Journal of Clinical Epidemiology 69 (2016) 96–106

Journal of Clinical Epidemiology

The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs

Bianca Kollhorst[a,*], Michal Abrahamowicz[b,c], Iris Pigeot[a,d]

[a]*Leibniz Institute for Prevention Research and Epidemiology—BIPS, Department Biometry and Data Management, Achterstr. 30, Bremen 28359, Germany*
[b]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal H3A 1A2, Québec, Canada*
[c]*Division of Clinical Epidemiology, McGill University Health Centre Royal Victoria Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Québec, Canada*
[d]*Faculty of Mathematics and Computer Science, University of Bremen, Bibliothekstraße 1, Bremen 28359, Germany*

Accepted 24 August 2015; Published online 2 September 2015

Abstract

Objectives: To investigate whether physician's prescribing preference is a valid instrumental variable (IV) for patients' actual prescription of selective cyclooxygenase-2 (COX-2) inhibitors in the German Pharmacoepidemiological Research Database (GePaRD).

Study Design and Setting: We compared the effect of COX-2 inhibitors vs. traditional nonsteroidal anti-inflammatory drugs (tNSAIDs) on the risk of gastrointestinal complications using physician's preference as IV. We used different definitions of physician's preference for COX-2 inhibitors. A retrospective cohort of new users was built which was further restricted to subcohorts. We compared IV-based risk difference estimates, using a two-stage approach, to estimates from conventional multivariate models.

Results: We observed only a small proportion of COX-inhibitor users (3.2%) in our study. All instruments, in the full cohort and in the subcohorts, reduced the imbalance in most of the covariates. However, the IV treatment effect estimates had a highly inflated variance. Compared to the most recent prescription, the proportion of previous patients was a stronger instrument and reduced the variance of the estimates.

Conclusion: The proportion of all previous patients is a potential IV for comparing COX-2 inhibitors vs. tNSAIDs in GePaRD. Our study demonstrates that valid instruments in one health care system may not be directly applicable to others.

Keywords: Confounding by indication; COX-2 inhibitors; German Pharmacoepidemiological Research Database; Instrumental variables; Physician's preference; Two-stage least squares

1. Introduction

Observational studies are necessary to assess the effectiveness and safety of drugs after marketing. Claims databases are frequently used for this purpose. However, because claims data are mainly collected for reimbursement of patients' costs, they lack important confounder information, which in turn leads to biased effect estimates [1]. Instrumental variable (IV) analysis is a "generic" approach to deal with unmeasured confounding [2]. Applications of this method in observational studies of the effectiveness and safety of drugs exploit random variation in treatment assignment to define the IV that influences treatment but does not have an independent effect on the outcome [3]. Using an IV instead of the actual treatment is equivalent to pseudorandomizing the patients to alternative treatments [3]. However, IV analysis can reduce bias in effect estimates due to unmeasured confounding, only if a valid instrument can be identified [4,5]. An observable variable is a valid instrument provided that all three following assumptions are met. First, the IV is associated with the treatment. Second, the IV is independent of unobserved confounders and third, conditionally on unmeasured confounders and treatment, the IV and the outcome are independent, implying that the IV association with the outcome is fully mediated by the observed treatment (exclusion restriction) [6].

Conflict of interest: B.K. is working and I.P. is head of an institute that occasionally performs studies for pharmaceutical industries. The companies include Mundipharma, Bayer-Schering, Stada, Sanofi-Aventis, Sanofi-Pasteur, Novartis, Purdue, Celgene, and GSK, but none of them were involved in this study.

Funding: This research was partly supported by the grant MOP 81275 from the Canadian Institutes for Health Research (CIHR).

* Corresponding author. Tel.: +49-421-2185-6980; fax: +49-421-2185-6941.

E-mail address: kollhorst@bips.uni-bremen.de (B. Kollhorst).

http://dx.doi.org/10.1016/j.jclinepi.2015.08.008

What is new?

- Previous research has not evaluated the feasibility and validity of using prescribing preference-based instrumental variables (IVs) in German administrative health databases. We demonstrated that the proportion of all previous patients, in an individual physician practice, who were prescribed a given drug, meets the criteria for an IV for the patients' actual prescriptions of cyclooxygenase-2 inhibitors or traditional nonsteroidal anti-inflammatory drugs in Germany.
- Although sample size was large and the instruments met the standard criteria for "strong" instruments, the IV-based estimates of treatment effect suffered from a highly inflated variance and varied substantially depending on the definition of the instrument.
- Our study has shown that valid instruments in one health care system may not be directly applicable to other settings. IV assumptions should be carefully checked for each particular research question and for the relevant study population.

In 2006, Brookhart et al. [7] proposed that the physician's prescribing preferences, which can be quantified based on the physician's prior prescriptions, might be used to define a valid IV. They applied this approach to compare the risk of gastrointestinal (GI) complications associated with nonsteroidal anti-inflammatory drugs (NSAIDs) selective for cyclooxygenase-2 (COX-2) vs. traditional NSAIDs (tNSAIDs). Because then several other studies evaluated different definitions of provider prescribing preference-based IVs and gave ambiguous results regarding the "optimal" IV definition. For example, Henessey et al. [8] reported that physician's preference based on the most recent NSAID prescription was a stronger IV than the IV based on several recent prescriptions. In contrast, Ionescu-Ittu et al. [9] found that IVs depending on the proportion of all previous patients, in a given physician's practice, who were prescribed a specific drug were stronger and had smaller variance than estimates based on the most recent prescription. Abrahamowicz et al. [10] adapted this approach to settings where physician's preferences may change over time and demonstrated through simulations that the change-time method reduced the variance of the IV estimates relative to the IV based on physician's prior prescriptions. However, Davies et al. [11] concluded that the physician's preference based on the most recent prescription had weaker associations with observed confounders and, hence, might be expected to be less related to unobserved confounders than IVs based on multiple prescriptions, but the latter led to treatment effect estimates with smaller standard errors. Finally, Rassen et al. [12] increased the strength of their instruments by restricting the cohort to physicians who treated many patients.

