PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2016; **25**(Suppl. 1): 114–121
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3864

ORIGINAL REPORT

Methodological comparison of marginal structural model, time-varying Cox regression, and propensity score methods: the example of antidepressant use and the risk of hip fracture†

M. Sanni Ali[1,2], Rolf H. H. Groenwold[1,2], Svetlana V. Belitser[1], Patrick C. Souverein[1], Elisa Martín[3], Nicolle M. Gatto[4,5], Consuelo Huerta[3], Helga Gardarsdottir[1,6], Kit C. B. Roes[2], Arno W. Hoes[2], Antonius de Boer[1] and Olaf H. Klungel[1,2]*

[1]*Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands*
[2]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands*
[3]*BIFAP Research Unit, Division of Pharmacoepidemiology and Pharmacovigilance, Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain*
[4]*Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA*
[5]*Epidemiology, Worldwide Safety and Regulatory, Pfizer Inc., New York, NY, USA*
[6]*Department of Clinical Pharmacy, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, the Netherlands*

ABSTRACT

Background Observational studies including time-varying treatments are prone to confounding. We compared time-varying Cox regression analysis, propensity score (PS) methods, and marginal structural models (MSMs) in a study of antidepressant [selective serotonin reuptake inhibitors (SSRIs)] use and the risk of hip fracture.

Methods A cohort of patients with a first prescription for antidepressants (SSRI or tricyclic antidepressants) was extracted from the Dutch Mondriaan and Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP) general practice databases for the period 2001–2009. The net (total) effect of SSRI versus no SSRI on the risk of hip fracture was estimated using time-varying Cox regression, stratification and covariate adjustment using the PS, and MSM. In MSM, censoring was accounted for by inverse probability of censoring weights.

Results The crude hazard ratio (HR) of SSRI use versus no SSRI use on hip fracture was 1.75 (95%CI: 1.12, 2.72) in Mondriaan and 2.09 (1.89, 2.32) in BIFAP. After confounding adjustment using time-varying Cox regression, stratification, and covariate adjustment using the PS, HRs increased in Mondriaan [2.59 (1.63, 4.12), 2.64 (1.63, 4.25), and 2.82 (1.63, 4.25), respectively] and decreased in BIFAP [1.56 (1.40, 1.73), 1.54 (1.39, 1.71), and 1.61 (1.45, 1.78), respectively]. MSMs with stabilized weights yielded HR 2.15 (1.30, 3.55) in Mondriaan and 1.63 (1.28, 2.07) in BIFAP when accounting for censoring and 2.13 (1.32, 3.45) in Mondriaan and 1.66 (1.30, 2.12) in BIFAP without accounting for censoring.

Conclusions In this empirical study, differences between the different methods to control for time-dependent confounding were small. The observed differences in treatment effect estimates between the databases are likely attributable to different confounding information in the datasets, illustrating that adequate information on (time-varying) confounding is crucial to prevent bias.

KEY WORDS—bias; collider stratification; confounding; Cox model; inverse probability of treatment weighting; time-dependent propensity score; time-varying treatment; pharmacoepidemiology

Received 31 October 2014; Revised 31 July 2015; Accepted 31 July 2015

*Correspondence to: O. H. Klungel, University of Utrecht, Utrecht Institute for Pharmaceutical Sciences, Division of Pharmacoepidemiology and Clinical Pharmacology,PO Box 80082, 3508 TB Utrecht, the Netherlands. E-mail: O.H.Klungel@uu.nl

†The abstract has been presented in the 30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE), 24–27 October 2014, Taipei, and published in *Pharmacoepidemiology and Drug Safety*, 2014; 23: (Issue S1): 1–519 as "Antidepressant Use and the Risk of Hip Fracture: A Comparison of Marginal Structural Models and Conventional Regression Methods."

On behalf of Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) WP2 (framework for pharmacoepidemiology studies). The PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium) is a private–public partnership coordinated by the European Medicines Agency.

INTRODUCTION

Antidepressants, notably selective serotonin reuptake inhibitors (SSRIs), have been associated with increased risk of femur or hip fracture.[1,2] In observational studies, patients' exposure to SSRI medication may change over time; that is, physicians may stop SSRI therapy or switch to other classes of antidepressants because of adverse effects such as sexual dysfunction and drowsiness. Patients may not adhere to the prescribed drug regimen.[3] Furthermore, the severity of the depression,

co-medication use (e.g., benzodiazepine), and the presence of other co-morbidities (e.g., anxiety) might change over time and may influence the use of antidepressants at later time points, hence, confounding the observed relationship between antidepressant use and hip fracture. On the other hand, antidepressant medication treatment may improve subsequent depression severity or co-medications such as benzodiazepine use. As a result, unbiased estimation of the (net) effect of SSRI use over time on the risk of hip fracture requires that the time-varying nature of both SSRI treatment and confounders (in particular, the potential for some confounders to simultaneously act as a mediator and confounder of the treatment effect, e.g., severity of depression and benzodiazepine use) is accounted for.[4,5] Here, the effect of interest is the net (total) effect of SSRI use (defined as SSRI use during the current and previous periods) on hip fracture.

