PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2016; **25**(Suppl. 1): 122–131
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3863

ORIGINAL REPORT

Instrumental variables analysis using multiple databases: an example of antidepressant use and risk of hip fracture†

Md Jamal Uddin[1,2], Rolf H. H. Groenwold[1,3], Anthonius de Boer[1], Helga Gardarsdottir[1,4], Elisa Martin[5], Gianmario Candore[6], Svetlana V. Belitser[1], Arno W. Hoes[3], Kit C. B. Roes[3] and Olaf H. Klungel[1,3]*

[1] *Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands*
[2] *Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh*
[3] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands*
[4] *Department of Clinical Pharmacy, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, the Netherlands*
[5] *BIFAP Research Unit. Division of Pharmacoepidemiology and Pharmacovigilance, Medicines for Human Use Department, Spanish Agency for Medicines and Medical Devices (AEMPS), Madrid, Spain*
[6] *European Medicines Agency (EMA), London, UK*

ABSTRACT

Purpose Instrumental variable (IV) analysis can control for unmeasured confounding, yet it has not been widely used in pharmacoepidemiology. We aimed to assess the performance of IV analysis using different IVs in multiple databases in a study of antidepressant use and hip fracture.

Methods Information on adults with at least one prescription of a selective serotonin reuptake inhibitor (SSRI) or tricyclic antidepressant (TCA) during 2001–2009 was extracted from the THIN (UK), BIFAP (Spain), and Mondriaan (Netherlands) databases. IVs were created using the proportion of SSRI prescriptions per practice or using the one, five, or ten previous prescriptions by a physician. Data were analysed using conventional Cox regression and two-stage IV models.

Results In the conventional analysis, SSRI (vs. TCA) was associated with an increased risk of hip fracture, which was consistently found across databases: the adjusted hazard ratio (HR) was approximately 1.35 for time-fixed and 1.50 to 2.49 for time-varying SSRI use, while the IV analysis based on the IVs that appeared to satisfy the IV assumptions showed conflicting results, e.g. the adjusted HRs ranged from 0.55 to 2.75 for time-fixed exposure. IVs for time-varying exposure violated at least one IV assumption and were therefore invalid.

Conclusions This multiple database study shows that the performance of IV analysis varied across the databases for time-fixed and time-varying exposures and strongly depends on the definition of IVs. It remains challenging to obtain valid IVs in pharmacoepidemiological studies, particularly for time-varying exposure, and IV analysis should therefore be interpreted cautiously. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS—unmeasured confounding; instrumental variables; physician's prescribing preference; multiple general practice databases; antidepressant; hip/femur fracture; pharmacoepidemiology

Received 17 September 2014; Revised 28 July 2015; Accepted 29 July 2015

INTRODUCTION

Electronic healthcare databases are being used to detect unintended and intended effects of drugs in comparative safety and effectiveness research. However, such databases sometimes have very limited or inaccurate information on potential confounding variables, such as alcohol consumption and functional health status (e.g. activities of daily living),[1,2] which may impair the validity of study results. For example, several observational studies[3–8] indicated an association between antidepressants use (mainly tricyclic antidepressants and selective serotonin re-uptake inhibitors) and the risk of hip fracture. However, results from these studies are heterogeneous for many reasons.[3,5,6,8] Among them, confounding by unmeasured patient characteristics such as frailty, severity of depression, and

*Correspondence to: O. H. Klungel, Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Universiteitsweg 99, PO Box 80082, 3584CG Utrecht, the Netherlands. E-mail: O.H.Klungel@uu.nl

†The manuscript in this form has not been published and is not being considered for publication elsewhere, in whole or in part, in any language. The abstract has been presented at the 30th International Conference on Pharmacoepidemiology and Therapeutic Risk Management, October 24–27, 2014, Taipei, Taiwan.

lifestyle factors (e.g. smoking and alcohol consumption) have usually not been accounted for in most of the studies.[3] Instrumental variable (IV) analysis has been proposed to control for unmeasured confounding in pharmacoepidemiologic studies.[2,9–15]

An IV is a variable that can be considered to mimic the treatment allocation process in a randomized trial.[2] That means an IV (i) is associated with the exposure, (ii) affects the outcome only through the exposure, and (iii) is independent of confounders.[16–18] If these key assumptions and some additional assumptions (e.g. monotonicity) are satisfied, IV analysis may consistently estimate the average causal effect of an exposure on an outcome.[18,19]

Several studies have applied IV analysis using electronic healthcare databases on various pharmacoepidemiological issues.[10,11,13–15,20–29] As far as we know, there is no study that assessed the performance of IV analysis across multiple databases in different countries using both time-fixed and time-varying exposures. We therefore aimed to assess the performance of IV analysis using different IVs in general practice (GP) databases from three European countries (Spain, UK, and The Netherlands) in a study of antidepressants use and risk of hip/femur fracture.

