Eur J Clin Pharmacol (2013) 69:549–557
DOI 10.1007/s00228-012-1334-2

PHARMACOEPIDEMIOLOGY AND PRESCRIPTION

High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications

E. Garbe · S. Kloss · M. Suling · I. Pigeot · S. Schneeweiss

Received: 28 February 2012 / Accepted: 6 June 2012 / Published online: 5 July 2012

Abstract
Purpose High-dimensional propensity score (hd-PS) adjustment has been proposed as a tool to improve control for confounding in pharmacoepidemiological studies using longitudinal claims databases. We investigated whether hd-PS matching improved confounding by indication in a study of Cox-2 inhibitors (coxibs) and traditional nonsteroidal anti-inflammatory drugs (tNSAIDs) and their association with the risk of upper gastrointestinal complications (UGIC).
Methods In a cohort study of new users of coxibs and tNSAIDs we compared the effectiveness of these drugs to reduce UGIC using hd-PS matching and conventional propensity score (PS) matching in the German Pharmacoepidemiological Research Database.
Results The unadjusted rate ratio (RR) of UGIC for coxib users versus tNSAID users was 1.21 [95 % confidence interval (CI) 0.91–1.61]. The conventional PS matched cohort based on 79 investigator-identified covariates resulted in a RR of 0.84 (0.56–1.26). The use of the hd-PS algorithm based on 900 empirical covariates further decreased the RR to 0.62 (0.43–0.91).
Conclusions A comparison of hd-PS matching versus conventional PS matching resulted in improved point estimates for studying an intended treatment effect of coxibs versus tNSAIDs when benchmarked against results from randomized controlled trials.

Keywords NSAIDs · Upper gastrointestinal complications · Confounding by indication · Propensity score · High-dimensional propensity score

E. Garbe (✉) · S. Kloss · M. Suling · I. Pigeot
Department of Clinical Epidemiology, BIPS–Institute for Epidemiology and Prevention Research,
Achterstr. 30,
28359 Bremen, Germany
e-mail: garbe@bips.uni-bremen.de

E. Garbe
Faculty of Human and Health Sciences, University of Bremen,
Bremen, Germany

S. Schneeweiss
Division of Pharmacoepidemiology, Department of Medicine,
Brigham and Women's Hospital and Harvard Medical School,
Boston, MA, USA

Introduction

Electronic healthcare databases have become an increasingly important tool by which to study the safety and effectiveness of drugs and vaccines [1]. Such secondary data reflect medical care as it is practiced outside of controlled research environments, unlike most randomized controlled trials (RCTs) [2, 3]. These databases are of sufficient size to study rare health outcomes, and they do not suffer from the delays that result from primary data collection. However, since these data are not ascertained for research purposes, relevant confounder information may remain unobserved or incompletely measured. This is particularly problematic when studying intended treatment effects where treatment selection factors are often strong, as is the case in the selection of coxibs based on their upper gastrointestinal (GI) protective effects [4]. Comparative effectiveness research studies based on electronic healthcare databases have therefore been criticized to suffer from residual confounding [5].

Even though certain confounders may not be directly recorded in electronic healthcare databases, these databases contain considerably more information than is usually included in pharmacoepidemiology database studies. Some of this information may be used as a surrogate for unobserved

confounder information, but it is unknown which observable surrogate information is relevant for the control of confounding in a specific study. A close correlation of one or more surrogate variables to an unobserved or imperfectly measured confounder is needed to achieve control for confounding by this means [6]. This challenge to identify the relevant surrogate information meants that the wealth of information contained in large healthcare utilization databases has remained largely unused.

Recently, high-dimensional propensity score (hd-PS) adjustment has been proposed as an approach to empirically identify relevant surrogate information for improved confounder adjustment [7, 8]. The approach was illustrated with the examination of several well-known drug effects, including the effect of non-selective traditional nonsteroidal anti-inflammatory drugs (tNSAIDs) and Cox-2 selective NSAIDS (coxibs) on the risk of upper GI bleeding in Medicare databases.

The example studies introducing hd-PS adjustment were based on an empirical comparison of the new and traditional PS method, but they lacked a gold standard. Further validation of the algorithm based on other data sources has been proposed in order to achieve a better understanding of its value in comparative effectiveness research [9]. In the study reported here, we sought to further evaluate the hd-PS algorithm in a large German population-based healthcare utilization database by investigating the risk of upper GI complications (UGIC) associated with tNSAIDs and coxibs.

Methods

Data source

We used information from the German Pharmacoepidemiological Research Database (GePaRD), which consists of claims data from four German statutory health insurance (SHI) providers and includes more than 14 million enrollees covering all regions in Germany. The study was conducted with data collected between 2004 and 2008, since more recent data were not available at the time of the analyses. This longitudinal patient-level database contains demographic information and information on hospital discharges, physician visits, and outpatient prescription dispensings. The hospital data include admission and discharge dates, diagnoses associated with the admission, all discharge diagnoses, and diagnostic and therapeutic procedures with their respective dates. Claims of outpatient physician visits include codes for all procedures and diagnoses. Since physician visits are reimbursed on a quarterly basis, diagnoses associated with a visit can only be attributed to a calendar quarter and not an exact date. All diagnoses are coded according to the German modification of the International Statistical Classification of Diseases and Related Health Problems (ICD-10 GM). Prescription drug information is recorded for all outpatient dispensings that are reimbursable by the SHI providers and includes the date of prescription, the date of the dispensing, the quantity, strength, formulation, generic and trade name, the anatomical–therapeutic–chemical (ATC) code, the defined daily dose (DDD), as well as information on the prescribing physician and the physician's specialty. Age and sex distributions, the number of hospital admissions, and the quantity of drug use have been shown to be representative of the German general population [10–12].