Cox proportional hazard models with time-varying coefficients are capable of addressing the time-varying nature of both treatment and covariates.[6,7] However, when potential confounders are themselves affected by the previous SSRI use, the time-varying Cox model can no longer provide unbiased estimates of the treatment effect.[8,9] This is because the time-dependent confounding factors are also intermediates on the causal path from treatment to outcome, and conditioning on such factors will artificially dilute (or 'adjust-away' part of) the treatment, SSRI use, effect.[8–11] In addition, when such time-dependent confounders are common effects (i.e., colliders) of previous treatment and unmeasured factors that are also predictors of outcome (in this example, the risk of hip fracture), time-varying Cox models[8,9] and propensity score (PS) methods that condition on time-dependent confounders (colliders) induce a non-causal association between previous treatment and unmeasured risk factors, thereby introducing collider-stratification bias.[8,9,12] The potential associations between treatment (SSRI use), covariates, outcome (hip fracture), and the impact of adjustment methods are depicted using causal diagrams in the Supporting Information (Appendix 1).

In inverse probability of treatment weighting (IPTW) estimation of marginal structural models (MSMs), a weight is assigned to each observation that is proportional to the inverse of the probability of treatment received given time-dependent confounders and previous treatment. This method provides unbiased estimates of causal treatment effects under the following assumptions: (i) exchangeability, that is, no unmeasured confounding or informative censoring; (ii) positivity; that is, both treated and untreated subjects exist at each level of confounders; (iii) correct model specification; (iv) the intervention is well defined; that is, there are no unrepresented versions of treatment; and (v) no measurement error in exposure, covariates, and outcome.[8,9,13] Moreover, inverse probability of censoring weighting in MSMs enables investigators to control for bias because of informative censoring, that is, non-random or systematic loss to follow-up.[8,9]

Often ignored but important in comparing the different methods is the key difference between the methods under consideration: The parameters are typically not equal, even in the absence of (time-dependent) confounding. The MSM addresses a marginal hazard ratio (HR) while the Cox model and other PS methods estimate a conditional HR.[8–12] The conditional estimates from conventional Cox model and PS models are often different in the absence of a null treatment effect because of non-collapsibility. The estimate is conditional on all the covariates included in the outcome model and the summary of the covariates, the PS, in the conventional Cox and the PS methods, respectively.

Our primary objective was to assess the sensitivity of the estimated net (total) effect of SSRI use on the risk of hip fracture to different approaches of controlling time-dependent confounding, including time-varying Cox model, PS methods, and MSMs using inverse probability of treatment and/or censoring weights. The second objective was to assess whether this sensitivity differed between two observational databases, with varying information on covariates.

METHODS

Data sources and study population

The Mondriaan databases include the Netherlands Primary Care Research Database and the Almere Health Care database.[14] The Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP) database, a computerized database of medical records of primary care, is a non-profit research department operated by the Spanish Medicines Agency (Agencia Española de Medicamentos y Productos Sanitarios).[14] The BIFAP database includes clinical and prescription data from around 3.1 million patients covering around 6.8% of the Spanish population. In both databases, the International Classification in Primary Care (ICPC) is used for coding diagnoses, and the Anatomical Therapeutic Chemical classification system for coding drugs. Further details of the Mondriaan and BIFAP databases can be found elsewhere.[14] The study population included a cohort of patients with a first prescription for antidepressants [SSRI or tricyclic antidepressants (TCAs)] extracted from both the Dutch Mondriaan and Spanish BIFAP general practice (GP) databases for the period from 1 January 2001 to 31 December 2009.

DOI: 10.1002/pds

Exposure, outcome, and potential confounders

Only patients who received at least one prescription of antidepressants, either SSRIs or TCAs, were included in this study. For each patient, all prescriptions for SSRI or TCA were identified, and treatment episodes were constructed. A treatment episode was defined as a series of subsequent prescriptions, irrespective of changes in dosage regimen or switching between antidepressants (SSRI or TCA). The theoretical duration of each prescription was estimated based on the number of tablets prescribed and the prescribed dosage regimen (BIFAP). In Mondriaan, prescription length was set at 90 days as information on the dosage regimen was not available. The choice of the 90-day prescription length was based on the maximum allowed duration of an antidepressant prescription issued by GPs in the Netherlands.[12,15] Patients were considered to have discontinued therapy if 30 days or more elapsed between the theoretical end date of an SSRI prescription and the subsequent SSRI prescription. In the original cohort, the study was designed in such a way that both SSRI and TCA were exposure variables; however, in the current study, the exposure of interest was only SSRI. Baseline TCA use, apart from inclusion of patients into the cohort, was considered as confounding variable and was adjusted for in each analysis.