METHODS

Data sources and study population

This study was conducted using data from the health improvement network, UK (THIN); the Spanish Base de datos para la Investigación Farmacoepidemiológicaen Atención Primaria (BIFAP); and the Dutch Mondriaan GP database.[30–32] Detailed information on these databases and data specification can be found in a common study protocol, which is available online.[33]

We used information on 587 637 patients from THIN, 252 203 patients from BIFAP, and 22 954 patients from Mondriaan who were all aged ≥18 years and had at least one prescription of selective serotonin re-uptake inhibitors (SSRI) or tricyclic antidepressants (TCA) during the study period (1 January 2001 to 31 December 2009). Moreover, at least one year of enrolment in the database without any antidepressants prescription in the six months and 1 year without a hip fracture preceding the enrolment in the cohort were required to enter the study. The index date (date of cohort entry) was defined for each patient as the date of first SSRI or TCA prescription within the study period. The observation period for each patient lasted from the index date to the end of data collection (31 December 2009), the date of the first fracture, the date of death, or loss to follow-up, whichever occurs first. We excluded patients (17 498 in THIN, 1338 in BIFAP, and 480 in Mondriaan) from the study population in case of combined use of SSRI and TCA or subsequent use within the same treatment episode. Furthermore all subjects from practices that contributed less than 50 patients were excluded, because a considerable number of prescriptions is needed to accurately measure the IV.[20] The final number of practices was 502 for THIN, 280 for BIFAP, and 133 for Mondriaan.

Exposures and outcome

We considered two types of exposure: (i) exposure that was determined at the index date and considered constant over time; and (ii) exposure that was time-varying. Throughout the manuscript we will use the abbreviations "$SSRI_{fixed}$" and "$SSRI_{time}$" for the exposure that was considered constant during follow-up and time-varying exposure, respectively. The exposure $SSRI_{fixed}$ was defined based on the first SSRI (vs. TCA) prescription, and consequently the follow-up time was the time between the first prescription and the end of follow-up. During follow-up, we ignored any changes in medication status (e.g. switch from TCA to SSRI or stopping SSRI use).

For the exposure $SSRI_{time}$, antidepressant use was re-categorized into two groups: current SSRI users and non-SSRI users. Non-SSRI users could be current, recent, or past TCA users or recent/past SSRI users. Follow-up time for time-varying exposure, $SSRI_{time}$, was divided into periods of "current use" to SSRI (from the beginning of the prescription up to the calculated end date of the last prescription supply in a continuous treatment episode, i.e. <30 days between the end of a prescription supply and the beginning of the next one) and non-current SSRI use (all other time periods). The drug codes are provided in the study protocol.[33]

A first fracture of the hip/femur (HF) during the study period was the outcome of interest, regardless of whether a subject had a fracture before the index date. In BIFAP, hip fractures that, based on review of automated clinical records, were considered to be the result of a major trauma (e.g. car accident) were excluded. A detailed description of the outcome is provided in the study protocol.[33]

Potential confounding factors

The potential confounding factors (i.e. co-medications, co-morbidities, and lifestyle factors) are listed in Table 1. Because lifestyle factors such as smoking, BMI, and alcohol consumption were not available in all databases, we did not add them in the conventional analyses and considered them as unmeasured, to allow for a proper comparison between databases. For the

DOI: 10.1002/pds

Table 1. Characteristics of patients stratified by database and type of antidepressant drugs (SSRI vs. TCA) assigned at index date

	THIN		BIFAP		Mondriaan	
	TCA	SSRI	TCA	SSRI	TCA	SSRI
Sample size	220 562	349 577	44 599	207 604	8033	14 441
Number of cases (HF)	1694	1937	240	1288	32	49
Confounders (%)						
Mean age (SD)	56.0 (17.5)	46.0 (18.06)	53.0 (16.8)	50.0 (17.8)	54.0 (16.9)	46.0 (16.8)
Sex	37.2	35.6	25.9	27.6	35.9	36.6
Stroke	4.7	3.5	3.6	3.8	2.7	2.1
Ischaemic heart disease	9.1	5.6	4.6	4.9	5.6	3.8
Dementia	0.3	0.9	0.6	1.3	—	—
History of fractures	18.9	18.5	4.7	4.6	5.5	4.4
Mental disorders	1.7	2.8	1.5	2.2	2.7	4.8
Osteoporosis	3.7	1.8	6.5	5.1	2.7	1.3
Anaemia	7.8	6.5	7.7	7.3	5.0	4.4
DMARD	1.6	0.7	0.8	0.5	1.7	0.7
Benzodiazepines	14.1	17.5	29.4	41.8	29.3	37.0
Antiepileptic drugs	4.0	2.0	7.1	3.6	5.9	1.6
Antidiabetic drugs	6.4	3.5	5.9	5.3	8.6	4.2
Glucocorticoids	6.3	3.7	3.5	2.6	7.1	3.3
Antihypertensive	29.2	18.9	20.1	18.6	24.7	16.1
ACE inhibitors	—	—	9.0	8.4	8.7	5.7
Angiotensin II antagonists	—	—	4.9	5.1	5.4	2.6
Diuretics	17.8	10.3	9.8	9.1	13.3	7.5
Opioids (including Morphine)	19.3	6.2	13.1	5.4	16.8	4.1
Hormone replacement therapy	7.2	5.5	1.9	1.5	3.5	3.0
Other antidepressant	1.9	2.9	3.0	3.5	2.2	3.9
Antipsychotics	4.6	5.2	4.8	5.6	1.7	3.1
Vitamin-D	0.3	0.1	5.3	3.9	2.4	1.2
Antiparkinson	0.6	0.6	0.7	0.9	0.8	0.6
Statins	15.7	8.9	9.9	9.1	12.3	7.3
Other information						
SSRI users (%)	61.3		82.2		64.0	
Total follow-up time in years	2 118 320		757 403		63 806	
Median follow-up time in years	3.37		2.71		2.25	
Number of practices	502		280		133	

Abbreviations: SSRI: selective serotonin re-uptake inhibitor; TCA: tricyclic antidepressant; THIN: The Health Improvement Network, UK; BIFAP: the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan: Dutch general practices databases.