In Germany, the utilization of health insurance data for scientific research is regulated by the Code of Social Law (SGB X). This study was conducted with permission from the Federal Ministry of Health according to this code. The study was based on pseudonymous data.

Study design

This was a cohort study in which a new user design was applied. The aim was to estimate the effect of tNSAIDs and coxibs on the risk of UGIC. New users of tNSAIDs or coxibs were defined as patients who were continuously enrolled in their SHI provider for at least 12 months without any notation of NSAID use, including coxibs, during this time period. Cohort entry was the first notation of a prescription for a tNSAID or a coxib. Cohort exit was defined as discontinuation or switch of the initial NSAID, disenrollment from the SHI provider, hospitalization for UGIC, hospital diagnosis of cancer, death, or the end of the study period, whichever came first. Patients were required to be at least 16 years of age at the time of first use and not to have a diagnosis of cancer in the 12 months preceding cohort entry.

Exposure

The tNSAIDs included aceclofenac (ATC code M01AB16), acemetacin (M01AB11), azapropazone (M01AX04), butylpyrazolidines (M01AA), dexibuprofen (M01AE14), dexketoprofen (M01AE17), diclofenac (M01AB05), diclofenac+ combination (M01AB55), fenamates (M01AG), flurbiprofen (M01AE09), glucosamine (M01AX05), ibuprofen (M01AE01), indomethacin (M01AB01), indomethacin and combination (M01AB05), kebuzone (M01AA06), ketoprofen (M01AE03), lonazolac (M01AB09), lornoxicam (M01AC05), mefenamic acid (M01AG01), meloxicam (M01AC06), mofebutazone (M01AA02), nabumetone (M01AX01), naproxen (M01AE02), oxaceprol (M01AX24), oxaprozine (M01AE12), phenylbutazone (M01AA01), piroxicam (M01AC01), proglumetacine (M01AB14), tenoxicam (M01AC02), and tiaprofenic acid (M01AE11). The coxibs included celecoxib (M01AH01),

etoricoxib (M01AH05), lumiracoxib (M01AH06), parecoxib (M01AH04), and valdecoxib (M01AH03).

Definition of UGIC

Upper gastrointestinal complications were defined as hemorrhage, perforation, or obstruction located in the stomach, duodenum, or gastrojejunal part of the GI tract. The following ICD-10 codes included in the subdivisions hemorrhage and perforation were ascertained for the outcome: gastric ulcer (K25), duodenal ulcer (K26), peptic ulcer (K27), gastrojejunal ulcer (K28), hemorrhage of anus and rectum (K62.5), hematemesis (K92.0), melena (K92.1), and GI hemorrhage unspecified (K92.2). High positive predictive values of site- and lesion-specific codes (between 80 and 97 %) and somewhat lower predictive values of non-specific codes (between 57 and 70 %) have been reported for GI ulcers and complications in the ICD coding system in several studies [13, 14].

Covariates included in the PS model

The conventional PS model included 79 covariates that were considered to be risk factors for the outcome or to possibly influence physician prescribing of coxibs or tNSAIDs. Comorbidities were ascertained in the 6 months before cohort entry. With the exception of medications which inhibited cytochrome P450 (CYP) 2C9, each drug was considered as currently used (exposed at the date of cohort entry or within 10 days before cohort entry) or recently used (exposure ending within 11 days and 6 months before cohort entry). For CYP2C9 inhibitors, only current use was considered.

The following diagnoses and covariates were considered in the PS model (ICD-10-GM codes available upon request): age at cohort entry, sex, year, history of gastric or duodenal ulcer or UGIC, esophagitis and esophageal disease, noninfective gastroenteritis and colitis, coagulation disorders, alcohol abuse, chronic liver disease, kidney failure, osteoarthritis, rheumatoid arthritis, ischemic heart disease, acute myocardial infarction, old myocardial infarction, hypertension, heart failure, myocarditis, pericarditis, hyperlipidemia, cardiomyopathies, valvular heart disease and endocarditis, cardiac conduction disorders and cardiac arrest, atrial fibrillation and flutter, transitory ischemic attack, stroke, peripheral arterial disease, arterial thrombosis and embolism, diabetes mellitus, obesity, chronic respiratory disease, and iron deficiency anemia.

The PS model included the following medications (ATC codes available upon request): proton pump inhibitors, H_2-receptor antagonists, other drugs for acid-related disorders (e.g., prostaglandins or sucralfate), combination treatment for *Helicobacter pylori* eradication, antithrombotic agents, including vitamin K antagonists and heparins, platelet aggregation inhibitors, glucocorticoids, nitrates, calcium channel blockers, angiotensin-converting enzyme (ACE) inhibitors, angiotensin II antagonists, beta blocking agents, diuretics, other antihypertensive drugs, combination antihypertensive drugs, cardiac glycosides, bisphosphonates, sulfonylureas, selective serotonin reuptake inhibitors, serotonin–noradrenaline reuptake inhibitors, CYP2C9 inhibitors, methotrexate, oral contraceptives, postmenopausal hormones, and bile acid sequestrants.

Statistical analysis

We used a Poisson regression analysis to estimate the rate ratio (RR) of UGIC for coxib initiation versus tNSAID initiation and its 95 % confidence interval (CI). A conventional approach and the hd-PS approach were used to estimate the PS. The PS was estimated as the probability of initiating a coxib in a logistic regression model. In the conventional approach, the PS was estimated via a logistic regression model for coxib initiation that included all 79 pre-specified covariates described above, which were ascertained during the 6-month period before cohort entry.

