

A Randomized, Double-Blind, Study of Rofecoxib in Patients with Mild Cognitive Impairment

Leon J Thal¹, Steven H Ferris², Louis Kirby³, Gilbert A Block⁴, Christopher R Lines^{*4}, Eric Yuen⁴, Christopher Assaid⁴, Michael L Nessly⁴, Barbara A Norman⁴, Christine C Baranak⁴ and Scott A Reines⁴, on behalf of the Rofecoxib Protocol 078 study group⁵

¹University of California, San Diego, CA, USA; ²New York University School of Medicine, New York, NY, USA; ³Pivotal Research Centers, Peoria, AZ, USA; ⁴Merck Research Laboratories, West Point, PA, USA

Inflammatory mechanisms have been implicated in Alzheimer's disease (AD) and might be mediated via the COX-2 enzyme. Previous studies with the selective COX-2 inhibitors, rofecoxib and celecoxib, have shown that they do not alter the progression of AD. We conducted a double-blind study to investigate whether rofecoxib could delay a diagnosis of AD in patients with mild cognitive impairment (MCI), a group with an expected annual AD diagnosis rate of 10–15%. MCI patients ≥ 65 years were randomized to rofecoxib 25 mg ($N = 725$) or placebo ($N = 732$) daily for up to 4 years. The primary end point was the percentage of patients with a clinical diagnosis of AD. The estimated annual AD diagnosis rate was lower than the anticipated 10–15%: 6.4% in the rofecoxib group vs 4.5% in the placebo group (rofecoxib:placebo hazard ratio = 1.46 (95% CI: 1.09, 1.94), $p = 0.011$). Analyses of secondary end points, including measures of cognition (eg the cognitive subscale of the AD Assessment Scale (ADAS-Cog)) and global function (eg the Clinical Dementia Rating (CDR)), did not demonstrate differences between treatment groups. There was also no consistent evidence that rofecoxib differed from placebo in *post hoc* analyses comparing ADAS-Cog and CDR-sum of boxes scores in overlapping subgroups of patients who had Mini Mental State Exam scores of 24–26 in the present MCI study and in a previous AD treatment study with a similar design. The results from this MCI study did not support the hypothesis that rofecoxib would delay a diagnosis of AD. In conjunction with the lack of effects observed in previous AD studies, the findings suggest that inhibition of COX-2 is not a useful therapeutic approach in AD.

Neuropsychopharmacology (2005) 30, 1204–1215, advance online publication, 2 March 2005; doi:10.1038/sj.npp.1300690

Keywords: rofecoxib; nonsteroidal anti-inflammatory drugs; mild cognitive impairment; Alzheimer's disease; COX-2 inhibitors

INTRODUCTION

Neuropathological and epidemiological data suggest that inflammatory mediators and immune mechanisms may play a role in the pathogenesis of Alzheimer's disease (AD) (McGeer *et al*, 1996; Cagnin *et al*, 2001; Ho *et al*, 2001; in 't Veld *et al*, 2002; Etminan *et al*, 2003). It has been proposed that these effects may be mediated via the COX-2 enzyme (Pasinetti, 2001). These observations have led researchers to investigate whether nonsteroidal anti-inflammatory drugs (NSAIDs) might slow the progression of dementia. Initially, two small studies with the NSAIDs indomethacin (Rogers

et al, 1993) and diclofenac (Scharf *et al*, 1999) provided preliminary evidence for efficacy over 6 months in patients with AD. However, interpretation of the results from both studies was confounded by high dropout rates, mainly due to gastrointestinal side effects thought to be mediated via inhibition of COX-1 (nonselective NSAIDs inhibit both COX-1 and COX-2) (Warner *et al*, 1999). More recently, three larger randomized, controlled, 1-year clinical studies with the selective COX-2 inhibitors rofecoxib or celecoxib failed to show any effects of treatment on the progression of AD (Sainati *et al*, 2000; Aisen *et al*, 2003; Reines *et al*, 2004). One of these studies included the nonselective NSAID naproxen, which also failed to show any effects (Aisen *et al*, 2003). In addition, the anti-inflammatory agent hydroxychloroquine did not show any benefits in an 18-month trial (Van Gool *et al*, 2001).

A possible explanation for the lack of efficacy in the above studies is that the underlying pathology might be too advanced in patients with an established diagnosis of AD for an anti-inflammatory treatment to alter the course of the disease. Epidemiological evidence indicates that there may

*Correspondence: Dr CR Lines, Merck Research Laboratories, 10 Sentry Parkway, Blue Bell, PA 19422, USA, Tel: +1 484 344 2595, Fax: +1 484 344 4940, E-mail: chris_lines@merck.com

⁵Members of the Rofecoxib Protocol 078 study group are listed in Acknowledgments

Received 8 March 2004; revised 16 December 2004; accepted 29 December 2004

Online publication: 12 January 2005 at <http://www.acnp.org/citations/NPP011205040106/default.pdf>

be a critical period, 2 or more years before the onset of dementia, during which exposure to NSAIDs protects against AD (in 't Veld *et al*, 2001). Selective COX-2 inhibitors offer the potential for treatments that may have beneficial protective effects in AD while being well tolerated in long-term use. We therefore conducted a study to determine whether treatment with rofecoxib could delay a diagnosis of AD. Rather than evaluating a general elderly population with a relatively low incidence rate of AD, we sought to recruit elderly patients with mild cognitive impairment (MCI) (Petersen *et al*, 2001a, b). These patients were expected to have an annual AD diagnosis rate of 10–15% vs a rate of 1–2% for the general elderly population (Petersen *et al*, 2001a, b). The study also provided the opportunity to gather important placebo-controlled, long-term safety data on rofecoxib in an elderly population.

METHODS

Patients

Patients aged 65 years or older who had completed at least 8 grades of education, and had a reliable informant who could accompany them to each clinic visit, were recruited at 46 study sites in the United States from April 1998 to March 2000. Potentially eligible patients were initially identified by investigators (by any means available) or via a centralized telephone prescreening process (Lines *et al*, 2003). Patients were screened at the study sites to determine if they met all the following criteria for MCI: patient reports memory problem, or informant reports that patient has memory problem; informant reports that patient's memory has declined in the past year; Mini Mental State Exam (MMSE) (Folstein *et al*, 1975) score ≥ 24 ; Clinical Dementia Rating (CDR) (Morris, 1993) global score = 0.5 with memory domain score ≥ 0.5 ; Blessed Dementia Rating Scale (BDRS) (Morris *et al*, 1988) total score ≤ 3.5 , with no part 1 item score > 0.5 ; Auditory Verbal Learning Test (AVLT) (Schmidt, 1996) total score ≤ 37 . The cut score on the AVLT corresponds to a score ≤ 1 standard deviation (SD) below the mean for normal elderly subjects; for the first 6 months of study enrollment, age-adjusted cut scores 1.5 SDs below the means for separate age bands were used (see Appendix A1). Further details of these tests of cognition and function are given in Appendix A1. The CDR assessment was completed on the basis of interviews with the patient and informant by a rater who was blinded to the results of psychometric tests (AVLT and MMSE).

Patients were excluded if they had: dementia; inadequate motor or sensory capacities to comply with testing; a modified Hachinski Ischemic Scale (Rosen *et al*, 1980) score > 4 (to exclude patients whose cognitive impairment may have been related to a vascular condition); a Hamilton Depression Scale (17-item version) (Hamilton, 1960) score > 13 (to exclude patients whose cognitive impairment may have been related to depression); a history of angina or congestive heart failure with symptoms that occurred at rest; uncontrolled hypertension; a history within the past year of myocardial infarction, coronary artery bypass, angioplasty, or stent placement; a history within the past 2 years of stroke, multiple lacunar infarcts, or transient ischemic events; a history within the past 3 months of

gastrointestinal bleeding; an expected therapeutic need for chronic NSAID or estrogen replacement therapy during the study. Patients taking NSAIDs on a chronic basis (≥ 7 days/month for the 2 months prior to study entry), estrogen replacement therapy (excluding topical ointments) within 2 months of study entry, or cholinesterase inhibitors within 1 month of study entry were also excluded. No more than 20% of patients at each study site could be taking vitamin E > 400 IU at the time of study entry. Concomitant use of the above medications during the study was discouraged, but patients who did take them were not discontinued for this reason. Patients who developed a need for cardio-protective doses of aspirin after randomization were permitted to use aspirin ≤ 100 mg/day; clopidogrel was also allowed. Each site received the approval of its local institutional review board to perform the study, and informed consent was obtained from each subject.

