STOMACH

Open

GI-REASONS: A Novel 6-Month, Prospective, Randomized, Open-Label, Blinded Endpoint (PROBE) Trial

Byron Cryer, MD[1], Chunming Li, PhD[2], Lee S. Simon, MD[3], Gurkirpal Singh, MD[4], Martin J. Stillman, MD, JD[5] and Manuela F. Berger, MD[2]

OBJECTIVES: **Because of the limitations of randomized controlled trials (RCTs) and observational studies, a prospective, randomized, open-label, blinded endpoint (PROBE) study may be an appropriate alternative, as the design allows the assessment of clinical outcomes in clinical practice settings. The Gastrointestinal (GI) Randomized Event and Safety Open-Label Nonsteroidal Anti-inflammatory Drug (NSAID) Study (GI-REASONS) was designed to reflect standard clinical practice while including endpoints rigorously evaluated by a blinded adjudication committee. The objective of this study was to assess if celecoxib is associated with a lower incidence of clinically significant upper and/or lower GI events than nonselective NSAIDs (nsNSAIDs) in standard clinical practice.**

METHODS: **This was a PROBE study carried out at 783 centers in the United States, where a total of 8,067 individuals aged ≥55 years, requiring daily NSAIDs to treat osteoarthritis, participated. The participants were randomized to celecoxib or nsNSAIDs (1:1) for 6 months and stratified by *Helicobacter pylori* status. Treatment doses could be adjusted as per the United States prescribing information; patients randomized to nsNSAIDs could switch between nsNSAIDs; crossover between treatment arms was not allowed, and patients requiring aspirin at baseline were excluded. The primary outcome was the incidence of clinically significant upper and/or lower GI events.**

RESULTS: **Significantly more nsNSAID users met the primary endpoint (2.4% (98/4,032) nsNSAID patients and 1.3% (54/4,035) celecoxib patients; odds ratio, 1.82 (95% confidence interval, 1.31–2.55); $P=0.0003$). Moderate to severe abdominal symptoms were experienced by 94 (2.3%) celecoxib and 138 (3.4%) nsNSAID patients ($P=0.0035$). Other non-GI adverse events were similar between treatment groups. One limitation is the open-label design, which presents the possibility of interpretive bias.**

CONCLUSIONS: **Celecoxib was associated with a lower risk of clinically significant upper and/or lower GI events than nsNSAIDs. Furthermore, this trial represents a successful execution of a PROBE study, where therapeutic options and management strategies available in clinical practice were incorporated into the rigor of a prospective RCT.**

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/ajg

INTRODUCTION

Nonsteroidal anti-inflammatory drugs (NSAIDs) are widely prescribed for the relief of arthritis pain; however, they have been shown to increase the risk of gastrointestinal (GI) ulcer bleeding (1–5). Previous blinded, prospective randomized controlled trials (RCTs) have provided evidence of NSAID-induced GI toxicities and have suggested a lower rate of upper GI complications with cyclooxygenase (COX)-2–selective NSAIDs compared with nonselective NSAIDs (nsNSAIDs) (6–9); however, the majority of these trials have focused on damage to the upper GI tract only (6,7,9–12). In addition, because of the relative inflexibility of management options in their protocols, prior RCTs of NSAIDs have not captured the common characteristics of NSAID utilization in clinical practice, such as switching among NSAIDs (13), modifications of dosing, drug holidays, and concomitant therapies used to potentially reduce GI symptoms and complications.

[1]University of Texas Southwestern Medical Center, Dallas, Texas, USA; [2]Pfizer Inc, New York, New York, USA; [3]SDG, LLC, Cambridge, Massachusettes, USA; [4]Stanford University, Palo Alto, California, USA; [5]Hennepin County Medical Center, Minneapolis, Minnesota, USA. **Correspondence:** Byron Cryer, MD, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. E-mail: Byron.Cryer@UTSouthwestern.edu

Additional safety information that considers the potential for NSAID damage in both the upper and lower GI tract and better reflects standard clinical practice, could help clinicians make improved treatment decisions.

Blinded RCTs are often regarded as the "gold standard" for assessing treatment effects of medications (14). However, when attempting to extrapolate findings from RCTs to clinical practice, the ability to generalize these results is commonly limited by factors, such as strict inclusion and exclusion criteria, inability to switch among comparator medications, restriction of dose adjustment, or the inability to institute a drug holiday as would happen in routine clinical practice. Furthermore, the protocol-driven study requirements of RCTs may lead to informative censoring of patient participation because of perceptions of medication adverse effects, which may lead to patient discontinuation from the study. Informative censoring may result in an imbalance of exposure time and a compromised ability to accrue clinically relevant or valid information regarding the primary clinical study outcomes (15).

Noninterventional observational studies, compared with RCTs, may allow a better assessment of medication effects in typical clinical situations. However, as there is no randomization in these studies, it is difficult to draw firm conclusions about causation. Channeling, selection, and other biases inherent in nonrandomized studies may confound outcomes. Therefore, these types of studies often generate hypothesis rather than evidence of causality.

Prospective, randomized, open-label, blinded endpoint (PROBE) studies are designed to address some of the potential limitations of RCTs and observational studies. A PROBE trial design assesses clinical outcomes in large simple studies that allow a broad patient population, in this case patients who require the long-term use of NSAIDs owing to chronic pain, which better reflects clinical practice but with the advantage of randomization and a rigorous evaluation of endpoints by blinded expert adjudication committees (16).