For the hd-PS algorithm version 1, we used the SAS (SAS Institute, Cary, NC) program (available at: www.drugepi.org). Briefly, the hd-PS algorithm (1) requires the identification of the different data dimensions (e.g., hospitalization data, outpatient care data, outpatient drug dispensation data) in the database, (2) identifies the top *n* most prevalent codes (e.g., ICD codes, ATC codes) in each data dimension as candidate covariates, (3) ranks candidate covariates based on their recurrence (the frequency that the codes are recorded for each individual during the baseline period), (4) ranks covariates across all data dimensions by their potential for control of confounding based on the bivariate associations of each covariate with the treatment and with the outcome, (5) selects a pre-specified number of covariates from step 4 (e.g., 500) for PS modeling, and (6) estimates the propensity score with multivariable logistic regression using the selected covariates plus any pre-specified covariates.

We structured the data collected from the 6-month period before cohort entry as six datasets corresponding to the following data dimensions: (1) inpatient diagnoses (ICD-10 GM); (2) outpatient diagnoses (ICD-10 GM); (3) inpatient procedures [OPS (Operationen- und Prozedurenschlüssel) codes, i.e., the German modification of the International Classification of Procedures in Medicine]; (4) outpatient procedures [EBM (Einheitlicher Bewertungsmaßstab) codes, i.e., claim codes for outpatient services and procedures]; (5) outpatient surgery (OPS codes); (6) outpatient drug use (ATC codes). Inpatient and outpatient diagnoses were based on three

digits of the ICD-10-GM codes, OPS procedure codes were based on four digits, and EBM codes wee based on five digits. The full ATC code was used to identify a single drug substance.

We calculated the hd-PS for each patient four times by including 200, 500, 700, or 900 empirically identified and prioritized covariates into a logistic regression model of treatment choice. Each model additionally included age, sex, calendar year, and the number of distinct drugs during the last 6 months before cohort entry as a proxy for healthcare intensity.

Initiators of coxibs were 1:1 matched to initiators of tNSAIDs by the conventional PS or the hd-PS derived from each of the four models using a different number of empirically identified covariates. PS matching was conducted using nearest neighbor matching. Methods for confounder adjustment with the PS are displayed in the text box.

Methods for Confounder Adjustment with a Propensity Score (PS)

Regression adjustment	The PS is used as a covariable in an outcome regression model to adjust the association of treatment effect on study outcome. This approach assumes that exposed subjects and controls with the same PS have the same distribution of baseline characteristics and that the functional relationship between propensity score and outcome is correctly specified.
Matching	The PS is used to match exposed subjects to unexposed subjects with similar values of the PS. This method assumes that within the matched sample, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Stratification	The PS is used to stratify subjects into (often quintiles or deciles) strata. Treatment effects are estimated separately within each stratum and then combined into an overall estimate of treatment effect. This method assumes that within each stratum, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Inverse Probability Weighting	The PS is used to create weights based on the inverse probability which is defined as: E*/PS + (1-E)/(1-PS). This assumes that baseline characteristics are similar in the exposed and unexposed group.

* E: exposure

The RR of UGIC associated with coxibs or tNSAIDs were calculated with Poisson regression analysis in the full cohort and in each of the PS and hd-PS 1:1 matched cohorts, with further adjustment for age, sex, and calendar year in the Poisson regression outcome model [15].

Results

We identified 2,141,895 new users of NSAIDs between 2005 and 2008 (Fig. 1). Of these, 37,281 were coxib users and 2,104,614 tNSAID users. The most frequently used coxib was etoricoxib with 62.0 % followed by celecoxib (30.4 %) and valdecoxib (4.8 %). Among tNSAIDs, diclofenac was with 51.4 % most frequently used, followed by ibuprofen (41.7 %) and dexketoprofen (1.5 %). The average number of days (SD) on treatment was 31.8 (19.8) days in the tNSAID cohort and 40.1 (42.7) days in the coxib cohort.

A total of 1,670 patients suffered an UGIC in the tNSAID cohort and 48 in the coxib cohort, resulting in an unadjusted RR of 1.21 (95 % CI 0.91–1.61) for coxib versus tNSAID initiators. The conventional and hd-PS matched cohort study included 37,281 coxib and 37,281 tNSAID initiators. The distribution of covariates in the unmatched and in the PS matched cohort is shown in (Table 1). The conventional

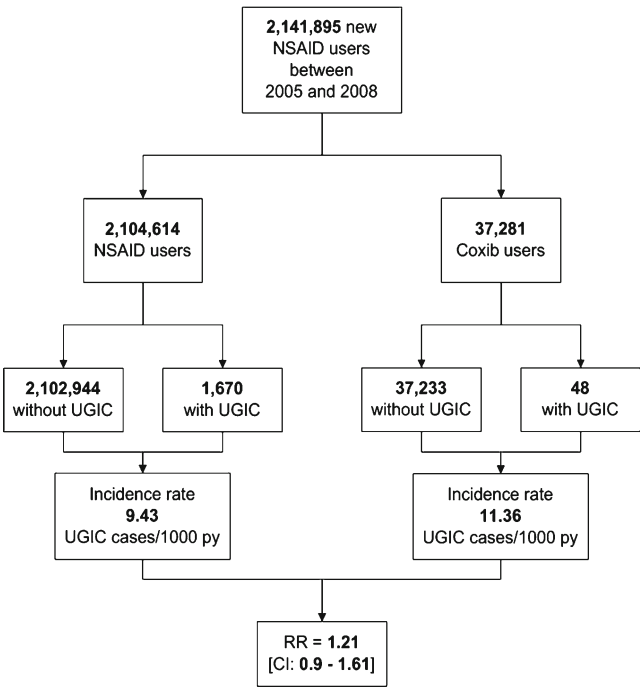

Fig. 1 Patient flow chart for the study cohort and incidence rates of upper gastrointestinal complications (*UGIC*). *NSAIDs* Nonsteroidal anti-inflammatory drugs, *Coxib* Cox-2 selective NSAIDS, *RR* rate ratio, *CI* confidence interval

