Increase in Lifetime Adverse Drug Reactions, Service Utilization, and Disease Severity Among Patients Who Will Start COX-2 Specific Inhibitors: Quantitative Assessment of Channeling Bias and Confounding by Indication in 6689 Patients with Rheumatoid Arthritis and Osteoarthritis

FREDERICK WOLFE, NANCY FLOWERS, THOMAS A. BURKE, LESTER M. ARGUELLES, and DAN PETTITT

ABSTRACT. ***Objective.*** Nonrandom assignment of therapy in observational studies and clinical practice can be accompanied by channeling bias and confounding by indication. This in turn can lead to unreliable conclusions about treatment effectiveness. Although widely acknowledged as important, no studies in rheumatology have measured the extent of these biases. We identified variables contributing to confounding and investigated the strength of the confounding effect. Analytical methods (propensity scores) are available to mitigate the effect of nonrandom assignment if the full extent of confounding can be understood.

Methods. A population of 6637 patients with rheumatoid arthritis (RA) and osteoarthritis (OA) from the practices of 433 US rheumatologists completed 2 sets of detailed questionnaires concerning (1) the last 6 months in 1998 and (2) the first 6 months of 1999, generally prior to and after the release of celecoxib and rofecoxib. Patients who received the COX-2 specific inhibitors in period 2 were identified (n = 1517), and their characteristics were compared to the 5120 who did not start a new COX-2 specific inhibitor during Period 1.

Results. Patients starting a new COX-2 specific inhibitor had a greater lifetime history of adverse reactions of all kinds, but particularly gastrointestinal adverse drug reactions. They also had more severe scores for pain, functional disability, fatigue, helplessness, and global severity, and they used more inpatient and outpatients services than patients who would not switch to COX-2 specific inhibitors.

Conclusion. Confounding by indication and channeling bias result in an overall increase in severity of about 25% for the above measures. Observational studies should account for these biases by a broadly defined propensity score that includes the variables identified in this report. These observations are germane to observational studies of disease modifying antirheumatic drugs and biologics, as well, and suggest the need for careful control of confounders when assessing treatment effects in rheumatic disease observational studies. (J Rheumatol 2002;29:1015–22)

Key Indexing Terms:
BIAS
CHANNELING BIAS
RHEUMATOID ARTHRITIS
CONFOUNDING BY INDICATION
COX-2
OSTEOARTHRITIS

In chronic illnesses such as rheumatoid arthritis (RA) and osteoarthritis (OA), time plays a crucial role in establishing the extent of benefit associated with specific treatments. What seems to work is retained; that which does not work is abandoned. But it is not always so simple, and clinical judgment can be faulty, particularly as to the extent of efficacy or toxicity, and particularly during times when drugs are going into and out of favor. Observational studies have the potential for providing key information regarding outcomes and how they are affected by treatment. But because treatment allocation in observational studies is nonrandom, it is exceedingly difficult to come to accurate and unbiased conclusions regarding treatment effect.

From the National Data Bank for Rheumatic Diseases – Arthritis Research Center Foundation and University of Kansas School of Medicine, Wichita, Kansas; Pharmacia Global Health Outcomes, Peapack, New Jersey; University of Illinois at Chicago, Chicago, Illinois; and Pfizer Outcomes Research, New York, New York, USA.

Supported by a grant from Pharmacia and Pfizer Outcomes Research.

F. Wolfe, MD; N. Flowers, MPH, National Data Bank for Rheumatic Diseases – Arthritis Research Center Foundation and University of Kansas School of Medicine; T.A. Burke, PharmD, Pharmacia Global Health Outcomes; L.M. Arguelles, MS, University of Illinois at Chicago; D. Pettitt, DVM, MSc, Pfizer Outcomes Research.

Address correspondence to Dr. F. Wolfe, Arthritis Research Center Foundation, 1035 N. Emporia, Suite 230, Wichita, KS 67214. E-mail: fwolfe@arthritis-research.org

Submitted March 27, 2001; revision accepted November 6, 2001.

Two similar biases relating to nonrandom assignment predominate in observational studies. Channeling bias is a form of allocation bias, and occurs when drugs with similar therapeutic indications are prescribed to groups of patients with prognostic differences[1,2]. For example, in the early days of methotrexate (MTX) usage, MTX was prescribed to RA patients with the worst prognosis. Although MTX improved such patients, the underlying severity or their illnesses outweighed MTX effectiveness; MTX appeared not to work well and was a marker for poor outcome. Channeling may lead to another form of bias, confounding by indication[3-6]. This occurs when the indication for the drug prescription results in preferential identification of the patients with the condition and, at the same time, increases the risk of the outcome under study.