The GI Randomized Event and Safety Open-Label NSAID Study (GI-REASONS; NCT00373685) is a novel PROBE study conducted in the United States, which measured clinical outcomes throughout the GI tract using blinded adjudication (17). Designed to address many of the potential limitations of RCTs and observational studies, the GI-REASONS tested the hypothesis that celecoxib use in patients with osteoarthritis of at least moderate GI risk (aged ≥55 years) would be associated with a lower incidence of clinically significant upper and/or lower GI events than nsNSAIDs in standard clinical practice. A prior double-blind RCT of celecoxib vs. diclofenac slow release plus omeprazole in patients with osteoarthritis and rheumatoid arthritis, the Celecoxib vs. Omeprazole and Diclofenac in Patients With Osteoarthritis and Rheumatoid Arthritis (CONDOR) trial, used a predefined primary endpoint of clinically significant upper and/or lower GI events to capture the spectrum of NSAID-associated GI damage throughout the entire GI tract (18). Although a similar endpoint is used in GI-REASONS, the novel aspect of this study is its PROBE design, which allowed for the possibility of switching among multiple nsNSAID comparators, dose adjust-

ment, and the discretionary use of concomitant proton-pump inhibitors (PPIs) in both arms. Inclusion and exclusion criteria in the GI-REASONS were simple compared with conventional RCTs, providing a patient population more reflective of those encountered in typical clinical practice.

METHODS
Study design
GI outcomes were evaluated in patients with osteoarthritis in need of daily NSAIDs (i.e., using celecoxib or nsNSAIDs) as a result of chronic pain, at moderate GI risk (defined as aged ≥55 years (12)), who were treated in a manner typical of US clinical patient care in 783 primary care and specialty practices. We hypothesized celecoxib would be associated with lower incidences of clinically significant upper and/or lower GI events compared with nsNSAIDs, with or without PPIs. Aspirin users were excluded to select a population with lower cardiovascular risk and to determine whether celecoxib has a greater GI benefit than nsNSAIDs, without the confounding use of aspirin. It has been reported that in patients taking COX-2 selective NSAIDs and low-dose aspirin, the GI advantages of COX-2 selective NSAIDs were reduced (19).

We estimated that 6,400 patients (3,200 per group) were needed to provide 90% power to detect a 56% reduction in the primary endpoint (from 1.6% for nsNSAIDs to 0.7% for celecoxib, observed in a similar population), (9) based on a χ^2-test with a two-sided $\alpha = 0.05$. Assuming a 25% discontinuation rate, the target enrollment was 8,000. Eligible patients were randomized via an interactive voice response system (1:1) in block sizes of 4 to either open-label celecoxib or an nsNSAID for 6 months. The nsNSAID in the comparator group was any nsNSAID of the investigator's choice, prescribed within the dosages allowed in the US package insert. Patients with osteoarthritis, who previously had been taking an NSAID, were switched to their assigned study medication after randomization without a washout period. Patients were provided a pharmacy card that allowed each patient to fill prescriptions for their allocated treatment only. This pharmacy card was also used to capture study medication utilization behavior, such as dose adjustment and switching among nsNSAIDs. Patients were stratified at baseline by *Helicobacter pylori* status (assessed by serological testing at a central laboratory). *H. pylori*-positive patients were not treated for their infection during the trial. During the 6-month study, patients were evaluated every 2 months and at end of study in office visits that included assessment of hemoglobin. Celecoxib or nsNSAID dosage could be adjusted within the US prescribing guidelines. Patients randomized to the nsNSAID arm could switch between nsNSAIDs, but crossover between the nsNSAID and celecoxib treatment arms was not allowed. PPIs and histamine-2 receptor antagonists (H_2-RA) were allowed in either arm at the provider's discretion. This study was conducted in accordance with the International Conference on Harmonisation Good Clinical Practice guidelines and local regulatory requirements. The protocol was modified twice after study initiation: first to clarify the inclusion/exclusion criteria regarding estimation of

STOMACH

Table 1. Primary endpoint analysis

	Celecoxib (*n*=4,035) Patients with events, no. (%)	nsNSAID (*n*=4,032) Patients with events, no. (%)
Clinically significant upper and/or lower GI events	54 (1.3)	98 (2.4)
Gastroduodenal hemorrhage[a]	0	2 (<0.1)
Gastric outlet obstruction[b]	1 (<0.1)	0
Gastroduodenal, small bowel or large bowel perforation[c]	1 (<0.1)	1 (<0.1)
Small bowel hemorrhage[d]	0	0
Large bowel hemorrhage[e]	3 (<0.1)	6 (0.1)
Clinically significant anemia of defined GI origin[f]	4 (0.1)	6 (0.1)
Symptomatic ulcers[g]	0	5 (0.1)
Small bowel obstruction[h]	0	0
Acute GI hemorrhage of unknown origin, including presumed small bowel hemorrhage[i]	1 (<0.1)	3 (<0.1)
Clinically significant anemia of presumed occult GI origin including possible small bowel blood loss[j]	44 (1.1)	75 (1.9)
Odds ratio (95% CI)	1.82 (1.31–2.55)	
Hypothesis testing	*P*=0.0003	
Helicobacter pylori status		
Positive	25/1,401 (1.8)	34/1,386 (2.5)
Negative	29/2,634 (1.1)	64/2,646 (2.4)

CI, confidence interval; EGD, esophagogastroduodenoscopy; GI, gastrointestinal; nsNSAID, nonselective, nonsteroidal anti-inflammatory drug.

Definitions of endpoint components

[a]Endoscopic evidence of gastroduodenal ulceration (mucosal break with definite depth) or erosion (mucosal break without depth), or other likely causative lesion and clinical evidence of hemorrhage (hematemesis or melena, or evidence of recent hemorrhage on EGD—e.g., clot, blood in stomach, or visible vessel).

[b]Clinical, surgical, endoscopic, or radiographic evidence with symptoms consistent with gastric outlet obstruction.

[c]Clinical, surgical, or radiographic confirmation associated with symptoms consistent with perforation.

[d]Melena or hematochezia with likely causative lesion on small bowel investigation.

[e]Melena or hematochezia with no evidence of source on EGD and likely causative lesion on colonoscopy.