Design

This was a randomized, double-blind, placebo-controlled study with parallel groups. After screening, eligible patients were randomly assigned to receive rofecoxib 25 mg once daily or placebo once daily for up to 4 years. Rofecoxib 25 mg was the maximum recommended dose for approved chronic indications (Merck & Co. Inc., 2003; on September 30, 2004, Merck & Co. Inc., announced the voluntary worldwide withdrawal of rofecoxib from the market). Randomization of patients at each study site was determined by a computer-generated allocation schedule and was stratified according to MMSE score (24–26, > 26). The allocation schedule was generated by a statistician at Merck Research Laboratories according to in-house blinding conditions. The rofecoxib and placebo tablets were visually identical. The original intention was that the study would be event-rate driven and would continue until 220 patients had received an AD diagnosis, which it was anticipated would take approximately 2 years. The study was extended to 4 years due to lower than expected AD diagnosis rates. A decision was made in October 2002 to terminate the study in April 2003, 11 months earlier than the scheduled March 2004 termination (ie 4 years after the last patient was enrolled). Based on the diagnosis and discontinuation rates observed as of July 2002, it was estimated that 219 end points should have been observed by April 2003. In fact, only 189 diagnoses of AD had been made at this time, and discontinuation rates had continued to increase. The combination of the low diagnosis rates and increasing discontinuation rates made it unlikely that the target number of events could have been achieved and it was therefore considered that continuation of the study until its planned termination would not have been productive. The decision was made prior to unblinding.

Procedure

After randomization, patients were scheduled to attend the clinic at months 1 and 4, and then every 4 months until the study was completed or the patient was diagnosed with dementia. The study was conducted using an intent-to-treat approach; that is, patients who discontinued treatment, but who had not developed dementia, were asked to return to

the clinic for all remaining visits and assessments. The following assessments were administered at baseline and every 4 months, or at discontinuation from the study: CDR, MMSE, Selective Reminding Test (SRT; see Appendix A1 for details) (Buschke, 1973). The cognitive subscale of the AD Assessment Scale (ADAS-Cog; see Appendix A1 for details) (Rosen *et al*, 1984) was administered at baseline and then every 12 months, or at discontinuation from the study. The BDRS was administered at baseline and at 24, 36, and 48 months, or at discontinuation from the study. Any patient who received a global CDR score ≥ 1 at a routine clinic visit was suspected of having converted to dementia and administered a CT or MRI scan of the brain and any psychometric tests that were not already scheduled for that visit. The patient was continued on treatment and returned for an end point confirmation visit 2 months later, at which time all the psychometric evaluations were repeated. If the global CDR score was still ≥ 1 at this visit, the patient was considered to have reached the end point of dementia and was discontinued from the study (regardless of the outcome of the adjudication process described below). In some cases, a patient was determined by an investigator to have developed dementia despite maintaining a global CDR score of 0.5 at the trigger and confirmation visits, and these patients were also counted as end points and were discontinued from the study. The investigator determined the type of dementia: possible or probable AD according to NINCDS-ADRDA criteria (McKhann *et al*, 1984) or other, for example, vascular dementia. For patients who reached the end point of clinically diagnosed dementia, all relevant data were sent to an independent blinded adjudication committee consisting of three experts. Each adjudicator reviewed the data independently and indicated whether or not they concurred with the investigator's diagnosis. In order to qualify as an event for the primary analysis, a majority decision in ≥ 2 of the three adjudicators that the patient met criteria for possible or probable AD was required (ie ≥ 2 adjudicators classified the event as probable AD, ≥ 2 adjudicators classified the event as possible AD, or one adjudicator classified the event as possible AD and one adjudicator classified it as probable AD).

Any adverse experiences occurring during the study were recorded and rated by the investigator, while still blinded to the treatment that the patient was receiving, as to seriousness (death, life threatening, resulting in persistent or significant disability, resulting in hospitalization, prolonging an existing hospitalization, any other important medical event), drug-relatedness (possibly, probably or definitely drug-related, probably not or definitely not drug-related), and intensity (mild, moderate, or severe). All serious vascular events (including cardiac, peripheral vascular, and cerebrovascular events) and upper gastrointestinal perforations, ulcers, and bleeds were reviewed by independent blinded adjudication committees, who determined if they were confirmed events according to prespecified case definitions (confirmed events) (Bombardier *et al*, 2000; Konstam *et al*, 2001).

Statistical Analysis

The primary efficacy analysis compared the cumulative incidence of possible or probable AD according to NINCDS-

ADRDA criteria (McKhann *et al*, 1984); patients with dementia of other cause were censored in the analysis at the time of diagnosis. AD diagnoses confirmed by the end point adjudication committee were the only end points included in the primary analysis. The analysis was based on a Cox proportional hazards model of time-to-event data (based on the initial diagnosis of AD) using an intention-to-treat approach, which included all randomized patients regardless of whether or not they were taking study medication. The model included terms for treatment, region within the United States (North East, South East, South, Midwest, West), and baseline MMSE strata (baseline MMSE score ≤ 26 vs > 26). Region was intended to be a surrogate for investigating the influence of study site, since there were too few events at individual study sites to include that as a factor. An additional prespecified on-drug analysis was restricted to patients who converted to AD within 14 days of being on study medication.

The calculation for the power statement assumed that the incidence of possible or probable AD over 2 years in the placebo group would be 30% and that the discontinuation rate would be 20%. Based on these considerations and a planned sample size of 520 evaluable patients per group, the study had 90% power to detect a one-third reduction in the incidence of AD in the rofecoxib group vs the placebo group with two-tailed $\alpha = 0.05$.

Analyses of prespecified secondary measures (SRT-summed recall score, SRT-delayed recall score, MMSE score, ADAS-Cog score, CDR-sum of boxes score, BDRS score) were based on available data for evaluable patients (ie patients who had a baseline score and at least one postrandomization score). The annual rates of change from baseline to a given time point (slopes), and slope differences between groups, were estimated using an intention-to-treat approach (ie including all data regardless of whether a patient discontinued from therapy or not) and analyses were performed using longitudinal repeated measures models for the comparison, under the assumption that missing data were missing at random (ie ignorable missingness). Additional sensitivity analyses were also performed using a last-observation-carried forward approach, as well as with other models with different assumptions about the missing data structure.

The present study had a broadly similar design (randomized, double-blind, placebo-controlled) and utilized some similar assessments (ADAS-Cog and CDR) to a previous 1-year AD treatment study of rofecoxib 25 mg in patients who had MMSE scores of 14–26 (Reines *et al*, 2004). Since there were overlapping subgroups of patients who had MMSE scores of 24–26 in both studies, indicating a similar level of cognitive impairment despite the difference in diagnosis, we also conducted a *post hoc* analysis to compare change from baseline scores on measures of cognition (ADAS-Cog) and global function (CDR-sum of boxes) in the overlapping subgroups in the two studies. The analysis looked at estimated annual slope differences using the same methods as described above. For the subgroup from the previous AD treatment trial, the annual slope estimates were derived from data over 12 months. For the subgroup from the present study, the annual slope estimates were derived from data over 48 months.

The assessment of tolerability included on-drug adverse experiences with onset up to 14 days after patients stopped taking test medication, and was based on the population of patients who took at least one dose of study medication. Prespecified groupings of adverse experiences (eg the number of patients with one or more adverse experience) were analyzed using Fisher's exact test.

RESULTS

The study profile is shown in the study flowchart (Figure 1). A total of 1457 patients were randomized. The study was terminated after 189 confirmed diagnoses of AD had been made. The median duration of study participation was 115 weeks in the rofecoxib group and 130 weeks in the placebo group. The median duration patients took study medication was 94 weeks in the rofecoxib group and 105 weeks in the

placebo group. Estimates of treatment compliance based on counts of the number of tablets in medication bottles at each clinic visit and calculated as ((number of days on therapy/number of days in study) × 100) indicated reasonable compliance; 61.0% of the 725 patients randomized to rofecoxib and 70.8% of the 732 patients randomized to placebo had ≥80% compliance during the time they were in the study. Approximately 45% of patients discontinued the study prematurely, while 40% completed the study on-drug (including patients who had not completed 48 months of treatment but were still in the study at the time of termination), and 15% completed the study off-drug. Reasons for discontinuation were generally similar across the treatment groups, although there were some small differences; for example, a greater proportion of patients in the placebo group discontinued due to withdrawal of consent (see Figure 1). The time course of discontinuations was similar between the treatment groups over the duration