exposure $SSRI_{fixed}$, confounders were assessed at the index date (i.e. comorbidities were identified as "ever before" the index date and all other confounders were identified during the 182 days prior to index date) and for the exposure $SSRI_{time}$ confounders were considered time-varying and assessed at baseline and updated whenever a patient's exposure status changed (current, recent, and past) or every six months (if exposure status did not change). For co-morbidities, patients were classified as having the disease from the first date of diagnosis onwards. More details on confounding factors are available online.[33]

Construction of instrumental variables

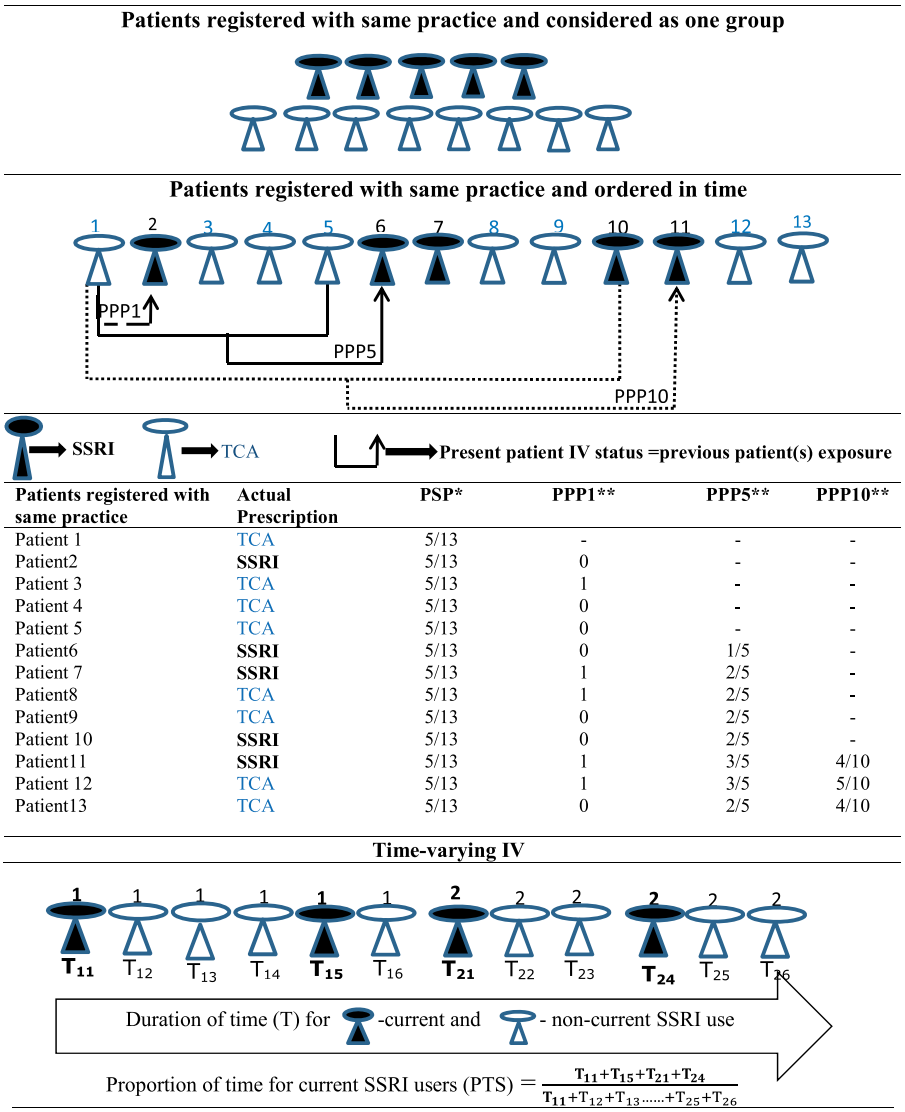
We considered several IVs based on the physician's prescribing preferences (PPP) of SSRI or TCA (Figure 1).[2,20] For $SSRI_{fixed}$, the PPPs were based on either the single last prescription (PPP_1), the last five consecutive prescriptions (PPP_5), or the last ten consecutive prescriptions (PPP_{10}). Furthermore, we used the proportion of SSRI prescriptions (PSP) per practice to create an alternative preference based IV. We dichotomized the PSP at its median to create a binary IV (PSP_{dich}). Finally, we also considered other IVs based on the last 20 or 50 prescriptions issued by physicians within a practice.

For the exposure $SSRI_{time}$, the IVs were measured using the proportions of time of "current" SSRI use (PTS) in a practice. Specifically, the PTS is the ratio of the follow-up time contributed by patients who are current SSRI users compared to the total follow-up time contributed by patients in the same practice. We also dichotomized the PTS (at the median) to create a binary IV (PTS_{dich}).

Assessment of instrumental variable assumptions

IV assumption 1: We used point bi-serial correlation (r) for binary exposure and continuous IV and odds

DOI: 10.1002/pds



Patients registered with same practice	Actual Prescription	PSP*	PPP1**	PPP5**	PPP10**
Patient 1	TCA	5/13	-	-	-
Patient2	**SSRI**	5/13	0	-	-
Patient 3	TCA	5/13	1	-	-
Patient 4	TCA	5/13	0	-	-
Patient 5	TCA	5/13	0	-	-
Patient6	**SSRI**	5/13	0	1/5	-
Patient 7	**SSRI**	5/13	1	2/5	-
Patient8	TCA	5/13	1	2/5	-
Patient9	TCA	5/13	0	2/5	-
Patient 10	**SSRI**	5/13	0	2/5	-
Patient11	**SSRI**	5/13	1	3/5	4/10
Patient 12	TCA	5/13	1	3/5	5/10
Patient13	TCA	5/13	0	2/5	4/10