[f]No overt clinical evidence of acute GI hemorrhage, but with fall in hematocrit ≥10% points and/or hemoglobin ≥2 g/dl from baseline, with likely causative lesion on upper GI or lower GI endoscopic investigation, and with no non-GI source of anemia identified.

[g]Ulcers without complications, which present with dyspepsia and have endoscopic or X-ray evidence of a gastric and/or duodenal ulcer.

[h]Clinical, surgical, endoscopic or radiographic evidence with symptoms consistent with small bowel obstruction.

[i]Hematemesis, melena, or hematochezia with no evidence of likely causative lesion on endoscopic investigation.

[j]No overt clinical evidence of acute GI hemorrhage, but with fall in hematocrit ≥10% points and/or hemoglobin of ≥2 g/dl without a GI lesion endoscopically identified and no non-GI source of anemia.

cardiovascular risk, and second to include additional prespecified sensitivity analyses. These modifications did not result in any change in study conduct.

Study eligibility

Patients aged ≥55 years with a clinical diagnosis of osteoarthritis, who required daily NSAID therapy for the management of osteoarthritis symptoms, were eligible. Key exclusions were active GI ulcer hemorrhage or gastroduodenal ulceration within 90 days of screening, known established cardiovascular disease, any patient who, in the opinion of the investigator, was at sufficiently high cardiovascular risk to require low-dose aspirin, or positive fecal occult blood test at screening. Acetaminophen for occasional treatment of nonarthritis pain (up to 4 g per day for no more than 2 consecutive days), narcotic analgesics, and gastroprotective agents were allowed.

Study endpoints

The primary endpoint was the incidence of clinically significant upper and/or lower GI events over 6 months, individual components of which are shown in the Results section (**Table 1**). Potential GI endpoint events were adjudicated by an independent, blinded, expert GI events adjudication committee that evaluated whether the suspected GI event met predefined criteria for a component of the composite GI endpoint. The primary endpoint component, clinically significant anemia of presumed occult GI origin, was defined as a hemoglobin decrease of more than 2 mg/dl from baseline, which had no apparent non-GI cause to explain the decrease. Patients could have only one primary endpoint. Secondary endpoints included moderate to severe abdominal symptoms, withdrawal owing to GI adverse events, changes in hemoglobin and hematocrit from baseline, study and nonstudy drug utilization, gastroprotective drug use, and fecal occult blood test results.
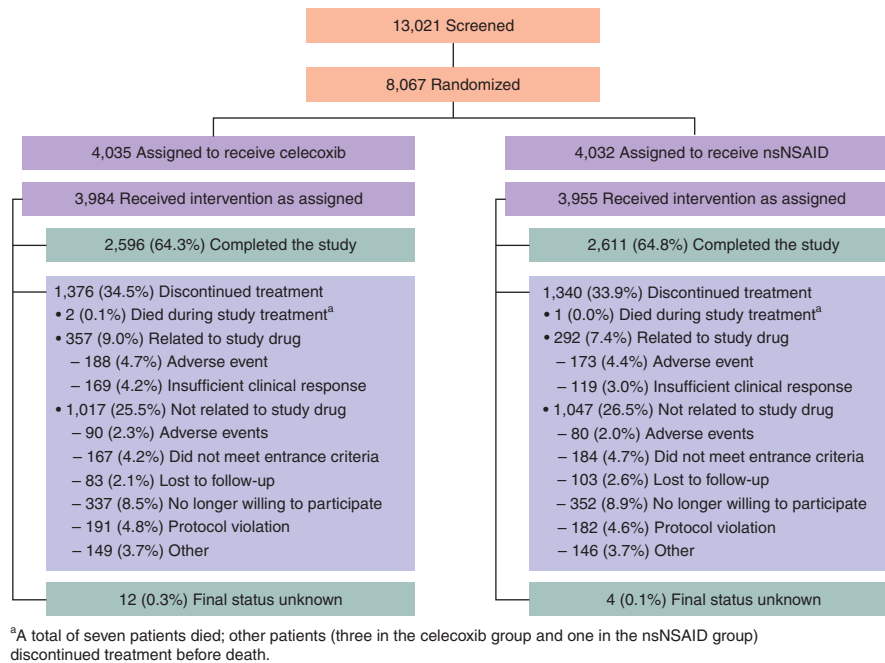
^aA total of seven patients died; other patients (three in the celecoxib group and one in the nsNSAID group) discontinued treatment before death.

Figure 1. Patient disposition. nsNSAID, nonselective nonsteroidal anti-inflammatory drug.

Patient satisfaction was measured with a validated instrument (20) as a secondary endpoint to assess potential treatment benefit. Cardiovascular events were also assessed and adjudicated by an independent, blinded, expert cardiovascular adjudication committee.

Statistical analysis

The primary and secondary endpoint analyses were based on the intent-to-treat population, defined as all randomized subjects. Statistical tests were significant at the 0.05 level. The primary endpoint was analyzed by comparing the incidence proportions between the treatment arms using a life-table (actuarial) extension of the Mantel–Haenszel method, stratified by *H. pylori* status at screening. As a secondary analysis to confirm robustness, the primary endpoint was analyzed using the Kaplan–Meier method and the log-rank test. Two sensitivity analyses were planned: a worst-case analysis, in which all patients who were lost to follow-up were conservatively assumed to have had a GI event meeting the criteria for a primary endpoint, and an analysis limited to patients with no protocol deviation. Incidences of hemoglobin and hematocrit drops were compared using a Cochran–Mantel–Haenszel test adjusted for *H. pylori* status at screening. Analysis of covariance was used to analyze continuous secondary endpoint variables, such as patient satisfaction. Secondary endpoint analyses were performed using the last observation carried forward, and were not adjusted for multiple comparisons. Safety data, physical examination, vital signs, laboratory data, and treatment-emergent adverse events were summarized. During the study, 36 patients were inadvertently randomized more than once. These patients were evenly distributed between groups and were excluded from analysis.