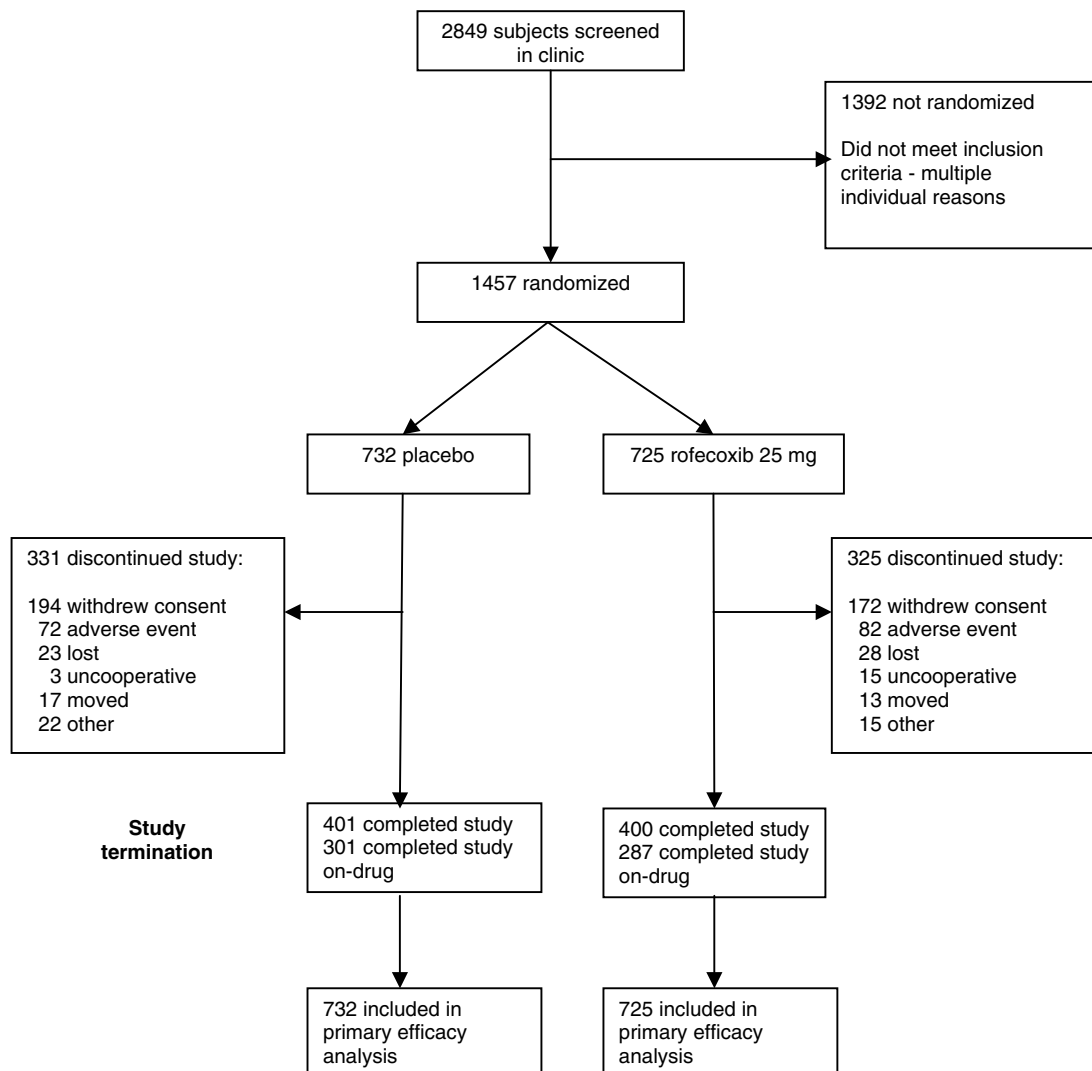


Figure 1 Study flowchart. Patients were prescreened for eligibility before being screened in the clinic. Data on the precise number of subjects prescreened were not collected, but the number was > 17 000. The total who completed the study includes patients who completed 48 months of study participation ($N = 115$ for rofecoxib, $N = 146$ for placebo), patients who were diagnosed with dementia of any cause ($N = 112$ for rofecoxib, $N = 83$ for placebo), patients who completed less than 48 months of study participation but were still in the study at the time it was terminated ($N = 121$ for rofecoxib, $N = 123$ for placebo), and patients who were discontinued because the study site closed ($N = 52$ for rofecoxib, $N = 49$ for placebo).

of the study (data not shown). Of those who discontinued, placebo patients had higher baseline ADAS-Cog scores, indicative of greater cognitive impairment and possibly greater risk of progression to AD, than patients randomized to rofecoxib ($p = 0.034$ in a logistic regression model); the mean baseline ADAS-Cog score was 9.6 in placebo discontinuers vs 9.1 in rofecoxib discontinuers. The change from baseline to last evaluation time point scores on the ADAS-Cog, MMSE, and CDR-sum of boxes in the subgroup of discontinued patients who had at least one on-treatment evaluation was similar between the treatment groups (data not shown).

Table 1 shows the baseline demographic characteristics and baseline test scores for randomized patients. The groups were generally comparable with regard to baseline characteristics and test scores. The distribution of secondary diagnoses at baseline was generally similar between the groups. The most common pre-existing medical conditions were hypertension (37.7% in the rofecoxib group and 34.3% in the placebo group) and osteoarthritis (22.2% in the

rofecoxib group and 24.3% in the placebo group). The percentages of patients with reported prior use (for any length of time in the 2 months prior to study entry for NSAIDs and estrogen, and 1 month prior to study entry for other drugs) of drugs claimed to have an influence on dementia were generally similar in the rofecoxib vs placebo groups for NSAIDs (8.4 vs 10.2%), ginkgo (12.6 vs 11.2%), statins (14.9 vs 13.0%), estrogen (0.4 vs 1.4%), and vitamin E >400 IU daily (5.8 vs 5.6%). The percentages of patients with reported concomitant use (for any length of time during the study) of drugs claimed to have an influence on dementia were lower in the rofecoxib group than the placebo group for NSAIDs (30.9 vs 34.7%) and estrogen (2.6 vs 4.8%), higher for cholinesterase inhibitors (11.2 vs 8.6%), and similar for statins (24.8 vs 24.9%), ginkgo (13.7 vs 12.8%), and vitamin E >400 IU daily (7.3 vs 7.5%). The median duration of reported concomitant NSAID use was lower in the rofecoxib group (5.6 weeks) than the placebo group (7.4 weeks). The percentages of patients with reported aspirin use in the 2 months prior to the study were 10.1% in the rofecoxib group and 9.0% in the placebo group. The percentages of patients with reported aspirin during the study were 31.7% in the rofecoxib group and 29.9% in the placebo group.

A total of 195 investigator diagnoses of dementia (190 diagnoses of possible or probable AD and five diagnoses of non-AD dementia) were evaluated by the end point adjudication committee. The primary end point of clinically diagnosed AD included the 189 patients who were confirmed to have developed possible or probable AD by ≥ 2 members of the committee. Two events adjudicated to be non-AD dementia, two events adjudicated to be nondementia, and two events in which there was no agreement between adjudicators (one adjudicator classified the event as AD, one classified it as non-AD dementia, and one classified it as nondementia) were not included as end points in the primary analysis. In the 189 patients with adjudicator-confirmed clinically diagnosed AD, global CDR scores at the last clinic visit were 0.5 (questionable dementia) in 21 patients, one (mild dementia) in 159 patients, two (moderate dementia) in eight patients, and three (severe dementia) in one patient.

In the rofecoxib group, 107 of 725 (14.8%) patients had clinically diagnosed AD over the 4-year study period vs 82 of 732 (11.2%) placebo patients over 4 years. The estimated hazard ratio (rofecoxib : placebo), adjusting for the effects of baseline MMSE stratum and region, was 1.46 (95% CI: 1.09, 1.94), which was statistically significant ($p = 0.011$, Wald χ^2 test) in favor of placebo. The treatment-by-time interaction was not significant ($p = 0.260$), indicating that the proportional hazards assumption was reasonably met. Baseline MMSE stratum (score ≤ 26 ; score > 26) had a highly statistically significant effect on outcome with an estimated hazard ratio of 3.23 (95% CI: 2.42, 4.32) ($p < 0.001$), indicating that patients in the lower stratum were much more likely to progress to AD irrespective of treatment assignment. Treatment effects were consistent across MMSE stratum levels, as well as geographic regions. Kaplan-Meier estimated proportions of patients with clinically diagnosed AD at 4-month increments are shown in Table 2. A separation in event rates between treatment groups was evident at the earliest time point (4 months), a gap that was

Table 1 Baseline Patient Characteristics

Characteristic	Placebo (N = 732) ^a	Rofecoxib 25 mg (N = 725) ^a
% Women	31.1%	34.3%
% Family history of AD	29.5%	31.2%
% Apolipoprotein $\epsilon 4$	35.8%	34.6%
Years of education		
11 or less	9.8%	11.2%
12–17	77.7%	74.7%
18 or more	12.4%	13.4%
Mean (SD) years of age	74.8 (6.0)	75.1 (6.0)
Mean (SD) AVLT score ^b	28.6 (5.8)	28.6 (5.7)
Mean (SD) MMSE score ^c	27.3 (1.7)	27.4 (1.7)
Mean (SD) CDR-sum of boxes score ^d	1.4 (0.8)	1.4 (0.8)
Mean (SD) ADAS-Cog score ^e	9.4 (3.8)	9.2 (4.0)
Mean (SD) SRT-summed recall score ^f	33.6 (9.1)	33.4 (9.0)
Mean (SD) SRT-delayed recall score ^g	4.5 (2.4)	4.4 (2.5)
Mean (SD) BDRS score ^h	1.1 (0.7)	1.1 (0.8)

^aSample sizes varied for some measures due to missing data. For apolipoprotein genotyping, sample sizes were 669 for placebo and 652 for rofecoxib.