Figure 1. Definition of instrumental variables based on proportion of SSRI prescriptions and previous prescription written by physician

ratio (OR) for both binary exposure and IV to check the association between IV and exposure.[11,18] Although the partial F-statistic value from the first-stage IV model is often used to assess this assumption, we did not use it because it is highly influenced by sample size.[11,18]

IV assumptions 2 and 3: We assumed that a physician's preference did not influence the risk of HF of a patient (assumption 2) nor that it was correlated with confounders (assumption 3).[20] In addition, a falsification test was applied to assess assumption 3 using the standardized difference (SDif) to assess balance of measured confounders between IV categories (when an IV was continuous, balance was assessed across the quintiles).[34] We also measured the balance on several confounders simultaneously using the multivariate Mahalanobis distance (MD).[23,35] Smaller values of SDif (e.g. <0.10)[36] and MD indicate better balance.[34,35] If measured confounders are imbalanced between IV categories, this may also imply imbalance of unmeasured confounders (in which case assumption 3 is violated) and thus IV analysis is inappropriate.[34] Nevertheless, if measured confounders are balanced, we assumed that such balance could be carried-over to unmeasured confounders.[2,27]

Statistical analyses

We analysed the three datasets (THIN, BIFAP, and Mondriaan) separately. We also studied the exposure $SSRI_{fixed}$ using the pooled data. A conventional Cox

DOI: 10.1002/pds

proportional hazards model and two-stage IV models were used to analyse the data. The first-stage model of the IV analysis was a linear regression model,[37,38] where the exposure ($SSRI_{fixed}$ or $SSRI_{time}$) was the dependent variable and the IV was the independent variable. The second-stage model was a Cox proportional hazard model, in which predicted treatment rather than actual treatment was the independent variable.[18] We estimated the unadjusted and adjusted hazard ratios with and without including confounders (confounders measured at the index date for the $SSRI_{fixed}$ and the time-varying confounders for $SSRI_{time}$) in both conventional and IV analyses. Bootstrapping (1000 bootstrap samples) was applied to estimate the confidence intervals (CI) of IV estimates. All analyses were performed using the statistical software R version 2.15.2.[39]

The effects of SSRI/TCA on the risk of HF may not be homogeneous in our study population. Hence, the point estimate of the average causal effect of exposure for the whole study population could not be identified, yet other effects (e.g. complier average causal effects [CACE] or local average treatment effect [LATE]) could be identified.[16,18] In our example, the compliers are subjects who would be prescribed an SSRI had they seen a physician who preferred SSRI, but would be prescribed a TCA had they seen a physician who preferred TCA.[13,19] For continuous IVs, the CACE is a weighted average of the effect in multiple subgroups where the more compliant subjects obtain a larger weight.[16,19,40,41] To identify the CACE, a fourth assumption, i.e. monotonicity (the IV affects the exposure deterministically in one direction meaning there are no defiers), was considered.[16,19,40,41] Roughly speaking, this assumption implies that there are no subjects who would be prescribed an SSRI by a TCA preferring physician, while they would be prescribed a TCA by a SSRI preferring physician.

RESULTS

Table 1 shows the patients characteristics, stratified by exposure status (SSRI vs. TCA) and database ($n = 570\,139$ for THIN, $n = 252\,203$ for BIFAP, and $n = 22\,474$ for Mondriaan). In total, 3631 persons in THIN and 1528 persons in BIFAP and 81 persons in Mondriaan experienced a HF during the study period. As expected, measured confounders were more balanced between IV categories than between exposure groups (Table A2a–A2b, Appendix).

Several methods were applied to evaluate the validity of IVs particularly the association between exposure and IV (assumption 1) and independence between IV and confounders (assumption 3). The summary results are presented in Table 2. In THIN, the IVs, PSP and PPP_{10}, were strongly associated with $SSRI_{fixed}$ ($r \geq 0.15$, Table A1 in the Appendix), and confounders were balanced across the quintiles of these IVs (SDif < 0.10, Table A2a, Appendix). Similar performance was observed for the IV PSP in BIFAP and the IVs PPP_5 and PPP_{10} in Mondriaan. In many situations in which an IV was weakly associated with the exposure, the confounders were balanced with that IV or vice versa. For instance, the IV PPP_1 was weakly associated with $SSRI_{fixed}$ (OR < 2 in all databases, Table A1, Appendix), but the measured confounders were balanced across the categories of the IV PPP_1 (SDif < 0.10 in all databases, Tables A2a–b, Appendix). Alternatively, PSP and PSP_{dich} were strongly associated with $SSRI_{fixed}$ ($r = 0.26$ and OR $= 2.44$, respectively, Table A1, Appendix) in Mondriaan, but confounders were imbalanced (e.g. difference in mean age, Table A2a, Appendix). For the pooled data, only the IVs PPP_5 and PPP_{10} were strongly associated with the $SSRI_{fixed}$ ($r = 0.18$ and 0.21, respectively) and confounders were balanced across the categories of these IVs (SDIF < 0.10). For time-varying exposure, both IVs (PTS and PTS_{dich}) were weakly associated with $SSRI_{time}$ in all databases (Table A1, Appendix). In Mondriaan and BIFAP (PTS_{dich}) confounders were also imbalanced across levels of these IVs (Table A2b, Appendix).