Table 1 Comparison of the distribution of the covariates included in the propensity score (PS) model before and after PS matching for traditional nonsteroidal anti-inflammatory drug (tNSAID) users and Cox-2 selective NSAIDS

Covariates	Unmatched			Propensity score matched		
	tNSAID initiators (*n*=2,104,614)	Coxib initiators (*n*=37,281)	Difference	tNSAID initiators (*n*=37,281)	Coxib initiators (*n*=37,281)	Difference
Age (mean years & SD)	45.3 (16.9)	52.7 (15.6)	−7.4	53.0 (15.6)	52.7 (15.6)	0.3
Number of distinct drugs (mean & SD)	3.4 (3.0)	4.4 (3.9)	−1	4.4 (3.9)	4.4 (3.9)	0
Gender (female)	1,134,617 (53.9)	22,349 (59.9)	−6	22,490 (60.3)	22,349 (59.9)	0.4
Arterial embolism and thrombosis	4,550 (0.2)	216 (0.6)	−0..4	192 (0.5)	216 (0.6)	−0.1
Arterial fibrillation and flutter	24,360 (1.2)	1,010 (2.7)	−1.5	992 (2.7)	1,010 (2.7)	0
Alcohol abuse	19,357 (0.9)	395 (1.1)	−0.2	414 (1.1)	395 (1.1)	0
Cardiac conduction disorders and arrest	81,434 (3.9)	2,231 (6.0)	−2.1	2,329 (6.2)	2,231 (6.0)	0.2
Cardiomyopathies	8,230 (0.4)	228 (0.6)	−0.2	238 (0.6)	228 (0.6)	0
Chronic respiratory disease	230,984 (11.0)	4,706 (12.6)	−1.6	4,870 (13.1)	4,706 (12.6)	0.5
Chronic liver disease	100,057 (4.8)	2,366 (6.3)	−1.5	2,471 (6.6)	2,366 (6.3)	0.3
Coagulation disorders	13,736 (0.7)	539 (1.4)	−0.7	432 (1.2)	539 (1.4)	−0.2
Diabetes mellitus	129,093 (6.1)	3,320 (8.9)	−2.8	3,322 (8.9)	3,320 (8.9)	0
Esophageal disease	56,027 (2.7)	2,116 (5.7)	−3	2,086 (5.6)	2,116 (5.7)	−0.1
Heart failure	38,155 (1.8)	1,241 (3.3)	−1.5	1,183 (3.2)	1,241 (3.3)	−0.1
Hyperlipidemia	273,677 (13.0)	6,718 (18.0)	−5	6,889 (18.5)	6,718 (18.0)	0.5
Hypertension	422,327 (20.1)	10,787 (28.9)	−8.8	10,929 (29.3)	10,787 (28.9)	0.4
GI inflammatory	81,812 (3.9)	1,704 (4.6)	−0.7	1,539 (4.1)	1,704 (4.6)	−0.5
Iron deficiency anemia	25,934 (1.2)	609 (1.6)	−0.4	582 (1.6)	609 (1.6)	0
Ischemic heart disease	105,577 (5.0)	3,051 (8.2)	−3.2	3,095 (8.3)	3051 (8.2)	0.1
Kidney failure	23,597 (1.1)	672 (1.8)	−0.7	718 (1.9)	672 (1.8)	0.1
Acute MI	16,915 (0.8)	443 (1.2)	−0.4	440 (1.2)	443 (1.2)	0
Myocarditis and pericarditis	3,993 (0.2)	87 (0.2)	0	98 (0.3)	87 (0.2)	0.1
Obesity	126,607 (6.0)	2,750 (7.4)	−1.4	2,722 (7.3)	2,750 (7.4)	−0.1
Past MI	3,407 (0.2)	83 (0.2)	0	89 (0.2)	83 (0.2)	0
Osteoarthritis	82,286 (3.9)	2,911 (7.8)	−3.9	2,947 (7.9)	2,911 (7.8)	0.1
Peripheral arterial disease	45,914 (2.2)	1,368 (3.7)	−1.5	1,360 (3.6)	1,368 (3.7)	−0.1
Rheumatoid arthritis and inflammatory polyarthritis	83,960 (4.0)	3,334 (8.9)	−4.9	3,056 (8.2)	3,334 (8.9)	−0.7
Stroke	40,094 (1.9)	1,073 (2.9)	−1	1,190 (3.2)	1,073 (2.9)	0.3
TIA	4,988 (0.2)	140 (0.4)	−0.2	146 (0.4)	140 (0.4)	0
Past DU, GU, UGIC	8,128 (0.4)	374 (1.0)	−0.6	292 (0.8)	374 (1.0)	−0.2
Valvular disease and endocarditis	37,092 (1.8)	1,099 (2.9)	−1.1	1,047 (2.8)	1,099 (2.9)	−0.1
ACE inhibitors (past use)	116,422 (5.5)	2,787 (7.5)	−2	2,775 (7.4)	2,787 (7.5)	−0.1
ACE inhibitors (current use)	29,339 (1.4)	621 (1.7)	−0.3	665 (1.8)	621 (1.7)	0.1
Anticoagulants (past use)	34,274 (1.6)	1,624 (4.4)	−2.8	1,486 (4.0)	1,624 (4.4)	−0.4
Anticoagulants (current use)	38,687 (1.8)	921 (2.5)	−0.7	974 (2.6)	921 (2.5)	0.1
ATII antagonists (past use)	37,769 (1.8)	1,265 (3.4)	−1.6	1,133 (3.0)	1,265 (3.4)	−0.4
ATII antagonists (current use)	8,357 (0.4)	303 (0.8)	−0.4	248 (0.7)	303 (0.8)	−0.1
Beta blocker (past use)	184,652 (8.8)	4,749 (12.7)	−3.9	4,930 (13.2)	4,749 (12.7)	0.5
Beta blocker (current use)	45,878 (2.2)	1,131 (3.0)	−0.8	1,138 (3.1)	1,131 (3.0)	0.1
Bile-acid-sequestants (past use)	591 (0.0)	27 (0.1)	−0.1	21 (0.1)	27 (0.1)	0
Bile-acid-sequestants (current use)	130 (0.0)	5 (0.0)	0	5 (0.0)	5 (0.0)	0
Bisphosphonates (past use)	10,055 (0.5)	472 (1.3)	−0.8	446 (1.2)	472 (1.3)	−0.1
Bisphosphonates (current use)	2,856 (0.1)	173 (0.5)	−0.4	141 (0.4)	173 (0.5)	−0.1
Calcium channel blocker (past use)	81,025 (3.8)	2,227 (6.0)	−2.2	2,271 (6.1)	2,227 (6.0)	0.1
Calcium channel blocker (current use)	21,147 (1.0)	566 (1.5)	−0.5	604 (1.6)	566 (1.5)	0.1
Cardiac glycosides (past use)	16,320 (0.8)	618 (1.7)	−0.9	638 (1.7)	618 (1.7)	0
Cardiac glycosides (current use)	3,193 (0.2)	111 (0.3)	−0.1	126 (0.3)	111 (0.3)	0