Study governance

The data and safety monitoring committee met regularly during the study to review the safety outcomes. An interim futility analysis was performed to re-estimate sample size after 50% of patients completed, and no adjustment to sample size or alpha was needed. An executive committee (BC, LSS, MJS, and GS) oversaw study conduct, analysis of outcomes, and interpretation of results.

RESULTS

Patient disposition

Between October 2006 and November 2010, a total of 4,035 celecoxib and of 4,032 nsNSAID patients were randomized and included in the intent-to-treat analyses (**Figure 1**). Baseline demographics were similar between treatment arms. Mean age (s.d.) was 63 (6) years, 76% were female, 259 patients (3.2%) had a history of GI ulcer or ulcer bleeding, 34% were *H. pylori* positive, and 15 patients (0.19%) had a history of coronary artery disease or myocardial infarction (**Table 2**).

In celecoxib-treated patients, 90% were initially prescribed celecoxib 200 mg daily. In nsNSAID-treated patients, the most common initially prescribed comparator nsNSAIDs were meloxicam (38%), naproxen (17%), nabumetone (11%), diclofenac (15%), ibuprofen (5%), and etodolac (5%). Because of switching, the number of patients who took these nsNSAIDs increased during the study, and the most common nsNSAIDs taken were as follows: meloxicam (42%, average total daily dose 13.0 mg); naproxen (21%, average total daily dose 819.8 mg); diclofenac (20%, average total daily dose 124.4 mg); nabumetone (14%, average total daily dose 1089.0 mg); ibuprofen (7%, average total daily dose

Table 2. Demographic characteristics

	Celecoxib (*n*=4,035)	nsNSAID (*n*=4,032)
Female, no. (%)	3,049 (75.6)	3,064 (76.0)
Age, mean (s.d.)	63.3 (6.3)	63.3 (6.4)
Weight (kg), mean (s.d.)	83.5 (20.1)	83.5 (19.9)
Duration of OA (years), mean (s.d.)	8.0 (7.7)	8.0 (7.6)
H. pylori positive, no. (%)	1,365 (33.8)	1,350 (33.5)
History of coronary artery disease or myocardial infarction, no. (%)	7 (0.2)	8 (0.2)

nsNSAID, nonselective, nonsteroidal anti-inflammatory drug; OA, osteoarthritis.

1453.2 mg); and etodolac (7%, average total daily dose 709.2 mg); and, as with celecoxib, the nsNSAIDs were taken at the usual US Food and Drug Administration-recommended dosages for management of osteoarthritis.

Similar percentages of celecoxib and nsNSAID patients completed the study (2,596 (64.3%) and 2,611 (64.8%), respectively). Overall, 1,376 (34.5%) and 1,340 (33.9%) patients discontinued treatment in the celecoxib and nsNSAID treatment groups, respectively. Reasons for discontinuations were similar between the two treatment arms, and patients discontinued because of adverse events were 7.0 % and 6.4% of the celecoxib- and nsNSAID-treated patients, respectively. Approximately, 2.8% and 3.0% of the celecoxib and nsNSAID treated patients, respectively, withdrew from the study owing to GI adverse events. Discontinued patients were not censored from the analysis; 186 patients were lost to follow-up (2.1% celecoxib and 2.6% nsNSAID).

Primary endpoint

Significantly more nsNSAID users met the primary endpoint (2.4% (98/4,032) nsNSAID patients and 1.3% (54/4,035) celecoxib patients; odds ratio, 1.82 (95% confidence interval (CI), 1.31–2.55); *P*=0.0003; **Table 1** and **Figure 2**). Of the patients who were *H. pylori* positive, 1.8% met the primary endpoint in the celecoxib group and 2.5% in the nsNSAID group; of the patients who were *H. pylori* negative, 1.1% met the primary endpoint in the celecoxib group and 2.4% in the nsNSAID group. In a sensitivity analysis, attributing the primary endpoint to all patients lost to follow-up (worst case; 2.1% of celecoxib and 2.6% of nsNSAID patients), the difference between treatment arms remained significant (odds ratio, 1.46; 95% CI, 1.18–1.82; *P*=0.0006). In another sensitivity analysis, excluding patients with protocol deviations, the difference between treatment arms also remained significant (odds ratio, 2.51; 95% CI, 1.35–4.65; *P*=0.0025).

Secondary endpoints

A smaller proportion of patients receiving celecoxib (94 (2.3%)) experienced moderate to severe abdominal symptoms than those receiving nsNSAIDs (138 (3.4%); *P*=0.0035). However, withdrawals owing to GI adverse events were similar between groups (112 (2.8%) and 120 (3.0%), respectively; *P*=not significant).
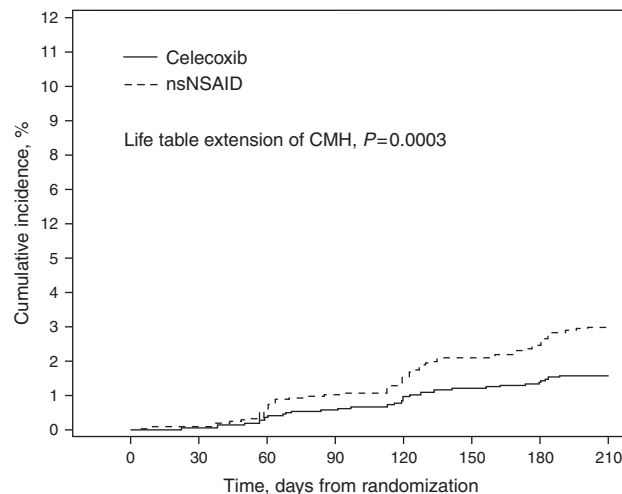
Celecoxib	4,035	4,033	3,720	3,406	3,096	2,892	2,684	1,332
nsNSAID	4,032	4,029	3,740	3,425	3,088	2,884	2,701	1,275

Figure 2. Cumulative incidence of clinically significant upper and/or lower gastrointestinal events. CMH, Cochran–Mantel–Haenszel; nsNSAID, nonselective, nonsteroidal anti-inflammatory drug. Note: any potential event occurring during the 180 days of treatment plus 28 days after last dose would have been reviewed and adjudicated by design. Hence, the Kaplan–Meier (KM) plot is presented up to 210 days here. The KM estimate beyond that duration became unreliable owing to censoring.