On a practical level, these effects are well known in rheumatology. New disease modifying antirheumatic drugs (DMARD) go preferentially to those who have failed previous DMARD; new nonsteroidal antiinflammatory drugs (NSAID) to those who have not tolerated the available NSAID. The advent of the cyclooxygenase (COX-2) specific inhibitors, however, has added an additional dimension. Not only are the COX-2 specific inhibitors prescribed to treatment failures, but reimbursement is often limited to those who are at high risk for gastrointestinal (GI) events and adverse drug reactions (ADR). While these biases to our understanding are acknowledged, there are no quantitative data on their extent: we know about them, but can't tell when they are operative and how important they are.

Within the last decade attempts have been made to control for confounding through the use of propensity scores[7-17]. Propensity scores represent the likelihood of receiving one therapeutic alternative versus another by reflecting the composite effect of multiple predictors of therapy, which are often confounding factors for treatment effect. An effective propensity score allows one to control for up to 90% of channeling bias and confounding by indication[11,12]. Two steps are necessary for the development of a propensity score: identification of the confounders and an understanding of the strength of the confounding effect. With this information, propensity scores can be generated by simple logistic regression in which the dependent grouping variable is regressed on the chosen set of confounder variables. This will result in a single score for each patient or observation that can be used to adjust for confounding.

We evaluated 6637 patients with RA and OA in the 6 month period prior to the US release of celecoxib and rofecoxib. Of these patients, 1517 were started on a COX-2 agent during the following 6 months. We studied these patients for a wide variety of demographic, disease severity, concomitant therapy, ADR, and service utilization variables. As the primary aim of this study, we describe the extent of confounding among patients starting on COX-2 specific inhibitors, and we identify variables appropriate for constructing a rheumatic disease propensity score.

MATERIALS AND METHODS

Study population. Patients in this study were 6637 RA and OA patients who completed 2 sets of detailed questionnaires concerning (1) the last 6 months in 1998 and (2) the first 6 months of 1999. The first period occurred prior to the general use of COX-2 specific inhibitors, and the second period occurred during the earliest prescription of these drugs. Patients who received the COX-2 specific inhibitors in Period 2 were identified and tagged in the computer file. The second period data were then deleted and the first period data analyzed to compare those patents who would and those who would not receive COX-2 specific inhibitors in the second period.

Patients in this study are participants in the National Data Bank for Rheumatic Diseases enrolled by 433 US rheumatologists[18]. In this particular project 1706 patients were recruited from the practices of US rheumatologists during a 30 day enrollment period[19]; 2221 patients were enrolled from community rheumatologists who made their patient populations available to us; 750 patients were enrolled at the time they started taking leflunomide as part of their ordinary medical care and as part of a leflunomide research project; and 1960 patients were followed in the Wichita data bank. The characteristics of the Wichita data bank have been described[20-22]. RA was diagnosed in 4754 patients and OA in 1883. Diagnoses were made by the referring rheumatologists.

Demographic and clinical data. At each questionnaire assessment, demographic variables were recorded including sex, age, ethnic origin, education level, and current marital status. Study variables included the Stanford Health Assessment Questionnaire functional disability index (HAQ disability)[23,24], a visual analog pain scale (VAS), a VAS for global disease severity, VAS sleep and fatigue scales[25], the Arthritis Impact Measurement Scales (AIMS)[26,27], anxiety and depression scales[28,29], the SF-36 mental and physical component scales (MCS and PCS)[30], the WOMAC pain, stiffness and function scale[31,32], and Likert scales that assessed current satisfaction with health and current perceived health. To measure "health quality of life" we used the VAS from the EuroQol[33,34]. To assess GI symptom severity we used a VAS with the instructions, "How much trouble have you had with your stomach (i.e., nausea, heartburn, bloating, pain, etc.)? Place a mark on the line that best describes the severity of your stomach problems on the scale of 0–100."

Service utilization was measured from patient self-reports[35]. By specific questions we inquired about the lifetime history of adverse events in each body system. The count of somatic symptoms asked patients about 37 specific symptoms that occurred in the week prior to completing the questionnaire[35,36].

Event rates, including service utilization, are calculated and described for the 6 month period from July through December 1998.