Table 3 shows the hazard ratios (HR) of the association between SSRI use and the risk of HF. In this setting, we used the first prescription ($SSRI_{fixed}$) of eligible patients. In all databases, conventional analysis showed that SSRI use was associated with an increased risk of HF compared to TCA use (though not statistically significant in Mondriaan); the adjusted HR was about 1.35 (Table 3). The results from IV analysis based on the apparently valid IVs (i.e. those IVs that appeared to satisfy the IV assumptions) showed that SSRI use was not associated with an increased risk of HF in any of the databases; the adjusted HRs ranged from 0.55 to 2.75 (Table 3). In that case, the length of confidence intervals of the IV estimates was quite large though the point estimates were far from the null. All IVs, except PSP in BIFAP, were weakly associated with the $SSRI_{fixed}$, and exposure effect estimates based on the weak IVs had very wide confidence intervals (Table 3). When the number of previous prescriptions increased for estimating the IV PPP (e.g. previous 20 or 50 prescriptions), the IV estimates were approximately similar to those observed for the IV PPP_{10} (data not shown).

DOI: 10.1002/pds

Table 2. Summary of results of instrumental variables analysis of the relation between SSRI vs. TCA use and risk of hip fracture

Database	Instrumental variables	IV assumption 1	IV assumption 3	Unadjusted hazard ratio	Adjusted hazard ratio	Length CI of adjusted HR
THIN	**PSP**			**0.51**	**1.09**	**0.85**
	PSP_{dich}		X	0.52	1.07	1.29
	PPP1	X		1.00	2.02	5.74
	PPP5	X		0.62	1.31	1.75
	PPP10			**0.57**	**1.16**	**1.22**
	PTS*	X		0.97	1.90	4.50
	PTS_{dich}*	X		0.50	0.97	4.34
BIFAP	**PSP**			**4.51**	**2.75**	**6.13**
	PSP_{dich}	X		3.18	0.86	7.10
	PPP1	X		22.2	42.2	5258
	PPP5	X		3.07	3.44	26.6
	PPP10	X		2.57	1.89	9.02
	PTS*	X		79.0	4.04	15.6
	PTS_{dich}*	X	X	45.8	4.21	46.5
Modriaan	PSP		X	0.28	1.19	8.05
	PSP_{dich}		X	0.24	1.10	15.3
	PPP1	X		0.09	0.20	93.5
	PPP5			**0.20**	**0.55**	**16.2**
	PPP10			**0.44**	**1.67**	**27.5**
	PTS*	X	X	0.44	0.83	25.8
	PTS_{dich}*	X	X	0.03	0.27	32.4

"X" indicates instrumental variables assumptions are violated. **Bold** numbers indicate that the potential instrumental variables at least appeared to satisfy assumptions 1 and 3. Unadjusted and adjusted estimates are based on IV models without and with confounders, respectively. Confidence intervals were estimated using bootstrapping.
*IVs for time-varying settings.
Abbreviations: SSRI: selective serotonin re-uptake inhibitor; TCA: tricyclic antidepressant; THIN: The Health Improvement Network, UK; BIFAP: the Spanish Base de datos para la Investigación Farmacoepidemiológicaen Atención Primaria; Mondriaan: Dutch general practices databases; IV: instrumental variable; PSP: proportion of SSRI prescriptions per practice; PSP_{dich}: proportion of SSRI prescriptions per practice (dichotomized at the median); PPP1: single previous prescription by a physician; PPP5: previous five prescriptions by a physician; PPP10: previous ten prescriptions by a physician; PTS: proportion of time of SSRI prescriptions; PTS_{dich}: proportion of time of SSRIprescriptions (dichotomized at the median); HR: hazard ratio; CI: confidence interval.

Table 3. Hazard ratios of the associations between SSRI vs. TCA use and risk of hipfracture based on conventional and IV analysis

		Conventional Cox model	Instrumental variable analysis				
			PSP	PSP_{dich}	PPP_1	PPP_5	PPP_{10}
Database	Model	HR [CI]	HR [CI*]	HR [CI*]	HR [CI*]	HR [CI*]	HR [CI*]
THIN	Unadjusted	0.72 [0.67–0.77]	**0.51 [0.36–0.70]**	0.51 [0.31–0.80]	1.00 [0.39–2.69]	0.62 [0.37–1.08]	**0.57 [0.36–0.92]**
	Adjusted	1.35 [1.26–1.44]	**1.09 [0.75–1.60]**	1.23 [0.72–2.01]	2.02 [0.61– 6.35]	1.31 [0.73– 2.48]	**1.16 [0.70–1.92]**
BIFAP	Unadjusted	1.21 [1.06–1.39]	**4.51 [1.68–11.1]**	3.18 [0.91–11.6]	22.2 [0.44–1917]	3.07 [0.47– 23.0]	2.57 [0.59–12.0]
	Adjusted	1.35 [1.18–1.56]	**2.75 [0.97–7.10]**	1.86 [0.58–7.68]	42.2 [0.41–5259]	3.44 [0.56– 27.1]	1.89 [0.33–9.35]
Mondriaan	Unadjusted	0.75 [0.48–1.17]	0.28 [0.07–1.33]	0.24 [0.02–1.84]	0.09 [0.001– 18.3]	**0.20 [0.01– 3.19]**	**0.44 [0.04–5.43]**
	Adjusted	1.36 [0.84–2.15]	1.19 [0.16– 8.21]	1.10 [0.09–15.4]	0.20 [0.001–93.5]	**0.55 [0.03–16.3]**	**1.67 [0.15–27.7]**

Exposure status is based on the first prescription of a patient. Confounders in the adjusted models are listed in the Table 1. **Bold** numbers indicate that 2 of the assumption for IV analysis (assumptions 1 and 3) appear to be satisfied.
*Confidence intervals are estimated by bootstrap method.
Abbreviations: SSRI: selective serotonin re-uptake inhibitor; TCA: tricyclic antidepressant; THIN: The Health Improvement Network, UK; BIFAP: the Spanish Base de datos para la Investigación Farmacoepidemiológicaen Atención Primaria; Mondriaan: Dutch general practices databases; IV: instrumental variable; PSP: proportion of SSRI prescriptions per practice; PSP_{dich}: proportion of SSRI prescriptions per practice (dichotomized at the median); PPP1: single previous prescription by a physician; PPP_5: previous five prescriptions by a physician; PPP10: previous ten prescriptions by a physician; HR: hazard ratio; CI: confidence interval.