Exposure to SSRI was further divided into episodes of current, recent, and past uses. Current use was considered as the calculated treatment episode plus 30 days after the estimated theoretical end date of the last prescription, to account for carry-over effects. Recent use included the period between 1 and 60 days after the period of current use. Past use included the period following recent use until a new SSRI prescription was filled or until the end of follow-up. In the cohort for this study, episodes of recent and past uses were considered a reference group, "no SSRI use," and episodes of current SSRI use were considered as "SSRI use." Each patient was followed from the first prescription until the occurrence of the first hip fracture or loss to follow-up (because of unregistration with the GP or death) or until the end of data collection (31 December 2009), whichever date came first. Hip fractures were identified by ICPC-2 codes and specific string texts in BIFAP and by ICPC-2 codes in Mondriaan.[12] Hip fractures were manually reviewed in BIFAP but not in Mondriaan. In the time-varying analysis, exposures only in the current and previous periods were considered relevant.

Potential confounders (co-medications and co-morbidities including baseline TCA use) were measured at baseline and updated whenever patients switched between exposure episodes (Table 1 footnote). When a patient was in the same exposure episode (current or past use), confounding factors were updated every 6 months (i.e., 182 days). The status of co-medication use as a confounding variable for the current period was defined as the use in the prior 182 days. Only chronic diseases such as diabetes mellitus and chronic obstructive pulmonary disease were considered as co-morbidities. If a patient is diagnosed with chronic condition (e.g., diabetes mellitus), he or she is considered as having the co-morbidity from the first date of diagnosis for the rest of the time to the end of the study period. More details on study design, exposure, confounding, and outcome are available online at the ENCePP e-register of studies.[16]

Statistical analysis

First, time-fixed analysis accounting only for baseline covariates was conducted using conventional Cox proportional hazard models to give a benchmark for the degree of confounding by baseline variables included in the model. This analysis is based on the key assumption of the Cox model: the issue of *non-informative censoring*, meaning the mechanisms giving rise to censoring of individual subjects during follow-up are not related to the probability of an event, hip fracture, occurring. The estimates from adjusted models are "ever" versus "never" use ("intention-to-treat analysis") of SSRI during follow-up conditional on baseline covariates included in the models.

Next, three methods were applied to estimate the risk of hip fracture associated with the use of SSRI: (i) time-varying Cox regression models; (ii) PS methods; and (iii) MSMs. Time-varying Cox models were applied with and without adjustment for the demographic and clinical variables listed in Table 1. Two PS techniques were applied: covariate adjustment using the PS (i.e., including the PS as independent variable in the regression model) and PS stratification. The PS was estimated using ordinary logistic regression including the demographic and clinical variables listed in Table 1. The PS was defined as the probability of exposure to SSRI in a specific period, conditional on measured covariates in the previous period. The periods were based on exposure episodes as described earlier. Hence, for each patient, the PS could change over time (i.e., it was considered time varying). We considered a PS model in which all measured covariates were included as main terms without any interaction or higher order terms. PS stratification was based on quintiles as well as deciles of the PS. The interaction between SSRI use and PS strata was tested in order to compare differences in treatment effect between strata. All PS methods used a Cox model as the model for

DOI: 10.1002/pds

Table 1. Baseline characteristics of patients stratified by selective serotonin reuptake inhibitor treatment status and cohort

	Mondriaan cohort $N = 22\,903$ (169 948 person times)		BIFAP cohort $N = 251\,884$ (2 332 487 person times*)	
	SSRI users	Non-SSRI users	SSRI users	Non-SSRI users
Number of person moments[†]	46 175	123 773	745 922	1 586 565
Number of cases (Hip fracture), *N* (%)	35 (%)	47 (%)	756 (%)	763 (%)
Age (years), mean ± SD	47.1 ± 15.8	50.2 ± 16.2	54.1 ± 17.3	51.8 ± 16.8
Range	18–105	18–105	18–106	18–107
Men, *N* (%)	16 009 (34.7)	44 850 (36.2)	181 875 (24.4)	408 568 (25.8)
TCA use, %	0.09	12.5	0.07	7.7
Benzodiazepine use, %	34.1	24.9	59	43.8
Bone-related medications*, %	20.2	22.6	18.7	15.6
Anti-inflammatory medications[§] %	8.5	11.7	20.9	21.3
Gastrointestinal medications[††], %	18.3	20.2	33.4	31.0
Cardiovascular morbidities[¶], %	25.7	29.9	38.1	32.3
Neurological co-morbidities[‖], %	24.1	25.9	49.8	45.7
Respiratory co-morbidities**, %	10.89	12.33	10.8	12.3
Previous history of fractures[‡‡], %	9.8	9.0	8.5	8.2

BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; SSRI, selective serotonin reuptake inhibitor; SD, standard deviation; TCA, tricyclic antidepressant.