Statistical methods and interpretation. The primary methods of analysis used in this study were univariate logistic regression and univariate Poisson regression. These analyses express the association between switching/nonswitching to a new NSAID and various demographic and clinical variables (Tables 1–3) in the form of odds ratios and their 95% confidence intervals. Confidence intervals that cross 1 are not statistically significant. Within these tables the variables are organized by scale type (e.g., present/absent, 0–3, 0–10) and then sorted by the odds ratios (OR) so that the strength of the association can be seen easily. To understand the strengths and clinical significance of the associations, the group values shown in the "mean or %" columns of Tables 1–4 offer additional insight into treatment group differences.

Poisson analyses are similar to those of logistic regression except that the continuous variable takes the form of a count (e.g., number of hospitalizations). Poisson regression reports incidence rate ratios (IRR), and they have an interpretation similar to OR, as strength of association.

The Bayesian information criterion (BIC) was used to understand and

Table 1. Demographics and medical status variables and the association with COX-2 prescription (univariate analysis).

Variable	Mean or % Continue Taking NSAID	SD	Mean or % Switch to COX-2	SD	OR	Lower CI	Upper CI
Age, yrs	60.84	12.86	62.07	11.93	1.01	1.00	1.01
Disease duration, (yrs)	12.81	10.86	12.32	10.33	1.00	0.98	1.01
Total Income ($1000)	42155.56	27.46	42282.84	27.41	1.00	1.00	1.00
Comorbid conditions	1.27	1.31	1.52	1.40	1.14	1.10	1.18
Count of somatic symptoms	6.83	5.31	8.43	5.76	1.05	1.04	1.06
Medicare insurance, %	38.93		43.73		1.22	1.05	1.41
Medicare disability insurance, %	3.87		4.49		1.17	0.84	1.61
High School graduate, %	88.87		89.97		1.12	1.04	1.22
Medicare HMO insurance, %	6.94		7.19		1.04	0.85	1.27
Married, %	70.83		71.52		1.03	0.93	1.15
White, %	93.61		93.67		1.00	0.98	1.01
Medicaid insurance, %	4.67		4.62		0.99	0.73	1.33
Private insurance only, %	33.53		31.97		0.93	0.82	1.05
Sex % male, %	22.13		20.37		0.90	0.87	0.94
No insurance, %	2.09		1.71		0.82	0.70	0.95
HMO membership, %	15.22		12.01		0.76	0.57	1.02

Table 2. Disease status variables and the association with COX-2 prescription (univariate analysis).

Variable	Mean or % Continue Taking NSAID	SD	Mean or % Switch to COX-2	SD	OR	Lower CI	Upper CI
HAQ disability (0–3)	1.01	0.70	1.23	0.65	1.57	1.38	1.79
VAS pain (0–10)	3.78	2.64	4.88	2.61	1.16	1.13	1.20
Global severity (0–10)	3.26	2.39	4.14	2.44	1.16	1.11	1.21
AIMS depression (0–10)	2.40	1.58	2.73	1.66	1.13	1.08	1.18
VAS GI severity (0–10)	1.93	2.43	2.65	2.78	1.11	1.09	1.13
VAS fatigue (0–10)	4.33	2.87	5.17	2.82	1.11	1.07	1.14
AIMS anxiety (0–10)	3.49	1.86	3.85	1.84	1.11	1.06	1.16
WOMAC pain scale	14.24	11.86	19.39	12.51	1.03	1.03	1.04
WOMAC stiffness scale	6.91	5.28	8.85	5.40	1.07	1.05	1.08
WOMAC function scale	46.33	40.11	63.12	42.29	1.01	1.01	1.01
SF-36 physical component scale	30.85	8.58	27.56	8.32	0.96	0.95	0.96
SF-36 mental component scale	44.94	13.21	41.03	13.58	0.98	0.97	0.99
Rheumatology Distress Index (0–100)	34.11	18.49	40.43	18.34	1.02	1.01	1.02
Health status (1–4)	2.41	0.72	2.58	0.73	1.39	1.22	1.59
Health satisfaction (–2–2)	–0.28	1.20	0.11	1.22	1.30	1.21	1.40
Helplessness (5–25)	11.39	4.74	13.03	4.83	1.07	1.06	1.09

rank the relative fit of comparative models. The BIC is a relative measure of the comparative fit of the models[37]. However, selection of the best model and most important predictors of future COX-2 specific inhibitor use was made primarily using classification and regression tree methods (CART)[38]. CART is a nonparametric tree-based modeling method that can be used to identify the "best" predictors of NSAID/non-NSAID switch (Table 5). CART performs better than conventional multiple logistic regression when the data contain nonlinear features, colinearity, and interactions. In the CART analyses of Table 5, the primary splitters are those variables that best identify group differences. Primary splitters and surrogates include best splitters and surrogates, variables that could substitute for the primary splitter if it were missing. Variable importance is a relative measure of the importance of the variable in group classification in the current analysis. As with the BIC, CART splitters and surrogates are only approximate guides, and must be interpreted within the clinical setting.