^bAuditory Verbal Learning Test. Range = 0–75; lower score indicates greater impairment.

^cMini Mental State Exam. Range = 0–30; lower score indicates greater impairment.

^dClinical Dementia Rating. Range = 0–18; higher score indicates greater impairment.

^eCognitive subscale of the AD Assessment Scale. Range = 0–70; higher score indicates greater impairment.

^fSelective Reminding Test (immediate recall component). Range = 0–72; lower score indicates greater impairment.

^gSelective Reminding Test (delayed recall component). Range = 0–12; lower score indicates greater impairment.

^hBlessed Dementia Rating Scale. Range = 0–17; higher score indicates greater impairment.

Table 2 Kaplan–Meier Estimated Proportions of Clinically Diagnosed AD at 4-Month Intervals

Month after treatment	Placebo		Rofecoxib	
	N ^a	Estimated proportion	N ^a	Estimated proportion
0	732	0.0	725	0.0
4	707	0.3	682	0.9
8	666	1.4	642	3.1
12	625	3.1	582	5.1
16	581	4.2	535	7.4
20	528	6.4	486	9.2
24	458	9.2	415	12.3
28	390	10.3	344	14.2
32	359	11.0	311	15.7
36	310	12.5	258	17.9
40	240	13.7	214	19.3
44	177	15.6	159	21.0
48	86	17.4	68	23.0

^aNumber of patients at risk (number who had not been censored nor had an event up to, but not including, that time point).

maintained through the last follow-up time point of 48 months as shown by the nonsignificant *p*-value for the test of proportionality of treatment-specific hazard rates (see above). Estimated annual diagnosis rates were 6.4% (95% CI: 5.3%, 7.7%) in the rofecoxib group and 4.5% (95% CI: 3.6%, 5.6%) in the placebo group. The above analysis included patients whether or not they were taking study medication. In the prespecified analysis looking at events that occurred on-drug (*N* = 723 for rofecoxib and *N* = 728 for placebo), there was no evidence for an increased hazard ratio (1.49 (95% CI: 1.08, 2.05), *p* = 0.014) compared with the hazard ratio of 1.46 observed in the primary analysis, which included patients who had stopped taking study treatment.

To further explore the unexpected finding favoring placebo in the primary analysis, we conducted a *post hoc* analysis to adjust for factors that showed an effect, at a significance level of *p* < 0.10, on progression to AD. Those factors correlated with greater likelihood of progression to AD were lower baseline MMSE score stratum (24–26), female gender, age > 75, and prior ginkgo use. Factors associated with a decreased risk of progressing to AD were longer duration of concomitant NSAID use, and concomitant use of statins. In the analysis that adjusted for these factors, the statistical significance of the hazard ratio was reduced (1.31 (95% CI: 0.98, 1.75) *p* = 0.065). Presence of the apolipoprotein ε4 allele was a risk factor but was not included in the model because information was missing for a substantial proportion of patients. Based on a significance level of *p* < 0.10, family history of AD did not appear to be a risk factor in this study.

Because rofecoxib and other NSAIDs can cause small mean increases in blood pressure (Gertz *et al*, 2002; Schwartz *et al*, 2002), and there is some evidence to suggest

that increased blood pressure might be associated with an increased risk of dementia (Skoog, 1997; Birkenhager *et al*, 2001), we also performed two *post hoc* analyses to evaluate whether the rofecoxib:placebo risk ratio for diagnosis of AD increased as a function of increased blood pressure change. In the first analysis, change from baseline in mean arterial blood pressure (defined as (2 × diastolic blood pressure + systolic blood pressure)/3) at month 4 was calculated for each patient. The rofecoxib:placebo odds ratios for diagnosis of AD were then calculated for three categories of patients: those with no change or a decrease (odds ratio = 1.43), those with an increase ≤ 5 mmHg (odds ratio = 1.18), and those with an increase > 5 mmHg (odds ratio = 1.47). The Breslow–Day test of homogeneity of the odds ratios across categories indicated no significant differences (*p* = 0.895). In the second *post hoc* analysis, we looked at a predefined limit of change in systolic blood pressure, which was prespecified as a postrandomization value that was ≥ 180 mmHg and showed a ≥ 20 mmHg increase from baseline. The rofecoxib:placebo hazard ratios for diagnosis of AD were similar in those patients who did not meet the predefined limit of change criteria (hazard ratio = 1.42 (95% CI: 1.06, 1.92)), compared with those who did meet the criteria (hazard ratio = 1.53 (95% CI: 0.49, 4.81)).

The least squares mean scores for the secondary end points of cognition and function at 1-year intervals using the repeated measures models approach described in Methods are summarized in Table 3. The prespecified analysis looked at the estimated annual slope difference (placebo minus rofecoxib) for each measure based on data over the entire 4-year period. In contrast to the primary end point, there were no significant differences between treatment groups on estimated annual slope differences for SRT-summed recall score (slope difference = 0.026 (95% CI: -0.307, 0.359), *p* = 0.878), SRT-delayed recall score (slope difference = -0.012 (95% CI: -0.104, 0.081), *p* = 0.806), MMSE score (slope difference = -0.002 (95% CI: -0.095, 0.090), *p* = 0.959), ADAS-Cog score (slope difference = -0.098 (95% CI: -0.287, 0.091), *p* = 0.311), or BDRS score (slope difference = 0.079 (95% CI: -0.081, 0.238), *p* = 0.333). The slope difference for the CDR-sum of boxes score showed a nonsignificant trend in favor of placebo (slope difference = -0.068 (95% CI: -0.139, 0.002), *p* = 0.058); this would be expected to be highly correlated with the primary end point since the global score on the CDR was used to trigger the diagnosis of AD. Additional sensitivity analyses for the secondary end points were also performed using a last-observation-carried forward approach. The results, indicating a lack of treatment differences, were similar across these analyses, which were based on different assumptions about the missing data structure.

The *post hoc* analysis comparing test scores in overlapping subgroups of patients with MMSE scores of 24–26 in the present study (the subgroup with the worst prognosis) and a previous AD treatment study (Reines *et al*, 2004) included a total of 405 patients (rofecoxib *N* = 189, placebo *N* = 216) from the present MCI study, and 205 patients (rofecoxib *N* = 91, placebo *N* = 114) from the previous AD treatment study. There was no consistent evidence of a differential treatment effect of rofecoxib in the

Table 3 Secondary End Points: Least Squares Mean (Standard Error) Scores with Difference and 95% Confidence Interval

Measure	1-Year		2-Year		3-Year		4-Year					
	Placebo ^a (N ^c = 659)	Rofecoxib ^a (N ^c = 624)	Difference ^{a,b} (95% CI)	Placebo ^a (N ^c = 509)	Rofecoxib ^a (N ^c = 460)	Difference ^{a,b} (95% CI)	Placebo ^a (N ^c = 344)	Rofecoxib ^a (N ^c = 302)	Difference ^{a,b} (95% CI)	Placebo ^a (N ^c = 195)	Rofecoxib ^a (N ^c = 176)	Difference ^{a,b} (95% CI)
SRT-summed	35.7 (0.3)	35.6 (0.4)	-0.2 (-1.1,0.7)	36.5 (0.4)	35.7 (0.4)	-0.8 (-1.9,0.3)	36.1 (0.5)	35.7 (0.5)	-0.3 (-1.7,1.1)	36.0 (0.7)	35.6 (0.7)	-0.4 (-2.3,1.5)
SRT-delayed	4.8 (0.1)	4.7 (0.1)	-0.0 (-0.3,0.2)	4.9 (0.1)	4.8 (0.1)	-0.1 (-0.5,0.2)	4.9 (0.1)	4.7 (0.1)	-0.2 (-0.6,0.1)	4.5 (0.2)	4.5 (0.2)	0.1 (-0.5,0.6)
ADAS-Cog	8.9 (0.1)	9.3 (0.1)	0.3 (-0.0,0.7)	9.4 (0.2)	9.7 (0.2)	0.3 (-0.3,0.8)	9.8 (0.3)	10.1 (0.3)	0.3 (-0.4,1.0)	10.3 (0.3)	10.7 (0.3)	0.3 (-0.4,1.1)
MMSE	27.3 (0.1)	27.2 (0.1)	-0.1 (-0.3,0.1)	27.2 (0.1)	27.3 (0.1)	0.1 (-0.2,0.4)	27.1 (0.1)	27.1 (0.1)	-0.1 (-0.4,0.3)	26.7 (0.2)	26.5 (0.2)	-0.2 (-0.7,0.3)
CDR-sum of boxes	1.5 (0.0)	1.6 (0.0)	0.1 (-0.0,0.2)	1.8 (0.1)	2.0 (0.1)	0.2 (-0.0,0.4)	2.1 (0.1)	2.2 (0.1)	0.2 (-0.0,0.4)	2.2 (0.1)	2.4 (0.1)	0.2 (-0.0,0.5)
BDRS	—	—	—	1.4 (0.1)	1.4 (0.1)	0.0 (-0.1,0.2)	1.4 (0.1)	1.5 (0.1)	0.1 (-0.1,0.3)	1.4 (0.1)	1.6 (0.1)	0.1 (-0.1,0.4)