At baseline, mean hemoglobin levels were similar between treatment groups (13.6 g/dl in each group). At study completion, the hemoglobin decrease from baseline was significantly greater in those taking nsNSAIDs (treatment difference, 0.132; 95% CI, 0.10–0.16; *P*<0.0001). Over the course of the 6-month study, 1.8% of celecoxib patients compared with 2.9% of nsNSAID patients had a ≥2-g decrease in hemoglobin or 10% or greater decrease in hematocrit (relative risk, 1.6; 95% CI, 1.2–2.2; *P*=0.0023).

The proportion of patients taking gastroprotective agents (PPI or H$_2$-RA) for ≥75% of the time under study was 22.4% in the celecoxib group and 23.8% in the nsNSAID group. The proportion of patients with clinically significant upper and/or lower GI events among those who took a PPI was approximately 1.4% in the celecoxib group and 3.0% in the nsNSAID group, and among those who did not take a PPI, the proportion was 1.3% and 2.3%, respectively. Furthermore, among patients who took a PPI, the proportion of patients with clinically significant anemia (defined as either clinically significant anemia of defined GI origin or clinically significant anemia of presumed occult GI origin, including possible small-bowel blood loss) was approximately 1.3% in the celecoxib group and 2.4% in the nsNSAID group. This rate changed to 1.2% and 1.9%, respectively, for those who did not use a PPI.

The proportion of patients taking nonstudy analgesics for ≥75% of the time was comparable between treatment groups: 13.8% of patients in the celecoxib group (6.8% acetaminophen, 12.8% over-the-counter NSAID, 14.2% opioid) and 14.9% of patients in the nsNSAID group (6.5% acetaminophen, 13.3% over-the-counter NSAID, 15.6% opioid). Approximately 3.5% of celecoxib and 3.0% of

Table 3. Treatment-emergent adverse events

	Celecoxib (n=4,035) no. (%)	nsNSAID (n=4,032) no. (%)
Total number of patients evaluable for AEs	4,018 (99.6)	4,022 (99.8)
Patients with AEs	1,663 (41.4)	1,869 (46.5)
Patients with serious AEs	100 (2.5)	96 (2.4)
Patients with dose reduction or temporary discontinuation due to AEs	144 (3.6)	202 (5.0)

AEs, adverse events; nsNSAID, nonselective, nonsteroidal anti-inflammatory drug.

Table 4. Adjudicated cardiovascular events

	Celecoxib (n=4,035) no. (%)	nsNSAID (n=4,032) no. (%)
Total CV events	17 (0.4)[a]	13 (0.3)
APTC-like events	8 (0.2)	6 (0.1)
Acute myocardial infarction	2 (<0.1)	3 (0.1)
Stroke	3 (0.1)	3 (0.1)
Cardiovascular death	3 (0.1)	0
CV events of special interest	9 (0.2)	7 (0.2)
Unstable angina	2 (<0.1)	1 (<0.1)
Coronary revascularization	4 (0.1)	0
Transient ischemic attack	2 (<0.1)	2 (<0.1)
Venous and peripheral arterial thrombotic event	1 (<0.1)	3 (0.1)
Congestive heart failure	0	1 (<0.1)

APTC, Antiplatelet Trialists' Collaboration Combined Endpoint; CV, cardiovascular; nsNSAID, nonselective, nonsteroidal anti-inflammatory drug.
[a]One patient in the celecoxib group had an acute myocardial infarction and coronary revascularization.

nsNSAID patients used aspirin, which was noted as a protocol deviation.

The most commonly used nsNSAIDs were meloxicam (42%), naproxen (21%), diclofenac (20%), and nabumetone (14%), but comparison of celecoxib with any of the individual nsNSAIDs was not conducted, as the study was neither designed nor powered for such comparisons. About one quarter (24.4%) of nsNSAID patients modified their drug regimen (dose changes, switching to another NSAID); 5.4% of celecoxib patients changed their dose. Treatment satisfaction improved from baseline in both groups. However, patients taking celecoxib reported greater treatment satisfaction than patients taking nsNSAIDs (*P*<0.0001).

Overall and cardiovascular safety evaluation

Overall, adverse events and serious adverse events were similar between groups (**Table 3**). Adjudicated cardiovascular events were similar in both groups (celecoxib 0.4% vs. nsNSAID 0.3%). These included both Antiplatelet Trialists' Collaboration Combined Endpoint events (acute myocardial infarction, stroke, or cardiovascular death; eight celecoxib and six nsNSAID events) and cardiovascular events of special interest (unstable angina, coronary revascularization, transient ischemic attacks, venous and peripheral arterial vascular thrombotic events, and congestive heart failure; nine celecoxib and seven nsNSAID events; see **Table 4** for more detail). There were seven deaths in total, five in the celecoxib group (three adjudicated as cardiovascular-related) and two in the nsNSAID group (none adjudicated as cardiovascular-related).