CART uses a method called cross-validation to produce more accurate estimates. In this method, after obtaining results using the full sample, the sample is divided into 10 equal subsamples and the results obtained by using the entire sample are compared with results obtained on 10 subsamples. The error rates generated by the 10 test subsamples are then used to determine the overall error rate for the entire sample. This complex methodology allows us to estimate how well any classification tree will perform on an independent sample. Additional information on CART can be found at http://www.salfordsystems.com/.

In the logistic and Poisson regression analyses, standard errors and CI were corrected for clustering within the 4 patient groups[39]. In addition, correction for clustering also implies robust analyses, or the use of the Huber/White/sandwich estimator of variance in place of the traditional calculation. Statistical significance was set at 0.05 and all tests were 2 tailed.

RESULTS

Demographic factors and general medical symptoms. As

Table 3. History of lifetime adverse drug reactions (ADR) and GI drug treatment variables and the association with COX-2 prescription (univariate analyses).

Variable	Remain on NSAID, %	Switch to COX-2, %	OR	Lower CI	Upper CI
All proton pump inhibitors	15.57	29.20	2.24	2.11	2.37
All H2, PPI, and gastroprotective agents	34.32	52.21	2.09	1.90	2.31
Any ADR ever	62.30	74.75	1.79	1.59	2.02
GI ADR ever	36.46	49.97	1.74	1.47	2.06
Cardiopulmonary ADR ever	8.00	11.98	1.56	1.32	1.85
Musculoskeletal ADR ever	17.12	24.02	1.53	1.29	1.82
H2 blockers ever	22.25	30.32	1.52	1.27	1.83
CNS ADR ever	15.24	21.24	1.50	1.30	1.74
ENT ADR ever	36.70	44.74	1.40	1.20	1.63
Skin ADR ever	29.97	35.74	1.30	1.00	1.69
Other ADR ever	6.83	8.60	1.28	1.15	1.43
Hematologic ADR ever	11.79	14.16	1.23	1.10	1.39
Mean no. of NSAID in 6 mo	1.03 (SD 0.84)	1.22 (SD 1.01)	1.25	1.17	1.34

Table 4. Prior service utilization and the association with COX-2 prescription (univariate analyses).

Utilization Variables	Mean or % Continue Taking NSAID	SD	Mean or % Switch to COX-2	SD	IRR*	Lower CI	Upper CI
GI diagnostic tests	0.17	0.71	0.31	0.92	1.81	1.51	2.16
Expensive tests	0.32	0.87	0.52	1.11	1.63	1.45	1.84
All diagnostic procedures and tests	2.36	4.74	3.72	6.10	1.58	1.32	1.88
Radiographic examinations	1.87	3.88	2.89	4.90	1.55	1.27	1.88
Specialist visits	1.89	1.83	2.53	2.14	1.34	1.15	1.55
All medical visits **	4.18	3.47	5.53	4.24	1.32	1.23	1.42
Primary care physician visits	1.19	1.30	1.54	1.50	1.29	1.24	1.34
Hospitalizations	0.28	0.72	0.35	0.79	1.26	1.04	1.54

* 6 month period incidence rate ratios for patients who will switch to COX-2 agents compared to patient who will not switch, as determined by Poisson regression. ** Includes physician and nonphysician outpatient visits.

shown in Table 1, COX-2 specific inhibitors were more often prescribed to the elderly, women, high school graduates, and those receiving Medicare. Future COX-2 prescription was less common among health maintenance organization (HMO) members: OR 0.76 (0.57, 1.02), and these values changed slightly when adjusted for age: OR 0.80 (0.62, 1.05). Among general medical conditions, 2 factors stood out. A one-unit increase in the number of reported comorbid conditions was associated with a risk of COX-2 prescription of 1.14, and persons with any comorbid condition compared to those without a comorbid condition had an OR of 1.32 (1.19, 1.49), and had an OR 1.31 (1.18, 1.45) after adjustment for age. The count of somatic systems, on a review of systems-like checklist for the week prior to completing the questionnaires, showed that future COX-2 patients had 8.43 symptoms compared to 6.83 for non-future COX-2 patients, and that a one-unit increase in somatic symptoms was associated with a 1.05 OR. Although it is difficult to tell the relative strength of these effects from the table, the Bayesian information criterion statistic provides strong evidence (BIC 58.32) that the symptom count is the most important predictor, followed by the count of comorbid symptoms.