*Person times refer to the number of data points (records) a patient may contribute to several periods of SSRI use and no SSRI use.

[†]Person moments refer to the number of observation times contributed by the patients in the cohort (a single patient may contribute several person times).

[‡]Bone-related medications: previous use of bisphosphonate or any of the other bone-protecting drugs: raloxifene, strontium ranelate, parathyroid hormone, calcium and vitamin D, calcitonin, calcitriol, thyroid hormones, and antithyroid drugs.

[§]Anti-inflammatory medications: inhaled glucocorticoids, non-steroidal anti-inflammatory drug, and disease-modifying anti-rheumatic drug.

[¶]Cardiovascular morbidities and medications: antihypertensive drugs (including angiotensin-converting enzyme inhibitors, angiotensin II antagonists, beta blocking agents, calcium channel blockers, and other antihypertensive), diuretics, anti-arrhythmics, statins, and ischemic heart disease.

[‖]Neurological co-morbidities and medications: mental disorders and dementia and/or Alzheimer, seizures, syncope, cerebrovascular disease, malignant neoplasms, medications such as anti-Parkinson drugs, antipsychotics/lithium, anticonvulsants, and sedating antihistamines.

**Respiratory co-morbidities and medications: chronic obstructive pulmonary disease, bronchodilators (including beta-2-adrenoceptors agonist and anticholinergics).

[††]Gastrointestinal-related medications and morbidities: proton pump inhibitors, antiemetic (metoclopramide), inflammatory bowel disease and liver disease.

[‡‡]Previous history of fractures and history of other bone diseases (Paget's disease and osteogenesis imperfect).

treatment effect estimation. These analyses assume that censoring is non-informative within each period. The estimates from time-varying Cox and PS models are the net (total) effect of SSRI use (defined as SSRI use during the current and previous periods) on hip fracture conditional on the time-varying covariates and PS in the absence of informative censoring at each time point during follow-up. However, this net (total) effect of SSRI does not include the effect of SSRI use in the previous period that is mediated through subsequent time-varying covariates included in the Cox or PS model. This effect is "adjusted away" by conditioning on time-varying covariates or PS.

Propensity scores were also used to construct IPTW for the MSM. In the IPTW approach, the estimated PS was used to assign weights to all observations resulting in an altered composition of the study population, also referred to as a "pseudo-population." The weight for each patient was the inverse of the probability that the patient had the treatment that she or he actually received given a set of time-fixed and time-dependent covariates as well as previous treatment. Hence, the weights for SSRI users and non-SSRI users were 1/PS and 1/(1-PS), respectively. Then, cumulative weights were calculated for each patient by multiplying the weights up to that observation period. Finally, a Cox model was fitted using SSRI use as the only covariate in the pseudo-population created using the cumulative weights. To assess the possible impact of informative censoring, MSM with and without censoring weights, which are also cumulative over observation periods, was applied. In addition, stabilized weights were estimated by replacing the numerator of IPTW by the probability of SSRI use conditional on previous SSRI use or previous SSRI use, age, and gender. Stabilized treatment weights (STW_i) and censoring weights (SCW_i) were calculated using the method described by Hernan *et al.*[7] For detail on constructing the inverse probability of treatment and censoring weights, we refer to the Supporting Information (Appendix 2). The estimates from MSMs are the net (total) effect of SSRI use (defined as SSRI use during the current and previous periods) on hip fracture including the effect of SSRI mediated through subsequent time-varying covariates included in denominator of the

DOI: 10.1002/pds

weights. The estimates from MSMs are often marginal but can also be conditional (when age and gender are included in the outcome model), depending on the adjustment set, yet they are still marginal across other covariates. MSMs with censoring weights account for informative censoring related to covariate sets included in the denominator of the weights. Additional analyses adjusting for baseline TCA use, benzodiazepine use, and other covariates were also conducted. In MSMs, 95% confidence intervals were estimated using bootstrapping (number of bootstraps = 10 000). Furthermore, analysis using truncated (treatment and censoring) weights at 0.5th and 99.5th percentiles was performed. All analyses were performed in R, version 2.15.2, and correlation between observations was taken into account in both PS and Cox analyses using the cluster function in R.[17] In all analyses, we assumed exchangeability. The intervention is well defined; that is, there are no unrepresented versions of treatment, positivity, correct model specification, and no measurement error of exposures, covariates, or outcomes. For further details on these assumptions, we refer to the literature.[8,9,11,13,18]

RESULTS

There were 22 903 new antidepressant (SSRI and TCA) users included in the Mondriaan cohort and 251 884 in the BIFAP cohort (Table 1). The mean ages were 49.4 years (±16.27) in Mondriaan and 52.6 years (±17.0) in BIFAP. The proportion of patients using SSRIs was 65.5% in the Mondriaan and 84.8% in the BIFAP cohort. The baseline characteristics of the two cohorts are shown in Table 1.