It is plausible that these divergent findings are partly because the validity and the relative strengths of alternative definitions of prescribing preference-based IVs depend on both the assessed drug and the characteristics of the health system, especially those related to prescribing habits. In the present study, we aim to identify a valid IV in the German Pharmacoepidemiological Research Database (GePaRD) [13], to compare the risk of GI complications between users of COX-2 inhibitors and tNSAIDs. Thus, we assess the alternative IVs for the same association as the one studied in the original IV article by Brookhart et al. [7] but applied in a different health system context. We consider three definitions of the physician's preference by using, first, the most recent prescription, second, the proportion of previous patients, and third, a set of indicator variables for the physician's seven prior prescriptions. To increase the strength of the IVs and to create subcohorts with lower variation in unmeasured confounders, we restrict the cohort to subgroups that are more homogenous with respect to either patients or physicians characteristics. We then compare the instruments in terms of strength and ability to balance the distributions of observed covariates both in the full cohort and in the subcohorts and explore a possible violation of the exclusion restriction. Furthermore, we compare IV effect estimates, obtained using a two-stage approach [5], with (1) estimates obtained from the conventional analysis that adjusts only for observed covariates and (2) results of randomized controlled trials and previously published database studies.

2. Methods

2.1. Data source

The study was based on claims data (2004–2009), extracted from GePaRD, from four German statutory health insurances (SHIs). The source population consisted of more than 14 million insurance members and is nationally representative with respect to sex, age, and region of residence. Membership in an SHI is compulsory in Germany for employees below an annual income threshold (approximately 49,000€ in 2009). Although individuals with higher incomes may switch to private health insurances, around 75% of them remain voluntary members of SHIs. About 70 million people (85% of the German population) are SHI members, including about five million voluntary members, children, and patients who are retired or unemployed. For each insurance member, the database contains information on demographics as well as on hospital admissions, outpatient physician visits, and prescriptions [13]. The hospital data comprise the dates of hospitalization, diagnoses,

reasons for admission and discharge, as well as diagnostic and therapeutic procedures, together with their respective dates. Claims data on all outpatient physician visits include outpatient treatments, procedures, and diagnoses. All diagnoses are coded according to the German Modification of the International Statistical Classification of Diseases (ICD-10 GM). Prescription data are available for all outpatient prescriptions which are reimbursed by SHIs and include the dates of prescription and drug dispensation at the pharmacy, the amount of substance prescribed (total dose), and a pseudonymous identifier and the specialty of the prescribing physician. Prescription data are linked via the central pharmaceutical reference number to a reference database, which contains information on the anatomic-therapeutic-chemical code, the defined daily dose, packaging size, strength, formulation, generic, and trade name.

In Germany, the utilization of health insurance data for scientific research is regulated by the Code of Social Law. All involved SHIs as well as federal and regional authorities approved the use of the data for this study. Informed consent was not required because the study was based on pseudonymous data.

2.2. Study design and measurements

The study was designed as a retrospective cohort of new NSAIDs users in 2004−2009. NSAIDs are a heterogeneous group of agents that comprises traditional (t) NSAIDs and NSAIDs selective for COX-2, so-called COX-2 inhibitors. Specifically, insurants were included in the cohort if they were ≥65 years, were continuously insured for at least 365 days before their first NSAID prescription, and had no diagnosis of a malignant cancer, except nonmelanoma skin cancer, in this period. Patients with missing information on the prescribing physician were excluded because they could not be included in the IV analyses [7]. Cohort entry was defined as the first prescription of an NSAID during the study period, and patients were followed until either end of insurance time, death, a diagnosis of malignant cancer except nonmelanoma skin cancer, hospitalization for GI complications, end of follow-up (180 days after cohort entry), or end of the study period (December 31, 2009).

The outcome was defined as either hospitalization for upper GI complications using the main discharge diagnosis or an outpatient diagnosis with assured diagnosis certainty of ulcus ventriculi (K25.0-K25.2, K25.4-K25.6), ulcus duodeni (K26.0- K26.2, K26.4-K26.6), ulcus pepticum (K27.0-K27.2, K27.4-K27.6), ulcus pepticum jejuni (K28.0-K28.2, K28.4-K28.6), or gastritis (K29.0-K29.2), at any time during the 180-day follow-up after the first NSAID prescription. The outcomes have not been validated for this study, but in several studies, high positive predictive values (81−100%) have been reported for specific discharge diagnoses for GI complications in the ICD coding system, whereas lower values (53−68%) have been shown for nonspecific diagnoses [14−16].

Comorbid conditions were considered as confounders if they occurred in the 365-day baseline period. Information on relevant conditions was obtained from inpatient and outpatient diagnoses of complicated (with hemorrhage), uncomplicated (without hemorrhage) and other GI disease (including gastroesophageal reflux disease, diverticulosis, esophageal disease), alcohol abuse, cardiovascular disease (including myocardial infarction, chronic ischemic disease, heart failure, stroke, hypertension, atrial fibrillation, flutter, peripheral arterial diseases), diabetes mellitus, rheumatoid arthritis, and osteoarthritis. Comedications such as aspirin, glucocorticoids, nitrates, platelet aggregation inhibitors, anticoagulants, proton pump inhibitors, other gastroprotective drugs (including antacids, H2-receptor antagonists, prostaglandins, other drugs for peptic ulcer), and cardiovascular drugs (including angiotensin-converting enzyme inhibitors, angiotensin II receptor antagonists, beta blockers, calcium channel blockers, diuretics) were assessed in the 120 days before cohort entry. Additionally, we assessed proton pump inhibitors, H2-receptor antagonists, and misoprostol that are prescribed at the day of cohort entry as gastroprotective agents possibly related to the NSAID prescription.