Disease severity factors. A wide series of disease severity variables were significantly different among those who would and those who would not receive future COX-2 drugs, and all predicted future COX-2 usage (Table 2). Of the most significant variables, the strongest association with COX-2 future use (in order of strength as determined by the BIC) occurred with WOMAC pain, VAS pain, WOMAC function, SF-36 physical component score, WOMAC stiffness, VAS global severity, and HAQ.

Previous ADR and GI protective agents. As shown in Table 3, a history of a previous ADR of almost any type and the use of GI protective agents predict future prescription of COX-2 specific inhibitors. The strongest predictors, however, were the current use of the GI agents. These data clearly show that those with previous adverse reactions and those with current GI symptoms are preferentially selected for COX-2 prescription.

Utilization of services. Persons who would receive COX-2

specific inhibitors consumed more services prior to therapy than NSAID users, as shown in Table 4. The incidence rate ratio for prior medical outpatient visits ranged from 1.29 to 1.34 for the 6 month study period. Over the one-year preceding period, patients who would receive COX-2 specific inhibitors historically had 2.7 more medical visits compared to those who would not receive COX specific inhibitors. In addition, COX-2 patients had almost one additional hospitalization during the year prior to NSAID or COX-2 therapy.

These findings, of course, relate to other covariates, and estimates and CI change significantly after adjusting for covariates. For example, after adjusting for age, sex, and comorbidity, future COX-2 patients had 5.3 (4.90, 5.91) previous medical visits versus 4.16 (3.57, 4.89) for non-COX-2 patients compared to values presented in Table 4. For hospitalization, the historical hospitalization rate after adjustment was 0.31 (0.28, 0.35) compared to 0.27 (0.21, 033) for future COX-2 patients versus patients continuing current treatment.

Modeling COX-2 prescription. Because of overlapping and colinear variables, we explored multivariable predictors of COX-2 prescription using the nonparametric classification and regression tree methodology (CART). In using this method our concern was not so much how to best predict COX-2 as it was to identify those variables that contributed to the prediction, and to understand their contribution to COX-2 prescription. In this model we omitted the WOMAC variables, since the WOMAC was designed for OA, not RA, and some investigators are uncertain about its use in RA.

The cross-validated CART model correctly predicted 61.1% of cases correctly using VAS pain, GI drugs (any), GI ADR ever, Age, All OP visits, Specialist visits, Helplessness, SF-36 PCS, Any ADR ever, SF-36 MCS, Smoking now, Global severity, Skin ADR ever, HAQ disability, and Total income.

Table 5 shows the relative importance[38] of the variables. The 3 right hand columns report the importance of the variables in the final model. The 3 columns on the left include both the primary splitters (variables upon which tree splitting was based) and also surrogates. Surrogates are variables that may be almost as effective as the primary splitter. They may be used when the primary splitter variable is missing. In ordinary logistic regression only variables that are "significant" are included in the model. Therefore the CART importance list on the left provides additional information about what variables contribute to COX-2 selection. In this particular instance we used only the top 2 surrogates.

Table 5 gives insight into the classes of variables that contribute to COX-2 selection in this study. These classes are disease severity (pain, global severity, fatigue), GI protective drugs, age, number of medical visits, GI ADR, any ADR, functional disability, mental status, and total income.

A multivariable logistic model yields a slightly different model, and one in which HAQ is dropped because of colinearity. The standardized coefficients (a measure of variable importance) for this model are GI drugs (any) (0.132), VAS pain (0.116), all medical visits (0.115), age (0.099), total income (0.095), SF-36 PCS (–0.070), global severity (0.052), GI ADR ever (0.040), helplessness (0.039), and VAS QOL (0.035). The logistic model predicts 77.4% of cases correctly. Logistic models, however, often overestimate the accuracy of prediction, and it is likely that the cross-validated CART model (61.1% prediction accuracy) is a better representation of the predictive ability of these variables.