Table 1 (continued)

Covariates	Unmatched			Propensity score matched		
	tNSAID initiators (*n*=2,104,614)	Coxib initiators (*n*=37,281)	Difference	tNSAID initiators (*n*=37,281)	Coxib initiators (*n*=37,281)	Difference
Combined hypertension drugs (past use)	119,491 (5.7)	3,302 (8.9)	−3.2	3,413 (9.2)	3,302 (8.9)	0.3
Combined hypertension drugs (current use)	30,639 (1.5)	800 (2.1)	−0.6	782 (2.1)	800 (2.1)	0
Diuretics (past use)	85,838 (4.1)	2,474 (6.6)	−2.5	2,443 (6.6)	2,474 (6.6)	0
Diuretics (current use)	20,728 (1.0)	604 (1.6)	−0.6	577 (1.5)	604 (1.6)	−0.1
Other drugs for acid related disorders (past use)	1,833 (0.1)	75 (0.2)	−0.1	63 (0.2)	75 (0.2)	0
Other drugs for acid related disorders (current use)	490 (0.0)	18 (0.0)	0	10 (0.0)	18 (0.0)	0
Glucocorticoids (past use)	47,759 (2.3)	1,770 (4.7)	−2.4	1,602 (4.3)	1,770 (4.7)	−0.4
Glucocorticoids (current use)	34,317 (1.6)	1,007 (2.7)	−1.1	793 (2.1)	1,007 (2.7)	−0.6
H2 antagonists (past use)	16,519 (0.8)	480 (1.3)	−0.5	524 (1.4)	480 (1.3)	0.1
H2 antagonists (current use)	13,852 (0.7)	169 (0.5)	0.2	304 (0.8)	169 (0.5)	0.3
Helicobacter pylori eradication (past)	2,152 (0.1)	83 (0.2)	−0.1	73 (0.2)	83 (0.2)	0
Helicobacter pylori eradication (current)	180 (0.0)	12 (0.0)	0	4 (0.0)	12 (0.0)	0
Methotrexate (past use)	2,782 (0.1)	241 (0.6)	−0.5	170 (0.5)	241 (0.6)	−0.1
Methotrexate (current use)	900 (0.0)	132 (0.4)	−0.4	49 (0.1)	132 (0.4)	−0.3
Nitrates (past use)	23,036 (1.1)	744 (2.0)	−0.9	696 (1.9)	744 (2.0)	−0.1
Nitrates (current use)	5,526 (0.3)	176 (0.5)	−0.2	170 (0.5)	176 (0.5)	0
Oral contraceptives (past use)	49,716 (2.4)	289 (0.8)	1.6	269 (0.7)	289 (0.8)	−0.1
Oral contraceptives (current use)	7,875 (0.4)	33 (0.1)	0.3	50 (0.1)	33 (0.1)	0
Other hypertension drugs (past use)	15,882 (0.8)	507 (1.4)	−0.6	480 (1.3)	507 (1.4)	−0.1
Other hypertension drugs (current use)	4,121 (0.2)	133 (0.4)	−0.2	121 (0.3)	133 (0.4)	−0.1
Platelet aggregation inhibitor (past use)	40,603 (1.9)	1,114 (3.0)	−1.1	1,126 (3.0)	1114 (3.0)	0
Platelet aggregation inhibitor (current use)	8,381 (0.4)	217 (0.6)	−0.2	237 (0.6)	217 (0.6)	0
Postmenopausal hormones (past use)	47,767 (2.3)	1,381 (3.7)	−1.4	1,409 (3.8)	1,381 (3.7)	0.1
Postmenopausal hormones (current use)	6,606 (0.3)	179 (0.5)	−0.2	172 (0.5)	179 (0.5)	0
Proton pump inhibitor (past use)	88,968 (4.2)	3,789 (10.2)	−6	3,534 (9.5)	3,789 (10.2)	−0.7
Proton pump inhibitor (current use)	77,418 (3.7)	1,879 (5.0)	−1.3	2,087 (5.6)	1,879 (5.0)	0.6
SNRIs (past use)	7,035 (0.3)	193 (0.5)	−0.2	200 (0.5)	193 (0.5)	0
SNRIS (current use)	1,444 (0.1)	49 (0.1)	0	31 (0.1)	49 (0.1)	0
SSRIs (past use)	28,881 (1.4)	674 (1.8)	−0.4	654 (1.8)	674 (1.8)	0
SSRIs (current use)	5,371 (0.3)	174 (0.5)	−0.2	134 (0.4)	174 (0.5)	−0.1
Sulfunylureas (past use)	21,199 (1.0)	511 (1.4)	−0.4	548 (1.5)	511 (1.4)	0.1
Sulfunylureas (current use)	4,970 (0.2)	127 (0.3)	−0.1	117 (0.3)	127 (0.3)	0
CYP2C9 inhibitor (current use)	16,763 (0.8)	248 (0.7)	0.1	337 (0.9)	248 (0.7)	0.2