Table 2 shows the crude and adjusted HRs for hip fracture associated with SSRI use from time-fixed Cox models. The crude HRs were 0.60 (95%CI: 0.37, 0.99) in Mondriaan and 1.17 (1.01, 1.35) in BIFAP. After adjustment for gender and age using the time-fixed Cox model ("intention-to-treat analysis"), the HR increased to 1.11 (0.67, 1.85) in Mondriaan and 1.23 (1.06, 1.42) in BIFAP.

Table 3 shows the crude and adjusted HRs for hip fracture associated with SSRI use from time-varying Cox models. The crude HRs were 1.75 (95%CI: 1.12, 2.72) in Mondriaan and 2.09 (1.89, 2.32) in BIFAP. After adjustment for gender and age using the time-varying Cox model, the HR increased to 2.36 (1.52, 3.69) in Mondriaan but decreased to 1.51 (1.37, 1.68) in BIFAP. Additional adjustment for baseline TCA use only marginally changed the risk of hip fracture associated with SSRI use: HR 2.59 (1.63, 4.12) in Mondriaan and 1.56 (1.40, 1.73) in BIFAP. In both cohorts, further adjustment for other covariates did not materially alter the HR: When fully adjusted for all covariates in Table 1, the HRs were 2.62 (1.63, 4.19) and 1.52 (1.37, 1.69) in Mondriaan and BIFAP, respectively. As we previously reported,[19] there seems to be an indication for interaction between SSRI use and age in Mondriaan [*p* value for the interaction term was 0.11, and effect of SSRI use taking into account interaction was 1.22 (0.42, 3.58)] but not in BIFAP [*p* value for the interaction term was 0.77, and effect of SSRI use taking into account interaction was 1.74 (1.56, 1.94)]. Again, the estimates from adjusted models are conditional on the covariates included in the models (Table 3).

Table 4 shows the results of time-dependent PS-based Cox analyses. Using PS covariate adjustment, the adjusted HR for hip fracture of SSRI use versus no SSRI use was 2.82 (95%CI: 1.63, 4.25) and 1.61 (1.45, 1.78) in Mondriaan and BIFAP, respectively. When quintile and decile stratifications on the PS were used, the HRs were 2.64 (1.63, 4.25) and 2.72 (1.63, 4.54) in Mondriaan and 1.54 (1.39, 1.71) and 1.53 (1.38, 1.70) in BIFAP, respectively. This analysis did

Table 2. Associations between selective serotonin reuptake inhibitor use and the risk of hip fracture using conventional time-fixed Cox models

	Mondriaan		BIFAP	
Adjusted for	HR	95%CI	HR	95%CI
None (crude)	0.60	0.37, 0.99	1.17	1.01, 1.35
Gender	0.61	0.38, 1.00	1.17	1.02, 1.35
Gender + Age	1.06	0.65, 1.74	1.23	1.06, 1.41
Gender + Age + TCA	1.06	0.65, 1.74	1.23	1.06, 1.41
Gender + Age + TCA + Benzo.	1.10	0.67, 1.80	1.23	1.06, 1.41
All confounders*	1.11	0.67, 1.85	1.23	1.06, 1.42

BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; HR, hazard ratio; TCA, tricyclic antidepressant.
*Age, gender, TCA use, benzodiazepine use (Benzo), bone-related medications, anti-inflammatory medications, cardiovascular co-morbidities, neurological co-morbidities, respiratory co-morbidities, previous history of fractures, and gastrointestinal medications as listed in the Table 1 footnote.

Table 3. Associations between selective serotonin reuptake inhibitor use and the risk of hip fracture using time-varying Cox models

	Mondriaan		BIFAP	
Adjusted for	HR	95%CI	HR	95%CI
None (crude)	1.75	1.12, 2.72	2.09	1.89, 2.32
Gender	1.73	1.10, 2.69	2.07	1.87, 2.30
Gender + Age	2.36	1.51, 3.68	1.51	1.37, 1.68
Gender + Age + TCA_t	2.59	1.63, 4.12	1.56	1.40, 1.73
Gender + Age + TCA_t + $Benzo_t$	2.60	1.63, 4.16	1.54	1.38, 1.71
All confounders*	2.62	1.63, 4.19	1.52	1.37, 1.69

BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; HR, hazard ratio; TCA, tricyclic antidepressant.
*Age, gender, TCA_t use, benzodiazepine use ($benzo_t$), bone-related medications, anti-inflammatory medications, cardiovascular co-morbidities, neurological co-morbidities, respiratory co-morbidities, previous history of fractures, and gastrointestinal medications as listed in the Table 1 footnote.