^aResults based on repeated measures model.

^bDifference is rofecoxib–placebo.

^cSample size (number of patients with any secondary efficacy measure by time point).

See Table 1 for key to test abbreviations and interpretation of direction of scores. 1-year data are not provided for the BDRS because it was not scheduled to be administered until the 2-year time point.

overlapping subgroups in the two studies. The estimated annual slope difference for the ADAS-Cog score was -0.240 (95% CI: $-0.736, 0.255$), $p = 0.340$ in the MMSE 24–26 subgroup from the present study, and 0.447 (95% CI: $-1.284, 2.178$), $p = 0.611$ in the MMSE 24–26 subgroup from the previous AD treatment study. The estimated annual slope difference for the CDR-sum of boxes score was -0.180 (95% CI: $-0.350, -0.009$), $p = 0.039$ in the MMSE 24–26 subgroup from the present study, and -0.053 (95% CI: $-0.560, 0.454$), $p = 0.837$ in the MMSE 24–26 subgroup from the previous AD treatment study.

A total of 1451 patients were included in the on-drug safety analysis (723 in the rofecoxib group and 728 in the placebo group). There were slightly fewer patients in the safety analysis than in the efficacy analysis because six patients who never took study medication were excluded from the safety analysis. The adverse experience profile is summarized in Table 4. The rofecoxib and placebo groups were similar with regard to the percentages of patients with any adverse experience, any serious adverse experience, and who discontinued treatment due to an adverse experience. There was a significant increase for rofecoxib in the number of patients with adverse experiences that were considered possibly, probably, or definitely drug-related by the investigators. The most common drug-related adverse experiences are shown in Table 4. There was no particular individual adverse experience that contributed to the overall difference between rofecoxib and placebo for drug-related adverse experiences, and relatively few patients discontinued study treatment due to drug-related adverse experiences (58 or 8.0% for rofecoxib and 41 or 5.6% for placebo). The number of patients with confirmed upper gastrointestinal perforations, ulcers, or bleeds was 14 in the rofecoxib group and four in the placebo group. A total of 39 deaths occurred in patients who were taking study treatment or from fatal adverse events that started within 14 days of the last dose (24 or 3.3% for rofecoxib and 15 or 2.1% for placebo). Patients died from a range of causes that were consistent with expectations for an elderly population, and there was no specific pattern as to the cause of death in either treatment group. The only specific fatal adverse events with more than one patient per treatment group were myocardial infarction (four patients on rofecoxib and three on placebo), cardiac arrest (two patients on rofecoxib and none on placebo), pneumonia (two patients on rofecoxib and none on placebo), and renal failure (one patient on rofecoxib and two on placebo). (An individual patient may have had more than one adverse event associated with death.) Off-drug follow-up mortality data were available for less than half of the patients ($N = 356$ for rofecoxib, $N = 307$ for placebo); the median duration of off-drug follow-up in these patients was 29 weeks in the rofecoxib group and 20 weeks in the placebo group. There were an additional 22 deaths in the off-drug period (17 in patients assigned to rofecoxib and five in patients assigned to placebo); 12 of these (11 in the rofecoxib group and one in the placebo group) occurred more than 48 weeks after treatment discontinuation. Compared with the on-drug period, there were an additional five patients with off-drug myocardial infarction fatal adverse events (five rofecoxib, none placebo), and an additional two patients with each of cardiac arrest (two rofecoxib, none placebo), pneumonia

Table 4 Summary of Clinical Adverse Experiences

Number (%) of patients with	Placebo (N = 728)	Rofecoxib 25 mg (N = 723)
≥ 1 adverse experience	670 (92.0%)	651 (90.0%)
≥ 1 drug-related adverse experience ^a	173 (23.8%)	211 (29.2%)*
Serious adverse experience ^b	236 (32.4%)	217 (30.0%)
Discontinued treatment due to adverse experience	147 (20.2%)	156 (21.6%)
<i>Most common drug-related adverse experiences^c</i>		
Diarrhea	16 (2.2%)	11 (1.5%)
Dyspepsia	17 (2.3%)	24 (3.3%)
Nausea	15 (2.1%)	14 (1.9%)
Dizziness	14 (1.9%)	18 (2.5%)
Peripheral edema	11 (1.5%)	23 (3.2%)
Increased blood pressure	14 (1.9%)	16 (2.2%)
Hypertension	25 (3.4%)	29 (4.0%)

^aExperiences that were rated as possibly, probably, or definitely drug-related by the investigator.

^bDeath, cancer, any adverse event that was life threatening, resulted in persistent or significant disability, resulted in or prolonged an existing hospitalization, or any other important medical event.

^cIncidence >2.0% in either treatment group.

*Significantly different from placebo, $p < 0.05$.

(two rofecoxib, none placebo), and renal failure (two rofecoxib, none placebo) off-drug fatal adverse events. The number of patients with confirmed serious thrombotic vascular events on-drug was similar in the two groups (38 in the rofecoxib group and 36 in the placebo group). There were a total of six patients with confirmed ischemic strokes and one patient with hemorrhagic stroke in the rofecoxib group compared to 13 patients with confirmed ischemic strokes and two patients with hemorrhagic strokes in the placebo group. Thirteen patients in the rofecoxib group had confirmed nonfatal myocardial infarctions vs 10 in the placebo group.

DISCUSSION

In this 4-year study of 1457 patients with MCI, there was no evidence that rofecoxib delayed a diagnosis of AD. A treatment difference in favor of placebo was observed on the primary end point of time to clinically diagnosed AD. This finding was not confirmed by secondary measures of cognition (ADAS-Cog, SRT, MMSE) or global function (BDRS, CDR-sum of boxes), which found no statistically significant or clinically meaningful differences between treatment groups. The possibility that rofecoxib might be inferior to placebo was also not supported by data from two large previous AD treatment studies, which included patients with an MMSE score up to 26 (thereby partly overlapping with patients in the present study who had an MMSE score ≥ 24), and found that rofecoxib had no significant effect on the progression of cognitive or functional decline (Aisen et al, 2003; Reines et al, 2004).

Given the unexpected nature of the primary finding, it was thought important to compare the present results with those from an independent database. We therefore conducted a *post hoc* analysis of test scores in subgroups of patients with MMSE scores of 24–26 in the present study (ie those patients who were most likely to receive a diagnosis of AD) and in a previous AD treatment study (Reines et al, 2004). These studies had similar designs and utilized similar assessments. Since MCI is hypothesized to be on a continuum of cognitive impairment ranging from very mild impairment (ie MCI) through mild, moderate, and severe dementia, it is plausible that many patients with MMSE scores of 24–26 in the two studies may have been biologically similar, even though one group was diagnosed with dementia and the other was not. There was no evidence to suggest differences between treatments on assessments of cognition (ADAS-Cog) in this subgroup in either study. In the subgroup analysis from the present study, there was a significant difference between treatments for the CDR-sum of boxes score. This finding was not surprising given that the global score on the CDR was the trigger for diagnosing AD. Indeed, because of its close relation with the primary end point of AD diagnosis, the CDR sum-of boxes score can be viewed as a surrogate for the primary end point. There was no difference between rofecoxib and placebo on the CDR sum-of-boxes score in the subgroup analysis from the previous AD treatment study.