GI, Gastrointestinal; MI, myocardial infarction; SD, standard deviation; TIA, transient ischemic attack; UGIC, upper gastrointestinal complication; ACE, angiotensin-converting enzyme; SNRIs, serotonin–norepinephrine reuptake inhibitors; SSRIs, selective serotonin reuptake inhibitors; CYP, cytochrome P450

PS resulted in a RR of 0.84 (95 % CI 0.56–1.26). Using the hd-PS derived from models with 200, 500, 700, or 900 empirically identified covariates, we obtained a RR of 0.68 (95 % CI 0.49–1.01), 0.67 (0.49–0.97), 0.64 (0.44–0.94), and 0.62 (0.43–0.91), respectively (Fig. 2).

Discussion

Coxibs were developed with the intention to improve the GI tolerability of NSAIDs while maintaining their analgetic and anti-inflammatory efficacy. Large RCTs and meta-analyses of RCTs have shown a 50–60 % lower risk of upper GI bleeding among coxib initiators compared with tNSAID initiators [16, 17]. Because of this intended treatment effect of coxibs, they are preferentially prescribed to patients with existing GI problems or complications [18–21], and it is not surprising that our unadjusted analysis showed a 21 % increased risk of UGIC for coxibs due to confounding bias. Such confounding by indication has also been shown in other observational studies comparing coxibs and tNSAIDs [22, 23].

The results of the conventional PS matched cohort analysis showed a numerical risk reduction of UGIC by 16 % for initiators of coxibs versus tNSAIDs. Although this result is in line with the expectation of a gastroprotective effect for coxibs, the magnitude of the risk reduction is lower than the 50–60 % risk reduction that would be expected based on the results from the RCTs [16, 24, 25]. Results of the hd-PS matched cohort moved the risk estimate further towards a gastroprotective effect that would be expected from the randomized trials. Inclusion of a greater number of empirically identified covariates in the PS model slightly emphasized the protective effect of coxibs but did not qualitatively change results. The inclusion of 200 empirically identified covariates in the PS model resulted in an estimated risk reduction of 32 %, and the use of 900 covariates further increased the risk reduction to 38 %. This was achieved without forcing any of the investigator pre-specified covariates apart from age, sex, and the number of distinct drugs during the 6 months before cohort entry into the model. These factors were in fact automatically identified by the hd-PS algorithm and then included.

We observed a greater risk reduction of UGIC for coxibs than two previous studies that applied the hd-PS algorithm in U.S. Medicare data of elderly patients [7] and in the UK THIN database [26]. In contrast to these two studies, which used an intention-to-treat analysis in coxib and tNSAID initiators with a fixed follow-up of 180 days, our study included only continuous users of the same NSAID compound without a predefined follow-up and defined discontinuation of the initial NSAID as a censoring event. This design will result in less misclassification of exposure than the intention-to-treat design used in the previous two studies [27]. In the study of the UK-based THIN database, 89 % of the coxib initiators and 95 % of the tNSAID initiators stopped using their medications or switched to the comparator NSAIDs before the end of the follow-up period [26]. The investigators themselves questioned the adequacy of the intention-to-treat design under these circumstances. The lower potential for misclassification of exposure in our study may explain the greater magnitude of the protective effect of coxibs that we observed [28].

The hd-PS algorithm recommends that important confounders which are not included in the data dimensions (see below) should be forced into the hd-PS model. Age, sex, calendar year, and the number of distinct drugs used during the 6 months before cohort entry were included in each of the models estimating the hd-PS. The number of distinct drugs used has been shown to perform as well or even better than other comorbidity scores in predicting the risks of clinical outcomes [29]. Summary measures of healthcare utilization are automatically created by version 2 of the hd-PS [30], including number of drugs used, number of physician visits, and number of hospitalizations. In the GePaRD it is not possible to obtain the number of physician visits to the same physician in a yearly quarter as information is lacking on whether a patient saw the same physician more than once in a yearly quarter for treatment of the same disease episode, whereas visits to different physicians or to the same physician for a different disease episode are captured in the database.

We set the granularity to three digits for ICD-10-GM data and to four digits for the OPS code, and we considered the full ATC code for drugs. The granularity needs to be decided for all data dimensions. Since candidate covariates are based on their prevalence in the respective data dimension, considering only the fourth digit of the ICD-10 code will reduce the prevalence of the code; however, it might be a better proxy for the underlying confounder [7]. The estimated hd-PS can be used to adjust for confounding through stratification, matching, regression, or inverse probability weighting. Whereas it is recommended to either match or trim extremes off a PS distribution [31], the U.S. and THIN database studies investigating the same association adjusted for it in deciles of the hd-PS in the logistic regression analysis [7, 26].