The analysis of pooled data, showed an unadjusted HR of 0.83 [0.78–0.88] and adjusted HR of 1.35 [1.27–1.43] based on the conventional analysis, while IV analysis based on the IVs PPP_5 and PPP_{10} produced unadjusted HRs of 1.10 [0.78–1.53] and 1.07 [0.80–1.40] and adjusted HRs of 1.70 [1.16–2.50] and 1.57 [1.13–2.15], respectively.

Table 4 shows the HRs for the time-varying exposure ($SSRI_{time}$, current SSRI users vs. non-SSRI users). The conventional analysis showed that the current

DOI: 10.1002/pds

Table 4. Hazard ratios of the associations between time-varying SSRI versus non-SSRI use and risk of hip fracture based on conventional and IV analysis

		Conventional*	Instrumental variables	
			PTS	PTS_{dich}
	Model	HR [CI]	HR [CI†]	HR [CI†]
THIN	Unadjusted	1.76 [1.64–1.89]	0.97 [0.43–2.16]	0.50 [0.20–2.78]
	Adjusted	1.67 [1.56–1.80]	1.90 [0.69–5.19]	0.97 [0.39–4.73]
BIFAP	Unadjusted	2.09 [1.89–2.31]	79.0 [17.5–270]	45.8 [27.1–2083]
	Adjusted	1.50 [1.35–1.66]	4.04 [0.74–16.4]	4.21 [0.54–47.0]
Mondriaan	Unadjusted	1.75 [1.12 –2.72]	0.44 [0.01–17.4]	0.03 [0.001–40.3]
	Adjusted	2.49 [1.59–3.90]	0.83 [0.02–25.8]	0.27 [0.001–32.4]

Exposure status is based on (time-varying) prescription data. Confounders in the adjusted models are listed in the Table 1.
*Time-dependent Cox Model.
†Confidence intervals are estimated by bootstrap method.
Abbreviations: SSRI: selective serotonin re-uptake inhibitor; TCA: tricyclic antidepressant; non-SSRI: non-SSRI users could be current, recent or past TCA users or recent/past SSRI users; HR: hazard ratio; THIN: The Health Improvement Network, UK; BIFAP: the Spanish Base de datos para la Investigación Farmacoepidemiológicaen Atención Primaria; Mondriaan: Dutch general practices databases; IV: instrumental variable; CI: confidence interval.

SSRI use was associated with the increase risk of HF compared to non-SSRI use in all databases; the adjusted HRs were 1.67 [1.56–1.80] (THIN), 1.50 [1.35–1.66] (BIFAP), and 2.49 [1.59–3.90] (Mondriaan). However, because of violation of one or more of the key IV assumptions (Table 2), these results could not be interpreted.

DISCUSSION

Conventional analysis showed an increased risk of HF for SSRI users versus TCA users, whereas IV analyses based on the apparently valid IVs (i.e. those IVs that appeared to satisfy the IV assumptions) showed that SSRI use was not associated with an increased risk of HF compared to TCA use, although the point estimates were quite far from the null. In that case, we observed that the width of confidence intervals of the IV estimates was quite large. However, IV analysis based on the pooled data showed opposite results that are consistent with the conventional analyses.

The exposure effects from IV analysis are less precise than those of conventional analysis. None of the IVs we considered appeared to be valid for time-varying exposure ($SSRI_{time}$), and thus the exposure effects are invalid. There is a trade-off between the strength of the IV and the balance of confounders across levels of the IV, which has been described by others.[42] We note that results from the conventional analysis seem consistent for time fixed and time-varying exposures and across the databases, which is not true for IV analysis mainly because of violations of the key IV assumptions. It should be noted that the exposure effects from conventional analysis and IV analysis are not directly comparable as the conventional analysis provides the ACE and the IV analysis provides the CACE.[43,44]

With respect to our assessment criteria, we did not identify any IV that is consistently valid across the databases. However, we found that at least one of the IVs for $SSRI_{fixed}$ appears valid in each database. For example, the IV PSP in THIN and BIFAP and the IV PPP_{10} in THIN and Mondriaan appear valid because they are strongly associated with the exposure $SSRI_{fixed}$, and measured confounders are balanced across levels of these IVs. On the other hand, neither of the IVs across the databases we considered appears valid for the exposure $SSRI_{time}$. The main limitation is the weak association between the IVs and the exposure $SSRI_{time}$. We noticed that the strength of the IVs and effect measure estimates vary between the databases. Particularly, we observed that most of the IVs in the THIN and BIFAP databases were weaker than those in the Mondriaan database. This may be because of little variation in prescribing patterns between practices or prescribing guidelines that limit variation in prescriptions between physicians in THIN and BIFAP. These findings indicate that it is challenging to define a generic IV for pharmacoepidemiologic studies, especially when exposure is time-varying. Therefore, future research could be carried out to define an alternative IV for time-varying settings.

