data. Therefore, it is important to identify and exclude IVs at the stage of PS construction, especially when comparing different modalities of treatment (e.g., infusion vs. oral therapy; surgery vs. nonsurgery). This can be done by assessing the correlation between each potential PS covariate, and (i) the exposure, and (ii) the outcome. Only covariate associated with both the exposure and the outcome should be included in a PS. This check may be implemented manually in classical PS construction, for instance relying on directed acyclic graphs [30,31], or automated in package such as for hdPS. A flaw of this approach is that incriminated variables will likely be excluded over the whole CAP, and not only on the period where it turns to be an IV, which may rule out potentially relevant information. For instance, in the presented example, the use of corticosteroids over the treatment implementation period appears to act as an IV, but the use of corticosteroids over the CAP is still a relevant information. An alternative solution is to exclude from the CAP the period running from treatment decision to treatment start. This can be achieved by excluding a fixed period of time consistent with the disease management process, as it was done in this example, or ideally when available, by setting the PS index date to the medical visit during which the treatment was decided. In the latter case, dates of each

treatment decision should be identified with the same precision in order to avoid introducing bias. Note that potential IV anchored prior to the theoretical treatment decision will not be affected by this alternative solution (e.g., distance to care, regional variation, and provider preference) [32,33]. Therefore, inputs from medical experts and database experts relying on clinical judgment and epidemiologic arguments are essential to assess the feasibility of the different options, as excluding time before decision to treat or specific variables may deprive the PS of key elements for the assessment of the true probability to be treated.

## 5. Conclusion

The current work illustrated that strong treatment predictors that act like IVs may be frequent in select conditions especially when comparing different treatment modalities like intravenous vs. oral oncology treatment, and their adjustment in PSs or other models may lead to biased findings. Excluding the immediate pre-exposure time from the pre-exposure CAP is one option to reduce the risk of bias. However, the suitability of this approach, and where appropriate, the period of time to be excluded, should be assessed in light of clinical context and empirical diagnostics.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2023.01.002>.

## References

- [1] Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2008:776.
- [2] Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;155:176–84.
- [3] Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* 2012;3:9–12.
- [4] Schneeweiss S, Paterno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr Rev* 2021;42:658–90.
- [5] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41–55.
- [6] Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiol* 2009;20:512–22.
- [7] Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.
- [8] Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res* 2012;21(3):273–93.
- [9] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
- [10] Blin P, Dureau-Pournin C, Cottin Y, Bénichou J, Mismetti P, Abouelfath A, et al. Comparative effectiveness and safety of standard or reduced dose dabigatran vs. Rivaroxaban in nonvalvular atrial fibrillation. *Clin Pharmacol Ther* 2019;105(6):1439–55.
- [11] Pearl J. In: *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, UK: Cambridge University Press; 2009.
- [12] Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127(8\_Part\_2):757–63.
- [13] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006;163(12):1149–56.
- [14] Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statist Med* 2007;26(4):734–53.
- [15] Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol* 2011;174:1213–22.
- [16] Westreich D, Cole SR, Funk MJ, Brookhart MA, Stürmer T. The role of the c-statistic in variable selection for propensity score models. *Pharmacoepidemiol Drug Safe* 2011;20(3):317–20.
- [17] Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol* 2011;173:1404–13.
- [18] Rassen JA, Blin P, Kloss S, Neugebauer RS, Platt RW, Pottegård A, et al. High-dimensional propensity scores for empirical covariate selection in secondary database studies: planning, implementation, and reporting. *Pharmacoepidemiol Drug* 2023;32:93–106.
- [19] Thurin NH, Rouyer M, Gross-Goupil M, Rebillard X, Soulié M, Haaser T, et al. Epidemiology of metastatic castration-resistant prostate cancer: a first estimate of incidence and prevalence using the French nationwide healthcare database. *Cancer Epidemiol* 2020;69:101833.
- [20] Thurin N, Rouyer M, Jové J, Gross-Goupil M, Haaser T, Rébillard X, et al. Abiraterone acetate versus docetaxel for metastatic castration-resistant prostate cancer: a cohort study within the French Nationwide Claims Database. *Expert Rev Clin Pharmacol* 2022;15:1139–45.
- [21] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf [Internet]* 2017;26(8):954–62.
- [22] Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: from the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Revue d'Épidémiologie et de Santé Publique* 2017;65:S149–67.
- [23] Thurin NH, Bosco-Levy P, Blin P, Rouyer M, Jové J, Lamarque S, et al. Intra-database validation of case-identifying algorithms using reconstituted electronic health records from healthcare claims data. *BMC Med Res Methodol* 2021;21(1):95.
- [24] Bross IDJ. Spurious effects from an extraneous variable. *J Chronic Dis* 1966;19(6):637–47.
- [25] Software Tools, Pharmacoepidemiology Toolbox including High-dimensional Propensity Score (hd-PS) Adjustment version 2. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School; 2019. Available at <https://www.drugapi.org/dope/software>. Accessed February 7, 2023.
- [26] Schulte PJ, Mascha EJ. Propensity score methods: theory and practice for anesthesia research. *Anesth Analgesia* 2018;127(4):1074–84.



- [27] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statist Med* 2009;28(25):3083–107.
- [28] Harder VS, Stuart EA, Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Methods* 2010;15(3):234–49.
- [29] Hernán M, Robins J. Outcome regression and propensity scores. In: *Causal Inference: What If*. Boca Raton: Chapman&Hall/CRC; 2020.
- [30] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
- [31] Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021;50(2):620–32.
- [32] Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med* 2014;161:131–8.
- [33] Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011;64(6):687–700.