DISCUSSION

In this study, designed to reflect the GI consequences of NSAIDs in a typical clinical practice setting, celecoxib was associated with an approximately twofold lower incidence of clinically significant upper and/or lower GI events than nsNSAIDs. Even in the "worst-case" sensitivity analysis, the treatment effect of COX-2-selective inhibition remained significant. The GI-REASONS had proportionally fewer upper GI ulcer complications than historical comparator studies (6,8,9), and changes in hemoglobin and/or hematocrit had a greater influence on the primary composite GI endpoint. A decline in clinically significant upper GI events and

an increase in lower GI events have also been identified in some observational studies (21,22). This phenomenon of suspected lower GI events taking on a greater proportion of the total GI tract events seen in the GI-REASONS is an important contribution that adds to our current understanding of the proposed burden of the effects of NSAIDs throughout the entire GI tract. In addition, the concomitant allowance of GI-protective therapies in both arms may have decreased the observed incidence of upper GI ulcers, contributing to the relative finding of increased suspected lower GI events.

We have provided the data on the total number of adjudicated events based on individual NSAIDs, even though the authors do not believe that these data can be analyzed in any scientifically valid fashion, for the following reasons. To begin, the study was designed and powered to detect potential differences between celecoxib and nsNSAIDs as a collective group. The study was not designed to examine potential outcome differences among nsNSAIDs or between any single nsNSAID and celecoxib, and is therefore insufficiently powered to conduct valid subgroup analysis of any isolated nsNSAID comparator. Equally important, such analyses would be significantly biased. The major strengths of the GI-REASONS are (a) randomization and (b) the open-label PROBE design, which allows physicians to select any NSAID they want in the usual care arm, thus mimicking clinical practice. However, these strengths also make it uniquely ill-advised to conduct subgroup analyses of individual NSAIDs, because the selection of any individual nsNSAID in the usual care arm is not randomized. In fact, the decision of what nsNSAID to prescribe is made after randomization, in an open-label fashion. This introduces a strong channeling bias, similar to that often seen in observational studies. It is entirely possible that patients perceived to be at higher risk for GI complications in the usual-care arm will be selectively

STOMACH

prescribed specific nsNSAIDs, which the physicians believe are "safer," based on prior perceptions or marketing (e.g., meloxicam). No such channeling will occur in the celecoxib arm. The channeling bias would then result in a significantly higher complication rate in those receiving the channeled "presumed safer" NSAID (e.g., meloxicam) compared with the non-channeled arm (in this case, celecoxib). Therefore, the authors believe that it is not possible to compare specific nsNSAIDs with celecoxib, or with each other in a quantitative and scientifically valid fashion. Not only are these analyses not prespecified, the authors had agreed during the conduct of the study that subgroup analyses would not be conducted because of their questionable scientific validity. Therefore, although the authors feel that no conclusions can be drawn from the following data, it is provided at the editor's request.

The proportion of patients with clinically significant upper and/or lower GI events in the individual nsNSAID group (based on first prescription only) with ≥1 event was 9.1% (1/11 patients) in diflunisal-, 6.7% (7/105 patients) in oxaprozin-, 6.7% (1/15 patients) in indomethacin-, 3.3% (20/609 patients) in diclofenac-, 2.8% (5/181 patients) in etodolac-, 2.7% (18/679 patients) in naproxen-, 2.7% (3/112) in piroxicam-, 2.2% (34/1514 patients) in meloxicam-, 1.4% (6/418 patients) in nabumetone-, and 1.4% (3/211 patients) in ibuprofen-treated patients.

Recognizing the limitations of both RCTs and observational studies, recent discussions of the US Food and Drug Administration on the use of COX-2-selective NSAIDs and nsNSAIDs have proposed alternative clinical trial strategies, including the PROBE design as a novel method to be pursued for assessing the comparative effectiveness of these medications (23). The strengths of the PROBE design are randomization, prospective follow-up, and a rigorous evaluation of study endpoints by a blinded, independent, expert adjudication committee (16). Using a PROBE design, the GI-REASONS used simple inclusion and exclusion criteria to enroll a broad osteoarthritis patient population of moderate GI and low cardiovascular risk. Switching among nsNSAIDs, allowing dose adjustments, and drug holidays, along with use of PPIs and H_2-RA as needed, more closely reflected daily clinical practice. Another strength of this study was providing pharmacy cards to each patient; this allowed tracking of prescriptions, evaluation of medication compliance, and monitoring of drug utilization and drug switching behavior, while preventing treatment crossover.

The relative merits of the various methods used to assess NSAID-induced GI damage (e.g., endoscopy of the upper GI tract, a composite endpoint evaluation of the entire GI tract, or assessment of ulcer complications) have been the focus of the recent US regulatory discussions (24). NSAID toxicity in GI-REASONS was assessed by a composite primary endpoint of clinically significant upper and/or lower GI events (17). Historically, the most commonly employed investigational method in the study of NSAID-related GI damage is endoscopy. Although upper endoscopy is useful to assess the upper GI tract, it is not appropriate for the entire small intestine. In the small intestine, capsule endoscopy studies have shown differences in mucosal damage between COX-2-selective and nonselective NSAIDs (25,26); however, the clinical

relevance of these endoscopic findings is still unclear. In addition, a common gastroprotective strategy, PPI use, is not anticipated to have pharmacological effects extending beyond the duodenum. We believe a primary endpoint that assesses damage through the entire GI tract provides valuable safety data to guide management of arthritis patients, as well as methodological and regulatory discussions.

Prior data indicate that the GI advantage of COX-2-selective NSAIDs may be compromised in patients taking concomitant aspirin (19). Aspirin use was an exclusion criterion in this study. Although we avoided the potential confounding of the GI endpoint, this exclusion criterion means that our observations may not be generalizable to patients on prophylactic aspirin therapy.

The largest proportion of the composite GI endpoint can be attributed to decreases in hemoglobin ≥2 g/dl and/or ≥10% hematocrit, and the clinical relevance of a 2-g/dl drop in hemoglobin is currently the subject of active debate. We believe that the difference reflects those taking nsNSAIDs, even with concomitant PPI use, are at increased risk of mucosal damage throughout the entire GI tract compared with the relative GI mucosa protection offered by COX-2 inhibition via celecoxib. On a large scale, such as that examined in this study, it is possible that this difference of protection between nsNSAIDs and celecoxib is likely enough to expose the smaller, nonscoped bleeding changes that can occur throughout the entire GI tract with nsNSAID use. These smaller bleeding changes would then be reflected in the hemoglobin differences seen in the two arms.