2.3. Actual treatment and definition of IVs

Patients were classified as users of either COX-2 inhibitors or tNSAIDs, depending on the first prescription, at cohort entry. The binary instrument (PP1) was an indicator variable, assigned the value of 1 or 0 if the prescription written to the most recent patient by the same physician was for a COX-2 inhibitor or a tNSAID, respectively (Table 1). The continuous instrument (PP2) was defined as the proportion of all previous patients of the same physician who were prescribed COX-2 inhibitors in the study period. If 2 or more NSAID prescriptions were filled the same day, 1 was randomly picked to determine the prescribing preference.

To increase the strength of both instruments, we restricted the full cohort in the analyses to create three, partly overlapping subcohorts that were more homogenous with respect to either patient or prescribing physician characteristics. This resulted in three separate analyses, each restricted to a specific subcohort (Table 1): patients with a diagnosis of osteoarthritis or rheumatoid arthritis (R1), only patients prescribed by general practitioners (GPs) (R2), and only patients prescribed by physicians who treated at least 20 patients (R3). In a sensitivity analysis, we used a third IV in R3 constructing indicators for each physician's seven prior prescriptions for COX-2 inhibitors as more recent prescriptions are likely to be stronger associated with the true physician's preference.

2.4. Statistical analysis

In all analyses, the association between the treatment (COX-2 inhibitor vs. tNSAIDs) and the probability of upper

Table 1. Description of the instrumental variables and the three subcohorts

Name	Definition
Instrumental variable definition	
PP1	Treatment prescribed to the previous patient of the same physician; binary variable (0 = tNSAID; 1 = COX-2 inhibitor)
PP2	Proportion of all previous patients who are treated with COX-2 inhibitors by the same physician; continuous variable (0–1)
…PP3	Indicator of the number of COX-2 inhibitors prescriptions in each physician's seven previous prescriptions; seven binary variables (0 = tNSAID; 1 = COX-2 inhibitor) (only defined in subcohort R3)
Subcohort based on patient characteristics	
R1	Patients have a diagnosis of rheumatoid arthritis or osteoarthritis
Subcohorts based on physician characteristics	
R2	Physician is a general practitioner (GP)
R3	Physician treated at least 20 patients during the study period

Abbreviations: tNSAID, traditional nonsteroidal anti-inflammatory drug; COX-2, cyclooxygenase-2.

GI complications was quantified by adjusted risk differences, estimated by multivariate linear regression models [5]. We estimated four models: three IV models, which relied on either one of the binary (PP1 or PP3) instruments or the continuous (PP2) instrument (Table 1), and the conventional model, which was adjusted for all measured confounders. All models were adjusted for all a priori selected potential confounders (Section 2.2).

2.4.1. IV analyses

For the IV analyses, a two-stage least square regression was implemented [17]. For the first stage, we fitted a multivariate linear model that predicted for each patient i the probability of receiving a COX-2 inhibitor (T_i=1):

$$Pr(T_i = 1|Z_i, X_{1i}, \ldots, X_{mi}) = \beta_0 + \beta_1 z_i + \sum\nolimits_{j=1}^{m} \beta_j x_{ji} \quad (1)$$

where $x_{1i},\ldots,x_{mi}$, i=1,…,n, denote the values of the covariates and z_i the values of the respective IV.

In the second stage, the probability of GI complications (Y_i=1) was modeled in a multivariate linear model, conditional on (1) the stage 1 estimate of the probability of the COX-2 treatment and (2) all measured covariates, as follows:

$$Pr\left(Y_i = 1|\widehat{Pr}(T_i = 1|Z_i, X_{1i}, \ldots, X_{mi}),\, X_{1i}, \ldots, X_{mi}\right) = \alpha_0 + \alpha_1 \widehat{Pr}(T_i = 1|Z_i, X_{1i}, \ldots, X_{mi}) + \sum_{j=1}^{m} \alpha_j x_{ji} \quad (2)$$

The estimator of α_1 in Equation (2) represents the IV estimate of the adjusted risk difference for the treatment effect of COX-2 inhibitors vs. tNSAIDs.

2.4.2. Conventional analyses

For comparison, we fitted a conventional multivariate linear "risk difference" regression model using ordinary least squares that regressed the probability of the outcome on the actual treatment (T_i) and the same covariates as those adjusted for in the IV model (2). Thus, to increase the comparability of the results, we repeated all conventional analyses on the same subsamples as used for the IV analyses.

For all treatment effect estimates, we provided 95% confidence intervals (95% CIs) using robust White's standard errors, which account for the clustering of patients by physician and for the heteroskedasticity of the residuals [18].

Finally, we applied the Durbin-Wu-Hausman test [19] to test whether there is a difference between the conventional and the IV estimates.