DOI: 10.1002/pds

Table 4. Associations between selective serotonin reuptake inhibitor use and the risk of hip fracture using propensity score-based Cox analyses

		Mondriaan		BIFAP	
Adjusted for		HR	95%CI	HR	95%CI
PS adjustment		2.82	1.63, 4.25	1.61	1.45, 1.78
PS stratification	Quintiles	2.64	1.63, 4.25	1.54	1.39, 1.71
	Deciles	2.72	1.63, 4.54	1.53	1.38, 1.70

BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; HR, hazard ratio; PS, propensity score.

not account for censoring because it was only minimally informative given the results in Tables 5 and 6.

Table 5 shows the results from MSMs with and without accounting for potential informative censoring. The mean (range) of stabilized treatment weights was 0.97 (0.02–218) in Mondriaan and 0.96 (0.06–110) in BIFAP. Similarly, the mean (range) of stabilized weights for censoring was 0.99 (0.05–111) in Mondriaan and 0.98 (0.21–8.24) in BIFAP. Estimates from MSMs using combined treatment and censoring weights were similar to those using only treatment weights [HR: 2.15 (95% CI: 1.30, 3.55) versus 2.13 (95%CI: 1.32, 3.45) in Mondriaan and 1.63 (95%CI: 1.28, 2.07) versus 1.66 (95%CI: 1.30, 2.12) in BIFAP, respectively].

When stabilized weights were trimmed at 0.5th and 99.5th percentiles, the treatment effect estimates in Mondriaan changed, particularly when adjustment was made for age and gender. On the other hand, weight truncation in BIFAP resulted in improved precision of the effect estimate without a substantial change in the point estimates even after adjustment was made for additional confounders (Table 6). The range of the weights after truncation was narrower: 0 to 80 in Mondriaan and 0.33 to 1.65 in BIFAP. When weights were trimmed at 2.5th and 97.5th percentiles or at 1st and 99th percentiles, results were similar in both Mondriaan and BIFAP.

Table 5. Association between selective serotonin reuptake inhibitor use and the risk of hip fracture using inverse probability of treatment weighting estimation of marginal structural models

		Mondriaan		BIFAP	
	Adjusted for	HR	95%CI	HR	95%CI
Not accounting for informative censoring*	None (crude)	1.60	1.00, 2.55	2.32	1.82, 2.96
	Gender	1.60	1.00, 2.54	2.31	1.81, 2.94
	Gender + age	2.13	1.32, 3.45	1.66	1.30, 2.12
Accounting for informative censoring†	None (crude)	1.65	1.02, 2.67	2.32	1.83, 2.95
	Gender	1.65	1.02, 2.67	2.31	1.81, 2.93
	Gender + age	2.15	1.30, 3.55	1.63	1.28, 2.07

Both treatment and censoring weights were stabilized.
BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; HR, hazard ratio.
*Only inverse probability of treatment weights was used.
†Combined inverse probability of treatment and censoring weights was used.

Table 6. Association between selective serotonin reuptake inhibitor use and the risk of hip fracture using trimmed inverse probability of treatment weighting estimation of marginal structural models without and with accounting for censoring

		Mondriaan		BIFAP	
	Adjusted for	HR	95%CI	HR	95%CI
Not accounting for informative censoring*	None (crude)	1.69	1.05, 2.67	2.14	1.91, 2.39
	Gender	1.68	1.06, 2.67	2.11	1.89, 2.38
	Gender + Age	2.46	1.55, 3.99	1.54	1.37, 1.72
Accounting for informative censoring†	None (crude)	1.73	1.08, 2.77	2.05	1.83, 2.30
	Gender	1.71	1.07, 2.74	2.03	1.81, 2.28
	Gender + Age	2.47	1.53, 3.98	1.51	1.35, 1.70

Both treatment and censoring weights were stabilized and trimmed at 0.5th and 99.5th percentiles.
BIFAP, Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; HR, hazard ratio.
*Only inverse probability of treatment weights was used.
†Combined inverse probability of treatment and censoring weights was used.

DISCUSSION

Our study shows that use of selective serotonin receptor inhibitor (SSRI) is associated with an increased risk of hip fracture in both cohorts (Mondriaan and BIFAP) when comparing with no SSRI use. In addition, this increased risk was consistently found for different analytical approaches: time-varying Cox model, PS methods, and MSMs. However, the magnitude of the association clearly differed between Mondriaan and BIFAP. For example, HR was 2.61 versus 1.52 using a time-varying Cox model and 2.13 versus 1.58 when applying MSMs in Mondriaan and BIFAP, respectively.