We recognize the challenge of determining the significance of hemoglobin drops, which counted as a primary outcome but did not render a participant anemic by laboratory definition (e.g., a hemoglobin drop from 15.5 to 13.5 g/dl). Although the relative merits of using reference ranges to define anemia rather than relative change from baseline is not yet settled, we believe that drops in hemoglobin ≥2 g/dl represent a change from the patient's usual baseline that is clinically informative and may warrant further clinical examination.

Two major risk factors for GI mucosal damage are *H. pylori* and NSAID use (27). Historically, *H. pylori* has been the major cause of peptic ulcer disease. Many previous studies of NSAID GI injury excluded *H. pylori*-positive patients. By stratifying for *H. pylori* status in the current study, we believe that our analysis provides clinically relevant information on the use of NSAIDs in a patient population of mixed *H. pylori* status as found in typical clinical practice.

In this study we did see a numeric but nonstatistically significant imbalance in rates of cardiovascular events between the two study groups. As the GI-REASONS trial is designed as a trial to evaluate GI endpoints, this study was not sufficiently powered to assess rates of cardiovascular events between celecoxib and NSAIDs. A larger trial is currently ongoing to evaluate cardiovascular events between celecoxib and NSAIDs (Prospective Randomized Evaluation of Celecoxib Integrated Safety vs. Ibuprofen or Naproxen [PRECISION]) (28).

PROBE-designed studies have several limitations (29). The open-label design presents the possibility of bias. That is, patients

or investigators may add concomitant treatments to address lack of efficacy, or manage symptoms or risk based on their knowledge and beliefs of treatment allocation. However, although medications were open label, determination of endpoints was blinded and conducted by expert committees. In addition to these previously recognized limitations, our experience identified hurdles in the execution of a PROBE-designed study. Although assumed to be simple by design, in practice, incorporating the variety of therapeutic options and management strategies that are available in clinical practice, compared with the limited possibilities in an RCT, was challenging. However, greater investigator flexibility allowed outcomes reflective of actual clinical practice, and we believe that GI-REASONS has proven that clinically relevant comparative treatment data can be captured through a trial of PROBE design.

In summary, the GI-REASONS provides valuable GI safety data relevant to clinical practice. A greater understanding of NSAID risk throughout the entire GI tract should lead to more effective patient management and identification of improved risk-reduction strategies. Finally, we think this trial will be historically important with respect to clinical trial design in NSAID-related GI bleeding, as it represents the successful execution of a PROBE study, incorporating the variety of therapeutic options and management strategies that are available in clinical practice, with the rigor of a prospective RCT (**Supplementary Information**).

CONFLICT OF INTEREST

Guarantor of the article: Byron Cryer, MD.

Specific author contributions: B.C., L.S.S., G.S., and M.J.S. constituted the Executive Committee overseeing the study. M.F.B. designed the study with contributions from the other authors. C.L. performed the statistical analysis, and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors participated in drafting and revising the manuscript.

Financial support: This study was sponsored by Pfizer. Editorial support was provided by William Watkins, of PAREXEL, and was funded by Pfizer.

Potential competing interests: Byron Cryer has served as a consultant for Pfizer, Astra-Zeneca, PLx Pharma, McNeil Consumer Products, and Ritter Pharmaceuticals, and has received consulting fees, honorarium, and travel support from Pfizer. Chunming Li and Manuela Berger are full-time employees of Pfizer and own stocks and shares in Pfizer. Lee S. Simon has served as a consultant for Pfizer, Astra-Zeneca, PLx Pharma, Logical Therapeutics, Bayer Healthcare, Horizon Pharmaceuticals, Takeda, and Posen, and has received consulting fees and travel support from these companies. Gurkirpal Singh has received consulting fees and honorarium from Pfizer and has received grants from Pfizer, Novartis, Pozen, and Astra-Zeneca. Martin Stillman has served as a consultant for Pfizer, Alpharma, and NicOx S.A., and has received consulting fees and honorarium from Pfizer.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

✓ The ability to extrapolate findings from randomized controlled trials (RCTs) is restricted owing to the limitations associated with the study design.

✓ Nonsteroidal anti-inflammatory drug (NSAID) use has been shown to increase the risk of gastrointestinal (GI) ulcer bleeding; thus, changes in dose and drug switching are not uncommon in standard clinical practice as done in this study.

✓ RCTs have shown a lower rate of GI complications with cyclooxygenase (COX)-2-selective NSAIDs compared with nonselective NSAIDs (nsNSAIDS).

WHAT IS NEW HERE

✓ The successful execution of a prospective, randomized, open-label, blinded endpoint (PROBE) study, where therapeutic options and management strategies available in clinical practice were incorporated into the rigor of a prospective RCT.

✓ Celecoxib was associated with lower risks of clinically significant upper and/or lower GI events than nsNSAIDs in a large study that allowed for clinician-driven changes in drug dosing, nsNSAID choice, and discretionary use of proton-pump inhibitors.