DISCUSSION

The data of this study demonstrate channeling bias and confounding by indication in the prescription of COX-2 specific inhibitors immediately after their release for use in the United States. Patients who received these drugs had more severe disease symptoms, more severe GI symptoms, greater use of GI protective drugs, and more ADR generally, among other symptoms.

The extent of these differences is clinically important. The most important predictor of COX-2 use, VAS pain, was increased by 29.1% in future COX-2 users. Increases were also noted for HAQ (21.8%), WOMAC pain (36.2%), WOMAC function (36.2%), and SF-36 PCS (12.0%). These data indicate that patients receiving COX-2 specific inhibitors have more severe symptoms then those who do not switch. Such patients will not do as well on COX-2 specific inhibitors as predicted by randomized controlled trials because of the adverse selection process.

Clinical trials of COX-2 specific inhibitors showed reduction in GI ADR compared to patients treated with non-selective NSAID[40-46]. The data from this study indicate that patients prescribed COX-2 specific inhibitors in clinical use had increased rates of ADR in general, and specifically in GI ADR, prior to prescription. Fifty percent of future COX-2 users compared to 36.5% of patients who did not switch to COX-2 specific inhibitors reported lifetime GI ADR. In addition, previous proton pump inhibitor use was almost doubled for future COX-2 users (29.2 vs 15.6%), and when all GI protective agents were used the difference was 52.2 vs 34.3%. These data indicate substantial confounding by indication. Patients taking GI protective agents have higher rates of GI adverse events, and it can be expected, therefore, that these future COX-2 users will also be a greater risk for such events.

We also found higher rates of historical service utilization among future COX-2 users. They had more previous outpatient medical visits, hospitalizations, and testing. Over a one-year period, COX-2 versus NSAID use was preceded by 0.88 additional hospitalizations, 1.2 additional specialist visits, and 2.7 more outpatient visits.

Table 5. Relative importance of variables in CART model to predict change versus no change to COX-2 agent.

PRIMARY SPLITTERS & SURROGATES			PRIMARY SPLITTERS ONLY		
Variable	Importance		Variable	Importance	
VAS Pain	100.00		VAS Pain	100.00	
Global Severity	89.21		GI Drug (any)	39.77	
VAS Fatigue	58.56		GI ADR ever	27.38	
GI Drug (any)	39.77		Age	26.73	
GI PPI	38.69		All OP visits	20.22	
GI ADR ever	30.60		Specialist visits	14.01	
Age	26.73		Helplessness	12.52	
All OP visits	26.69		SF-36 PCS	12.44	
Any ADR ever	25.65		Any ADR ever	11.72	
Specialist visits	19.63		SF-36 MCS	10.91	
Helplessness	19.46		Smoking now	9.49	
SF-36 PCS	12.83		Global Severity	7.87	
SF-36 MCS	10.91		Skin ADR ever	6.60	
HAQ Disability	9.52		HAQ Disability	5.46	
Smoking now	9.49		Total income	3.75	
Skin ADR ever	6.60				
ENT ADR ever	6.47				
GI H2 blockers	6.45				
On Medicare	6.34				
Primary care MD visits	4.00				
Total income	3.75				
CNS ADR	1.75				
Depression	1.58				
Private insurance	0.91				
Anxiety	0.61				
Symptom count	0.36				
VAS QOL	0.11				
Cardio-pulmonary ADR	0.01				

That pain and other disease severity measures are the most important determinants in switching to a COX-2 selective agent suggests that inadequate disease control (disease severity) is the major primary determinant in the switching process. Given the possibility of switching, the patient's age, current and previous GI problems, and psychosocial status may influence the decision to start a COX-2 agent. A second primary pathway to COX-2 prescription seems to involve primarily current or past GI symptoms. As with the disease severity pathway, the GI pathway is modulated by age, psychosocial factors, and, of course, disease severity.

These data, then, identify the group of future COX users as those with more severe disease symptoms, more GI problems, and greater utilization of services. Although the likelihood of confounding by indication and channeling bias has long been known to exist, this is the first study to demonstrate quantitatively the extent of these biases. Observational studies are particularly prone to such biases. Studies outside of rheumatology have indicated that this type of bias often can be controlled by the use of propensity scores[7-17].

Using the data of this study: understanding the extent of channeling bias and confounding by indication. Tables 1–4 present data on the various demographic, disease severity, adverse events, and utilization variables for each NSAID group in detail sufficient that the strength of the effect of channeling bias and confounding by indication on each variable can be understood.