We found that when the IVs explain only a small proportion of the variance of the exposure (i.e. is weakly associated with the exposure), the estimated HRs are unstable with wide confidence intervals.[18] Moreover, these weak IVs may amplify biases because of small violations of the assumptions 2 and 3, and thus IV analysis produces more biassed results compared to conventional analysis.[11,16,19] We used an arbitrary cut-off (i.e. $r < 0.15$ and $OR < 2$) to identify a weak

DOI: 10.1002/pds

association between IV and exposure. Appropriate cut-off values may differ between studies.[18]

In the analysis of the pooled data, the variation in preferences between practices increased (compared to the analysis per database), resulting in a stronger IV, and hence IV estimates were more precise. Moreover, as IV analysis produces unstable estimates when the outcome is rare,[18] pooling databases may overcome this limitation too. Therefore, pooling databases in IV analysis seems effective provided that study protocols are consistent across the different databases.

When the measured confounders are balanced across IV categories (i.e. SDif < 0.10), unmeasured confounders could be balanced as well, in which case assumption 3 is fulfilled.[2,13,27] In that case, we argue that the assumption 2 (IV affects the outcome only through the exposure) might also be fulfilled as the assumption 2 and 3 are statistically similar.[19] Moreover, it is unlikely that the antidepressant prescription to a previous patient influences the risk of HF in a subsequent patient.[20] Additionally, our assumption was that physicians/practices act completely the same (equal standard of care) except for the preference to prescribe either SSRI or TCA. However, if concomitant treatments influence the risk of HF, this assumption may be violated.[2,29,41]

We assumed monotonicity (i.e. the IV affects the exposure deterministically in one direction meaning there are no defiers) to identify the CACE. Under this assumption, the CACE is the exposure effect for the subgroup of patients who actually adhere to the doctor's prescription.[13,40] However, it is difficult to identify those who are actual compliers. In addition, if there are defiers, the monotonicity assumption is violated, and effects estimates are biassed unless the effects are similar for both compliers and defiers.[41] Furthermore, the monotonicity may not be plausible for the preference based IVs because such IVs are surrogates for a likely continuous preference variable where complier status is not deterministic.[41]

Our study has several strengths. We studied a single drug–event pair (antidepressant drugs—HF) across multiple databases in different countries using a common study protocol to assess the performance of IV analysis with several plausible IVs.[45] Moreover, to our knowledge, this is the first study in which IVs are used to study time-fixed as well as time-varying exposure effects across multiple databases. Additionally, as we followed a common study protocol, our definitions, data extraction, and analysis were consistent across databases.

There are some limitations in our study. Although we applied a robust balance measure (standardized difference)[36,46] to falsify assumption 3, this could fail to identify an invalid IV even when the assumption is violated.[19] For example, we identified the IVs PSP in BIFAP and PPP_{10} in Mondriaan that appear valid for the exposure $SSRI_{fixed}$; however, the unadjusted and adjusted IV estimates are very different, which may indicate a possible association between the IV and unmeasured confounders.[47] Moreover, we provided some theoretical explanations in favour of assumption 2 as it is not possible to verify from the data;[16] however, this assumption may fail even when the assumptions 1 and 3 are fulfilled. Additionally, we have only information on GP prescriptions, but we do not know whether patients actually collected their medicines from the pharmacy and took them as prescribed. In addition, in order to create the preference based IV PPP we excluded the patients with combined use of SSRI and TCA. However, this exclusion may induce some selection bias in our study.

We present several recommendations based on our findings: (i) As there is a trade-off between the strength of the IV and the balance of the confounders across levels of the IV, to identify an optimal IV (i.e. strongly associated with exposure under study and independent of the confounders), it would be worthwhile to consider several plausible IVs and assess their validity; (ii) If an IV violates one of the assumptions, the IV should be treated as invalid and one should not proceed with IV analysis. On the other hand, if an IV satisfies all testable assumptions, still the IV assumptions should be justified and any inconsistency reported, which helps to understand the validity of the IV analysis. (iii) We suggest researchers to follow the checklists provided by Davies and by Swanson for reporting the results of IV analyses.[13,19]

In conclusion, we assessed the performance of IV analysis using several potential IVs in three GP databases using a common study protocol. The performance of IV analysis varied between time-fixed and time-varying exposures, across the databases, and strongly depends on the definition of IVs. Our multiple databases study shows that it is challenging to obtain a valid IV in pharmacoepidemiologic studies, especially for time-varying exposure, and thus the exposure effects from IV analysis should be interpreted cautiously.

CONFLICT OF INTEREST

Olaf Klungel has received unrestricted funding for pharmacoepidemiological research from the Dutch private-public funded Top Institute Pharma (TI Pharma Grant T6.101 Mondriaan).

DOI: 10.1002/pds

KEY POINTS

- This multiple database study shows that the performance of instrumental variables analysis varied across the databases for time-fixed and time-varying exposures and strongly depends on the definition of instrumental variables.
- Compared to the analysis of separate databases, in the analysis of pooled data there was more variation in the instrumental variable, resulting in a stronger instrumental variable and hence more precise and more stable estimates from instrumental variable analysis.
- It remains challenging to obtain valid instrumental variables in pharmacoepidemiological studies, particularly for time-varying exposure, and results of instrumental variable analysis should therefore be interpreted cautiously.

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 (Pharmacoepidemiologic Research on Outcomes of Therapeutics by a European ConsorTium) which is a public–private partnership coordinated by the European Medicines Agency.