2.4.3. Verification of the IV assumptions

To investigate whether the first IV assumption is satisfied, three measures of the strength of the association between the IV and the actual treatment were calculated based on the first stage model (1), separately for each of the two instruments. The three measures were as follows: the partial F-statistic for the adjusted IV effect [3,12], the squared partial correlation r^2 [3,12], and the estimated effect of the IV on the probability of treatment ($\widehat{\beta}_1$), quantified as the adjusted difference in prevalence (per 100 patients). The partial F-statistic reflects the statistical significance of the IV contribution to the first stage model. Staiger and Stock suggest that an F-statistic greater than 10 indicates that the IV is not weak [16]. The partial r^2, the square of the partial Spearman correlation coefficient, is the proportion of the variance in the actual treatment which is additionally explained by the inclusion of the IV in the model. Larger values indicate that the IV makes a more important contribution to the first-stage model. Finally, the effect of the IV on the treatment ($\widehat{\beta}_1$) estimates the adjusted difference in the probability of exposure to COX-2 inhibitors associated with an unit increase in the value of the respective instrument (from Z_i=0 to Z_i=1). For a binary IV (PP1) for instance, this corresponds to the difference in probability between study subjects who actually receive a COX-2 inhibitor for whom the previous patient got a COX-2 inhibitor (Z_i=1) vs. a tNSAID. Larger values indicate that the respective IV predicts more accurately the actual treatment [12].

Neither the second nor the third IV assumption can be empirically verified, but the plausibility of both can be explored [6]. To check whether the second assumption may be considered as being valid, we regressed each measured covariate on each of the three IVs and on the

Table 2. Comparison of the baseline characteristics of new users of tNSAIDs vs. COX-2 Inhibitors

Variable	tNSAID users (*N* = 681,470) (%)	COX-2 inhibitor users (*N* = 22,368) (%)
Female sex	395,156 (58.0)	13,955 (62.4)
70 years or older	399,776 (58.6)	13,856 (61.9)
Year of NSAID initiation[a]		
2004	489 (0.1)	24 (0.1)
2005	200,435 (29.4)	4,876 (21.8)
2006	160,669 (23.6)	5,085 (22.7)
2007	129,682 (19.0)	5,280 (23.6)
2008	105,743 (15.5)	3,926 (17.6)
2009	84,452 (12.4)	3,177 (14.2)
Comorbid conditions in the 365 days before cohort entry		
Complicated upper GI disease	13,312 (2.0)	668 (3.0)
Uncomplicated upper GI disease	75,326 (11.1)	3,624 (16.2)
Other GI disease	103,124 (15.1)	4,365 (19.5)
Alcohol abuse	8,536 (1.3)	254 (1.1)
Cardiovascular disease	509,845 (74.8)	17,454 (78.0)
Diabetes mellitus	144,606 (21.2)	4,969 (22.2)
Rheumatoid arthritis	22,273 (3.3)	1,273 (5.7)
Osteoarthritis	171,260 (25.1)	6,975 (31.2)
Concomitant medication in the 120 days before cohort entry		
Aspirin	38,618 (5.7)	1,255 (5.6)
Glucocorticoids	23,076 (3.4)	1,235 (5.5)
Nitrates	32,164 (4.7)	1,219 (5.4)
Platelet aggregation inhibitors	51,215 (7.5)	1,891 (8.5)
Anticoagulants	31,139 (4.6)	1,994 (8.9)
Proton pump inhibitors	61,668 (9.0)	3,394 (15.2)
Other gastroprotective agents	9,717 (1.4)	375 (1.7)
Cardiovascular drugs	327,128 (48.0)	11,501 (51.4)
Medication at cohort entry possibly related to the NSAID prescription		
Gastroprotective agents	41,011 (6.0)	1,201 (5.4)
Gastrointestinal complications (outcome)		
Events	6,038 (0.9)	235 (1.1)
Identified by inpatient diagnosis[b]	1,338 (22.2)	57 (24.3)
Identified by outpatient diagnosis[b]	4,700 (77.8)	178 (75.7)

Abbreviations: tNSAID, traditional nonsteroidal anti-inflammatory drug; COX-2, cyclooxygenase-2; GI, gastrointestinal.
[a] Low numbers occur due to the 365-day baseline period.
[b] Calculation of frequencies is based on the number of events.

actual treatment. We calculated and compared the corresponding partial *F*-statistics to assess the ability of the instruments to improve the balance in the distribution of the measured covariates. To ensure the comparability of the covariate balance between the binary vs. the continuous IV, we have chosen the *F*-statistic, as it applies to both analyses while at the same time reflecting the difference between the mean covariate values in the two treatment groups, which is a standard criterion for assessing binary instruments [3,12]. As the third assumption cannot be explored based on the data, we examined the association between the IVs and gastroprotective agents that are prescribed at the same day as the NSAID prescription using a logistic model, adjusted for age, sex, and year of the index NSAID prescription. To account for the clustering by physician, the parameters and standard errors were estimated using a robust generalized estimating equation approach and a working variance–covariance matrix with an exchangeable structure [20].

All statistical analyses were conducted in SAS, version 9.3.

3. Results

We identified 705,166 new NSAID users during the study period. Among these, 1,328 (0.2%) patients had to be excluded due to a missing identifier of the prescribing physician. This resulted in a final sample size for the full cohort of 703,838 patients. Because the prescription of the first patient of each physician was used to assess the NSAID preference, the first patient of each physician was excluded (94,331 patients) resulting in the final sample size of 609,527 patients for the IV cohort (eFigure 1/Appendix A at www.jclinepi.com). The 68,305 physicians included in the IV cohort prescribed an NSAID to nine patients on average (Q1–Q3: 1–11, eTable 1/Appendix G at www.jclinepi.com). The overall proportion of patients in the full cohort with a COX-2 inhibitor as the first NSAID prescription was 3.2% (IV cohort: 3.1%). Most frequently prescribed tNSAIDs were diclofenac (55.4%), ibuprofen (36.5%), piroxicam (1.1%), and others (7.0%). The three most commonly prescribed COX-2 inhibitors were etoricoxib (64.3%), celecoxib (29.5%), and lumiracoxib