In both Mondriaan and BIFAP, estimates from time-varying Cox, time-varying PS models, and MSMs with and without accounting for censoring were similar. Differences in treatment effect estimates between the two cohorts are unlikely because of the applied methods for two reasons. First, the performance of the different methods to control for confounding was very similar within the two cohorts, but the impact of confounding adjustment was in opposite directions in the two cohorts. Second, substantial effort has been made to harmonize the design of the study, protocol, and data specifications.[14,16]

Possible explanations for the observed differences in confounding adjustment across databases may

DOI: 10.1002/pds

include substantial differences in time-dependent confounding (particularly age)[8,9,12] and/or collider-stratification bias,[8,9,12,20] non-collapsibility of the HR,[12,21,22] differences in quality of confounder information between the datasets, or the small number of events in Mondriaan leading to unstable estimates. With regard to the first possible explanation, time-varying Cox model and PS models condition on time-dependent covariates, for example, severity of depression, which may also be intermediates in the causal pathway between treatment (SSRI use at time t-1) and hip fracture, whereas MSMs reweight the original population without conditioning on such time-dependent covariates. When such time-dependent confounders are affected by unmeasured factors that predict hip fracture risk (e.g., alcohol consumption, on which we had no information in Mondriaan), the analytic approaches considered, except MSMs, may induce collider-stratification bias. Although collider-stratification bias tends to be a less substantial source of bias than confounding,[21] under certain circumstances, it could result in a dramatic change not only in the magnitude of the effect estimate but also in the direction of the effect.[22,23] In addition, similarity of effect estimate between conventional time-varying Cox regression, time-varying PS, and MSMs should be interpreted with caution. It might be possible that confounding due to adjusting-away the effect of previous treatment (by conditioning on the intermediate, time-dependent confounding) and collider-stratification bias might have similar magnitude but opposite direction, hence canceling out and leading to similar effect estimate to those of conventional methods. Regarding the second explanation, the magnitude of non-collapsibility increases with the effect of the covariate on the outcome, the baseline risk, and the strength of the treatment effect (the latter seems more likely in Mondriaan than BIFAP because non-collapsibility causes the estimate to change away from the null effect, $HR = 1.0$).[22,24] However, non-collapsibility often results in an effect estimate away from the null,[21,22] and its impact in practice is often difficult to quantify. In contrast, marginal estimates tend to be closer to the null effect than conditional estimates.[25] Thus, in our study, one would expect the PS estimates to be closer to an $HR = 1$ than the conventional estimates. However, this was not the case in our study (PS: $HR = 2.82$ and conventional: $HR = 2.62$), which implies that reasons other than non-collapsibility explain the differences, for example, misspecification of the outcome model for conventional models or inadequate balance due to misspecification of the PS model. Both time-dependent confounding and collider-stratification bias but not non-collapsibility can be best identified using causal diagrams.[20,26]

Inverse probability-weighted estimation of MSMs not only controls for time-dependent confounding without any risk of collider-stratification bias but also can account for bias because of informative censoring and be modified to accommodate competing risks.[8,9] However, the impact of informative censoring seems minimal, if any, in both cohorts (assuming that the models for censoring are correctly specified and all predictors of informative censoring were observed). This was demonstrated by comparable treatment effect estimates with and without the use of censoring weights.

Weight truncation reduced variability of the weights and improved the precision, which has been described before.[27] However, treatment effect estimates were sensitive to weight truncation, particularly in Mondriaan, which could in part be due to strong covariate–treatment association and/or the small number of events in Mondriaan. Although the optimal level of truncation is difficult to determine, it is important to explore the sensitivity of effect estimates and precision to progressive weight truncations. Alternatively, other approaches proposed to deal with the positivity assumption could be employed.[28] Importantly, investigators should focus on the procedures leading to the generation of weights (i.e., proper specification of the PS model) rather than relying on ad hoc methods such as weight truncations.[27]

We conducted this study in two large cohorts with a reasonably long follow-up time. In addition, detailed information was collected on exposure, co-morbidities, co-medications, and the outcome (hip fracture). However, the level of detail on important information such as co-morbidities limited optimal adjustment and comparison.[14,16,19] Although, there might still be unmeasured confounding because of patient characteristics not recorded in both databases, such as body mass index and alcohol consumption, we previously demonstrated that additional adjustment for these variables had limited impact on effect estimates.[19]

In conclusion, this study indicates an increased risk of hip fracture associated with SSRI, which was consistently observed using different analytical approaches in two large electronic record healthcare databases. Although differences between methods to control for time-dependent confounding were small, relevant differences in treatment effect estimates between the two datasets were observed. These are possibly attributable to different confounder structures in the datasets, particularly of age.

DOI: 10.1002/pds

CONFLICT OF INTEREST

Olaf Klungel had received unrestricted funding for pharmacoepidemiological research from the Dutch private–public-funded Top Institute Pharma (TI Pharma grant T6.101 Mondriaan). Nicolle M. Gatto is an employee and shareholder of Pfizer Inc.