An intriguing observation in the present study was that the separation between treatment groups in rates of diagnosis of AD was apparent from 4 months (the earliest time point assessed) but did not increase over time, raising the possibility that there may have been an imbalance between the groups at baseline. Although there was no clear evidence of a major imbalance for measured baseline variables, including severity of impairment, the model which adjusted for covariates that were risk factors for receiving a diagnosis of AD showed the smallest treatment effect and was not statistically significant. The finding that the treatment difference was not further increased in the analysis restricted to the on-drug population was also not supportive of a true treatment effect.

Since this is the first report of a completed randomized controlled study examining progression to an AD diagnosis in MCI patients, it is important to consider aspects of the design or conduct of the study that could have had an influence on the results. An obvious concern in a study with a long duration is that differential discontinuation rates may have influenced the results. A discontinuation rate of approximately 45% occurred over the course of the study and only 40% of patients completed the study on-drug. These findings are not surprising given the duration of the study and the elderly population being investigated. Overall discontinuation rates, time course of discontinuations, and change from baseline test scores at the time of discontinuation were similar between the treatment groups, although there were some small differences between the groups with regard to specific reasons for discontinuation; for example, a greater proportion of patients in the placebo group discontinued due to withdrawal of consent. The relatively high proportion of patients who withdrew consent may have been a consequence of the fact that patients were required to sign an additional consent form after 2 years,

owing to the study being extended beyond the original timeline, and were unwilling to go beyond their initial commitment. There was also some evidence that patients who discontinued from the placebo group were more impaired at baseline on the ADAS-Cog than patients who discontinued from the rofecoxib group. If the more impaired patients who discontinued were at higher risk of conversion to AD, then differential dropouts may have had an influence on the results.

Another notable observation in the present study was that the overall annual AD diagnosis rate of 5–6% was lower than the anticipated 10–15% annual diagnosis rate reported in previous observational or natural history studies (Petersen *et al*, 2001a,b). There are several possible explanations for this disparity. Firstly, the diagnosis of MCI is heavily dependent on clinical judgment (Petersen, 2003), and the particular tests and score thresholds used to provide objective evidence of memory impairment are not standardized (Chertkow, 2002; Petersen, 2003). It is therefore possible that the population of patients included in the study was more heterogeneous than, or different from, MCI populations previously described. Indeed, the population contained a lower proportion of women and patients with the apolipoprotein $\epsilon 4$ allele than others have reported (Petersen *et al*, 1999; Jack *et al*, 1999; Morris *et al*, 2001). Another possible contributing factor to the low diagnosis rates was the change in the cut-score criteria on the AVLT (the memory test used to provide objective evidence of memory impairment) during the study. After 6 months of enrollment, the cutoff was changed from age-adjusted 1.5 SD cut scores to a single cut score 1 SD below the mean for normal elderly subjects (see Appendix A1). We retrospectively looked at diagnosis rates in patients who met the original 1.5 SD criteria ($N=203$ for rofecoxib and $N=203$ for placebo) and the estimated annual diagnosis rates (11.3% for rofecoxib and 8.3% for placebo) were more in line with previous data, although still lower than anticipated for the placebo group. Finally, the fact that the MCI concept was relatively new at the time the study was initiated, along with the source of the patients (advertising campaigns as well as direct clinical referrals), may have also resulted in greater heterogeneity with regard to underlying etiology in the recruited patients.

To summarize the efficacy data, the unexpected primary finding in the present study suggesting that rofecoxib might accelerate the rate of AD diagnosis in patients with MCI was not supported by data on secondary measures in the same study, nor by other data from previous studies that assessed cognitive and functional decline in AD patients (Aisen *et al*, 2003; Reines *et al*, 2004). These observations suggest that the finding may not be indicative of a true treatment effect. If not a true effect, then the most likely explanations are a chance occurrence, or an imbalance between the treatment groups that existed at baseline or arose during the study due to differential discontinuations.

We cannot exclude the possibility that rofecoxib might accelerate conversion of MCI patients to AD, although this possibility would be at variance with the prior epidemiological and clinical data cited above, as well as the body of experimental evidence suggesting that COX-2 inhibition might attenuate neuronal death in several disease states including Parkinson's disease, multiple sclerosis, and

amyotrophic lateral sclerosis, in addition to AD (Andreasson *et al*, 2001; Jain *et al*, 2002; Xiang *et al*, 2002; Giovannini *et al*, 2003; Scali *et al*, 2003; Teismann *et al*, 2003; Qin *et al*, 2003; Consilvio *et al*, 2004; Rose *et al*, 2004). The early onset of the treatment difference (and the lack of progressive widening of the difference over the course of the trial) further argues against a direct effect of rofecoxib on the underlying pathophysiology in AD. It could be speculated that the apparent increased conversion to AD might be secondary to a non-AD-specific aspect of rofecoxib's biological effects. For example, rofecoxib and other NSAIDs can cause small mean increases in blood pressure (Gertz *et al*, 2002; Schwartz *et al*, 2002). It has been suggested that increased blood pressure might be associated with an increased risk of dementia, although the evidence is inconsistent (Skoog, 1997; Birkenhager *et al*, 2001). We investigated this possibility by performing *post hoc* analyses to determine if the rofecoxib:placebo risk ratio for diagnosis of AD increased as a function of increased blood pressure change. There was no evidence for such a relationship. A proposed mechanism by which increased blood pressure could lead to dementia is due to an increased risk for cardiovascular adverse events such as strokes or infarcts. As noted below, there was no evidence for an increase in these types of events in patients taking rofecoxib. Furthermore, CT or MRI brain scans were performed in all patients diagnosed with dementia, to exclude the possibility of vascular dementia.

In addition to evaluating efficacy, the present study provided important placebo-controlled data on the safety of rofecoxib 25 mg over periods of up to 4 years in an elderly population. The median duration of exposure to study medication was approximately 2 years. The mean age of patients was 75 years and approximately 50% were at least 75 years old. Previous safety data on rofecoxib come largely from osteoarthritis studies looking at a relatively younger population with a mean age around 60 years (Langman *et al*, 1999; Reicin *et al*, 2002). Rofecoxib was generally well tolerated by the elderly patients in the study, consistent with results from prior clinical studies in osteoarthritis (Langman *et al*, 1999; Reicin *et al*, 2002) and AD (Reines *et al*, 2004). The overall incidence of adverse experiences, serious adverse experiences, and discontinuations due to adverse experiences were similar or only slightly increased for rofecoxib *vs* placebo. Not surprisingly, there was an increase in the number of adverse experiences thought to be drug-related by the investigators for rofecoxib *vs* placebo, but relatively few patients discontinued treatment due to these adverse experiences. The increase was mainly due to small increases in adverse experiences known to be associated with NSAID use and previously reported for rofecoxib such as dyspepsia, hypertension, and peripheral edema. Despite the high mean age of the study population, few patients had confirmed upper gastrointestinal perforations, ulcers, or bleeds, although there were numerically more in the rofecoxib group than in the placebo group. Elderly patients are at increased risk for serious vascular events. Rofecoxib did not appear to increase the risk in this study, since the number of confirmed serious thrombotic vascular events was similar in each treatment group. The procedure used for determining confirmed cardiovascular events, such as heart attack and stroke, was the same as for the recent

3-year APPROVe (Adenomatous Polyp Prevention on Vioxx) study in 2586 patients (62% male, mean age 59 years) with a history of colorectal adenomas, which found an approximately two-fold increased relative risk for these events in patients taking rofecoxib *vs* placebo, beginning after 18 months of treatment (Bresalier *et al*, 2004). The reason for the discrepancy between the present findings and those from APPROVe is unclear. No striking differences were noted between the treatment groups with regard to other, nonserious, vascular adverse events in the present study. The total number of deaths and causes of death on-drug were consistent with expectations for an elderly population, and there was no specific pattern as to the cause of death in either treatment group. The off-drug mortality data are difficult to assess since follow-up information was available for less than half the patients, there was an imbalance in the extent of follow-up data between the two groups, and the majority of deaths occurred more than 48 weeks after treatment had been discontinued.

In conclusion, the finding that rofecoxib did not delay a diagnosis of AD in the present study, or slow the progression of AD in previous studies (Aisen *et al*, 2003; Reines *et al*, 2004), suggests that inhibition of COX-2 is not a useful therapeutic approach in AD. It is possible that nonselective NSAIDs may still be found to have beneficial effects in treating or delaying the progression of symptoms in AD, due to factors other than COX-2 inhibition. However, the only long-term results from an adequately designed study with a nonselective NSAID (naproxen) available to date are not encouraging (Aisen *et al*, 2003). Trials evaluating alternative, non-anti-inflammatory, approaches to delaying a diagnosis of AD in patients with MCI are underway (Petersen, 2003).