Table 3. Assessing the strength of the instrumental variables in the full IV cohort and the three subcohorts

Instrument	*N*	Partial r^2	Partial *F*-statistic	Difference in prevalence of COX-2 per 100 (95% CI)
Previous prescription for COX-2 inhibitor (PP1)				
Full IV cohort	609,527[a]	0.014	8,652	12.0 (11.3–12.7)
R1	134,698	0.012	1,608	11.3 (10.2–12.4)
R2	293,315	0.012	3,985	11.7 (10.8–12.6)
R3	262,489	0.011	2,858	10.4 (9.2–11.7)
Proportion of all previous patients treated with COX-2 inhibitors (PP2)				
Full IV cohort	609,527[a]	0.015	17,105	44.3 (42.3–46.3)
R1	134,698	0.010	2,087	32.7 (29.5–35.9)
R2	293,315	0.016	7,468	41.3 (38.8–43.9)
R3	262,489	0.018	9,049	54.6 (50.3–59.0)
No. of COX-2 inhibitors in the seven previous prescriptions (PP3; only defined in R3)				
	215,665[b]	0.021	7,247	
1				4.6 (4.2–4.9)
2				10.1 (9.1–11.0)
3				18.1 (15.6–20.6)
4				27.7 (23.2–32.3)
5				37.7 (23.8–51.7)
6				69.7 (54.8–84.5)
7				70.8 (50.3–91.3)

Abbreviations: IV, instrumental variable; COX-2, cyclooxygenase-2; CI, confidence interval.

All estimates are adjusted for sex, age, year of index prescription, complicated, uncomplicated and other gastrointestinal disease, alcohol abuse, cardiovascular disease, diabetes mellitus, aspirin, glucocorticoids, nitrates, platelet aggregation inhibitors, anticoagulants, proton pump inhibitors, other gastroprotective drugs, and cardiovascular drugs.

[a] Because of the definition of the IVs, the first patient of each physician is excluded.

[b] Because of the definition of the IV, the first 7 patients of each physician are excluded.

(3.1%). Patients treated with COX-2 inhibitors were older, had more comorbidities, and were more often treated with concomitant medications, especially with gastroprotective drugs (Table 2).

3.1. Instruments' strength

Table 3 compares the values of the three statistics used to assess the instruments' strength in the full cohort and the three subcohorts. In general, the continuous physician's preference (PP2) was a stronger predictor of the actual treatment than the binary instrument (PP1), according to both the partial r^2 and the *F*-statistics. Both instruments met the Staiger and Stock criterion of an *F*-statistic greater than 10, but their high values reflected the very large sample sizes. In contrast, the low values of the partial r^2 indicated that the instruments explained only a very small proportion of the variance in the actual treatment assignments. Furthermore, the strength of the instruments varied across the subcohorts (Table 3). The binary instrument PP1 was weakest in R3, whereas the continuous instrument PP2 was weakest in R2. The PP1 was strongest in the full IV cohort, and PP2 was strongest in the R3, where the relevant proportions could be more accurately estimated. For the full IV cohort, the difference in prevalence for the binary instrument indicated that if the physician previously prescribed a COX-2 inhibitor, the next patient would be about 12% more likely to also be prescribed a COX-2 inhibitor (Table 3). However, the effect of the continuous instrument, estimated in the stage 1 model for R3, indicated that the next patient would be about 55% more likely to be prescribed a COX-2 inhibitor if the proportion of all previous patients of the same physician, who got COX-2 inhibitors, increased 100%. Including indicator variables for the

Table 4. Adjusted association between the instrumental variables and the coprescription of gastroprotective agents at the same day as the index NSAID prescription

	Previous prescription for COX-2 inhibitor (PP1)	Proportion of all previous patients treated with COX-2 inhibitors (PP2)	No. of COX-2 inhibitors in the seven previous prescriptions (PP3)
Cohort variation	Odds ratio[a] (95% CI)[b]	Odds ratio[a] (95% CI)[b]	Odds ratio[a] (95% CI)[b]
Full cohort	1.08 (1.02 – 1.14)	1.49 (1.28 – 1.73)	NA
R1	1.14 (1.03 – 1.27)	1.22 (0.91 – 1.62)	NA
R2	0.94 (0.87 – 1.02)	1.14 (0.93 – 1.38)	NA
R3	1.12 (1.01 – 1.23)	2.79 (2.12 – 3.68)	1.12 (1.08 – 1.15)

Abbreviations: NSAID, nonsteroidal anti-inflammatory drug; COX-2, cyclooxygenase-2; CI, confidence interval; NA, not applicable.

[a] Adjusted for sex, age, and year of index NSAID prescription.

[b] All CIs are clustered by physician.

Table 5. Multivariate conventional and instrumental variable regression estimates of risk differences of gastrointestinal complications per 100 patients for COX-2 inhibitor users compared to users of tNSAIDs

	Conventional analysis			Previous prescription for COX-2 inhibitor (PP1)			
Cohort variation	*N*	No. of events	Risk difference per 100 (95% CI)	*N*	No. of events	Risk difference per 100 (95% CI)	Hausman test *P*-value[a]
FC	703,838	6,273	−0.01 (−0.15, 0.12)	609,527	5,313	0.08 (−0.99, 1.15)	0.8823
R1	192,916	2,027	0.08 (−0.17, 0.32)	134,698	1,382	0.81 (−1.68, 3.31)	0.5799
R2	333,354	3,117	0.00 (−0.18, 0.18)	293,315	2,689	−0.06 (−1.56, 1.44)	0.9697
R3	270,293	2,410	0.08 (−0.16, 0.32)	262,489	2,309	0.46 (−1.58, 2.50)	0.7445

Abbreviations: tNSAID, traditional nonsteroidal anti-inflammatory drug; COX-2, cyclooxygenase-2; CI, confidence interval; FC, full cohort; NA, not applicable.