KEY POINTS

- In this study, differences between the different methods to adjust for time-dependent confounding (i.e., time-varying Cox and PS methods as well as inverse probability weighting of MSMs) were small.
- The observed differences in treatment effect estimates between the datasets are likely attributable to different confounding information in the datasets.
- Adequate information on (time varying) confounding is crucial to prevent confounding bias in observational studies of pharmacological exposures.

ETHICS STATEMENT

The authors state that no ethical approval was needed.

ACKNOWLEDGEMENTS

The research leading to these results was conducted as part of the PROTECT consortium, which is a public–private partnership coordinated by the European Medicines Agency.

The PROTECT project is supported by the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu) under grant agreement no. 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations companies' in-kind contribution. In the context of the Innovative Medicines Initiative Joint Undertaking, the Department of Pharmacoepidemiology, Utrecht University, also received a direct financial contribution from Pfizer. The views expressed are those of the authors only and not of their respective institution or company.

REFERENCES

1. Ginzburg R, Rosero E. Risk of fractures with selective serotonin-reuptake inhibitors or tricyclic antidepressants. *Ann Pharmacother* 2009; **43**: 98–103.
2. Van den Brand MWM, Samson MM, Pouwels S, *et al.* Use of anti-depressants and the risk of fracture of the hip or femur. *Osteoporosis Int* 2009; **20**: 1705–1713.
3. Hu XH, Bull SA, Hunkeler EM, *et al.* Incidence and duration of side effects and those rated as bothersome with selective serotonin reuptake inhibitor treatment for depression: patient report versus physician estimate. *J Clin Psychiatry* 2004; **65**: 959–965.
4. Tsai AC, Weiser SD, Petersen ML, *et al.* A marginal structural model to estimate the causal effect of antidepressant medication treatment on viral suppression among homeless and marginally housed persons with HIV. *Arch Gen Psychiatry* 2010; **67**: 1282–1290.
5. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *Am J Epidemiol* 2002; **155**: 1045–1053.
6. Cox DR. Regression models and life tables. *J R Statist Soc Series B* 1972; **34**: 187–220.
7. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* 1999; **20**: 145–157.
8. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
9. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**: 561–570.
10. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**: 143–155.
11. Robins JM. Marginal structural models. Proceedings or the American Statistical Association, Section on Bayesian Statistical Science, 1997; 1–10.
12. Ali MS, Groenwold RHH, Pestman WR, *et al.* Time-dependent propensity score and collider-stratification bias: an example of beta2-agonist use and the risk of coronary heart disease. *Eur J Epidemiol* 2013; **28**: 291–299.
13. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; **168**: 656–664.
14. Abbing-Karahagopian V, Kurz X, de Vries F, *et al.* Bridging differences in outcomes of pharmacoepidemiological studies: design and first results of the PROTECT project. *Curr Clin Pharmacol* 2014; **9**: 130–138.
15. De Vries F, Pouwels S, Bracke M, *et al.* Use of β2 agonists and risk of acute myocardial infarction in patients with hypertension. *Br J Clin Pharmacol* 2008; **65**: 580–586.
16. ENCePP guide on methodological standards in pharmacoepidemiology. EMA/95098/2010. Available at: www.encepp.eu/standards_and_guidances [21 March 2014].
17. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/, ISBN 3-900051-07-0
18. Robins JM. Causal inference from complex longitudinal data. In Latent Variable Modeling and Applications to Causality, Berkane M (ed.). Springer-Verlag: New York, 1997; 69–117.
19. Abbing-Karahagopian VSP, Martin E, Huerta C, *et al.* Understanding inconsistency in the results from observational pharmacoepidemiological studies: the case of antidepressant use and risk of hip fracture. *Pharmacoepidemiol Drug Saf* 2014; **23**: 36.
20. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004; **15**: 615–625.
21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003; **14**: 300–306.
22. Greenland S, Pearl J. Adjustments and their consequences–collapsibility analysis using graphical models. *Intl Stat Rev* 2011; **79**: 401–426.
23. Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW. Quantification of collider-stratification bias and the birthweight paradox. *Paediatr Perinat Epidemiol* 2009; **23**: 394–402.
24. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res* 2013; doi:10.1177/0962280213505804.
25. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Systematic differences in treatment effect estimates between propensity score methods and logistic regression. *Int J Epidemiol* 2008; **37**: 1142–1147.
26. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999; **10**: 37–48.
27. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS One* 2011; **6** e18174.
28. Xu S, Shetterly S, Raebel MA, *et al.* Estimating the effects of time-varying exposures in observational studies using Cox models with stabilized weights adjustment. *Pharmacoepidemiol Drug Saf* 2014. doi:10.1002/pds.3601.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.