ACKNOWLEDGEMENTS

This study was funded by Merck Research Laboratories. SH Ferris was a paid consultant for Merck Research Laboratories on this study. L Kirby received funding from Merck Research Laboratories to participate in this study. GA Block, CR Lines, E Yuen, C Assaid, ML Nessly, BA Norman, CC Baranak, and SA Reines were employees of Merck Research Laboratories at the time the study was performed (GA Block is currently an employee of Astra Zeneca, and SA Reines and E Yuen are currently employees of Johnson & Johnson). These data were previously presented at the 42nd annual meeting of the American College of Neuropsychopharmacology, Puerto Rico, December 7–11, 2003.

Participating investigators in the Rofecoxib Protocol 078 study group were as follows: Piero Antuono, Froedtert Lutheran Memorial Hospital, Milwaukee, WI; Barry Baumel, Baumel-Eisner Neuromedical Institute, Miami, FL; Karen Bell, Columbia University, New York, NY; B Ian Blatt, Southeastern Pennsylvania Medical Institute, Havertown, PA; Martin Conway, Lovelace Scientific Resources Inc., Albuquerque, NM; Howard S Cummins, The Institute for Advanced Clinical Research, Elkins Park, PA; Mary J Derbenwick, Atlantic Institute of Clinical Research, Daytona Beach, FL; P Murali Doraiswamy, Duke University Medical Center, Durham, NC; Eugene DuBoff, Summit Research Network Inc., Denver, CO; Keith Edwards, Bennington, VT;

Larry Eisner, Baumel-Eisner Neuromedical Institute, Jay Ellis, Neuroscience Research, Pittsfield, MA; Donald England, Radiant Research, Eugene, OR; James Ferguson, Pharmacological Research Corp, Salt Lake City, UT; Jerome Goldstein, San Francisco Headache Clinic, San Francisco, CA; James Hartford, Summit Research Network, Cincinnati, OH; Jon Heiser, Pharmacology Research Institute, Northridge, CA; Daniel Hier, University of Illinois at Chicago, Chicago, IL; Bruce Hodges, Midwest Center for Clinical Studies, Kansas City, MO; Richard Holub, Neurological Associates of Albany, Albany, NY; Lawrence R Jenkyn, Dartmouth-Hitchcock Medical Center, Lebanon, NH; F Cleveland Kinney, University of Alabama at Birmingham, Birmingham, AL; Louis Kirby, Pivotal Research Centers, Peoria, AZ; Peter D Londborg, Seattle, WA; Elliott Mancall, Headache Center, Philadelphia, PA; David Margolin, Margolin Brain Institute, Fresno, CA; James McCarthy, Comprehensive NeuroScience Inc., South Yarmouth, MA; William McEntee, Comprehensive NeuroScience Inc., Sarasota, FL; Margarita Nunez, Comprehensive NeuroScience Inc., St Petersburg, FL; Jorg Pahl, Pahl Brain Associates PC, Oklahoma City, OK; Anton Porsteinsson, University of Rochester, Rochester, NY; Eric Pfeiffer, USF Suncoast Gerontology Center, Tampa, FL; Norman Relkin, New York Presbyterian Hospital, New York, NY; Joel S Ross, Long Branch, NJ; Barry Rovner, Philadelphia, PA; Ward Tolbert Smith, Summit Research, Portland, OR; John A Stoukides, Comprehensive NeuroScience Inc., East Providence, RI; Joseph Tate, DeKalb-Gwinnett, OB/GYN, PC, Norcross, GA; Stephen Thein Jr, Pacific Research Network, San Diego, CA; Harvey Tilker, Four Rivers Clinical Research Center, Paducah, KY; Jerry R Tindel, Radiant Research, Austin, TX; Jack Tomlinson, Alzheimer's Diagnostic Clinic, Wichita Falls, TX; Richard Weisler, Raleigh, NC; Jeanette Wendt, Neurological Associates of Tucson, Tucson, AZ; Kerri Wilks, Innovative Medical Research, Towson, MD; Parvaneh Zolnouri, California Clinical Trials, Beverly Hills, CA.

Members of the Rofecoxib Protocol 078 Endpoint Ajudication Committee were as follows: Steven Ferris, New York University School of Medicine, New York, NY; Mary Sano, Mount Sinai School of Medicine, Bronx Veterans Medical Research Center, New York, NY; Peter Whitehouse, University Hospitals of Cleveland, Cleveland, OH.

REFERENCES

- Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, Davis KL *et al* (2003). Effects of rofecoxib or naproxen *vs* placebo on Alzheimer's disease progression: a randomized controlled trial. *JAMA* **289**: 2819–2826.
- Andreasson KI, Savonenko A, Vidensky S, Goellner JJ, Zhang Y, Shaffer A *et al* (2001). Age-dependent cognitive deficits and neuronal apoptosis in cyclooxygenase-2 transgenic mice. *J Neurosci* **21**: 8198–8209.
- Birkenhager WH, Forette F, Seux ML, Wang JG, Staessen JA (2001). Blood pressure, cognitive functions, and prevention of dementias in older patients with hypertension. *Arch Intern Med* **161**: 152–156.
- Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B *et al* (2000). Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* **343**: 1520–1528.