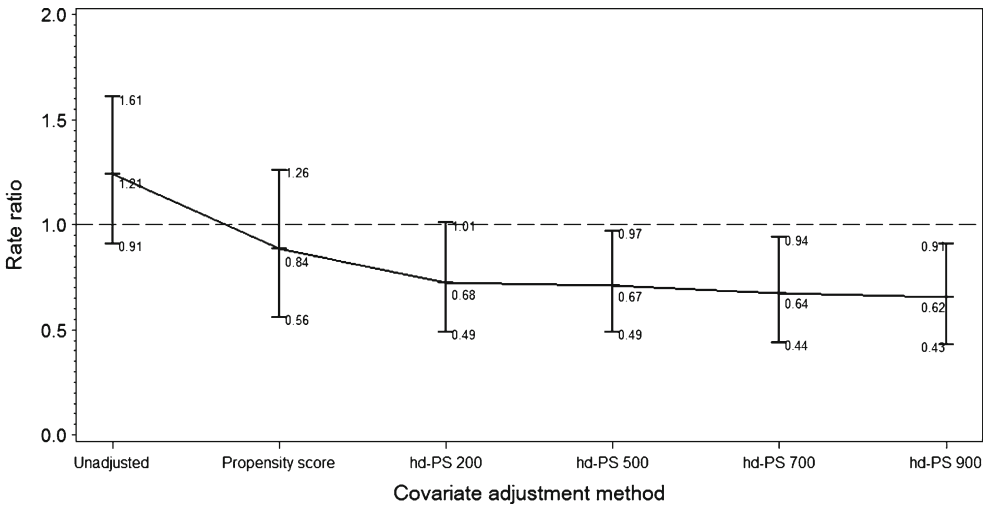

Fig. 2 Rate ratio of UGIC as a function of increasing covariate adjustment with predefined propensity scores and high-dimensional propensity scores (*hd-PS*)

Although the hd-PS algorithm enhanced adjustment for measured and unmeasured confounding in our example study through extensive proxy adjustment, some limitations need to be addressed. Typically PS analyses are limited to dichotomous exposures and are more complex to apply to multicategorical exposures (e.g., multiple doses), although this has been demonstrated [32, 33]. The current algorithm cannot adjust for time-varying confounding, since only baseline covariates are incorporated into the model estimating the hd-PS. The prioritization of covariates is based on the unconditional associations of potential empirical confounders with the exposure and separately with the outcome. Bias can also occur through the inclusion of a so-called "collider" variable, i.e., a variable in a causal system that is a shared effect of more than one cause, although simulation studies have shown this bias to be weak [34]. It cannot be ruled out with certainty that an empirically selected variable could increase variance and amplify the confounding of unobserved factors, although simulation studies by Myers have shown that more adjustment is a better strategy when there is doubt of whether a variable is a confounder or not [35]. The algorithm recommends the inspection and deletion of variables that are likely only associated with exposure, but not with the outcome. A screening tool has been developed to highlight such variables.

In this study we compared the results of the conventional and hd-PS method using the well-known association between NSAIDs and UGIC as an example. We do not know whether our findings with respect to the performance of the hd-PS algorithm will also apply to other treatment–outcome pairs. It has been suggested that the performance of the hd-PS ultimately depends on the information richness of the data source [7, 30]. We benchmarked our findings against results from RCTs which are usually conducted under tightly controlled research conditions. It is unclear whether the trial findings are applicable to our study population, which is less strictly controlled in terms of population composition and adherence. We assumed that a patient was exposed if she/he received a drug dispensation, since we lacked further information on whether the dispensed drugs were actually consumed, although several studies have shown that data on dispensed drugs are a more reliable exposure source than physician or patient reporting. Some NSAIDs may in low doses be dispensed over the counter (OTC) in Germany (diclofenac up to 25 mg, ibuprofen up to 400 mg, naproxen up to 200 mg, as well as oral or rectal aminosalicylic acid preparations), and these OTC dispensations are not captured in the GePaRD. Validation of the outcomes identified in the database against hospital charts has so far not been feasible because of German data protection legislation. However, coding is regularly checked in samples of patients in German hospitals by the Medical Review Board of the Statutory Health Insurance Funds so that coding quality is generally regarded to be high.

In conclusion, the results of our study showed that the hd-PS algorithm meaningfully improved the effect estimates above those derived from a large list of expert-defined confounders in comparative effectiveness research when benchmarked against the results of RCTs. The algorithm presents an important approach when dealing with the limited confounder information in information-rich large electronic healthcare databases. Instead of basing covariate selection on the recurrence of one covariate, the recurrence of specific covariate patterns could be formalized and explored. Given the potential of this new algorithm for confounder control in database studies, it is desirable to test its performance in simulation studies with known associations.

Acknowledgments This study was conducted as one subproject of The Safety of Nonsteroidal Anti-inflammatory Drugs (SOS) project (http://www.sos-nsaids-project.org/) which has received funding from the European Community's Seventh Framework Programme under grant agreement number 223495—the SOS project. The authors are grateful to all statutory health insurance providers that provided data for this study, namely the Allgemeine Ortskrankenkasse (AOK) Bremen/ Bremerhaven, the DAK–Gesundheit, the Techniker Krankenkasse (TK), and the hkk. S. Schneeweiss was supported by grants from the National Library of Medicine (RO1-LM010213), the National Center for Research Resources (RC1-RR028231), and the National Heart Lung and Blood Institute (RC4-HL106373).

Conflict of interest E. Garbe has received consulting fees by Novartis Pharma GmbH, Bayer AG and TEVA GmbH unrelated to this project and is member of a Scientific Advisory Board of Nycomed, unrelated to this project. S. Schneeweiss is Principal Investigator of the Brigham and Women's Hospital DEcIDE Center on Comparative Effectiveness Research and the DEcIDE Methods Center, both funded by AHRQ, and of the Harvard–Brigham Drug Safety and Risk Management Research Center, funded by the Federal Drug Administration. S. Schneeweiss is consultant to WHISCON LLC and Booz & Co, and his research is partially funded by investigator-initiated grants from Pfizer, Novartis, and Boehringer–Ingelheim unrelated to the topic of this study. The remaining authors declare no conflict of interest.