All estimates are adjusted for sex, age, year of index prescription, complicated, uncomplicated and other gastrointestinal disease, alcohol abuse, cardiovascular disease, diabetes mellitus, aspirin, glucocorticoids, nitrates, platelet aggregation inhibitors, anticoagulants, proton pump inhibitors, other gastroprotective drugs and cardiovascular drugs.

[a] The null hypothesis of the Hausman test is that there is no difference between the conventional and the instrumental variable estimates.

physician's seven previous prescriptions (R3) further increased the strength of the association with the actual prescription compared to the instruments PP1 and PP2 (Table 3).

3.2. Covariate balance

The two actual treatment groups were highly imbalanced for most of the patient- and physician-level covariates (Table 2). In the full IV cohort and all subcohorts, all instruments improved the balance for the patient-level and physician-level covariates (eFigures 2–6/Appendix B–F at www.jclinepi.com). Overall, the binary IVs, which measured physician's preference based on the previous prescription (PP1) and on the previous seven prescriptions (PP3), reduced the imbalance more than the continuous IV (PP2; eTable 2/Appendix H at www.jclinepi.com). Although all IVs reduced the imbalance in the covariates, some residual differences remained, for example, in the year of index prescription, which likely reflected the secular trends in prescribing of COX-2 inhibitors. In general, covariate balance in the patient-level characteristics was more improved than in the physician-level characteristics.

3.3. Exclusion restriction

In Table 4, we present the association of the IVs with a coprescription of gastroprotective agents that is issued at the same day as the index NSAID prescription. Regarding the IV PP1, physicians were more likely to coprescribe a gastroprotective agent with their next NSAID prescription in the full cohort and in R1 and R3 if their previous prescription was for a COX-2 inhibitor. In the full cohort, the estimate implies the odds of the physician to coprescribe a gastroprotective agent is 8% higher if the physician prescribed a COX-2 inhibitor, compared to a tNSAID, to the previous patient. Using the IV based on the proportion of all previous patients, physicians were more likely coprescribe a gastroprotective agent in the full cohort and R3. This also applies to the IV PP3.

3.4. Treatment effect estimates

Table 5 compares the adjusted risk differences, obtained from the conventional and the three IV models across different subcohorts, with positive values indicating higher risk of GI outcomes for COX-2 inhibitors. Unadjusted estimates of the risk differences are presented in eTable 3/Appendix I at www.jclinepi.com. Notice that the conventional analyses were based on a slightly higher number of patients than the IV analyses (Table 4), as the first patient (PP1 and PP2) and the first 7 patients (PP3 based on R3) of each physician had to be excluded from the IV analyses, respectively [7]. The subcohort R1 had the smallest sample size, resulting in wider CIs than for the full cohort and the subcohorts R2 and R3, which imposed restrictions only on the physician characteristics. The conventional analysis provided little evidence of an association of COX-2 inhibitors with GI complications in the full cohort and all subcohorts (Table 4). All estimated risk differences were close to zero with narrow CIs that included zero. In R1, the estimates imply 0.12 (95% CI: −0.03, 0.26) additional events of GI complications per 100 patients who were prescribed COX-2 inhibitors vs. tNSAIDs. In the subcohort R2 (PP1), the full cohort, and the subcohort R3 (PP2), respectively, IV estimates suggest a decreased risk for GI complications associated with COX inhibitors. Nevertheless, IV-based estimates had always much wider CIs, which included in each single case zero (Table 5). This was due to very weak associations between the instrument and the actual treatment as reflected in very low values of the partial r^2 in Table 3. There was no evidence of any difference between the conventional and the IV analysis (Hausman tests: *P*-values between 0.5799 and 0.9745). Thus, we may conclude that the results of the conventional analysis may not be impaired by residual or unmeasured confounding. Although we see minor changes in

Proportion of all previous patients treated with COX-2 inhibitors (PP2)			No. of COX-2 inhibitors in the seven previous prescriptions (PP3)		
Risk difference per 100 (95% CI)	Hausman test *P*-value[a]	*N*	No. of events	Risk difference per 100 (95% CI)	Hausman test *P*-value[a]
−0.02 (−0.80, 0.76)	0.9642	NA	NA	NA	NA
0.16 (−1.95, 2.27)	0.9745	NA	NA	NA	NA
0.05 (−1.05, 1.15)	0.8873	NA	NA	NA	NA
−0.10 (−1.24, 1.04)	0.7158	215,665	1,774	0.38 (−0.94, 1.71)	0.7925

the conventional estimates and their CIs when calculated on the same subsamples as used for the IV analyses (eTable 4/ Appendix J at www.jclinepi.com), the overall conclusions do not change.

4. Discussion

Our conventional analysis revealed no evidence that COX-2 inhibitors or tNSAIDs were associated with GI complications. However, it is well known that these estimates are biased and may thus yield misleading results as it is quite clear that there is unmeasured confounding by indication because COX-2 inhibitors are mostly prescribed to patients at higher risk of GI complications (Table 2). Because of the invalidity of our conventional analyses, it is highly recommended to apply IV estimation in situations where confounding by indication cannot be ruled out. If all assumptions are met, the estimates from IV methods such as the two-stage least squares regression are asymptotically unbiased [17]. Although the IVs were moderately strong and reduced the imbalance in the distribution of most of the measured covariates, especially with respect to the history of GI complications, the IV point estimates varied substantially depending on the definition of the instrument. We observed a highly inflated variance of the IV estimates of treatment effects, higher for the binary than for the continuous instruments, a finding which is consistent with a previous simulation study [4]. In comparison with the conventional analyses, some IV estimates suggest that patients who were prescribed COX-2 inhibitors might have had fewer GI complications. All three instruments yielded imprecise estimates and neither an increased nor a decreased risk for GI complications with COX-2 inhibitors could be ruled out.