- Bresalier RS, Sandler R, Riddell R, Lanus A, Morton D, Reicin A et al (2004). Preliminary Cardiovascular Safety Data from the APPROVe Study, Presented at the annual meeting of the American College of Rheumatology, San Antonio, Texas, October 18. Abstract available at: http://www.rheumatology.org/annual/press/APPROVesession_announce.asp.
- Buschke H (1973). Selective reminding for analysis of memory and learning. *J Verbal Learn Verbal Behav* 12: 543–550.
- Cagnin A, Brooks DJ, Kennedy AM, Gunn RN, Myers R, Turkheimer FE et al (2001). *In-vivo* measurement of activated microglia in dementia. *Lancet* 358: 461–467.
- Chertkow H (2002). Mild cognitive impairment. *Curr Opin Neurol* 15: 401–407.
- Consilvio C, Vincent AM, Feldman EL (2004). Neuroinflammation, COX-2, and ALS—a dual role? *Exp Neurol* 187: 1–10.
- Etmnan M, Gill S, Samii A (2003). Effect of non-steroidal anti-inflammatory drugs on risk of Alzheimer's disease: systematic review and meta-analysis of observational studies. *BMJ* 327: 128.
- Folstein MF, Folstein SE, McHugh PR (1975). 'Mini-Mental State': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res* 12: 189–198.
- Gertz BJ, Krupa D, Bolognese JA, Sperling RS, Reicin A (2002). A comparison of adverse renovascular experiences among osteoarthritis patients treated with rofecoxib and comparator non-selective non-steroidal anti-inflammatory agents. *Curr Med Res Opin* 18: 82–91.
- Giovannini MG, Scali C, Prosperi C, Belluci A, Pepeu G, Casamenti F (2003). Experimental brain inflammation and neurodegeneration as model of Alzheimer's disease: protective effects of selective COX-2 inhibitors. *Int J Immunopathol Pharmacol* 16(Suppl 2): 31–40.
- Hamilton M (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23: 56–62.
- Ho L, Purhoit D, Haroutunian V, Luteran JD, Willis F, Naslund J et al (2001). Neuronal cyclooxygenase 2 expression in the hippocampal formation as a function of the clinical progression of Alzheimer disease. *Arch Neurol* 58: 487–492.
- in 't Veld BA, Launer LJ, Breteler MMB, Hofman A, Stricker BH Ch (2002). Pharmacological agents associated with a preventive effect on Alzheimer's disease: a review of the epidemiologic data. *Epidemiol Rev* 24: 248–268.
- in 't Veld BA, Ruitenberg A, Hofman A, Launer LJ, van Duijn CK, Stinjen T et al (2001). Nonsteroidal antiinflammatory drugs and the risk of Alzheimer's disease. *N Engl J Med* 345: 1515–1521.
- Ivnik RJ, Malec JF, Tangalos EG, Petersen RC, Kokmen E, Kurland LT (1990). The auditory-verbal learning test (AVLT): norms for ages 55 years and older. *Psychol Assess: J Consult Clin Psychol* 2: 304–312.
- Jack Jr CR, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ et al (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology* 52: 1397–1403.
- Jain NK, Patil CS, Kulkarni SK, Singh A (2002). Modulatory role of cyclooxygenase inhibitors in aging- and scopolamine or lipopolysaccharide-induced cognitive dysfunction in mice. *Behav Brain Res* 133: 369–376.
- Konstam MA, Weir MR, Reicin A, Shapiro D, Sperling RS, Barr E et al (2001). Cardiovascular thrombotic events in controlled, clinical trials of rofecoxib. *Circulation* 104: 2280–2288.
- Langman MJ, Jensen DM, Watson DJ, Harper SE, Zhao PL, Quan H et al (1999). Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. *JAMA* 282: 1929–1933.
- Lines CR, McCarroll KA, Lipton RB, Block GA (2003). Telephone screening for amnesic mild cognitive impairment. *Neurology* 60: 261–266.
- McGeer PL, Schulzer M, McGeer EG (1996). Arthritis and anti-inflammatory agents as possible neuroprotective factors for Alzheimer's disease: a review of 17 epidemiological studies. *Neurology* 47: 425–432.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS/ADRDA work group under the auspices of the Department of Health and Human Services task force on Alzheimer's disease. *Neurology* 34: 939–944.
- Merck & Co. Inc. (2003). VIOXX[®] (rofecoxib tablets and oral suspension). In: *Physicians' Desk Reference*[®]. Thomson PDR: Montvale, NJ. pp 2120–2125.
- Morris JC (1993). The clinical dementia rating (CDR): current version and scoring rules. *Neurology* 43: 2412–2414.
- Morris JC, Mohs RC, Rogers H, Fillenbaum G, Heyman A (1988). Consortium to establish a registry for Alzheimer's disease (CERAD). Clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol Bull* 24: 641–652.
- Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH et al (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol* 58: 397–405.
- Pasinetti GM (2001). Cyclooxygenase and inflammation in the clinical progression of Alzheimer's disease. In: Vane JR, Botting RM (eds). *Therapeutic Roles of Selective COX-2 Inhibitors*. William Harvey Press: London. pp 191–205.
- Petersen RC (2003). Mild cognitive impairment clinical trials. *Nat Rev: Drug Discov* 2: 646–653.
- Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV et al (2001a). Current concepts in mild cognitive impairment. *Arch Neurol* 58: 1985–1992.
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999). Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 56: 303–308.
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST (2001b). Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56: 1133–1142.
- Qin W, Ho L, Pompl PN, Peng Y, Zhao Z, Xiang Z et al (2003). Cyclooxygenase (COX)-2 and COX-1 potentiate β -amyloid peptide generation through mechanisms that involve γ -secretase activity. *J Biol Chem* 278: 50970–50977.
- Reicin AS, Shapiro D, Sperling R, Barr E, Yu Q (2002). Comparison of cardiovascular thrombotic events in patients with osteoarthritis treated with rofecoxib versus nonselective nonsteroidal anti-inflammatory drugs (ibuprofen, diclofenac, and nabumetone). *Am J Cardiol* 89: 204–209.
- Reines SA, Block GA, Morris JC, Liu G, Nessly ML, Lines CR et al (2004). Rofecoxib: no effect on Alzheimer's disease in a 1-year, randomized, blinded, controlled study. *Neurology* 62: 66–71.
- Rogers J, Kirby LC, Hempelman SR, Berry DL, McGeer PL, Kaszniak AW et al (1993). Clinical trial of indomethacin in Alzheimer's disease. *Neurology* 43: 1609–1611.
- Rose JW, Hill KE, Watt HE, Carlson NG (2004). Inflammatory cell expression of cyclooxygenase-2 in the multiple sclerosis lesion. *J Neuroimmunol* 149: 40–49.
- Rosen WG, Mohs RC, Davis KL (1984). A new rating scale for Alzheimer's disease. *Am J Psychiatry* 141: 1356–1364.
- Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A (1980). Pathological verification of ischemic score in differentiation of dementias. *Ann Neurol* 7: 486–488.
- Sainati S, Ingram D, Talwalker S, Geiss G (2000). Results of a double-blind, randomized, placebo-controlled study of celecoxib in the treatment of the progression of Alzheimer's disease. Sixth International Stockholm/Springfield Symposium on Advances in Alzheimer Therapy, 180 [Abstract].
- Scali C, Giovannini MG, Prosperi C, Belucci A, Pepeu G, Casamenti F (2003). The selective cyclooxygenase-2 inhibitor rofecoxib suppresses brain inflammation and protects cholinergic neurons from excitotoxic degeneration *in vivo*. *Neuroscience* 117: 909–919.

- Scharf, Mander E, Ugoni A, Vajda F, Christophidis N (1999). A double-blind, placebo-controlled trial of diclofenac/misoprostol in Alzheimer's disease. *Neurology* 53: 197–201.
- Schmidt M (1996). *Rey Auditory Verbal Learning Test. A Handbook*. Western Psychological Services: Los Angeles.
- Schwartz JI, Vandormael K, Malice MP, Kalyani RN, Lasseter KC, Holmes GB et al (2002). Comparison of rofecoxib, celecoxib, and naproxen on renal function in elderly subjects receiving a normal-salt diet. *Clin Pharmacol Ther* 72: 50–61.
- Skoog I (1997). The relationship between blood pressure and dementia: a review. *Biomed Pharmacother* 51: 367–375.
- Teismann P, Vila M, Choi DK, Tieu K, Wu DC, Jackson-Lewis V et al (2003). COX-2 and neurodegeneration in Parkinson's disease. *Ann NY Acad Sci* 991: 272–277.
- Van Gool WA, Weinstein HC, Scheltens P, Walstra GJ, Scheltens PK (2001). Effect of hydroxychloroquine on progression of dementia in early Alzheimer's disease: an 18-month randomised, double-blind, placebo-controlled study. *Lancet* 358: 455–460.
- Warner TD, Giuliano F, Vojnovic I, Bukasa A, Mitchell JA, Vane JR (1999). Nonsteroid drug selectivities for cyclo-oxygenase-1 rather than cyclo-oxygenase-2 are associated with human gastrointestinal toxicity: a full *in vitro* analysis. *Proc Natl Acad Sci USA* 96: 7563–7568.
- Xiang Z, Ho L, Yemul S, Zhao Z, Qin W, Pompl P et al (2002). Cyclooxygenase-2 promotes amyloid plaque deposition in a mouse model of Alzheimer's disease neuropathology. *Gene Expr* 10: 271–278.

APPENDIX A1

Details of Tests Used

MMSE (Folstein et al, 1975): This is a brief test of overall cognitive function (range of scores = 0–30, lower score indicates greater impairment). Patients who scored <24 were excluded because of the possibility that they might have dementia.

AVLT (Schmidt, 1996): This is a 15-item, five-trial, word recall test (range of scores = 0–75, lower score indicates greater impairment). The interference trial and delayed recall trial, which usually follow the initial five-trial learning phase, were not administered. Patients had to score ≤37 words recalled correctly over five trials as objective evidence of memory impairment. This corresponds to a score ≤1 SD below the mean for normal elderly subjects aged 65–69 (Ivnik et al, 1990). For the first 6 months of study enrollment, age-adjusted cut scores 1.5 SDs below the means for separate age bands were used (eg there was a cut

score for those aged 67–69, another cut score for those aged 70–72, etc), but the criteria were modified after 6 months due to concerns that the original criteria may have been too strict and were conflicting with investigator's clinical judgment (see Petersen, 2003), resulting in low enrollment.

CDR (Morris, 1993): This is a global dementia rating scale, based on a semistructured interview with the patient and an informant. Patients are rated on dementia severity (0 = none, 0.5 = questionable dementia, 1 = mild dementia, 2 = moderate dementia, 3 = severe dementia) in six domains or 'boxes' (memory, orientation, judgment and problem solving, community affairs, home and hobbies, personal care). The global CDR score (range = 0–3, higher score indicates greater impairment) was determined according to a computerized version of the published scoring algorithm (<http://www.biostat.wustl.edu/~adrc/cdrpgm/index.html>). The sum of scores on the individual boxes was also calculated (range = 0–18, higher score indicates greater impairment). Patients had to have a global CDR score of 0.5 (questionable dementia) with a score of at least 0.5 in the memory domain to qualify for study entry.

ADAS-Cog (Rosen et al, 1984): This consists of 11 performance-based subtests of cognition, including measures of memory and praxis. The error scores on each subtest are added to give the total score (range = 0–70, higher score indicates greater impairment).

SRT (Buschke, 1973): This is a 12-item, six-trial, word list recall test with selective reminding of words not correctly recalled on the immediately preceding trial. Following the sixth trial, there is a 5 min delay, after which the patients are again asked to recall the list of 12 words. Two scores were derived—the sum of the number of words recalled over each of the six selective reminding trials (SRT-summed recall: range = 0–72, lower score indicates greater impairment) and the number of words recalled after the 5 min delay (SRT-delayed recall; range = 0–12, lower score indicates greater impairment).

BDRS (Morris et al, 1988): The version from the CERAD battery was used. This is an informant-based rating of a patient's ability to perform activities of daily living such as household tasks (Part 1) and self-care (Part 2) (range of scores = 0–17, higher score indicates greater impairment). Patients who scored >3.5 with any Part 1 item score >0.5 (indicating severe impairment) were excluded because of the possibility of dementia.