FUNDING

The PROTECT project is supported by Innovative Medicine 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 Programme (FP7/2007–2013) and EFPIA companies' in kind contribution. In the context of the IMI Joint Undertaking (IMI JU), 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. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: the example of influenza vaccination. *Int J Epidemiol* 2008; **37**: 1422–1429.
2. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**: 268–275.
3. Ginzburg R, Rosero E. Risk of fractures with selective serotonin-reuptake inhibitors or tricyclic antidepressants. *Ann Pharmacother* 2009; **43**: 98–103.
4. Thapa PB, Gideon P, Cost TW, Milam AB, Ray WA. Antidepressants and the risk of falls among nursing home residents. *N Engl J Med* 1998; **339**: 875–882.
5. 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.
6. Eom C, Lee H, Ye S, Park SM, Cho K. Use of selective serotonin reuptake inhibitors and risk of fracture: a systematic review and meta-analysis. *J Bone Miner Res* 2012; **27**: 1186–1195.
7. Liu B, Anderson G, Mittmann N, To T, Axcell T, Shear N. Use of selective serotonin-reuptake inhibitors or tricyclic antidepressants and risk of hip fractures in elderly people. *Lancet* 1998; **351**: 1303–1307.
8. Pouwels S, Van Staa T, Egberts A, Leufkens H, Cooper C, De Vries F. Antipsychotic use and the risk of hip/femur fracture: a population-based case–control study. *Osteoporosis Int* 2009; **20**: 1499–1506.
9. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* 2013; **24**: 363–369.
10. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; **169**: 273–284.
11. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; **65**: 155–162.
12. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care* 2007; **45**: S116–S122.
13. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: an instrumental variable analysis. *Epidemiology* 2013; **24**: 352–362.
14. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: an instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; **19**: 699–707.
15. Huybrechts KF, Gerhard T, Franklin JM, Levin R, Crystal S, Schneeweiss S. Instrumental variable applications using nursing home prescribing preferences in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014; **23**: 830–838.
16. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–372.
17. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental variables in influenza vaccination studies: mission impossible?! *Value Health* 2010; **13**: 132–137.
18. Uddin MJ, Groenwold RHH, de Boer A, *et al.* Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiol Drug Saf* 2014; **23**: 165–177.
19. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; **24**: 370–374.
20. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol* 2013; **66**: 1386–1396.
21. Suh HS, Hay JW, Johnson KA, Doctor JN. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensity-score and instrumental variable methods to adjust for treatment-selection bias. *Pharmacoepidemiol Drug Saf* 2012; **21**: 470–484.
22. Chen H, Mehta S, Aparasu R, Patel A, Ochoa-Perez M. Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents. *Pharmacoepidemiol Drug Saf* 2014; **23**: 299–308.
23. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; **62**: 1233–1241.
24. Fang G, Brooks JM, Chrischilles EA. Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication. *Am J Epidemiol* 2012; **175**: 1142–1151.
25. Ahern TP, Pedersen L, Sværke C, Rothman KJ, Sørensen HT, Lash TL. The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable. *Am J Epidemiol* 2011; **174**: 1382–1390.
26. Kivimäki M, Vahtera J, Kawachi I, *et al.* Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: an instrumental-variables analysis. *Am J Epidemiol* 2010; **172**: 167–172.
27. Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum* 2006; **54**: 3390–3398.

DOI: 10.1002/pds

28. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. *Can Med Assoc J* 2007; **176**: 627.
29. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol* 2008; **61**: 1285–1288.
30. Julius Clinical Primary Care Research Network. [cited 2014 January, 8]. Available from: http://www.juliusclinical.com/global-presence - cat10.
31. Top Institute Pharma TN. [cited 2014 January, 8]. Available from: http://www.tipharma.com/projects/efficiency-analysis-drug-discovery-process/the-mondriaan-project.html.
32. Stirbu-Wagner I, Dorsman S, Visscher S, *et al*. Landelijk Informatienetwerk Huisartsenzorg. Feiten en cijfers over huisartsenzorg in Nederland [Netherlands Information Network of General Practice. Facts and figures about general practice care the Netherlands]. *Utrecht/Nijmegen: NIVEL/IQ* 2010.
33. ENCePP-. Use of antidepressants and risk of hip/femur fracture. A methodological comparison across data sources and epidemiological design. Electronic Register of Studies, 2014. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/5847;jsessionid=mnbHT9QfbfCJSQLnPl91-LQkW3MhQ61rhRXG9gtx23tT8hvCLzcDF!-1931278576 (Date of access: 16 July 2014).
34. Ali M, Uddin MJ, Groenwold RHH, *et al.* Quantitative falsification of instrumental variables assumption using balance measures. *Epidemiology* 2014; **25**: 770–772.
35. Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med* 2013; **33**: 1685–1699.
36. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; **27**: 2037–2049.
37. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 2001; **15**: 69–85.
38. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; **19**: 2–28.
39. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2010. ISBN:3-900051-07-0.
40. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**: 467–475.
41. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014; **15**: 2297–2340.
42. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006; **17**: 260–267.
43. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; **175**: 60–65.
44. Boef AG, van Paassen J, Arbous MS, *et al.* Physician's preference-based instrumental variable analysis—is it valid and useful in a moderate-sized study? *Epidemiology* 2014. doi:10.1097/EDE.0000000000000151 00.
45. ENCePP. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance 2014 [2014/07/16]. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/3690.
46. Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Saf* 2011; **20**: 1115–1129.
47. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**: 537–544.

SUPPORTING INFORMATION

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