A possible explanation for the varying IV point estimates is that the estimates may be biased due to violations of the IV assumptions that would be further amplified by a weak IV [21]. Although the observed risk factors were more balanced across the levels of both instruments, the assumption that the IV is independent of the unobserved confounders could not be confirmed. As the estimates of the physician's preference based on the prescribing history are influenced by the physician's true preference, but also by the types of patients seen by each physician, it is possible that there is a case-mix of patients between physicians with different specialty [22]. Therefore, previous prescriptions of a physician can be related to patient characteristics and can further confound the IV analysis. In the subcohort of GP's patients, covariate balance was highly improved by both instruments, so that a case-mix of patients among GPs is not probable. Furthermore, we evaluated if the IV and the outcome were independent conditionally on the treatment and unobserved confounders. This assumption would for instance be violated if a physician with a preference for COX-2 inhibitors may frequently coprescribe gastroprotective agents. In this case, a protective effect of COX-2 inhibitors might be a result of the concomitant use of a gastroprotective agent. Our results indicate that there might be a violation of the exclusion restriction in the full cohort and the subcohorts R1 (only for PP1) and R3, but this assumption seems to be sufficiently satisfied when restricting the cohort to GP's patients.

Table 6 indicates that our results for the IV based on the previous prescription differ from the results of Brookhart et al. [7] and Davies et al. [23], who used similar instruments to estimate the same association. This may be at least partly explained by differences between study periods. In contrast to our study, these earlier studies were conducted before the withdrawal of rofecoxib in 2004 due to serious adverse effects and the subsequent reduced numbers of prescriptions of COX-2 inhibitors [24]. Indeed, Brookhart et al. [7] reported that COX-2 inhibitors were used more frequently than tNSAID. In our study, a very low proportion of COX-2 inhibitors users

was found, which may suggest that after 2004, German physicians have prescribed these drugs to patients with a higher risk for GI complications only. Moreover, a possible reason for the differences in the results might be that the effect of different medications was compared. Recent meta-analyses found that etoricoxib is less gastroprotective than celecoxib and rofecoxib, whereas diclofenac has a higher risk than, for example, ibuprofen [25,26]. Therefore, the difference in incidence of GI complications in our study might be smaller than in the study of Davies et al., as the proportion of users of diclofenac and etoricoxib was higher in our study. Furthermore, the study populations in Brookhart et al. and Davies et al. differed from our study population. Brookhart et al. included insurants of Medicare enrolled in the Pharmaceutical Assistance Contract for the Elderly which had more comorbidities and used more concomitant medications such as gastroprotective agents (COX-2 inhibitors: 17%, tNSAIDs: 20%) [7] than the patients in our study. Davies et al. [23] excluded patients with prior GI complications in the Clinical Practice Research Datalink (CPRD) that comprises only data from general practices. Furthermore, confounders that were adjusted for by Davies et al. were different from those we considered, which could also affect the comparability of the estimates. The use of gastroprotective agents was higher (COX-2 inhibitors: 47%, tNSAIDs: 25%) than in our study [23]. The CPRD also holds information on body mass index (BMI) and physical activity, so that the potential impact of residual confounding by unmeasured factors might have been reduced in the study by Davies et al., compared to our study. Our results also differ from the results of the VIOXX Gastrointestinal Outcomes Research Trial (VIGOR) [27] and Celecoxib Long-term Arthritis Safety Study (CLASS) [28] randomized controlled trials (Table 5). Only the IV estimate obtained on the subcohort of GP's patients suggests that patients who were prescribed COX-2 inhibitors had fewer GI complications. This can be explained by the fact that IV analyses, if based on relatively weak instruments, may fail to fully adjust for the confounding by indication [4].

To the best of our knowledge, this is the first study that has been conducted so far to investigate the physician's prescribing preference as an IV in GePaRD, but sample size can be a major limitation of the IV analysis as small sample sizes further inflate the variance of the estimates. Accordingly, in the case of weak IV and/or small to moderate sample size, the estimates obtained from the conventional analysis may be closer to the true treatment effect, in terms of smaller mean squared error, than the IV estimates [4]. Recently, Boef at al. [29] derived a formula to approximate the threshold sample size, above which IV estimates are expected to outperform the estimates obtained from the conventional analysis. To calculate the threshold sample size, the bias of the conventional estimate has to be known which will typically not the case because the true effect is unknown. Here, estimates obtained from randomized controlled trials are considered to represent the true effects. As one example, we considered our subcohort of patients with osteoarthritis or rheumatoid arthritis and assumed a risk difference of −0.96 obtained from the CLASS randomized controlled trial as the true effect [28]. Then, given that the IV assumptions are met, the threshold sample size for this specific setting can be calculated as approximately 1,803,000,000 patients which is an incredibly large sample size that typically would not be reached in pharmacoepidemiological studies. However, it becomes obvious that large sample sizes are needed to cope with weak to moderately strong instruments. In our case, our subcohort consisted of 134,698 patients which might partly explain the weak performance of the IV estimates. Additionally, the instruments were only moderately strong which may partly be explained by the very small proportion of COX-2 inhibitor users in our study. This further inflates the variance of the estimates.