Selecting variables for propensity scores. The CART analyses provide information concerning the relative importance of classes of variables in measuring the various biases. For example, for the primary splitters of Table 5, it can be seen that disease severity, as characterized by pain, is the most important predictor of NSAID switching. Next come GI drugs (any), a measure of current GI problems, followed by GI ADR ever, a measure of GI history. This, in turn, is followed by age, utilization variables, and other measures. The multivariate logistic models generally identify similar variables and variable classes, although in a different metric. It would be expected that the CART and logistic models would be in general agreement, as they are, but would differ at some points because the theoretical basis of the 2 analyses differ, particularly in the presence of colinearity and nonlinearity. Taking all the results into consideration, major classes of variables include (1) disease severity, (2) use of

GI protective drugs, (3) previous GI adverse events, (4) age, (5) service utilization, (6) quality of life measures, and (7) smoking status.

The univariate analyses of Tables 1–4 and the primary splitters and surrogated importance values of Table 5 provide data on which individual variables are most important within the larger classes of variables. For example, among the disease severity variables, pain is the most important predictor (CART importance = 100 and OR = 1.16). Among GI drugs, any GI drug or the use of proton pump inhibitors provides almost the same information regardless of whether CART analyses or univariate logistic regressions are consulted.

Propensity scores are unique to each patient and each study. A propensity score is usually generated by a logistic regression in which all the known confounding variables are included. In rheumatology, however, there have been 2 problems in implementing this strategy. First, the covariate adjustment variables have not been known, and second, the appropriate variables usually have not been collected. Our results indicate that a wide range of variables contribute to appropriate adjustment, and that such variables might include a number of disease severity variables, ADR variables, utilization variables, demographics, and insurance variables. Tables 1–5 can be helpful in understanding which classes of variables and which specific variable would best be included in the propensity score analysis.

The data of this report also suggest the effectiveness of using the WOMAC in patients with RA as well as OA, confirming our report about WOMAC in rheumatic diseases in general[31].

One factor that we were not able to measure here, but that is likely to be very important, is the general satisfaction with current therapy that must be true for many patients who, over the years, have self-selected themselves to continue their treatments for long periods of time.

In summary, patients beginning a new COX-2 specific inhibitor shortly after the release of that agent had a greater lifetime history of adverse reactions of all kinds, but particularly GI ADR. They also had more severe scores for pain, functional disability, fatigue, helplessness, and global severity; and they used more inpatient and outpatient services then patients who would not switch to COX-2 specific inhibitors. This confounding by indication and channeling bias results in an overall increase in severity of about 25% in the above measures. Observational studies should account for these biases by a broadly defined propensity score that includes the variables identified in this report. While this report only addresses the issue of COX-2 usage, it is likely that the set of variables identified here are those that should be used for propensity scores in general in RA and OA, perhaps with the addition of acute phase reactants and radiographic scores. Baseline adjustment is important, and adds a degree of validation to observational studies that is often missing when treatment effects are assessed.

REFERENCES

1. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.
2. Lanes SF, Garcia Rodriguez LA, Hwang E. Baseline risk of gastrointestinal disorders among new users of meloxicam, ibuprofen, diclofenac, naproxen and indomethacin. Pharmacoepidemiol Drug Saf 2000;9:113-7.
3. Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by indication in observational studies. J Am Geriatr Soc 1999;47:749-54.
4. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981-3.
5. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. BMJ 1997;315:1151-4.
6. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996;144:1161-9.
7. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol 1999;150:327-33.
8. Landrum MB, McNeil BJ, Silva L, Normand SL. Understanding variability in physician ratings of the appropriateness of coronary angiography after acute myocardial infarction. J Clin Epidemiol 1999;52:309-19.
9. D'Agostino RBJ. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-81.
10. Smith NL, Reiber GE, Psaty BM, et al. Health outcomes associated with beta-blocker and diltiazem treatment of unstable angina. J Am Coll Cardiol 1998;32:1305-11.
11. Rubin DB. Estimation of nonrandomized treatment comparisons using subclassification on propensity scores. In: Abel U, Koch A, editors. Nonrandomized comparative clinical studies: Proceedings of the International Conference on Nonrandomized Comparative Clinical Studies, Heidelberg. Symposium Publishing; 1997.
12. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.
13. Drake C, Fisher L. Prognostic models and the propensity score. Int J Epidemiol 1995;24:183-7.
14. Stone RA, Obrosky DS, Singer DE, Kapoor WN, Fine MJ. Propensity score adjustment for pretreatment differences between hospitalized and ambulatory patients with community-acquired pneumonia. Pneumonia Patient Outcomes Research Team. Med Care 1995;33:AS56-66.
15. Mark SD, Robins JM. Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. Stat Med 1993; 12:1605-28.
16. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics 1992;48:479-95.
17. Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. Biometrics 1991;47:1213-34.
18. Wolfe F, Flowers N, Anderson J. The National Rheumatic Disease Data Bank: case mix and severity characteristics of patients in rheumatological practice [abstract]. Arthritis Rheum 1998; Suppl 41:S132.
19. Wolfe F, Hawley DJ, Pincus T. Disease modifying antirheumatic drugs (DMARD) and DMARD combination use by 3604 patients