References

1. Schneeweiss S, Avorn J (2005) A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 58(4):323–337
2. Suissa S, Garbe E (2007) Primer: administrative health databases in observational studies of drug effects—advantages and disadvantages. Nat Clin Pract Rheumatol 3(12):725–732
3. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA (2011) Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther 90(6):777–790
4. Schneeweiss S (2007) Developments in post-marketing comparative effectiveness research. Clin Pharmacol Ther 82(2):143–156
5. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML (2009) Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research

Practices for Retrospective Database Analysis Task Force Report—Part I. Value Health 12(8):1053–1061
6. Wooldridge JM (2001) Econometric analysis of cross section and panel data. MIT Press, Cambridge
7. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA (2009) High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 20 (4):512–522
8. Schneeweiss S, Rassen J (2011) Re: Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. Pharmacoepidemiol Drug Saf 20(10):1110–1111
9. Joffe MM (2009) Exhaustion, automation, theory, and confounding. Epidemiology 20(4):523–524
10. Pigeot I, Ahrens W (2008) Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf 17 (3):215–223
11. Schink T, Garbe E (2010) Representativity of dispensations of non-steroidal anti-inflammatory drugs (NSAIDs) in the German Pharmacoepidemiological Research Database. Pharmacoepidemiol Drug Saf 19:S294
12. Schink T, Garbe E (2010) Assessment of the representativity of inpatient hospital diagnoses in the German Pharmacoepidemiological Research Database. Pharmacoepidemiol Drug Saf 19:S178–S179
13. Cattaruzzi C, Troncon MG, Agostinis L, Garcia Rodriguez LA (1999) Positive predictive value of ICD-9th codes for upper gastrointestinal bleeding and perforation in the Sistema Informativo Sanitario Regionale database. J Clin Epidemiol 52(6):499–502
14. Raiford DS, Perez GS, Garcia Rodriguez LA (1996) Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology 7(1):101–104
15. Glynn RJ, Schneeweiss S, Sturmer T (2006) Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 98(3):253–259
16. Rostom A, Muir K, Dube C, Jolicoeur E, Boucher M, Joyce J, Tugwell P, Wells GW (2007) Gastrointestinal safety of cyclooxygenase-2 inhibitors: a Cochrane Collaboration systematic review. Clin Gastroenterol Hepatol 5(7):818–28, 828
17. Moore RA, Derry S, Phillips CJ, McQuay HJ (2006) Nonsteroidal anti-inflammatory drugs (NSAIDs), cyxlooxygenase-2 selective inhibitors (coxibs) and gastrointestinal harm: review of clinical trials and clinical practice. BMC Musculoskelet Disord 7:79
18. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH (2005) A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. J Clin Epidemiol 58 (1):98–102
19. Solomon DH, Schneeweiss S, Glynn RJ, Levin R, Avorn J (2003) Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med 115(9):715–720
20. Depont F, Fourrier A, Merliere Y et al (2007) Channelling of COX-2 inhibitors to patients at higher gastrointestinal risk but not at lower cardiovascular risk: the Cox2 inhibitors and tNSAIDs description of users (CADEUS) study. Pharmacoepidemiol Drug Saf 16(8):891–900
21. Laharie D, Droz-Perroteau C, Benichou J et al (2010) Hospitalizations for gastrointestinal and cardiovascular events in the CADEUS cohort of traditional or Coxib NSAID users. Br J Clin Pharmacol 69(3):295–302
22. Morant SV, Pettitt D, MacDonald TM, Burke TA, Goldstein JL (2004) Application of a propensity score to adjust for channelling bias with NSAIDs. Pharmacoepidemiol Drug Saf 13(6):345–353
23. MacDonald TM, Morant SV, Goldstein JL, Burke TA, Pettitt D (2003) Channelling bias and the incidence of gastrointestinal haemorrhage in users of meloxicam, coxibs, and older, non-specific non-steroidal anti-inflammatory drugs. Gut 52(9):1265–1270
24. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ (2000) Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 343(21):1520–8, 2
25. Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, Makuch R, Eisen G, Agrawal NM, Stenson WF, Burr AM, Zhao WW, Kent JD, Lefkowith JB, Verburg KM, Geis GS (2000) 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 284(10):1247–1255
26. Toh S, Garcia Rodriguez LA, Hernan MA (2011) Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. Pharmacoepidemiol Drug Saf 20(8):849–857
27. Schneeweiss S (2010) A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 19(8):858–868
28. Hernan MA, Hernandez-Diaz S (2012) Beyond the intention-to-treat in comparative effectiveness research. Clin Trials 9(1):48–55
29. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ (2001) Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. Am J Epidemiol 154(9):854–864
30. Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S (2011) Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. Am J Epidemiol 173 (12):1404–1413
31. Sturmer T, Rothman KJ, Avorn J, Glynn RJ (2010) Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution–a simulation study. Am J Epidemiol 172(7):843–854
32. Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Sturmer T (2010) Confounder summary scores when comparing the effects of multiple drug exposures. Pharmacoepidemiol Drug Saf 19(1):2–9
33. Rassen JA, Solomon DH, Glynn RJ, Schneeweiss S (2011) Simultaneously assessing intended and unintended treatment effects of multiple treatment options: a pragmatic "matrix design". Pharmacoepidemiol Drug Saf 20(7):675–683
34. Wei L, Brookhart MA, Schneeweiss S, Setoguchi S (2012) Implications of collider-stratification bias in epidemiologic studies: simulation study. Am J Epidemiol (in press)
35. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ (2011) Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol 174(11):1213–1222