REFERENCES

1. García Rodríguez L, Williams R, Derby LE *et al.* Acute liver injury associated with nonsteroidal anti-inflammatory drugs and the role of risk factors. Arch Intern Med 1994;154:311–6.
2. Laine L. Approaches to nonsteroidal anti-inflammatory drug use in the high-risk patient. Gastroenterology 2001;120:594–606.
3. Laine L, Curtis SP, Cryer B *et al.* Assessment of upper gastrointestinal safety of etoricoxib and diclofenac in patients with osteoarthritis and rheumatoid arthritis in the Multinational Etoricoxib and Diclofenac Arthritis Long-term (MEDAL) programme: a randomised comparison. Lancet 2007;369:465–73.
4. Singh G, Ramey DR, Morfeld D *et al.* Gastrointestinal tract complications of nonsteroidal anti-inflammatory drug treatment in rheumatoid arthritis. A prospective observational cohort study. Arch Intern Med 1996;156:1530–6.
5. Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 1999;340:1888–99.
6. 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.
7. Cannon CP, Curtis SP, FitzGerald GA *et al.* Cardiovascular outcomes with etoricoxib and diclofenac in patients with osteoarthritis and rheumatoid arthritis in the Multinational Etoricoxib and Diclofenac Arthritis Long-term (MEDAL) programme: a randomised comparison. Lancet 2006;368:1771–81.
8. Schnitzer TJ, Burmester GR, Mysler E *et al.* Comparison of lumiracoxib with naproxen and ibuprofen in the Therapeutic Arthritis Research and Gastrointestinal Event Trial (TARGET), reduction in ulcer complications: randomised controlled trial. Lancet 2004;364:665–74.
9. 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. JAMA 2000;284:1247–55.
10. Silverstein FE, Graham DY, Senior JR *et al.* Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. A randomized, double-blind, placebo-controlled trial. Ann Intern Med 1995;123:241–9.
11. Singh G, Fort JG, Goldstein JL *et al.* Celecoxib versus naproxen and diclofenac in osteoarthritis patients: SUCCESS-I Study. Am J Med 2006;119:255–66.
12. Gutthann SP, Garcia Rodriguez LA, Raiford DS. Individual nonsteroidal antiinflammatory drugs and other risk factors for upper gastrointestinal bleeding and perforation. Epidemiology 1997;8:18–24.

STOMACH

13. Langman M, Kahler KH, Kong SX *et al.* Drug switching patterns among patients taking non-steroidal anti-inflammatory drugs: a retrospective cohort study of a general practitioners database in the United Kingdom. Pharmacoepidemiol Drug Saf 2001;10:517–24.

14. U.S.Preventive Services Task Force. U.S. Preventive Services Task Force Procedure Manual 2008. U.S.Preventive Services Task Force, AHRQ Publication No. 08-05118-EF, Rockville, MD.

15. Hernán MA, Hernández-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615–25.

16. Hansson L, Hedner T, Dahlöf B. Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Blood Press 1992;1:113–9.

17. Chan FK, Cryer B, Goldstein JL *et al.* A novel composite endpoint to evaluate the gastrointestinal (GI) effects of nonsteroidal antiinflammatory drugs through the entire GI tract. J Rheumatol 2010;37:167–74.

18. Chan FK, Lanas A, Scheiman J *et al.* Celecoxib versus omeprazole and diclofenac in patients with osteoarthritis and rheumatoid arthritis (CONDOR): a randomised trial. Lancet 2010;376:173–9.

19. Lanas A, García-Rodríguez LA, Arroyo MT *et al.* Risk of upper gastrointestinal ulcer bleeding associated with selective cyclo-oxygenase-2 inhibitors, traditional non-aspirin non-steroidal anti-inflammatory drugs, aspirin, and combinations. Gut 2006;55:1731–8.

20. Evans CJ, Trudeau E, Mertzanis P *et al.* Development and validation of the Pain Treatment Satisfaction Scale (PTSS): a patient satisfaction questionnaire for use in patients with chronic or acute pain. Pain 2004;112:254–66.

21. Lanas A, García-Rodríguez LA, Polo-Tomás M *et al.* Time trends and impact of upper and lower gastrointestinal bleeding and perforation in clinical practice. Am J Gastroenterol 2009;104:1633–41.

22. Zhao Y, Encinosa W. Hospitalizations for Gastrointestinal Bleeding in 1998 and 2006, HCUP Statistical Brief #65 2008. Agency for Healthcare Research and Quality: Rockville, MD. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb65.pdf. Accessed 15 August 2011.

23. US Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research. Proceedings of the Joint Meeting of the Arthritis Advisory Committee and the Drug Safety and Risk Management Advisory Committee 2005. US Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research, FDA Web site: Rockville, MD. Available at: http://www.fda.gov/ohrms/dockets/ac/cder05.html. Accessed 23 April 2011.

24. Food and Drug Administration. Gastrointestinal Drug Advisory Committee Meeting: Outcome Measures for Claims to Reduce NSAID-Associated Upper Gastrointestinal (UGI) Toxicity. Food and Drug Administration. Gastrointestinal Drug Advisory Committee Meeting, 2010 FDA Web site: Rockville, MD. Available at: http://www.fda.gov/downloads/Advisory Committees/CommitteesMeetingMaterials/Drugs/Gastrointestinal DrugsAdvisoryCommittee/UCM231990.pdf. Accessed 21 March 2011.

25. Goldstein JL, Eisen GM, Lewis B *et al.* Video capsule endoscopy to prospectively assess small bowel injury with celecoxib, naproxen plus omeprazole, and placebo. Clin Gastroenterol Hepatol 2005;3:133–41.

26. Goldstein JL, Eisen GM, Lewis B *et al.* Small bowel mucosal injury is reduced in healthy subjects treated with celecoxib compared with ibuprofen plus omeprazole, as assessed by video capsule endoscopy. Aliment Pharmacol Ther 2007;25:1211–22.

27. Ji K-Y, Hu F-L. Interaction or relationship between *Helicobacter pylori* and non-steroidal anti-inflammatory drugs in upper gastrointestinal diseases. World J Gastroenterol 2006;12:3789–92.

28. Becker MC, Wang TH, Wisniewski L *et al.* Rationale, design, and governance of Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen Or Naproxen (PRECISION), a cardiovascular end point trial of nonsteroidal antiinflammatory agents in patients with arthritis. Am Heart J 2009;157:606–12.

29. Kohro T, Yamazaki T. Cardiovascular clinical trials in Japan and controversies regarding prospective randomized open-label blinded end-point design. Hypertens Res 2009;32:109–14.