of 303 US rheumatologists [abstract]. Arthritis Rheum 1999; Suppl 42:S379.
20. Wolfe F, Kong SX, Watson DJ. Gastrointestinal symptoms and health related quality of life in patients with arthritis. J Rheumatol 2000;27:1373-8.
21. Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: Work disability: A prospective 18 year study of 823 patients. J Rheumatol 1998;25:2108-17.
22. Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: A 23-year prospective, longitudinal study of total joint replacement and its predictors in 1600 patients with rheumatoid arthritis. Arthritis Rheum 1998;41:1072-82.
23. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
24. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
25. Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. J Rheumatol 1996;23:1407-17.
26. Meenan RF, Gertman PM, Mason JH, Dunaif R. The Arthritis Impact Measurement Scales. Arthritis Rheum 1982;25:1048-53.
27. Meenan RF. The AIMS approach to health status measurement: conceptual background and measurement properties. J Rheumatol 1982;9:785-8.
28. Hawley DJ, Wolfe F. Depression is not more common in rheumatoid arthritis: a 10 year longitudinal study of 6,608 rheumatic disease patients. J Rheumatol 1993;20:2025-31.
29. Hawley DJ, Wolfe F. Anxiety and depression in patients with RA: A prospective study of 400 patients. J Rheumatol 1988;15:932-41.
30. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey SF-36. 1. Conceptual framework and item selection. Med Care 1992;30:473-83.
31. Wolfe F. Determinants of WOMAC function, pain and stiffness scores: evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. Rheumatology 1999;38:355-61.
32. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
33. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: Validity, responsiveness and reliability of EuroQol EQ-5D. Br J Rheumatol 1997;36:551-9.
34. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. Br J Rheumatol 1997;36:786-93.
35. Wolfe F, Anderson J, Harkness D, et al. A prospective, longitudinal, multicenter study of service utilization and costs in fibromyalgia. Arthritis Rheum 1997;40:1560-70.
36. Wolfe F, Hawley DJ. Evidence of disordered symptom appraisal in fibromyalgia: Increased rates of reported comorbidity and comorbidity severity. Clin Exp Rheumatol 1999;17:297-303.
37. Raftery AE. Bayesian model selection in social research. In: Marsden PV, editor. Sociological methodology. Oxford: Basil Blackwell; 2000:111-63.
38. Steinberg D, Colla P. CART: Tree-structured non-parametric data analysis. San Diego: Salford Systems; 1999.
39. Stata Corporation. Stata statistical software: release 6.0. College Station, TX: Stata Corp.; 1999.
40. Goldstein JL, Silverstein FE, Agrawal NM, et al. Reduced risk of upper gastrointestinal ulcer complications with celecoxib, a novel COX-2 inhibitor. Am J Gastroenterol 2000;95:1681-90.
41. Emery P, Zeidler H, Kvien TK, et al. Celecoxib versus diclofenac in long-term management of rheumatoid arthritis: randomised double-blind comparison. Lancet 1999;354:2106-11.
42. Bensen WG, Zhao SZ, Burke TA, et al. Upper gastrointestinal tolerability of celecoxib, a COX-2 specific inhibitor, compared to naproxen and placebo. J Rheumatol 2000;27:1876-83.
43. Simon LS, Weaver AL, Graham DY, et al. Antiinflammatory and upper gastrointestinal effects of celecoxib in rheumatoid arthritis: a randomized controlled trial. JAMA 1999;282:1921-8.
44. Langman MJ, Jensen DM, Watson DJ, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA 1999;282:1929-33.
45. Silverstein FE, Faich G, Goldstein JL, 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.
46. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 2000;343:1520-8.