Although we adjusted for a large number of risk factors, some probably important potential confounders, including the prescribed daily dose, BMI, and use of over-the-counter medication such as aspirin, were not available in the database and could therefore not be considered in the analysis. Finally, we did not validate the outcomes in our study which may introduce misclassification of the outcome.

Table 6. Comparison of adjusted instrumental variable estimates for physician's previous prescription compared to other observational studies and randomized controlled trials

Study	Patient population	Risk difference per 100 (95% CI)
IV estimate	Patients ≥65 years	0.08 (−0.99, 1.15)
	Patients ≥65 years with osteoarthritis or rheumatoid arthritis	0.81 (−1.68, 3.31)
	Patients ≥65 years treated by GPs	−0.06 (−1.56, 1.44)
	Patients >60 years old treated by GPs with more than 10 patients	0.81 (−0.90, 2.53)
Brookhart et al.	Patients ≥65 years	−1.21 (−2.46, 0.04)
	Patients ≥65 years old with osteoarthritis or rheumatoid arthritis	−1.52 (−3.74, 0.71)
	Patients ≥65 years old treated by GPs	−0.82 (−2.40, 0.75)
Davies et al.	Patients >60 years old treated by GPs with more than 10 patients	−0.46 (−1.07, 0.15)
VIGOR trial	Patients ≥50 years old with rheumatoid arthritis, rofecoxib vs. naproxen	−1.07 (−1.57, −0.57)
CLASS trial	Patients with osteoarthritis or rheumatoid arthritis, celecoxib vs. ibuprofen/diclofenac	−0.96 (−1.74, 0.18)

Abbreviations: CI, confidence interval; IV, instrumental variable; GP, general practitioner.

5. Conclusion

The proportion of all previous patients meets verifiable assumptions for a potential IV for the patients' actual prescriptions of COX-2 inhibitors vs. tNSAID in the GePaRD and, thus, may reduce the impact of unmeasured confounding. However, the instrument is only moderately strong, and the resulting variance inflation makes it difficult to derive robust conclusions about the treatment effect. We demonstrated that restricting the cohort to subgroups defined by patient or physician characteristics increases the strength of the instrument. We found that instruments improved the balance in the distribution of the observed confounders and, hence, may be also expected to be less associated with unobserved confounders. However, the IV estimates and their precision varied depending on the definition of the instrument, which may be partly due to the violation of the exclusion assumption, and also depending on the patient subcohort so that results should be interpreted with caution.

Acknowledgment

M.A. is a James McGill Professor of Biostatistics at McGill University. The authors are grateful to the SHIs, namely AOK Bremen/Bremerhaven, DAK-Gesundheit, HKK, and Techniker Krankenkasse, for providing data for this study.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.08.008.

References

[1] Walker AM. Confounding by indication. Epidemiology 1996;7:335−6.
[2] Greenland S. An introduction to instrumental variables for epidemiologists. Int Med Epidemiol 2000;29:722−9.
[3] Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. J Clin Epidemiol 2009;62:1226−32.
[4] Ionescu-Ittu R, Delaney JA, Abrahamowicz M. Bias-variance tradeoff in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. Pharmacoepidemiol Drug Saf 2009;18:562−71.
[5] Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17:260−7.
[6] Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res 2007;16:309−30.
[7] 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;17:268−75.
[8] 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;61:1285−8.
[9] 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;65:155−62.
[10] Abrahamowicz M, Beauchamp ME, Ionescu-Ittu R, Delaney JA, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. Am J Epidemiol 2011;174:494−502.
[11] 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;66:1386−96.
[12] 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;62:1233−41.
[13] Pigeot I, Ahrens W. Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf 2008;17:215−23.
[14] Castellsague J, Holick CN, Hoffman CC, Gimeno V, Stang MR, Perez-Gutthann S. Risk of upper gastrointestinal complications associated with cyclooxygenase-2 selective and nonselective nonsteroidal antiinflammatory drugs. Pharmacotherapy 2009;29:1397−407.
[15] Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology 1996;7:101−4.
[16] Cattaruzzi C, Troncon MG, Agostinis L, Garcia Rodriguez LA. Positive predictive value of ICD-9th codes for upper gastrointestinal bleeding and perforation in the Sistema Informativo Sanitario Regionale database. J Clin Epidemiol 1999;52:499−502.
[17] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91:444−55.
[18] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817−38.
[19] Hausman JA. Specification tests in econometrics. Econometrica 1978;46:1251−71.
[20] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988;44:1049−60.
[21] Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology 2006;17:360−72.
[22] Swanson SA, Miller M, Robins JM, Hernan MA. Definition and evaluation of the monotonicity condition for preference-based instruments. Epidemiology 2015;26:414−20.
[23] 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;24:352−62.
[24] Schussel K, Schulz M. Prescribing of COX-2 inhibitors in Germany after safety warnings and market withdrawals. Pharmazie 2006;61:878−86.
[25] Coxib and traditional NSAID Trialists' (CNT) Collaboration, Bhala N, Emberson J, Merhi A, Abramson S, Arber N, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. Lancet 2013;382:769−79.
[26] Castellsague J, Riera-Guardia N, Calingaert B, Varas-Lorenzo C, Fourrier-Reglat A, Nicotra F, et al. Individual NSAIDs and upper gastrointestinal complications: a systematic review and meta-analysis of observational studies (the SOS project). Drug Saf 2012;35:1127−46.

[27] Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, et al. Comparison of upper gastrointestinal toxicity of Rofecoxib and Naproxen in patients with rheumatoid arthritis. N Engl J Med 2000;343:1520−8.

[28] Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. JAMA 2000; 284:1247−55.

[29] Boef AGC, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. J Clin Epidemiol 2014;67:1258−64.